



UNIVERSITY OF
BIRMINGHAM

EXPLORING THE IMPACT OF TECHNOLOGY ON ENGLISH AS A SECOND
LANGUAGE WRITING: LINKING OUTCOMES, STUDENT ENGAGEMENT, AND
TEACHER GUIDANCE

by

NGUYEN HOAI LY NGA

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Department of Linguistics and Communication
School of English, Drama and Creative Studies
College of Arts and Law
University of Birmingham

December 2025

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive
e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

ACKNOWLEDGMENT

First and foremost, I would like to express my deepest gratitude to my family for their unwavering love, support, and encouragement throughout my life. I would not have been able to undertake, let alone complete, this PhD journey without them.

I am also profoundly grateful to my supervisors, Dr. Amanda Patten and Dr. Garry Plappert, for their invaluable guidance, consistent support, and insightful feedback. Their expertise and encouragement have been pivotal in shaping the development and completion of this thesis.

Finally, I extend my sincere thanks to my friends and colleagues in Vietnam for their continued support and reassurance during this journey.

ABSTRACT

This study investigates the impact of four digital writing tools—Sketch Engine (corpus), Google Translate (machine translation), QuillBot (automated paraphrasing), and Grammarly (automated writing evaluation)—on English as a second language (ESL) university students' IELTS Academic Writing Task 2 performance, tool-use strategies, and perceptions. It also examines the extent to which students' actual tool-use practices align with teachers' recommended strategies and explores how behavioural, cognitive, and emotional engagement ("critical engagement") mediates writing outcomes. Twenty ESL Vietnamese university students, with various levels of English proficiency, completed two timed IELTS Task 2 essays: Essay 1 without tools and Essay 2 with access to all four tools, followed by semi-structured interviews. Tool use strategies of all 20 students were directly observed and screen-recorded by the researcher. Essays were rated by three independent raters using the IELTS rubric. Ten ESL teachers provided lesson plans with recommendations of tool usage and participated in one-on-one interviews with the researcher.

Quantitative findings showed a statistically significant improvement in overall writing performance following tool use: 15 of 20 students increased their total scores, with the mean rising from 6.48 to 6.87 ($p = .0026$). Significant changes were observed in Lexical Resource (LR), and Grammatical Range and Accuracy (GA) criteria, but not in Coherence and Cohesion (CC) and Task Response (TR). Among the four tools, only Sketch Engine exhibited a moderate positive association with Essay 2 scores ($r = .529$, $p = .016$).

Qualitative findings further suggested that perceived language learning gains were shaped less by tool access or frequency of use than by students' cognitive engagement among higher-scoring students (mean Essay 2 score ≥ 6.5). Higher scorers showed stronger cognitive engagement by using multiple tools strategically, cross-checking and rejecting inaccurate outputs, and demonstrating awareness of tool limitations. They also reflected more deeply on vocabulary, collocation, and error noticing as learning opportunities during interviews. In contrast, lower scorers (mean Essay 2 score < 6.5) tended to rely on a single tool for basic functions and accepted suggestions with minimal contextual evaluation, showing weaker evidence of noticing and revision. Case study analyses further indicated that score declines

were often associated with uncritical adoption of tool outputs, producing awkward phrasing or errors, including occasional small declines among higher-proficiency learners when feedback missed subtle lexical or coherence problems.

Notable gaps were identified between teacher guidance and student practice. Teachers generally promoted critical evaluation and assigned stage-specific roles to tools (e.g., Google Translate and Sketch Engine for planning; QuillBot for composing; Grammarly for revising). However, many students bypassed planning and concentrated tool use in composing and revision, with Google Translate and Grammarly dominating overall usage. Alignment with teacher-recommended multi-tool strategies was most consistently observed among high scorers, who also extended strategies (e.g., back-translation) in flexible ways, whereas low scorers more often used tools as shortcuts, constraining language development even when scores improved.

The thesis contributes theoretically by linking engagement with Output and Noticing theories, showing that language learning depends on how learners actively produce texts, evaluate and refine tool outputs, rather than access to technology. Empirically, it demonstrates that technology access alone does not ensure learning, identifies an ethics gap regarding plagiarism and overreliance, and notes the growing convergence of tool functions as Generative AI evolves. Pedagogically, it calls for explicit training in cognitive engagement for tool use, proficiency-sensitive scaffolding, and ethical awareness of Generative AI use for essay writing. Methodologically, it shows the value of triangulating multiple data sources to capture behavioural, cognitive, and emotional engagement. The study concludes that digital tools can enhance L2 writing and L2 learning only when learners engage with them strategically, positioning technology as support rather than substitute.

Table of Contents

CHAPTER 1: INTRODUCTION	1
Background of the study	1
1.1 Theories in the writing process.....	1
1.2 Theories in second language acquisition and engagement	2
1.3 Digital writing tools	3
1.3.1 Corpus tools.....	4
1.3.2 Machine translation tools	5
1.3.3 Automated paraphrasing tools	5
1.3.4 Automated writing evaluation tools	5
1.3.5 Connection to the present study	5
1.4 Research objectives.....	6
1.5 Research questions	7
1.6 Methodology overview	8
1.7 Scope and delimitation.....	9
1.7.1 The IELTS test.....	9
1.7.2 Writing score outcomes	10
1.7.3 Language learning indicators.....	11
1.7.4 Cognitive engagement.....	11
1.8 Contribution.....	12
1.9 Thesis structure.....	13
CHAPTER 2: LITERATURE REVIEW	14
Overview of Chapter 2 sections	14
2.1. Theoretical frameworks	15
2.1.1 The Output Hypothesis	16
2.1.2 The Noticing Hypothesis	17
2.1.3 The concept of engagement and cognitive engagement in language learning.....	18
2.1.4 Linking the theoretical frameworks to the current study	19
2.2. Empirical studies on corpus tools in L2 writing	20
2.2.1 Corpus tools, corpora and data-driven learning approach	20
2.2.2 Sketch Engine.....	23

2.2.3 Merits and shortcomings of using corpus tools for L2 writing and language learning	25
2.2.4 Corpus tool usage in relation to language proficiency.....	27
2.2.5 The use of corpus tools by teachers	27
2.2.6 The use of corpus tools by students	29
2.2.7 Perceptions of teachers and students towards DDL and corpus tools ..	30
2.2.7.1 Perceptions of teachers.....	30
2.2.7.2 Perceptions of students	31
2.2.8 Summary of literature review on corpus tools	32
2.3. Empirical studies on machine translation tools in L2 writing	32
2.3.1 Machine translation tools and Google Translate	32
2.3.2 Merits and shortcomings of using machine translation tools for L2 writing and language learning	33
2.3.3 Machine translation usage in relation to language proficiency	34
2.3.4 The use of machine translation tools by teachers	35
2.3.5 The use of machine translation tools by students	36
2.3.6 Perceptions of teachers and students towards machine translation tools	37
2.3.6.1 Perceptions of teachers.....	37
2.3.6.2 Perceptions of students	39
2.3.7 Summary of literature review on machine translation tools	40
2.4. Empirical studies on automated paraphrasing tools in L2 writing	40
2.4.1 Automated paraphrasing tools and Quillbot	40
2.4.2. Merits and shortcomings of using automated paraphrasing tools for L2 writing and language learning	42
2.4.3 Automated paraphrasing tool usage in relation to language proficiency.....	43
2.4.4 The use of automated paraphrasing tools by teachers	44
2.4.5 The use of paraphrasing tools by students	45
2.4.6 Perceptions of teachers and students towards automated paraphrasing tools	45

2.4.6.1 Perceptions of teachers	45
2.4.6.2 Perceptions of students.....	46
2.4.7 Summary of literature review on automated paraphrasing tools	47
2.5. Empirical studies on automated writing evaluation tools.....	48
2.5.1 Introduction to automated writing evaluation tools	48
2.5.2 Grammarly	49
2.5.3. Merits and shortcomings of using automated writing evaluation tools for L2 writing	51
2.5.4. Automated writing evaluation tools usage in relation to language proficiency.....	53
2.5.5 The use of automated writing evaluation tools by teachers	54
2.5.6 The use of automated writing evaluation tools by students.....	54
2.5.7 Perceptions of ESL teachers and students towards AWE tools	55
2.5.7.1 Perceptions of teachers.....	55
2.5.7.2 Perceptions of students	57
2.5.8 Summary of literature review on the use of automated writing evaluation tools.....	58
2.6 Generative AI tools and L2 writing	59
2.6.1 Perceptions of teachers and students towards Generative AI	59
2.7. Research gaps and research questions.....	60
2.7.1 Research gaps related to the four tools' impact on IELTS writing scores and language development.....	60
2.7.2 Research gaps related to tool usage strategy alignment between teachers and students	62
2.7.3 Research gaps related to perceptions of tools between teachers and students	64
2.7.4 Research gaps in methodology and context.....	67
2.8 Chapter summary.....	68

CHAPTER 3: METHODOLOGY 71

3.1 Rationale for using the International English Language Testing System (IELTS) in this study.....	75
3.1.1 Detail and validity of the IELTS.....	75

3.2. The recruitment of student participants and the writing topic selection.....	86
3.3 The profiles of raters and the essay marking procedure	91
3.3.1 Summary of each rater's assessment.....	95
3.3.2 Summary of scores given by each rater.....	96
3.3.3. The inter-rater reliability rates	98
3.4. Quantitative analysis methodology for writing scores	100
3.4.1 Intraclass Correlation Coefficient (ICC).....	100
3.4.2 Paired T-tests	101
3.4.3 Pearson's Correlation Coefficient	101
3.4.4 Calculating overall scores and individual score changes	102
3.4.5 Quantifying students' tool usage instances.....	102
3.5. Qualitative analysis of students' tool usage strategy and interview responses.....	102
3.5.1 Interview questions for students	102
3.5.2 Qualitative analysis of student tool engagement and language learning	104
3.5.2.1 Classification of language proficiency levels for comparative engagement analysis.....	104
3.5.2.2 Coding for thematic analysis of students' interview responses ...	104
3.5.2.3. Analysis of case study	105
3.6 Teacher recruitment, interview questions, and lesson plan analysis.....	107
3.6.1 Recruitment process of teacher participants.....	107
3.6.2 Essay topic selection for teachers	110
3.6.3 Interview questions for teachers	111
3.6.4 Lesson plan analysis.....	113
3.7. Comparison of teachers' and students' tool usage, evaluation and recommendation.....	114
3.7.1 Comparison of quantified data from teachers' and students' tool usage	114
3.7.2. Comparison of teachers' and students' tool evaluation	116
3.7.3 Comparison of students' evaluations and writing performance.....	117

3.7.4 Comparison of teachers' and students' recommendations for additional tool features	117
3.8 Ethical considerations	118
CHAPTER 4: FINDINGS IN SCORE IMPACT OF THE FOUR TOOLS	120
4.1. Statistical analysis of writing scores and the relationship between tool usage and score changes	120
4.1.1 Summary of mean scores from three raters and the overall changes	121
4.1.2 T-test analysis of mean score changes across four criteria and overall scores	123
4.1.3 Correlation analysis between the mean overall scores of Essay 1 and Essay 2	125
4.2 The individual changes in mean scores	126
4.2.1 Relationship between the instances of tool usage and Essay 2 scores	128
4.2.2 Summary of findings in score changes after the four tools' intervention	131
4.3 Analysis of tool engagement	133
4.3.1 Classification of language proficiency levels	133
4.3.2 Thematic analysis of interview data	134
4.3.3 Comparison of thematic findings from the high-scoring and low-scoring	140
4.3.4 Summary of thematic findings	162
4.4 Factors contributing to decreased writing scores	163
4.4.1 The Case of Student 1	164
4.4.2 The Case of Student 3	176
4.4.3 The Case of Student 12	186
4.4.4 The Case of Student 17	195
4.4.5 Factors contributing to score decreases	206
4.5 Summary of findings in Chapter 4	208
CHAPTER 5: FINDINGS IN TOOL USAGE AND EVALUATIONS BETWEEN TEACHERS AND STUDENTS.....	210
5.1 Comparative analysis of tool usage between teachers and students	212

5.2 Comparative analysis of Sketch Engine	218
5.2.1 Planning.....	218
5.2.2 Composing.....	220
5.2.3 Revising	221
5.3 Comparative analysis of Google Translate	225
5.3.1 Planning.....	226
5.3.2 Composing.....	227
5.3.3 Revising	233
5.4 Comparative analysis of Quillbot.....	236
5.4.1 Planning.....	236
5.4.2 Composing.....	237
5.4.3 Revising	243
5.5. Comparative analysis of Grammarly	252
5.5.1 Planning.....	252
5.5.2 Composing.....	254
5.5.3 Revising	255
5.6 Summary of findings in tool usage between teachers and students.....	262
5.7 Findings in teachers' and students' evaluations of the four tools	267
5.7.1 Comparative analysis of teachers' and students' tool evaluations	268
5.7.2 Additional tool recommendations from teachers and students.....	278
5.7.2.1 Teachers' suggested alternatives to AI.....	282
5.7.2.2 Students' recommendations of other tools.....	284
5.7.3 Summary of findings in tools' evaluations and recommendations between teachers and students	291
CHAPTER 6: DISCUSSION.....	295
6.1. Discussion of key findings.....	295
6.1.1 Findings on students' writing scores and tool usage.....	296
6.1.2 Findings on students' tool engagement, writing score outcomes and language learning	298
6.1.2.1 Tool engagement and writing score outcomes	298
6.1.2.2 Tool engagement and language learning	301
6.1.2.3 Factors contributing to score decreases	302

6.2 Comparison of teachers' and students' tool usage, evaluations and recommendations.....	302
6.2.1 Comparison of teachers' and students' tool usage	302
6.2.2 Comparison of teachers' and students' tool evaluations	30
6.2.3 Comparison of students' evaluations and writing score outcomes.....	309
6.2.4 Comparison of teachers' and students' tool recommendations.....	310
6.3 Implications	311
CHAPTER 7: CONCLUSION.....	316
7.1. Purpose and research questions.....	316
7.2 Summary of key findings	317
7.3 Contributions	321
7.3.1 Theoretical contribution.....	321
7.3.2 Empirical contribution.....	322
7.3.3 Pedagogical contribution.....	322
7.3.4 Methodological contribution	323
7.4 Limitations of the study	323
7.4.1 Context limitations.....	323
7.4.2 Participant limitations	324
7.4.3 Methodological limitations.....	325
7.5 Future research directions	326
7.6 Concluding statement	327
REFERENCES.....	328
APPENDICES	389

List of Tables

Table 1. List of abbreviations in this study.....	xix
Table 2. Examples of the four tool types and their functions	3
Table 3. The participants with their coded names, their latest IELTS writing scores and their latest test year.	89
Table 4. Summary of the marking procedure.	92
Table 5. Descriptive statistics of scores by Rater 1	97
Table 6. Descriptive statistics of scores by Rater 2	97
Table 7. Descriptive statistics of scores by Rater 3	98
Table 8. Inter-rater reliability for Essay 1 across four criteria.....	99
Table 9. Inter-rater reliability for Essay 2 across four criteria.....	99
Table 10. Interview questions for students	102
Table 11. The teaching experience of 10 teacher participants	109
Table 12. The 19 interview questions for teachers	112
Table 13. An example of the lesson plan of Teacher 1.	113
Table 14. An example of quantified data of tool usage comparison between teachers and students in the planning stage	115
Table 15. Example comparison of teachers' and students' evaluations of Sketch Engine	116
Table 16. An example of teachers' and students' recommendations of other tools and their usage purposes	118
Table 17. Summary of mean scores among the four criteria from three raters ...	121
Table 18. Descriptive statistics of mean overall score changes between Essay 1 and Essay 2.....	122
Table 19. Paired t-tests on the mean score change from each criterion of three raters	124
Table 20. Pearson correlation coefficient between the mean overall scores of Essay 1 and Essay 2	125
Table 21. The mean overall scores of each essay and the mean overall score change of each student	126
Table 22. Essay scores, score changes, individual and total tool usage instances of each of 20 students.	128

Table 23. Pearson correlation coefficient between tool usage instances and Essay 2 scores.....	129
Table 24. The coding by colour and theme for the 20 interview transcripts and examples of responses.	135
Table 25. Example of the thematic findings from Student 1's interview transcript	136
Table 26. A summary of the number of theme instances and the specific tool mentioned in each theme between the low-scoring and high-scoring groups.	141
Table 27. The comparison of Essay 1 and Essay 2 scores of Student 1.....	171
Table 28. The interview transcript of Student 1	172
Table 29. The comparison of Essay 1 and Essay 2 scores of Student 3.....	181
Table 30. The interview transcript of Student 3	182
Table 31. Essay 1 and Essay 2 scores of Student 12	190
Table 32. The interview transcript of Student 12	192
Table 33. Essay 1 and Essay 2 scores of Student 17	201
Table 34. The interview transcript of Student 17	202
Table 35. An example of the lesson plan of Teacher 1.	211
Table 36. A summary of the recommended usage of the four tools presented at three main stages of the writing process by 10 teachers.....	212
Table 37. An example of transcriptions from Teacher 1.	213
Table 38. A summary of the usage of the four tools presented at three main stages of the writing process by 20 students	214
Table 39. A comparative summary of the recommended and actual usage of the four tools presented at three main stages of the writing process by teachers and students.....	215
Table 40. Comparative findings from teachers' and students' evaluations of Sketch Engine	269
Table 41. Comparative findings from teachers' and students' evaluations of Google Translate.	271
Table 42. Comparative findings from teachers' and students' evaluations of Quillbot	273
Table 43. Comparative findings from teachers' and students' evaluations of Grammarly.....	274

Table 44. A summary of teachers' recommendations of other tools and their usage purposes.....	278
Table 45. Summary of other tool recommendations from students and their purposes.....	284

APPENDICES

Table 5a. Detailed scores of 40 essays marked by Rater 1	423
Table 6a. Detailed scores of 40 essays marked by Rater 2	424
Table 7a. Detailed scores of 40 essays marked by Rater 3	425
Table 17a. Detailed mean scores of 40 essays from three raters	426
Table 24a. Full thematic coding for the interview transcripts of 20 students	428
Table 35a. Full lesson plans of ten teacher participants.....	472
Table 37a. Interview transcripts of ten teacher participants	482

List of Figures

Figure 1. Websites of corpus tools and their available corpora. From Charles (2018, p.2)	22
Figure 2. Examples of concordances for “research” and “suggest” in Sketch Engine. From Frankenberg-Garcia et al. (2019, p. 26)	24
Figure 3. An example of the “Word sketch” function results for the word “research” in Sketch Engine. From Frankenberg-Garcia et al. 2019, p. 26)	25
Figure 4. The interface of Quillbot as an automated paraphrasing tool	42
Figure 5. The interface of Grammarly feedback is presented in the free version of Grammarly (2023)	50
Figure 6. The research procedure and data processing sequence for students....	73
Figure 7. The research procedure and data processing sequence for teachers ...	74
Figure 8. The old version of the public scoring criteria for Academic IELTS Writing Task 2 by the British Council, retrieved from Uysal (2009).....	77
Figure 9. The updated public scoring criteria for Academic IELTS Writing Task 2 by the British Council (2023).	78
Figure 10. Scoring Profile (Jacob et al.,1981)	81
Figure 11. The Writing Rubric of Cambridge C1 Advanced English (CAE). Retrieved from Cambridge English (2020)	82
Figure 12. The Writing Rubric of Cambridge C2 Proficiency in English (CPE). Retrieved from Cambridge English (2023)	83
Figure 13. Rubric for Writing an academic discussion in TOEFL iBT. Retrieved from Educational Testing Service (n.d).	84
Figure 14. The essay questions in the main study	88
Figure 15. Examples of Rater 2’s comments on paper and Rater 3’ comments on Google Docs.....	95
Figure 16. Summary of the process for analysing the impact of the four tools and additional factors contributing to score declines.	106
Figure 17. The essay topic for teachers.	110
Figure 18. An example of how teachers can use Sketch Engine for lesson planning on the chosen topic.	111
Figure 19. Essay 2 of Student 1	165

Figure 20. Student 1 used Sketch Engine to review the concordance of “real-life interaction”	166
Figure 21. Student 1 used Sketch Engine to search for the word “remote learning” and chose the phrase “off-campus”	166
Figure 22. Student 1 typed in the word “top-notch” in Sketch Engine and replaced this word with the word “world-class”	167
Figure 23. Student 1 used Google Translate to check the meaning of the word “interruption” in Vietnamese.	168
Figure 24. Student 1 used Google Translate to check the meaning of the word “fully” in Vietnamese	168
Figure 25. Student 1 used Google Translate to check the meaning of the word “re-interact” in Vietnamese	169
Figure 26. Student 1 experiencing a technology issue with Quillbot.....	169
Figure 27. Screenshot of Student 1 checking the suggestion from Grammarly to correct article usage	170
Figure 28. The Essay 2 of Student 3	176
Figure 29. Student 3 used Sketch Engine to the collocation of “energy” and found the word “harness”	177
Figure 30. Student 3 used Google Translate to find the word “coal” in English. ...	177
Figure 31. Student 3 used Google Translate to find the phrase “burning coal wastes CO ₂ ”, but she did not use this result.	178
Figure 32. Student 3 used Google Translate to find the phrase “petrol-powered car” in English.....	178
Figure 33. Student 3 was looking for synonyms of “harmful by-product” in Quillbot	179
Figure 34. Student 3 reviewed suggestions from Grammarly before accepting all suggestions	180
Figure 35. The Essay 2 of Student 12	187
Figure 36. Student 12 used Google Translate to translate three phrases into English during the planning stage	187
Figure 37. Student 12 included the three phrases in her draft during the planning stage	188
Figure 38. Student 12 used Sketch Engine to check the concordance of the phrase “exhausted fumes”	189

Figure 39. Student 12 used Grammarly to correct her second essay but did not make any changes.	189
Figure 40. The Essay 2 of Student 17	196
Figure 41. Student 17 composed the first draft of Essay 2 independently in English.	197
Figure 42. Student 17 translated a clause from Vietnamese to English	197
Figure 43. Student 17 used some translated phrases and completed Sentence 14 independently	198
Figure 44. Student 17 revised the word “causes” to “cause” in Sentence 16 “	198
Figure 45. Student 17 translated two clauses of Sentence 16.....	198
Figure 46. Student 17 revised the phrase “mineral deposits” – a phrase suggested by Google Translate – to “mineral sources” in order to compose Sentence 16 ...	199
Figure 47. Student 17 used Quillbot to paraphrase Sentence 8	199
Figure 48. Grammarly identified an error with article usage in Sentence 1 and the preposition error in Sentence 17, but Student 17 only accepted the first change.	200
Figure 49. Student 13 searched for the concordance of the phrase “green energy” but she did not use this phrase.....	223
Figure 50. Student 13 used Sketch Engine to check the concordance of the phrase “CO2 emission”.	223
Figure 51. Screenshots from the computer of Student 16 using Google Translate during her second essay composition	231
Figure 52. Student 20 composed in Vietnamese and used Google Translate to translate the whole essay	231
Figure 53. Student 18 used Google Translate from English to Vietnamese to revise her phrases.	234
Figure 54. Student 19 used Google Translate from English to Vietnamese to revise her written essay.	234
Figure 55. Student 5 used Quillbot to find the synonyms of “distance learning” and chose the word “distance education” and “learning from afar” in his essay.	240
Figure 56. Student 7 paraphrased a sentence with Quillbot and selected the phrase “wholeheartedly endorse” from its output.	240
Figure 57. Student 9 used Quillbot to paraphrase the sentence that she wrote, paying attention to the word “mitigate”.	241

Figure 58. Student 10 used Quillbot to paraphrase the two sentences. He only chose the first sentence from the output.	241
Figure 59. Student 13 used Quillbot to paraphrase the phrase “renewable source” and she selected “alternative energy”	241
Figure 60. Student 6 paraphrased the first and second paragraphs, but he only chose the second paragraph output.	245
Figure 61. Student 7 used Quillbot to paraphrase a sentence and she chose the output.	245
Figure 62. Student 9 used Quillbot to paraphrase her introduction sentence and used it as the conclusion.	245
Figure 63. Student 11 used Quillbot to paraphrase the introduction and chose the output.	246
Figure 64. Student 15 used Quillbot to paraphrase the first body paragraph and made changes.....	246
Figure 65. Student 17 used Quillbot to paraphrase a sentence.....	246
Figure 66. Student 18 used Quillbot to paraphrase two paragraphs. She selectively modified the output and chose only two sentences	247
Figure 67. Student 19 used Quillbot to paraphrase two sentences and used the output.	247
Figure 68. Student 20 used Quillbot to paraphrase the first paragraph, which was the output of Google Translate from his Vietnamese text. He chose the output but did not use the last sentence.....	248

Glossary

Table 1. List of abbreviations in this study

L1	First language
L2	Second language
SLA	Second language acquisition
CEFR	Common European Framework of Reference for Languages
IELTS	International English Language Testing System
TR	Task Response
CC	Coherence and Cohesion
LR	Lexical Resource
GA	Grammatical range and Accuracy
TOEFL	Test of English as a Foreign Language
ESL	English as a second language
CALL	Computer-assisted language learning
AI	Artificial intelligence
CTs	Corpus tools

APTs	Automated paraphrasing tools
AWETs	Automated writing evaluation tools
MTTs	Machine translation tools

CHAPTER 1: INTRODUCTION

Background of the study

1.1 Theories in the writing process

The skill of transmitting ideas through written texts has enabled humans to communicate across space and time. By making language visible and permanent, writing is the foundation of much of human communication, and it is inseparably linked to tools, as Nietzsche stated when reflecting on his typewriter, “Our writing equipment takes part in the forming of our thoughts” (Kittler, 1990, p. 195, cited in Carr, 2008). From the typewriter to digital platforms, writing tools have transformed the ways in which texts are composed.

Since the 1980s, writing has been conceptualised as a complex, non-linear process shaped by multiple factors, with Flower and Hayes’ (1981) cognitive model serving as a foundational framework. Their framework identified three recursive stages: planning, translating, and reviewing, highlighting the dynamic and exploratory nature of text production. Later, Chenoweth and Hayes (2003) refined this model by introducing the ‘Transcriber,’ which captures the process of converting mental language into visible text. Hayes and Berninger (2012) further advanced this framework by explicitly incorporating transcription methods—typing or handwriting—as integral to the writing environment, thereby recognising technology as an additional factor in the composing process.

Building on these models, a considerable body of research, both in first language (L1) and second language (L2) writing, has explored how technology influences the writing process, especially in the digital era, such as think-aloud protocols, text analysis, naturalistic and video-based observation, stimulated recall, process logs, and computer-based tracking (Zamel, 1982, 1983; Pennington, 1993; Cumming, 1990; Manchon et al., 2000; Larios et al., 2008; Turner & Katic, 2009; Cohen & Morrison, 2017). Collectively, these approaches reflect the increasing research attention to how technology reshapes the composing process of writers. Table 1 present the list of abbreviations in this study.

1.2 Theories in second language acquisition and engagement

While models of the writing process have illustrated how technology is integrated in text production, it is equally important to consider how writing itself contributes to second language acquisition (SLA). Writing is not only a means of communication but also a learning activity through which learners can notice their language gaps, experiment with forms, and refine their linguistic knowledge. In this regard, digital writing tools create an environment that fosters cognitive engagement as students actively plan, compose, and revise their work. To capture this dual role of writing, as both process and acquisition, this study draws on two key SLA theories: Swain's Output Hypothesis (1995) and Schmidt's (1990, 2001) Noticing Hypothesis. Swain's (1995) Output Hypothesis posits that producing language compels learners to confront what they cannot yet express. Schmidt's (1990, 2001) Noticing Hypothesis emphasises that attending to such gaps is a prerequisite for language learning. Together, these two frameworks explain how engagement with digital tools can prompt reflection on language use, stimulate self-correction, and support language development.

In this thesis, such processes are conceptualised as 'cognitive engagement', reflecting learners' active involvement in noticing gaps, testing alternatives, and thinking critically about their language use during tool-mediated writing (Connell & Wellborn, 1991; Newmann et al., 1992). However, it should also be noted that the concept of 'engagement' usually extends beyond cognition alone and also includes behavioural and emotional dimensions. Behavioural engagement is reflected in students' observable actions, such as how often, when, and for what purposes they employ digital writing tools throughout the planning, composing, and revising stages (Birch & Ladd, 1997; Finn et al., 1995; Skinner & Belmont, 1993, cited in Fredricks et al., 2004, p.62). Emotional engagement captures learners' affective responses, such as positive feelings of confidence, motivation, and enjoyment, as well as negative emotions like confusion and frustration that arise during interactions with these tools (Finn, 1989; Skinner & Belmont, 1993, cited in Fredricks et al., 2004, p.63). Taken together, these three concepts of engagement provide a more comprehensive view of ESL learners' engagement with digital writing tools, which is also the central focus of this thesis.

1.3 Digital writing tools

In English language teaching, “technology” has typically encompassed areas such as computer-assisted language learning (CALL), mobile-assisted language learning (MALL), online resources, and collaborative platforms like Google Docs. More recently, however, attention has turned to the rise of Artificial Intelligence (AI)-driven writing tools, which form the focus of this study. Artificial Intelligence (AI), a concept introduced by John McCarthy in 1956, was originally envisioned as the creation of machines capable of reasoning, problem-solving, and language use (Russell & Norvig, 2010). In its current educational application, AI writing tools typically function as online platforms that provide assistance with collocations, concordances, translation, paraphrasing, and error corrections, helping users to produce more accurate and polished texts.

This study focuses on four types of AI-based tools grounded in natural language processing algorithms and specifically those applied to writing: corpus search tools (CSTs), machine translation tools (MTTs), automated paraphrasing tools (APTs), and automated writing evaluation tools (AWETs). Table 2 below summarises the functions and examples of each type.

Table 2. Examples of the four tool types and their functions

Type of tool	Functions	Examples of this tool
Corpus tools	Finding collocations of a word, displaying frequencies, and finding concordances of words or phrases within a selected corpus	Sketch Engine, AntConc, Ludwig, WordSmith Tools, UAM Corpus Tool
Machine translation tools	Translating texts from various languages to English and vice versa	Google Translate, Microsoft Translation, Amazon Translate, DeepL Translator, Reverso

Automated paraphrasing tools	Paraphrasing texts by changing sentence order and structure, and replacing words with synonyms	Quillbot, Spinbot, Ginger, Paraphraser.io
Automated writing evaluation tools	Identifying and correcting errors in punctuation, spelling, and grammar	Grammarly, Wordtune, WritetoLearn, Ginger, ProWritingAid, PaperRater, WhiteSmoke

1.3.1 Corpus tools

Corpus tools such as Sketch Engine, COCA, and AntConc are designed for tasks like finding collocations, displaying word or phrase frequencies, and generating concordance lines within a selected corpus—a collection of authentic texts (Gilquin & Laporte, 2021) (Gilquin & Laporte, 2021). In the language teaching context, they support prewriting, vocabulary development, and error correction (Flowerdew, 2015; Yoon, 2016). Their key strength is providing examples of authentic language but learners may also struggle with data overload (Boulton, 2009a).

In this study, it should be noted that Sketch Engine is a concordance package rather than a corpus. The corpus selected for participants to access in this study was English Web 2021 (enTenTen21). This corpus was chosen over specialised academic databases such as the British Academic Written English (BAWE) corpus primarily because of its massive scale of over 52 billion words and its reflection of contemporary, real-world language. This provides learners more concordance and collocation examples to work with. Moreover, English Web 2021 corpus provides a more diverse environment for English writing, which is better suited to investigating students' cognitive engagement with the tool's output in the context of this study. Therefore, in this thesis, the name "Sketch Engine" refers to the combination of the concordancer and the English Web 2021 (enTenTen21) corpus.

1.3.2 Machine translation tools

Machine translation tools such as Google Translate function by instantly converting text from one language to another, helping learners access vocabulary, meaning, and comprehension (Garcia & Pena, 2011; Lee, 2020). However, they often produce errors with idioms and context (Somers et al., 2006).

1.3.3 Automated paraphrasing tools

Automated paraphrasing tools such as Quillbot and Wordtune rephrase sentences or passages to produce alternative wordings in the same language (Rogerson & McCarthy, 2017). They can support vocabulary expansion and text variety (Chen, 2019), but outputs are often superficial, sometimes generating incoherent “word salad” or concealing plagiarism (Keck, 2006; Wahle et al., 2021).

1.3.4 Automated writing evaluation tools

Automated writing evaluation tools such as Grammarly analyse texts using NLP algorithms to detect errors, suggest corrections, and provide instant feedback (Allen et al., 2016). They can improve accuracy, enhance revision, and foster learner autonomy (Li et al., 2015; Palermo & Wilson, 2020; Wei et al., 2023). However, their feedback often remains surface-level, encouraging formulaic writing and overlooking cohesion or rhetorical quality (Dikli & Bleyle, 2014; Ranalli et al., 2017).

1.3.5 Connection to the present study

Aside from these aforementioned tools, there is an emerging new AI technology, which is referred as “Generative AI” that embedded all of these four functions. However, this technology was still in its early stage as it was officially introduced in November 2022, and the time of conduction of this study was January 2023. Therefore, the researcher decided to focus on the four specialised digital writing tools rather than a single, integrated GenAI tool (e.g. ChatGPT, Claude, Gemini, Copilot) because this approach allows for a clearer understanding of how each tool’s function uniquely contributes to students’ English language development. Building on that rationale, this study examines how four types of digital tools support ESL/EFL learners in composing an argumentative essay. This study employs a mixed-method approach. The quantitative

approach compares the writing scores of students in two essays—one written without tools (Essay 1) and another with access to the four tools representing four functions: corpus tool (Sketch Engine), machine translation tool (Google Translate), automated paraphrasing tool (Quillbot), and automated writing evaluation tool (Grammarly). Students' writing performance is assessed using the International English Language Testing System (IELTS) writing test format and rubric. The qualitative approach analyses students' tool usage strategy and interview responses to explain how behavioural, cognitive, and emotional engagement influence outcomes. The qualitative approach also analyses teachers' lesson plans and interview responses to examine how teacher guidance aligns or diverges from students' tool usage practice.

1.4 Research objectives

The central aim of this study is to address key gaps in research on digital writing tools in second language (L2) writing, with particular attention to standardised test preparation contexts such as IELTS. Although existing studies have demonstrated the pedagogical value of corpus tools, machine translation tools (MTTs), automated paraphrasing tools (APTs), and automated writing evaluation tools (AWETs), most have examined these technologies in isolation, overlooked their application to the IELTS essay writing format, or focused primarily on either teacher or student perspectives. This study, therefore, seeks to provide a more comprehensive approach by examining the actual tool usage strategies employed by students, comparing students' engagement with the tools and their writing outcomes. It further aims to evaluate the extent to which alignment or divergence between teacher-recommended strategies and student practices is reflected in students' writing outcomes. In addition, the study investigates how both teachers and students perceive the effectiveness and limitations of the four tool types, with particular attention to the role of emotional engagement—such as interest, confidence, and frustration—in shaping writing outcomes. Finally, by documenting teachers' and students' expectations for future improvements to writing technologies, the study situates its findings within the broader context of emerging tools such as Generative AI, while also offering a context-sensitive, methodologically robust analysis of Vietnamese IELTS learners that triangulates writing scores, observed usage patterns, and interview reflections.

1.5 Research questions

Building on those research objectives, this introductory chapter presents three main research questions and their corresponding sub-questions, as presented below:

RESEARCH QUESTIONS

RQ1: What is the impact of the four writing tools on students' IELTS essay writing scores?

RQ1a: Does the use of the writing tools result in an increase in the overall writing scores of students? If so, what aspects of IELTS writing assessment (i.e., Task Achievement, Coherence and Cohesion, Lexical Resource, Grammatical Range and Accuracy) are most affected by students' use of the tools?

RQ1b: How do high-scoring and low-scoring students differ in their ability to engage with the tools and use them to support their language development?

RQ1c: What factors contribute to any observed decrease in students' writing scores when using these tools?

RQ2: What similarities and differences exist between teachers' recommended strategies for using the four writing tools in IELTS essay preparation and students' actual usage patterns, and how are these similarities and differences reflected in students' essay writing performance?

RQ2a: What specific strategies do teachers recommend for using each of the four writing tools in IELTS essay preparation?

RQ2b: What strategies do students employ when using the four writing tools?

RQ2c: Did students whose tool usage shared similarities with teachers' strategies achieve higher Essay 2 scores compared to those whose tool usage differed?

RQ3: How do teachers and students perceive the effectiveness and limitations of the four writing tools

RQ3a: What are the strengths and limitations of each tool, as evaluated by teachers and by students?

RQ3b: How do students' evaluations relate to their writing performance?

RQ3c: What additional features do teachers and students expect beyond the tools' current capabilities?

1.6 Methodology overview

This study employs a quasi-experimental mixed-methods design to investigate the role of digital writing tools in second language (L2) writing, with a specific focus on IELTS essay preparation. The methodological framework combines quantitative and qualitative approaches to capture both measurable outcomes in students' writing performance and the nuanced strategies, perceptions, and engagement patterns of teachers and students. The quantitative analysis examines changes in IELTS writing scores across two essays written under controlled conditions: Essay 1, completed without access to any tools, and Essay 2, completed with access to four selected tools—Sketch Engine (CT), Google Translate (MTT), Quillbot (APT), and Grammarly (AWET). Focusing on these specialised tools rather than a single all-in-one Generative AI system allows for a more precise understanding of how specific tool functions influence various aspects of students' writing performance. Furthermore, data collected from distinct usages of these four tools are essential for understanding the role of behavioural, cognitive, and emotional engagement in shaping students' writing score outcomes and learning indicators, as well as the degree of alignment or divergence between teacher recommendations and student practices.

To address these research objectives, data were collected from 20 Vietnamese university students and 10 ESL teachers. The student participants completed two IELTS writing tasks, with their tool use closely observed and screen-recorded. Semi-structured interviews then used to explore tool-use strategies (cognitive and behavioural engagement), perceptions (emotional engagement), and perceived or evidenced language learning gains. In parallel, interviews with the 10 teachers

examined the instructional rationales and pedagogical strategies underpinning their recommendations for integrating these tools into IELTS writing instruction.

1.7 Scope and delimitation

This study focuses on the use of four digital writing tools—corpus tools (Sketch Engine), machine translation tools (Google Translate), automated paraphrasing tools (Quillbot), and automated writing evaluation tools (Grammarly)—in the context of IELTS essay preparation. Specifically, it examines how these tools are recommended and integrated into the writing process by teachers, how students employ them at different stages of the writing process—planning, composing, and revising—and how such patterns of use relate to students' essay writing performance.

1.7.1 The IELTS test

To examine the impact of the four aforementioned tools on writing performance in an argumentative essay, this study follows the writing test format of the International English Language Testing System (IELTS). The IELTS test was first introduced in 1980 and became compulsory for admission to courses in English-speaking countries in 1989. Its primary purpose is to evaluate a candidate's readiness for study or residence in an English-medium environment (Coleman et al., 2003). The test comprises four components: two core skills—listening and speaking—and two academic skills—reading and writing (Charge & Taylor, 1997). IELTS is jointly owned by the British Council, IDP, and Cambridge Assessment English (IELTS, n.d.) and is administered in over 120 countries, making it one of the most widely recognised international assessments (Uysal, 2009). Given its high-stakes nature and the growing demand for entry to English-speaking universities, the test has also fuelled a global market for IELTS preparation courses (Dang & Dang, 2021). In Vietnam, where IELTS was first introduced in 1994 (British Council, n.d.), it has since become a benchmark for academic and professional advancement. Therefore, this study adopts IELTS as the writing assessment framework because it is a widely recognised and standardised measure of academic writing proficiency. Within IELTS, Writing Task 2 was selected since it requires candidates to produce a 250-word argumentative essay under strict time constraints, reflecting the type of timed, independent writing expected in academic and professional contexts (Cooper, 2014; Moore & Morton, 2005). While the

researcher acknowledges that instances of tool-generated text may still occur, plagiarism is less feasible in this study due to the controlled, timed format of the writing tests. This format also enables a focused examination of how the four digital writing tools shape key dimensions of IELTS performance, including idea development and Task Response (TR), Coherence and Cohesion (CC), Lexical Resource (LR), and Grammatical Range and Accuracy (GA). Moreover, the international acceptance of the IELTS writing rubric enhances the relevance and potential transferability of the study's findings across diverse English language teaching contexts (see Chapter 3, Section 3.1). The writing component, in particular, assesses the quality of written language based on four criteria: Task Response, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy.

Task Response (TR): This criterion assesses how well candidates address the question, present clear viewpoints, and support them with relevant arguments and examples, while meeting the required essay length (Bagheri & Riasati, 2016).

Coherence and Cohesion (CC): This criterion assesses how effectively ideas are organised and linked (UCLES, 2004). Coherence refers to the logical flow of ideas, while cohesion concerns the use of linking devices (e.g., however, moreover). High scores require a clear structure and smooth progression of arguments.

Lexical Resource (LR) This criterion assesses the range, accuracy, and appropriacy of vocabulary. Candidates are expected to paraphrase effectively and use precise words to express ideas clearly.

Grammatical Range and Accuracy (GA): This criterion assesses the variety and accuracy of sentence structures. Candidates must demonstrate control of grammar and punctuation while avoiding errors.

1.7.2 Writing score outcomes

In this study, writing score outcomes refer to the assessed quality of students' two IELTS Task 2 essays. Each essay was evaluated independently by three raters using the official IELTS writing rubric. The rubric comprises four key criteria: Task Response (TR), Coherence and Cohesion (CC), Lexical Resource (LR), and Grammatical Range and Accuracy (GA) (see Figure 9 in Chapter 3, Section 3.1.1). Using this rubric allows

the study to capture both the overall performance of students and more specific aspects of their language use, such as vocabulary choice, grammatical control, and organisational clarity. This operationalisation of writing performance makes it possible to examine how digital writing tools influence different dimensions of essay quality. A more detailed account of the rating process and procedures for ensuring reliability is provided in Chapter 3: Methodology, Section 3.3.

1.7.3 Language learning indicators

In this study, language learning indicators are defined as evidence of students' linguistic development resulting from tool use, as detailed in Chapter 4, Section 4.3.2. Two indicators were identified to capture this process under two themes. The first, Vocabulary Acquisition (VA), refers to cases where students reported learning new words through interaction with the four tools. For example, students used Google Translate and Quillbot to discover, practise, and retain unfamiliar vocabulary. The second, Grammar Improvement (GI), describes instances where students verbalised their notice of errors and corrected errors with the assistance of the tools. This process helped them refine accuracy, adjust word choice, and enhance clarity. Taken together, these indicators show how digital tools can foster both vocabulary growth and grammatical development during writing.

1.7.4 Cognitive engagement

In this thesis, the notion of 'cognitive engagement' includes behavioural and cognitive dimensions, drawing primarily from Fredricks et al. (2004), Hiver et al. (2021a, 2021b) and Svalberg (2009). This offers a useful lens for analysing how learners work with digital tools, capturing not only what they do but also how they think and feel. Behavioural engagement refers to observable actions, such as how frequently and purposefully students use tools at different writing stages, including indicators like time on task or number of revisions. Cognitive engagement concerns the mental effort invested in evaluating tool outputs—comparing them with self-written texts, cross-checking suggestions, and reflecting on vocabulary and grammar gains. In this study, these two dimensions closely align with the two conceptual frameworks in second language acquisition being discussed in this study: The Output Hypothesis (Swain, 1995) and The Noticing Hypothesis (Schmidt, 1990, 2001).

The study is delimited in several ways. It focuses only on IELTS Academic Writing Task 2 and on four specific tools, excluding newer Generative AI applications such as ChatGPT. Participants are restricted to Vietnamese learners and teachers, limiting generalisability to other contexts. The data are primarily qualitative, supported by writing scores rather than large-scale quantitative testing. Finally, the study examines short-term effects in a test-preparation setting, so claims about long-term learning outcomes should be made cautiously.

1.8 Contribution

This study contributes to the growing body of research on digital writing tools in second language (L2) contexts by addressing several important gaps in the literature. First, it extends existing work by examining the use of corpus tools, machine translation tools (MTTs), automated paraphrasing tools (APTs), and automated writing evaluation tools (AWETs) in combination, rather than in isolation. By situating the investigation within the context of IELTS essay preparation, the study provides a more focused and practically relevant understanding of how these tools are integrated into test-oriented, time-restricted writing practices.

Second, the study makes a theoretical contribution by comparing teachers' recommended strategies with students' actual usage patterns, an area that has received limited empirical attention. More importantly, the study examines how these patterns are reflected in students' writing performance, thereby linking the role of engagement with pedagogical outcomes.

Third, the study advances knowledge on students' engagement with tool usage by incorporating behavioural, cognitive, and emotional dimensions. While prior research has often emphasised tool functionality or writing outcomes, this study also foregrounds the role of emotional engagement—such as confidence, interest, or frustration—in shaping perceptions of tool effectiveness. This perspective provides richer insights into how students and teachers evaluate digital writing tools and how these evaluations may influence writing outcomes.

Fourth, the study offers methodological significance through its triangulated design, which combines students' writing scores, observed tool usage, and interview responses. This multi-dimensional approach ensures a more robust and context-

sensitive analysis of how Vietnamese IELTS learners engage with digital tools, addressing the need for research that captures authentic writing behaviours in non-Western educational contexts.

Finally, the findings have practical implications for pedagogy, assessment, and technology design. For teachers and curriculum designers, the study highlights strategies for integrating digital tools into IELTS preparation in ways that support both language development and writing test performance. For tool developers, the documentation of teacher and student expectations provides valuable insights into how the four tools under study can be refined to support L2 learning more effectively, particularly in light of the growing prominence of Generative AI. Collectively, these contributions ensure that the study not only advances academic debates but also informs practice in language teaching and assessment.

1.9 Thesis structure

This thesis is organised into seven chapters, each building upon the previous to provide a coherent account of the study's rationale, methodology, findings, and contributions. Chapter 1 has introduced the study by outlining the background, identifying the research problem, and establishing its significance, objectives, scope, and guiding research questions. Chapter 2 reviews the literature on engagement frameworks and empirical studies of the four types of writing tools—corpus tools (CTs), machine translation tools (MTTs), automated paraphrasing tools (APTs), and automated writing evaluation tools (AWETs)—and synthesises these theoretical findings in second language acquisition, highlighting key areas, research gaps, and their connection to the present study. Chapter 3 details the methodology, including research design, context, participants, instruments, and procedures, as well as explaining both quantitative and qualitative data analysis. Chapter 4 presents the findings on the quantitative analysis of 20 students' essay score changes in four criteria (TR, CC, LR and GA), and individual score changes between use and non-use of the four writing tools. It also reports the thematic findings on students' tool usage strategies, exploring their cognitive engagement with the tools, and exploring language learning indicators. Moreover, Chapter 4 includes case study analyses to explore the factors underlying some students' decreased scores in Essay 2 (which was composed with tools). Chapter 5 presents the qualitative findings from teacher interviews and

student data, analysing similarities and differences between teachers' recommended strategies and students' actual practices across the planning, composing, and revising stages, and examines their impact on writing score outcomes and language learning indicators. Chapter 6 synthesises the findings from Chapters 4 and 5 and compares them with prior research on digital writing tools. The discussion in Chapter 6 also extends to the emerging role of Generative AI in L2 writing, considering its potential impact on future research and practice. Finally, Chapter 7 concludes the thesis by summarising the key findings, addressing the study's limitations and considering their implications and contributions to the field.

CHAPTER 2: LITERATURE REVIEW

Overview of Chapter 2 sections

Section 2.1 of Chapter 2 introduces the study's conceptual framework, drawing on the Output Hypothesis, the Noticing Hypothesis, and concepts about engagement to explain how using digital tools for writing may facilitate L2 development through behavioural, cognitive, and emotional engagement. In addition to the theoretical literature, Section 2.2 reviews empirical research on corpus tools in L2 writing, examining their contributions to vocabulary, grammar, and collocation development while also considering learner proficiency, usage patterns, and pedagogical challenges. Section 2.3 then turns to machine translation tools, particularly focusing on Google Translate, highlighting their influence on writing quality, learner autonomy, and proficiency-based usage, as well as the ways teachers and students perceive their value amid concerns about accuracy and academic integrity. Section 2.4 explores automated paraphrasing tools, such as QuillBot, analysing their impact on paraphrasing skills and writing quality, and addressing issues of learner proficiency, instructional integration, and some writing concerns like patchwriting. Section 2.5 synthesises findings on automated writing evaluation tools, with a focus on Grammarly, to assess their role in supporting revision, feedback, and learner autonomy, while also identifying engagement patterns and teacher perspectives on their effectiveness and limitations. Building on this discussion of established digital tools, Section 2.6 turns to emerging research on Generative AI in L2 writing, highlighting its rapid rise, the contrasting perceptions of teachers and students, and the limited attention on proficiency and perceptions. Section 2.7 then identifies the research gaps that this study seeks to address, leading to the development of the research questions. Finally, Section 2.8 concludes the chapter with a summary.

2.1. Theoretical frameworks

This section reviews the theoretical foundations that guide the present study, drawing primarily on the Output Hypothesis (Swain, 1985, 1995) and the Noticing Hypothesis (Schmidt, 1990, 2001), alongside the multidimensional concept of learner engagement (Fredricks et al., 2004; Hiver et al., 2021a; Svalberg, 2009). These frameworks are selected because they collectively provide both the theoretical rationale and the

analytical lens for examining how L2 learners' use of the four writing tools can push them to produce output (in the form of writing), prompt the noticing of language features and gaps, thereby facilitating language learning.

2.1.1 The Output Hypothesis

Swain's (1995) Output Hypothesis positions language output (i.e. speaking and writing) as a central mechanism in second language acquisition. Swain identified three core functions that language output facilitates learning and promotes linguistic development. The first is the 'noticing function', through which learners become aware of gaps in their linguistic knowledge when attempting to express meaning. This awareness can prompt deeper cognitive processing and attention to language form (grammar). The second function is 'hypothesis testing', through which learners actively experiment with language structures and receive feedback that helps refine their understanding. Finally, the 'metalinguistic function' of learners' output encourages learners to reflect on the relationship between form and meaning, thereby supporting the internalisation of language rules. In short, language output does not operate in isolation but actively triggers noticing, and making it a key mechanism for learning.

Bringing these three core functions of the Output Hypothesis to the context of this study, the researcher argues that the use of the four tool functions in this study for the essay writing task — corpus search tool (Sketch Engine), machine translation tool (Google Translate), paraphrasing tool (Quillbot), and correction tool (Grammarly)— can collectively act as a 'catalyst' for language learning. First and foremost, the act of using these tool functions pushes learners to express meaning and produce output in the form of writing. Unlike Generative AI tools that can generate texts independently with just a few words, these tool functions rely entirely on learners' output to make its input. More importantly, interaction with these tools' output triggers the 'noticing function' that Swain proposed. In other words, when learners compare their own writing with tool-generated feedback, their gaps in language uses become more noticable. This simultaneously reflects hypothesis testing and noticing: learners test assumptions about language use and, through feedback, notice mismatches between their output and the tools' output. For instance, the collocation data shown by Sketch Engine may challenge and change a learner's collocation choice. Google Translate's output may reveal differences in meanings. QuillBot and Grammarly's output prompt

learners to evaluate alternative phrasings and corrections. The use of these tools also prompt the ‘metalinguistic function’ in the Output Hypothesis, as learners are pushed to continuously reflect on why changes are suggested and decide whether to accept them, which can eventually lead to language learning.

2.1.2 The Noticing Hypothesis

Closely aligned with the ‘noticing function’ that Swain proposed, the Noticing Hypothesis (Schmidt, 1990, 2001) also posits that conscious attention to linguistic features is a necessary precondition for language acquisition. Schmidt (1990) defines “noticing” as the conscious attention to specific language form in the “input”— the language that learners are exposed to through reading, listening, or interaction. However, not all input leads to learning. Only the aspects of input that learners consciously attend to are transformed into intake, that is, the portion of input that is processed, internalised, and made available for subsequent language use. In this sense, noticing acts as a fundamental cognitive process that helps learners turn input into intake, allowing them to understand and use new language features (i.e. vocabulary or grammar).

In the context of this study, the use of the four writing tools can actively prompt such “noticing” by exposing learners to specific language through comparison and revision. For instance, Sketch Engine exposes learners to authentic concordance lines, prompting them to notice patterns in collocations and contexts within a large reference corpus like English Web 2021. Google Translate encourages noticing by allowing learners to compare their intended meaning with the translated output. QuillBot encourages noticing by presenting alternative phrasings, drawing learners’ attention to the input of variations in expression and lexical choice. Similarly, Grammarly encourages noticing by highlighting errors and suggests revisions, directing learners’ attention to specific issues in spelling and grammar. By using these tool functions, learners are not only exposed to input but are also pushed to consciously attend to mismatches between their own output and the target language, thereby facilitating the noticing necessary for language development.

In short, these two theories are the main framework for this study because they explain how language learning happens when students use these four writing tool functions.

As learners write, compare their own language with tool feedback, and decide whether to revise it, they notice gaps, test alternatives, and reflect on language use. The Output Hypothesis explains the role of producing and revising in writing in L2. The Noticing Hypothesis explains how attention to the tools' output as a form of feedback can support language development.

To establish a more comprehensive theoretical basis for understanding learners' interaction with digital writing tools, this study integrates the concept of learner engagement—a construct grounded in diverse educational and psychological frameworks, yet still without a universally standardised definition.

2.1.3 The concept of engagement and cognitive engagement in language learning

In the context of language learning, engagement is commonly conceptualised as multidimensional, comprising behavioural, cognitive, and emotional dimensions (Baralt et al., 2016; Fredricks et al., 2004; Hiver et al., 2021a; Lambert et al., 2017; Svalberg, 2009). However, there is no one definition for each dimension, and it varies across studies. Behavioural engagement was often defined as students' positive conduct and active involvement in academic tasks (Birch & Ladd, 1997; Finn et al., 1995; Skinner & Belmont, 1993, cited in Fredricks et al., 2004, p.62). Cognitive engagement often refers to students' psychological investment in learning, including their preference for challenging tasks, use of self-regulated learning strategies, and persistence (Connell & Wellborn, 1991; Newmann et al., 1992), as well as the quality of their participation and their degree of active involvement in the learning process (Sang & Hiver, 2021). Emotional engagement involves emotional reactions of students such as interest, boredom, happiness, and anxiety (Finn, 1989; Skinner & Belmont, 1993, cited in Fredricks et al., 2004, p.63).

The measurement of each dimension of engagement also varies across studies. Behavioural engagement, in particular, has been measured using various observable metrics, such as word counts, turn counts, semantic content produced, time on task, and task-related actions like initiating speech or completing assigned activities (Baralt et al., 2016; Bygate & Samuda, 2009; Dörnyei & Kormos, 2000; Hiver et al., 2021a; Lambert & Zhang, 2019; Lambert et al., 2017; Philp & Duchesne, 2016; Platt & Brooks,

2002). Cognitive engagement has been measured through learners' use of elaborative clauses and meaning negotiation moves such as feedback and clarification requests, as well as through verbal indicators, language-related episodes, and even nonverbal cues like gestures and eye movements (Baralt et al., 2016; Fredricks & McColskey, 2012; Lambert et al., 2017; Svalberg, 2017). Emotional engagement has typically been measured through learners' self-reported perceptions and emotional responses during target language activities, often using questionnaires, interviews, or reflective accounts to capture their affective experiences (Dao et al., 2019; Henry & Thorsen, 2020; Reeve, 2012; Skinner et al., 2009). Emotionally engaged learners are commonly described as demonstrating positive perceptions and a purposeful, willing, and autonomous disposition toward language learning tasks, while the presence of negative emotions—such as anxiety, boredom, frustration, or anger—has been interpreted as evidence of emotional disengagement or disaffection (Dao et al., 2019; Mercer, 2019; Svalberg, 2009).

2.1.4 Linking the theoretical frameworks to the current study

Taken together, these theoretical perspectives inform the present study's conceptualisation of learner engagement in tool-mediated writing. The Output Hypothesis (Swain, 1995) and the Noticing Hypothesis (Schmidt, 1990, 2001) both offer foundational insights into the cognitive mechanisms called 'noticing', through which digital writing tools may facilitate second language development. These two frameworks highlight how active language production and attention can prompt cognitive engagement and facilitate language learning. In other words, when learners use digital tools to express their ideas in writing, they are likely to encounter gaps in their linguistic knowledge, leading them to notice deficiencies, test alternative forms, and reflect on language use. Moreover, the multidimensional concept of engagement informs this study in that using digital writing tools involves not only cognitive engagement but also learners' actions (behavioural engagement) and their feelings or perceptions (emotional engagement). Behavioural and cognitive engagement refers to learners' observable actions, such as how purposefully they use the tools during different stages of writing. Emotional engagement captures learners' affective responses—such as feelings of confidence, anxiety, motivation, or frustration—that arise during tool-supported writing tasks. Taken together, the three dimensions of

engagement provide a basis for analysing learners' interactions with digital writing tools and examining how such interactions may contribute to language development. These processes—noticing language gaps, trying out new forms, and thinking about how language works—are all part of what this study considers as cognitive engagement, which is coded as “Cognitive Engagement” (CE) for the thematic analysis in the methodology (see Chapter 3, Section 3.5.2.2).

As Hiver et al. (2021a) noted, there remains no standardised framework for measuring engagement across its dimensions. In response to this lack of consensus, recent studies have called for more context-sensitive approaches that move beyond self-report data and focus on actual learner behaviour during tasks (Hiver et al., 2021a; 2021b; Hiver & Dao, 2025; Svalberg, 2017). Furthermore, while engagement is widely acknowledged as a key factor linked to positive academic outcomes (Chan et al., 2024; Lei et al., 2018; Luo et al., 2023; Christenson et al., 2012; Schrader & Kalyuga, 2020; Skinner et al., 2009; Wu et al., 2020), no study to date has directly investigated learner engagement during writing tasks involving the combined use of digital writing tools. This gap highlights a need for further research that examines learner engagement in greater depth and within specific instructional contexts. By drawing on these dimensions and this gap in research, the study investigates not only how language learners use digital tools for writing, but also how their engagement across behavioural, cognitive, and emotional dimensions may influence their language development and writing performance.

In conclusion, while the Output Hypothesis, Noticing Hypothesis, and engagement framework offer theoretical foundations for understanding language learning, a closer review of empirical studies is also important to explore how these concepts are reflected in practical L2 writing contexts. The next sections review research on the teaching value of four types of digital writing tools—corpus tools, machine translation tools, automated paraphrasing tools, and automated writing evaluation tools—and how learners engage with them behaviourally, cognitively, and emotionally during writing tasks. This review aims to synthesise key findings, identify similar patterns and gaps in the literature, and details how the study can fill in these gaps.

2.2. Empirical studies on corpus tools in L2 writing

2.2.1 Corpus tools, corpora and data-driven learning approach

A corpus, or corpora, is a collection of pieces of language text in electronic form to represent a language (Sinclair, 2005). Corpora emerged in the 1960s but were used mainly as a source of data for applied linguistic research (Charles, 2018). Over time, corpora have been found to be useful in a variety of ways for English language teaching (ELT). The Data-Driven Learning (DDL) approach, introduced by Johns (1991), encourages learners to take an active role as language investigators by analysing authentic language data in corpora, while teachers act as guides rather than traditional knowledge providers. This approach draws on several influential theories, including the inductive approach (Piaget, 1952, 1970), which emphasises learning through pattern recognition and rule formation based on experience; discovery learning (Bruner, 1961, 1966), which promotes problem-solving and exploration; and the Noticing Hypothesis (Schmidt, 1990), which posits that conscious attention to language features is essential for acquisition. Within the DDL framework, these principles are reflected in learners' use of corpus tools to identify and focus on linguistic patterns, thereby enhancing their awareness of previously unnoticed structures.

Since the mid-1990s, language corpora and the DDL method have shown considerable promise in computer-assisted language learning, language research, and pedagogy (Cobb, 1997, 1999, 2007; Horst et al., 2005; Stevens, 1991; Thurstun & Candlin, 1998). Thompson (2006) emphasised that corpora can serve a wide range of users—including lexicographers, grammarians, and language testers—who shape pedagogy through their descriptions of language and decisions about what aspects of knowledge to assess. Römer (2006) further distinguished between 'direct use' of corpora by students as linguistic reference tools and 'indirect use' by educators to develop instructional materials. Supporting these pedagogical applications, corpus tools—specialised software programs designed to analyse language data—enable access to vast collections of texts, known as corpora. Prominent examples of corpus tools and corpora include the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). Charles (2018) provided an overview of widely used and freely available English corpora and corpus tools, as illustrated in Figure 1.

<i>Web site</i>	<i>URL</i>	<i>Corpora available</i>
AntConc	http://www.laurenceanthony.net/software.html	software only
BYU	http://corpus.byu.edu	COCA, BNC, etc.
CLT	http://www.lextutor.ca	BNC, learners, etc.
FLAX	http://flax.nzdl.org/greenstone3/flax	BAWE, BNC, law, etc.
HKPU	http://langbank.engl.polyu.edu.hk/index.html	business, education, etc.
HKPU	http://www.engl.polyu.edu.hk/rcpce	engineering, finance, etc.
ICLE	http://www.uclouvain.be/en-cecl-icle.html	learner corpora
Just-The-Word	http://www.just-the-word.com	BNC
MICUSP	http://micusp.elicorpora.info	MICUSP
Sketch Engine	http://www.sketchengine.co.uk	BAWE, BNC etc.
SkELL	https://skell.sketchengine.co.uk/run.cgi/skell	large general corpus
WebCorp	http://www.webcorp.org.uk/live	web as corpus
WordSmith Tools	http://www.lexically.net/wordsmith/index.html	software only

Figure 1. Websites of corpus tools and their available corpora. From Charles (2018, p.2)

Accordingly, several corpus tools have been developed to support language learning and writing instruction by offering access to authentic language data and pattern analysis, including concordances. A concordance is a short snippet of authentic text, allowing users to observe how the target word occurs with surrounding words and grammatical structures (Sinclair, 1991). For example, AntConc is a widely used tool that enables users to explore lexical and grammatical patterns through keyword analysis, collocate searches, and concordance lines. The Compleat Lexical Tutor (www.lextutor.ca) offers functions such as vocabulary profiling, concordancing, and frequency list generation. JustTheWord, based on the British National Corpus (BNC), allows users to examine collocations with intuitive visualisations. Another popular tool is Sketch Engine, which provides concordance lines and automatically generated word sketches to support vocabulary development. Building on the capabilities of widely used corpus tools, this study employs Sketch Engine, a robust platform that allows learners to explore large-scale corpora—including the British National Corpus (BNC)—to access authentic language examples, identify collocations, and analyse word usage in context. These features make it especially well-suited for supporting writing, particularly in enhancing lexical resources.

2.2.2 Sketch Engine

Sketch Engine (<https://www.sketchengine.eu>) is a widely used corpus tool in lexicography, translation, and language teaching (Frankenberg-Garcia et al., 2019). Its central feature is the ‘concordance’ function, which retrieves all instances of a particular word or phrase from a corpus (a large collection of texts) and displays them in context. In Sketch Engine, concordance lines are automatically generated through an algorithm that prioritises representative and frequent examples while filtering out rare items, technical terms, and typographical errors (Baisa & Suchomel, 2014). Through this function, learners can examine authentic language use, explore semantically related expressions, and test whether particular word combinations form acceptable collocations in English. Figure 2 presents examples of concordances for the two words “research” and “suggest” using Sketch Engine.

The screenshot shows the Sketch Engine interface with the search term 'research' in the top bar. The 'Examples' tab is selected, displaying a list of 22 concordance lines. Each line shows the word 'research' in red and 'suggests' in green, with the surrounding text in a light gray background. The concordance is titled 'research + suggest' with a frequency of '2.09 hits per million'.

research + suggest 2.09 hits per million

- 1 Some **research suggests** moods are related to overall job satisfaction.
- 2 Additional **research suggests** the social component of ritual.
- 3 Further **research suggests** cosmic dust might have skewed the results.
- 4 Other **research suggests** that later training is associated with problems.
- 5 Some **research** has **suggested** that exercise may have antidepressant effects.
- 6 This **research suggests** independence of episodic and semantic musical memory.
- 7 Some **research suggests** alternate estimates to the accepted workforce size.
- 8 Some **research suggests** an association between personality and job satisfaction.
- 9 The same **research suggests** they may have fought barefoot.
- 10 Recent **research suggests** that this is untrue.
- 11 Latest **research suggests** its audience may be falling.
- 12 Some **research suggests** that it is becoming more common.
- 13 But the **research suggests** the problem is complicated.
- 14 My **research suggests** that Creek is more probable.
- 15 However, **research suggests** that many banks felt heavily pressured.
- 16 While **research suggests** that sexual dysfunction is common ... 1.
- 17 Other **research suggest** that siblings born further apart achieve higher educational outcomes.
- 18 Recent **research suggests** that Humboldt squid are only aggressive while feeding.
- 19 It might also lead to weight gain, **research suggests** .
- 20 Some **research suggests** that it may run in families.
- 21 Some **research suggests** that bipolar I disorder runs in families.
- 22 Recent **research suggests** that this process is not currently underway.

Figure 2. Examples of concordances for “research” and “suggest” in Sketch Engine. From Frankenberg-Garcia et al. (2019, p. 26)

Another important function, which is called the “Word Sketch” function (Figure 3) of Sketch Engine, can facilitate the task of analysing collocations by sorting their grammatical relations with the keyword.

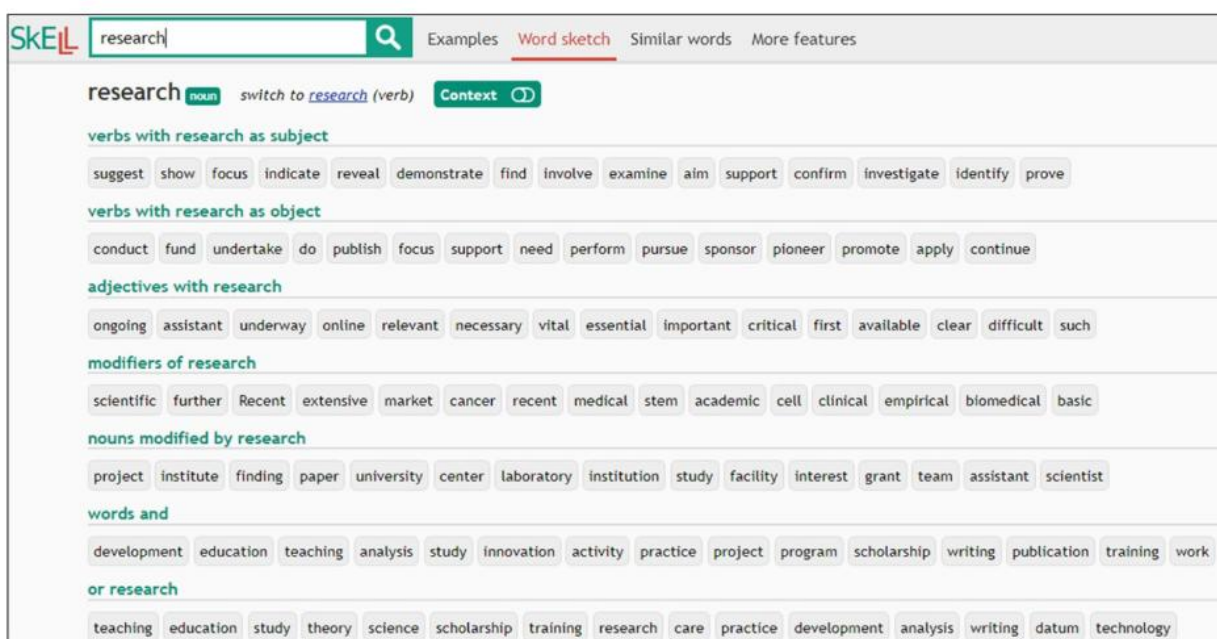


Figure 3. An example of the “Word sketch” function results for the word “research” in Sketch Engine. From Frankenberg-Garcia et al. 2019, p. 26)

In summary, Sketch Engine has been selected for this research due to its extensive data, diverse capabilities, and user-friendly accessibility, which make it a suitable resource for ESL learners.

2.2.3 Merits and shortcomings of using corpus tools for L2 writing and language learning

A substantial body of empirical research affirmed the pedagogical value of corpus tools in enhancing L2 writing, particularly in lexical, collocational, and grammatical development (Boulton, 2009a, 2009b; Chan & Liou, 2005; Gilmore, 2009; Hu & Deng, 2023; Li, 2017; Wu, 2016; Yoon & Jo, 2014). For instance, Chan and Liou (2005) found that Taiwanese college freshmen using the web-based corpus tool TOTALrecall achieved improvements in collocational knowledge, with lower-proficiency students showing more marked gains than their higher-proficiency counterparts. Similarly, Boulton (2009a) demonstrated that French ESL undergraduates who accessed corpus data outperformed peers using traditional dictionaries when completing adverbial linking tasks. Gilmore (2009) reported that intermediate Japanese students who revised essays using corpus tools showed improved lexical accuracy and naturalness in their writing, as validated by native speaker evaluations.

Corpus tools have also been shown to support grammar acquisition. Johns (1994) posits that when learners are introduced to concordances in sufficient quantity and variety, they can learn grammar. Gaskell and Cobb's (2004) longitudinal study of Chinese ESL students found that, although overall error reduction was not statistically significant, learners demonstrated a positive attitude toward corpus use and were able to correct specific errors through guided concordance consultation. Hong's study (2010) revealed that corpus-aided instruction improved learners' recognition of grammatical patterns, particularly determiners. Huang (2014) reported that students using corpora outperformed those using only dictionaries or conventional methods, especially in areas such as lexico-grammatical accuracy and verb usage. Collectively, these studies provide compelling evidence that corpus-based approaches, despite requiring careful scaffolding, can enhance both lexical and grammatical competencies in L2 writing. Lin (2021) further supported this claim in a study involving 179 Taiwanese ESL learners, showing that grammar instruction using concordances from COCA resulted in significant gains across all proficiency levels, with the greatest improvement observed among low-proficiency learners. These results suggest that Data-Driven Learning (DDL), when properly scaffolded, can effectively support grammar development even for less advanced students. Additional studies corroborate these findings.

Nevertheless, several scholars have highlighted the limitations of the DDL approach and challenges associated with corpus tool implementation. One commonly cited issue is the difficulty learners face in deriving generalisable grammatical rules from concordance lines, particularly those with limited linguistic proficiency or unfamiliarity with the topic at hand (Owen, 1996). Frankenberg-Garcia (2005) stressed the importance of corpus selection, cautioning that corpora such as the British National Corpus (BNC) or the British Academic Written English (BAWE), while high in quality, may be too complex for beginners or overly specialised for general academic learners. Students also need to develop skills in filtering irrelevant results and interpreting data effectively. Additionally, as Laufer (2011) noted, many L2 learners tend to overestimate their vocabulary knowledge and, as a result, do not frequently consult external linguistic resources such as corpora. Yoon (2016b) further observed that the sheer volume of concordance lines and frequency data can overwhelm learners, reducing the usability of corpus tools in the absence of adequate instructional support.

Thus, while corpus tools hold significant promise, their effectiveness depends heavily on learners' proficiency, training, and the pedagogical design of their implementation.

2.2.4 Corpus tool usage in relation to language proficiency

A growing body of research highlights that the effectiveness of corpus tools largely depends on L2 writers' proficiency levels and how they engage with these tools. Low-proficiency learners typically demonstrate superficial, search-engine-like behaviours, attending only to surface-level linguistic features without engaging deeply with concordance data (Crosthwaite et al., 2019; Fang et al., 2021; Pérez-Paredes et al., 2012). They often encounter difficulties in locating relevant concordance lines or extracting meaningful patterns, which results in frequent misinterpretations of usage (Fauziah et al., 2024; Lin, 2021; Zhao, 2025). In contrast, higher-proficiency learners display greater corpus literacy, navigating tools more effectively and applying complex grammatical patterns with higher accuracy (Crosthwaite et al., 2019; Fang et al., 2021; Lin, 2021; Pérez-Paredes et al., 2012). A key factor underlying this difference is vocabulary knowledge, as learners lacking lexical awareness tend to use corpus tools less frequently and with minimal strategic variation (Franken, 2012). Other studies further suggest that corpus tools are most beneficial for highly proficient learners, who can better manage the cognitive demands involved in interpreting large amounts of linguistic data (Crosthwaite & Steeples, 2022; Daskalovska, 2015), whereas lower-proficiency learners may experience higher cognitive load that inhibits productive engagement (Saeedakhtar, 2020).

2.2.5 The use of corpus tools by teachers

Researchers have promoted inductive, discovery-based approaches where students use corpora to identify language patterns. Flowerdew (2015a) highlighted how exposure to corpus examples enables learners to infer rules independently. This was demonstrated by Lee and Swales (2006), who guided students through hypothesis formation, corpus searches, and reflection to uncover academic patterns such as “appropriate for modelling” and “devices for storing.”

Building on this, other studies have shown how corpus tools enhance lexical precision in pre-writing tasks. Charles (2018) suggested comparing concordances of near-synonyms to reveal collocational differences, while Römer (2011) used pairs like

“speak” and “talk” to raise awareness of subtle word choices. Corpus-based approaches have also proved effective for grammar instruction. Lin (2019) described how learners analysed concordance lines, answered guiding questions, and created examples to consolidate understanding. Similarly, Ma et al. (2022a) reported on ESL teachers who used corpus-informed lessons to move students from error identification to independent exploration, and Ma et al. (2022b) found that most teachers adopted a four-step pre-writing sequence: activating prior knowledge, analysing concordances, summarising patterns, and applying them in practice.

Taken together, these studies demonstrate the pedagogical value of corpus-based pre-writing, which fosters inductive learning, builds lexical and grammatical awareness, and supports more informed language choices.

Beyond idea generation and keyword selection, corpus tools can also play a role in revision and error correction in L2 writing (Bridle, 2019; Chambers & O’Sullivan, 2004; Gaskell & Cobb, 2004; Mueller & Jacobsen, 2015; O’Sullivan & Chambers, 2006; Quinn, 2015; Todd, 2001). Prior research showed that teachers often integrate these tools into revision-focused instruction to reduce L1 interference and refine target language output. A common approach is for teachers to highlight problematic language and then guide students to consult corpora for self-correction, which Gaskell and Cobb (2004) termed as “feedback-driven learning”, distinguishing it from traditional data-driven learning. For instance, teacher participants in studies conducted by Chambers and O’Sullivan (2004) and O’Sullivan and Chambers (2006) applied corpora to a wide range of errors, including grammatical, lexical, and capitalisation mistakes. Consistently, findings showed that corpus consultation, when paired with feedback, improved the accuracy and appropriateness of student revisions. Quinn (2015) designed a training module using Collins WordBanks Online. Teachers in Quinn’s study (2015) first identified issues in student essays—such as word choice, form, prepositions, and articles—then provided coded feedback prompting corpus-based revisions. Students then need to investigate errors, explored alternatives, and selected contextually appropriate forms.

2.2.6 The use of corpus tools by students

Lai and Chen (2015) investigated 14 ESL students' use of four online corpus tools and dictionaries. After training with the tools and two dictionaries, participants completed three timed-writing tasks and follow-up interviews. They primarily used corpora for revision, particularly to check usage, collocations, grammar patterns, sentence structure, and spelling, especially when dictionaries offered insufficient clarity. Several other empirical studies have shown that ESL students use corpora deliberately as reference tools during writing and revision. Yoon (2008), for example, observed six English learners in Korea who frequently consulted corpus tools to resolve lexical and grammatical issues, particularly prepositions, contextual usage, and collocations. Students paid close attention to word usage during both drafting and editing, with one participant reporting increased confidence and perceived improvements in writing quality due to corpus use. In a follow-up study, Yoon (2016b) examined how two advanced ESL students used a combination of online tools—COCA, JustTheWord, Google, and bilingual dictionaries—to address linguistic challenges in academic writing. One learner developed a routine of using the translation tool for initial understanding, followed by using corpus tools for more nuanced writing expressions. The other student used the two corpus tools (COCA and JustTheWord) to refine lexical choices, focusing on collocates (43% of the total number of revision), simple alternatives (11.4%), and more sophisticated expressions (8.9%) to achieve more precise meaning. These findings from Yoon (2016b) highlighted the similar approach of using corpus tools during the revision stage by high-proficient learners. Conroy (2010) trained four groups of ESL students at an Australian university to use two online corpus tools (VLC and Compleat Lexical Tutor), Google, and dictionaries, then interviewed them about their practices, and 43% of students reported using the concordancers and Google during their writing of assignments, particularly to check lexico-grammatical issues such as prepositions, passive/active voice, and word choice, as well as to learn new vocabulary. The application of using corpora and a corpus tool for revision purposes was also noted in a recent study by Bridle (2019), who found that the 12 Chinese ESL students mainly utilised a corpus for correcting errors associated with selecting synonyms, and the revisions performed with the support of the corpus were largely effective.

2.2.7 Perceptions of teachers and students towards DDL and corpus tools

2.2.7.1 Perceptions of teachers

A wide range of studies have explored the perspectives of ESL teachers towards employing DDL and corpus technology in their teaching in different English language teaching contexts and have gathered mixed results. On the positive side, teachers expressed favourable attitudes and an intention to adopt corpus-based pedagogy after receiving corpus training in a wide range of studies (Breyer, 2009; Bunting, 2013; Chen et al., 2019; Lin & Lee, 2015; Ebrahimi & Faghih, 2016; Abdel Latif, 2021). For example, Lin and Lee (2015) interviewed six ESL teachers using the DDL approach in the classroom and found that they viewed DDL positively for teaching grammar, particularly appreciating its potential for active learning and classroom discussion.

Despite generally positive views of corpus technology, research highlights several teacher concerns. A key issue relates to teachers' changing roles in DDL. Hunston (2002) noted that teachers may feel a loss of authority when corpus searches fail or produce unexpected results, while Breyer (2009) found that teachers struggled to anticipate DDL exercise outcomes, prepare suitable examples, and ensure tasks matched learner needs. As a result, teachers often preferred closed tasks with predictable outcomes rather than open concordance activities. Similarly, Ma et al. (2022a) reported that teachers faced difficulties in selecting corpus tools aligned with objectives and motivating students who were reluctant to adopt new technologies.

A second issue involves managing the large volume of data in corpora. Bunting (2013), working with three ESL teachers in the USA, found mixed experiences with the Compleat Lexical Tutor (Lextutor.ca). One teacher doubted its value for typical students, another saw excitement but also frustration with tool malfunctions, while the third viewed it as effective for upper-intermediate learners in building vocabulary awareness. Yet all three expressed their discomfort with handling numerical data produced by the tool. Bridle (2019) reported that many studies have highlighted the challenge of managing large amounts of corpus data, which often leads to longer search times and difficulties in locating appropriate examples for specific errors. These

concerns have been consistently documented over the last two decades (e.g., Yoon & Hirvela, 2004; Mueller & Jacobsen, 2015; Abdel Latif, 2021).

2.2.7.2 Perceptions of students

On the side of language learners, multiple studies have reported positive attitudes when learners use corpus tools to support their language acquisition, especially for writing skills. Yoon and Hirvela (2004) conducted a study on the views of 22 intermediate and advanced ESL students about corpus usage through questionnaires and interviews. Feedback from the students revealed that using corpora was beneficial in improving their academic writing abilities and enhancing their confidence, despite it being a time-consuming process. In addition, learners value the authenticity of the language and the ability to see it in context (O'Sullivan & Chambers, 2006) and view the ability to independently correct their errors as an advantage (Gilmore, 2009; Quinn, 2015). In a study conducted by Mueller and Jacobsen (2015), students appeared to find the “frequency” and “chart” interface functions of the Corpus of Contemporary American English (COCA) particularly useful. Other studies have reported findings on learners’ long-term intentions to use corpus tools. Gaskell and Cobb (2004) found that over half of their 20 ESL participants perceived improved grammar proficiency after a corpus-based course, with eight expressing plans to continue using corpus tools independently. Conroy (2010), surveying 110 ESL students in Australia, reported that while a small number (9%) found concordancers frustrating, most had positive attitudes—36% found them very rewarding and 53% somewhat useful. Usage also increased from 4% before training to 28% after, with 36% intending to continue to use this tool in the future. Similarly, Charles (2014) surveyed 40 international graduate students a year after a corpus-based course and found that 70% continued using personal corpora—38% weekly, 33% monthly—mainly for grammar and vocabulary checks; 93% believed corpus use improved their academic writing.

By contrast, some studies found mixed results and perceptions that are slightly more negative. For instance, Leńko-Szymańska (2014) introduced 13 ESL students to a course in using and analysing corpora. After interviewing the participants, the researcher found that participants had a positive response to the course and recognised the benefits of using corpus-based materials and tools in language teaching. However, these participants also reported that they needed more time to

fully understand the resources and software presented and to receive more guidance on how to use corpora. This issue of complexity in corpus consultation is also mentioned by students in several other studies, such as Chambers (2005), Chambers and O'Sullivan (2004), Cheng, Warren, and Xun-feng (2003), Farr (2008), Gilmore (2009), and Yoon and Hirvela (2004). Furthermore, only one recent study by Qiu (2024) directly reported a lack of connection between students' attitudes toward corpus tools and their usage patterns or writing results, whereas other studies have yet to investigate this relationship.

2.2.8 Summary of literature review on corpus tools

Corpus tools have gained widespread recognition in L2 instruction for supporting lexical, grammatical, and collocational development (Boulton, 2009a; Yoon & Jo, 2014). Tools such as Sketch Engine, COCA, and AntConc provide access to authentic language patterns, aligning with data-driven learning (DDL) principles (Frankenberg-Garcia et al., 2019; Sinclair, 1991). However, effective use often depends on learners' proficiency and corpus literacy; higher-level students engage more strategically, while lower-level learners may struggle with concordance interpretation (Crosthwaite et al., 2019; Lin, 2021). Research also highlights differing practices between teachers and students: teachers tend to instruct students to use corpus tools during planning and revision (Charles, 2018; Lin, 2019; Ma et al., 2022a; Römer, 2011; Todd, 2001), whereas students often use this tool for grammar and vocabulary revision (Bridle, 2019; Yoon, 2008, 2016a, 2016b). Nonetheless, some studies report no clear link between perceived benefits and actual tool engagement or performance outcomes (Qiu, 2024). Despite their pedagogical value, issues such as data overload, technical complexity, and lack of training seem to hinder effective use of corpus tools in the language classroom (Chambers, 2005; Chambers & O'Sullivan, 2004; Cheng et al., 2003; Farr, 2008; Gilmore, 2009; Yoon & Hirvela, 2004).

2.3. Empirical studies on machine translation tools in L2 writing

2.3.1 Machine translation tools and Google Translate

Machine translation (MT) refers to the use of computer software to translate text between languages. Early MT systems, introduced in the 1970s, used rule-based machine translation (RBMT), which relied on grammatical rules for each language pair

and often produced literal, structurally flawed, or semantically inaccurate translations (Aiken & Wong, 2006; Hutchins, 2000, 2006; Kliffer, 2005, 2008; Niño, 2008, 2009). Google Translate (GT), launched in 2006, initially employed RBMT and faced similar criticisms for low translation quality. In 2016, Google adopted neural machine translation (NMT), a data-driven approach that generates statistically probable translations based on large bilingual corpora (Ducar & Schocket, 2018). This transition has been linked to notable improvements in translation accuracy and readability (Briggs, 2018; Ducar & Schocket, 2018; Sun, 2017). Today, GT is a widely used multilingual translation service offering text, speech, and image translation across numerous languages (Briggs, 2018; Ducar & Schocket, 2018).

2.3.2 Merits and shortcomings of using machine translation tools for L2 writing and language learning

In the context of learning English as a second language, several studies have highlighted the potential impact of machine translation tools (MTTs)—particularly Google Translate—to enhance writing outcomes, especially in grammar and vocabulary complexity (Cancino & Panes, 2021; Chon et al., 2021; Kol et al., 2018; Lee, 2020; Tsai, 2019). Kol et al. (2018) found that learners produced more advanced vocabulary when allowed to use GT, while Tsai (2019) noted that students generated longer texts with fewer grammatical and spelling errors when using GT compared to their self-written drafts. Lee (2020) reported that Korean learners at B1–B2 levels made fewer grammatical and lexical errors after revising their own English translations using outputs from GT and Papago, though gains in sentence complexity and vocabulary diversity remained minimal.

On the other hand, prior research also mentioned two broad shortcomings for using MTTs in L2 writing: variable translation accuracy across language pairs and unresolved questions about academic integrity. The accuracy of translated output of MTTs depends heavily on the quantity and quality of parallel data available, so languages with richer translation traditions tend to yield more reliable output than those with limited bilingual corpora or greater structural distance from English (Aiken & Balan, 2011; Groves & Mundt, 2015). For instance, empirical evaluations show wide performance variation: dominant lexical-semantic errors in Persian to English (Ghasemi & Hashemian, 2016); unreliable Arabic to English output (Jabak, 2019); low

accuracy and success mainly on simple sentences for Indonesian to English (Suhono et al., 2020); below-expected quality in English Afrikaans compared with human translators (van Rensburg & Lotz, 2012); mixed teacher judgments for Chinese to English, with average grammar but inconsistent vocabulary quality (Stapleton & Kin, 2019); and 64 errors/1,000 words for Malay to English, a level that roughly parallels grammar performance seen in IELTS 6.5 writing samples (Groves & Mundt, 2015; Müller, 2015). Evidence for Vietnamese to English accuracy remains scarce, yet one study revealed that student-mediated translations showed recurrent grammatical, clause, sentence, and lexical naming errors, but no quantified accuracy rates (Nguyen & Ngo, 2021). Furthermore, ethical concerns arise when students use MTTs to “back-translate” source texts, potentially obscuring plagiarism because standard detection software may not flag translated material as copied (Correa, 2011; Jones, 2009; Prentice & Kinden, 2018; Roe & Perkins, 2022). Several other empirical studies have also investigated how MTT access influences writing outcomes. Garcia and Pena (2011) compared two student groups: one wrote directly in the L2, while the other wrote in L1 and used MT. Findings revealed that MTTs helped low-proficiency learners communicate more effectively and produce higher-quality texts. Screen recordings showed that writing directly in L2 required more cognitive effort and editing, whereas MT users demonstrated greater fluency and engagement, though conclusive evidence on long-term language development remains limited. Other studies have linked MTT use to improved vocabulary acquisition and error reduction. Students with access to MTTs performed better in vocabulary tasks than those without (Kol et al., 2018; Fredholm, 2019; Lee, 2020). Lee (2020) also noted that students were able to correct their own grammatical and lexical errors after comparing their self-produced sentences with GT outputs. Chon et al. (2021) similarly suggested that MTTs support learners in resolving lexico-grammatical challenges during writing.

2.3.3 Machine translation usage in relation to language proficiency

Although prior studies have suggested that language proficiency influences students' engagement with machine translation tools (MTTs), the nature of this relationship appears to be complex and context-dependent. Some research indicates that higher-proficiency learners tend to input full sentences and critically evaluate MTT outputs before accepting them (Ahn & Chung, 2020; Chung & Ahn, 2021). In contrast, lower-

proficiency students are often more likely to accept machine-generated translations at face value, regardless of input length or quality (Ahn & Chung, 2020; Chung, 2020; Chung & Ahn, 2021; Lee, 2020; Xu, 2022). However, other studies present divergent patterns. For example, Üstünbaş (2025) reported no significant differences in MTT usage between pre-intermediate and upper-intermediate learners. Contrastingly, several studies have found that lower-proficiency learners may perceive greater benefits from MT tools, as they tend to produce longer texts with MTT support compared to their higher-proficiency counterparts (Cancino & Panes, 2021; Garcia & Pena, 2011), and may benefit more broadly from MTT (Liang, 2025; Xu & Zhang, 2022; Xu, 2022). Cancino and Panes (2021), for instance, demonstrated that beginner-level students (A1–A2) in Chile improved in syntactic complexity, accuracy, and text length when using Google Translate, particularly with guided instruction. Similarly, Xu (2022) found that although low-proficiency learners often engaged with MTTs at a surface level, they nonetheless reported gains in vocabulary acquisition and idea generation. More differentiated insights are provided by Lee and Briggs (2021), who showed that while both intermediate and high-intermediate learners benefited from MTT-assisted revision, those with higher proficiency were more likely to make accurate corrections, indicating more effective tool use. Chon et al. (2021) found that the use of machine translation tools (MTTs) helped reduce the performance gap between high- and low-proficiency Korean learners, as both groups produced comparable results on MTT-assisted writing tasks. Similarly, Lee (2021) reported that with teacher guidance, low-proficiency students began the writing process with planning and idea generation, using MTTs to clarify ambiguous expressions. Lo (2023) observed that high-proficiency learners effectively acquired new vocabulary through the use of MTT, while low-proficiency learners demonstrated short-term gains in vocabulary retention, though their delayed retention remained limited.

2.3.4 The use of machine translation tools by teachers

Early research in the 1990s recognised the potential of MTTs in language education. Ball (1989, cited in Somers, 2012) was among the first to suggest using MT in the classroom by presenting error-filled translations as learning tools for students to identify and correct. Subsequent studies also supported this approach, noting that such flawed outputs could foster revision skills and promote learner autonomy. French

(1991), for example, proposed using machine translation on short texts to help students detect ambiguities in the source language and better understand the limitations of MTTs. Shei (2002) asked Chinese learners of English to refine their English writing based on how natural the machine-translated Chinese texts appeared, with students documenting observed errors, most of which stemmed from limitations in the MTT itself. Similarly, Richmond (1994) encouraged students to adjust the L1 source text until the resulting L2 translation met expectations, while Anderson (1995) used MT to prompt Hebrew-speaking students to identify and revise MT errors using native speaker intuition and L2 reference materials.

In the subsequent years, MT was also frequently viewed as a source of errors in a wide range of studies (Somers, 2003; Kliffer, 2005; Niño, 2004, 2008), particularly lexico-grammatical errors that require post-editing correction. For instance, Niño (2004) assigned a group of advanced English students of Spanish an English source text and a raw MT output in Spanish, and asked them to underline and repair the errors. The emphasis of this teaching approach was placed on identifying common language and translation problems and correcting them in groups that proposed alternative modifications or corrections. In short, prior studies indicated that while teachers encouraged awareness of linguistic differences by examining machine-translated texts and discussing common translation pitfalls in the first language.

2.3.5 The use of machine translation tools by students

A substantial body of research shows that students frequently rely on machine translation tools (MTTs) for language learning, most notably for vocabulary support (Briggs, 2018; Fountain & Fountain, 2009; Luton, 2003; Steding, 2009; Williams, 2006). Many use MTTs as dictionaries to assist lexical choices during composing (Chandra & Yuyun, 2018; Clifford et al., 2013; Dahmash, 2020; Farzi, 2016; Larson-Guenette, 2013; Tsai, 2019). For instance, Clifford et al. (2013) found that 91% of 905 students used MTTs for vocabulary, with fewer for idiomatic expressions (36%), connectors (31%), verb tenses (29%), and word order (20%). Farzi (2016) likewise reported that ESL students from six countries used Google Translate to translate short segments and verify meanings. Other studies confirmed frequent use during composing (Chandra & Yuyun, 2018; Musk, 2022) and vocabulary retrieval during writing (Dahmash, 2020). MTTs have also been used in revision: Tsai (2019) noted

Chinese students used Google Translate for vocabulary and grammar awareness, while Chandra and Yuyun (2018) observed post-writing checks of vocabulary, spelling, and grammar. By contrast, Clifford et al. (2013, p. 111) reported a broader use of MTT across all stages of language learning. Among 905 online survey respondents, 54% used MTT to check vocabulary and grammar, 43% for writing, and 42% for pre-writing.

2.3.6 Perceptions of teachers and students towards machine translation tools

2.3.6.1 Perceptions of teachers

Early and recent research alike has documented a range of teacher perceptions toward machine translation tools (MTTs), with negative attitudes being particularly prominent. In the early 2000s, many educators viewed MTTs—particularly tools like Google Translate (GT)—as pedagogically harmful and ethically problematic. Criticisms focused on the frequency of literal, word-for-word translations, the mishandling of idioms, and the lack of contextual understanding in MTT outputs (Burton, 2003; Luton, 2003; Somers et al., 2006; Stapleton, 2005; Williams, 2006). Burton (2003), for instance, characterized the use of MTTs as “unethical,” while Somers et al. (2006) argued that evaluating MTT-assisted writing as equivalent to student-generated work was unacceptable, as it offered no instructional value: “a translation produced with no intellectual input has no instructional value, and it is, therefore, a waste of the teacher’s time to correct it” (p. 2). This critical stance led some researchers to propose detection or prevention measures to limit MTT misuse in classrooms (Steding, 2009; Fountain & Fountain, 2009).

In more recent years, teacher concerns have expanded to include academic integrity, student overreliance on translation tools, and doubts about MTT accuracy (Kirchhoff et al., 2011; Sheppard, 2011; van Rensburg et al., 2012). Clifford et al. (2013) reported that educators were particularly wary of students submitting machine-generated work as their own, raising concerns about plagiarism and exposure to flawed language models. Clifford et al. (2013, p.114) found that 42% of 40 instructors considered student use of GT to be cheating, while others believed its acceptability depended on the assignment type or usage context. Jolley and Maimone (2015) echoed these findings, noting that some instructors deemed translating entire paragraphs with GT unethical.

Scepticism about the pedagogical value of MTTs also persists. Many educators believe that MTTs can hinder the development of language skills by encouraging linguistic accuracy over complexity and self-expression (García & Pena, 2011; Correa, 2014; Van Praag & Sanchez, 2015; Briggs, 2018; Ducar & Schocket, 2018). For example, Stapleton and Kin (2019) found that teachers unanimously opposed using GT for full-text translation beyond single words, emphasising that it discourages students from producing original content. Stapleton & Kin (2019) expressed concerns that growing dependence on MTTs could eventually diminish the relevance of teachers' roles in the classroom.

Despite these concerns, a growing number of studies have highlighted more positive or balanced perspectives on the role of MTTs among language teachers and educators. Scholars such as Payne and Ross (2005), Niño (2009), Josefsson (2011), Clifford et al. (2013), and Tsai (2022) have pointed to potential pedagogical benefits of MTTs, including increased learner motivation, opportunities for language exploration, and post-editing as a learning activity. Payne and Ross (2005), for example, argued that MTTs can lower the cognitive burden for low-proficiency learners, allowing them to participate more confidently in writing tasks. Although such support may risk dependency, it can also mitigate anxiety and foster learner autonomy when used strategically. Stapleton and Kin (2019) found that when language teachers unknowingly evaluated Google Translate-generated compositions, most showed positive attitudes toward integrating machine translation tools (MTTs) into teaching. One teacher notably encouraged their use, stating, "if [students] have the motivation and the will to use...any tools available to them, then I will be more than happy" (pp. 25–26). However, teachers also recommended cautious implementation, viewing MTTs as supportive rather than replacements for active language learning.

In summary, teachers' perceptions of MTTs remain mixed. While early and continuing concerns focus on academic integrity, linguistic quality, and learner autonomy, more recent studies point to evolving views that acknowledge MTTs' potential as supportive tools, particularly when students are guided in their use. The tension between these perspectives underscores the importance of clearly defined instructional goals and appropriate pedagogical frameworks for integrating MTTs into language learning.

2.3.6.2 Perceptions of students

Student perceptions of machine translation tools (MTTs), particularly Google Translate (GT), are mixed and often reflect both critical evaluations and perceived benefits. Several studies have documented students' scepticism about the reliability of MTT outputs, especially for academic writing. For instance, Briggs (2018) surveyed 85 Korean university students regarding their views on GT and Naver, revealing that only 23.8% trusted MTT outputs for writing assignments, while 41.3% expressed distrust. The findings suggest that students do not necessarily use MTTs uncritically, even when they rely on them for language learning purposes. Concerns about the accuracy and limitations of GT were also prominent in Pham et al.'s (2022) study involving 250 Vietnamese university students. Overall, student perceptions were reported as neutral or ambivalent, with accuracy emerging as a central concern. Specifically, 70% of participants reported that Google Translate often produced confusing or inaccurate sentences, with disordered vocabulary and missing grammar, which is consistent with Nguyen's findings (2022). Across studies, students also viewed MTTs as ineffective for grammar learning (Jolley & Maimone, 2015; Lee, 2020). Nonetheless, no ethical or academic integrity concerns were reported in studies by Briggs (2018), Pham et al. (2022), or Nguyen (2022).

In contrast, several studies have highlighted more favourable student perceptions of machine translation tools, particularly regarding vocabulary support and writing efficiency (Alhaisoni & Alhaysony, 2017; Alm & Watanabe, 2021; Chung & Ahn, 2022; Garcia & Pena, 2011; Nguyen, 2022). Alhaisoni and Alhaysony (2017) found that Saudi ESL university students held generally positive attitudes toward GT, viewing it as an effective tool for determining the meanings of unfamiliar words in reading, writing, and translation tasks. Similarly, Garcia and Pena (2011) interviewed low-proficiency ESL learners at CEFR levels A1 and A2 and found that while some beginner students were skeptical, many reported tangible benefits from MT use. These included faster writing, fewer errors, less time spent searching for words, and opportunities to acquire new vocabulary. Chung and Ahn (2022) also reported largely positive perceptions among students using MT in English composition tasks. Most participants expressed satisfaction with MT use, acknowledging its limitations but highlighting its usefulness. Notably, over 90% indicated a willingness to continue using

MT tools in the future, despite recognising potential risks of lexical and grammatical inaccuracies. Alm and Watanabe's (2021) pilot study on university students and their teachers revealed that teachers were largely unaware of the varied strategies students employed when using MTTs in formal learning contexts. Contrary to common concerns about academic dishonesty, the study reported little evidence of cheating or misuse. Instead, students highlighted the need for clearer guidance on how to use such tools effectively and appropriately.

2.3.7 Summary of literature review on machine translation tools

Machine translation tools (MTTs), particularly Google Translate (GT), have significantly evolved with the advent of neural machine translation (NMT), improving translation accuracy (Briggs, 2018). Empirical research shows that MTTs can support L2 writing by enhancing vocabulary and reducing grammatical errors (Lee, 2020; Tsai, 2019), though concerns remain about translation inconsistency and ethical risks such as plagiarism via back-translation (Groves & Mundt, 2015; Roe & Perkins, 2022). Learner proficiency plays a complex role: while higher-proficiency users often engage more critically with MTT outputs, lower-proficiency learners may benefit more in terms of fluency and text expansion (Ahn & Chung, 2020; Xu, 2022). Teachers and students also differ in their use and perceptions. Teachers have traditionally used MTTs to raise language awareness during revision (Niño, 2004), whereas students frequently rely on them during composing, especially for vocabulary support (Chandra & Yuyun, 2018). While educators remain cautious due to issues of accuracy and overdependence (Stapleton & Kin, 2019), recent work points to the motivational potential of MTTs when used with pedagogical guidance (Tsai, 2022). Students, on the other hand, generally report positive experiences, citing gains in vocabulary, fluency, and writing confidence (Alm & Watanabe, 2021; Nguyen, 2022).

2.4. Empirical studies on automated paraphrasing tools in L2 writing

2.4.1 Automated paraphrasing tools and Quillbot

Automated paraphrasing tools (APTs) are web-based applications that apply algorithms similar to those used in machine translation to convert one text into another (Rogerson & McCarthy, 2017). Although the sophistication and effectiveness of machine translation vary, advancements in natural language processing (NLP) and

machine learning continue to enhance these tools. According to Bailey (2018), the development of paraphrasing technology originated from “article spinning”—a method of rewriting content by replacing words and phrases with synonyms to produce seemingly original text. Similarly, Zhang et al. (2014) noted that some early paraphrasing tools were created to support text spinning for boosting website rankings in search engines like Google.

APTs are widely accessible, with many free and paid versions available online (Somers, 2012). Most tools follow a straightforward process: users input text, activate the paraphrasing function, and receive rewritten output conveying the same meaning with altered wording. More advanced, subscription-based tools offer enhanced features, such as customizable dictionaries, synonym frequency settings, and options for modifying single words and phrases (Shahid et al., 2017, as cited in Prentice & Kinden, 2018). Examples of such tools include QuillBot (quillbot.com), Wordtune (wordtune.com), Paraphraser (paraphraser.io), Spinrewriter (spinrewriter.com), and SpinnerChief (spinnerchief.com). While most APTs are designed for English, some, like Wordtune, include translation support for languages such as Spanish, Mandarin, Arabic, Hindi, Korean, Hebrew, and Russian (Zhao, 2022). QuillBot, launched in 2017, is an AI-powered paraphrasing tool that utilizes NLP algorithms to help users rephrase text while retaining the original meaning. It offers a user-friendly interface and is accessible for free, with additional modes—Standard, Fluency, Creative, and Formal—available to subscribers. Figure 4 below shows the interface of Quillbot as a paraphrasing tool.

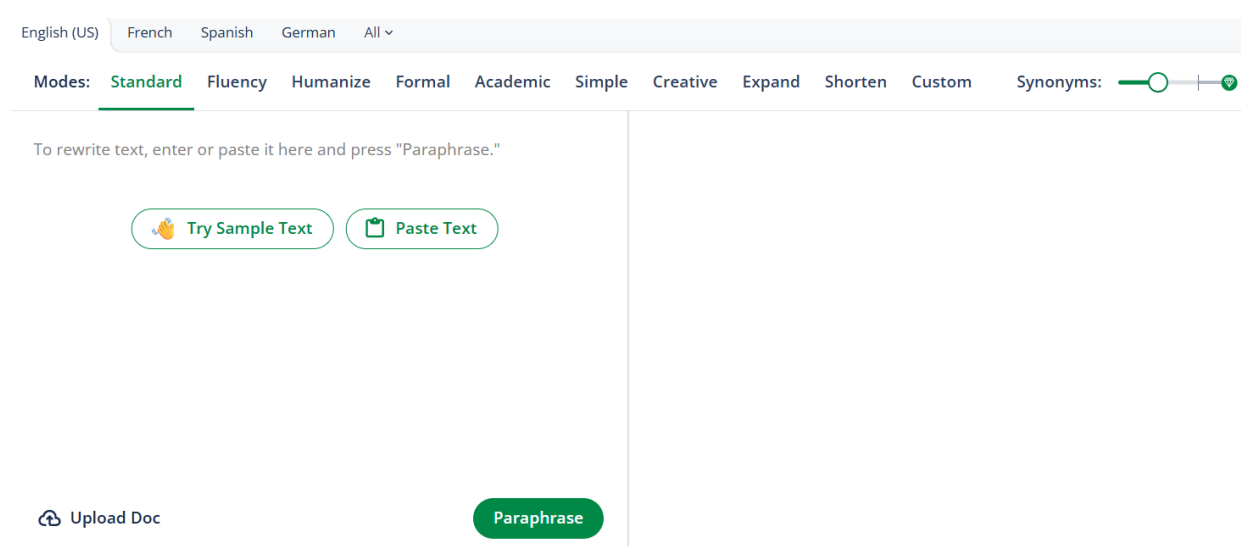


Figure 4. The interface of Quillbot as an automated paraphrasing tool

A recent survey by Alzubi et al. (2024) found that English teachers responded positively to QuillBot's features. The tool provides users with several options to refine their output, such as replacing words with synonyms, altering word order and word forms, and toggling between active and passive voice. These functionalities align closely with paraphrasing strategies identified in academic research (Bailey, 2018; Keck, 2006, 2014; Liao & Tseng, 2010; Na & Nhat, 2017). Despite its versatility, however, QuillBot and similar APTs have inherent limitations. As emphasised by Shi (2004) and Wette (2010), such tools are unable to fully grasp the meaning of source texts or determine key ideas. Effective paraphrasing still requires users to comprehend the original content and extract its core message before using these tools. This highlights the fact that APTs serve as linguistic aids, not substitutes for genuine understanding and cognitive engagement.

2.4.2. Merits and shortcomings of using automated paraphrasing tools for L2 writing and language learning

While APTs are increasingly available to L2 learners, relatively few studies have examined their impact on writing proficiency. Chen et al. (2013) evaluated PREFER, a web-based paraphrasing tool that offers paraphrases with Chinese equivalents, usage patterns, and example sentences for multi-word inputs. In an experiment with 55 ESL learners, the PREFER group outperformed groups using the Longman Dictionary or a Thesaurus, gaining 38.2% in post-test paraphrasing scores, while the other groups declined. Less proficient, more motivated, and conservative students benefited most; 90% were satisfied with the tool and 75% found it useful for writing.

More recent work has focused on QuillBot. Amyatun and Kholis (2023) reported that, after QuillBot training, 20 Indonesian eleventh graders' argumentative writing scores rose from 53.55/100 to 78.90/100, with significant gains in coherence, cohesion, grammar, and confidence. Truong (2023) likewise found that Vietnamese EFL students using QuillBot, compared with a control group receiving only teacher and peer feedback, showed significant improvements in grammar, vocabulary, cohesion, and coherence on an IELTS-based rubric, with many scoring 8–9/10 and reporting clearer sentence structure, more logical flow, and better organisation of ideas.

Despite these merits, APTs present notable shortcomings for users. Carter and Inkpen (2012) highlighted that many free online tools are infrequently updated, often relying on static algorithms maintained by non-specialists. This results in low-quality outputs or "word salad"—nonsensical, disjointed text—when the tool fails to analyze input comprehensively (Callison-Burch & Flounoy, 2001; Rogerson & McCarthy, 2017). Such outputs typically fall into the “minimal revision” category of Keck’s (2006) paraphrase taxonomy. Prentice and Kinden (2018) illustrated how APTs substituted technical terms with meaningless equivalents, and further cautioned that such substitutions may go undetected by plagiarism software (Bailey, 2018). As a result, APTs raise concerns regarding academic integrity, as students might produce different outputs from the same source material without being flagged for plagiarism (Socher et al., 2011; Wahle et al., 2021). Moreover, APTs can compromise assessment validity. Students who overuse these tools may fail to demonstrate actual comprehension of the source material or the assignment task (Rogerson & McCarthy, 2017). If source texts are paraphrased without citation, students may also violate academic integrity policies (Prentice & Kinden, 2018; Rogerson & McCarthy, 2017). Additionally, widespread APT use can impair teachers’ ability to accurately evaluate student performance and detect language gaps (Li & Casanave, 2012). In the long term, this may hinder learners’ progress and misrepresent their actual writing competencies.

In summary, experimental studies (Amyatun & Kholis, 2023; Chen et al., 2013; Truong, 2023) highlight the potential of automated paraphrasing tools (APTs) in supporting ESL/EFL writing. These studies show that APTs can improve learners’ paraphrasing accuracy as well as their use of grammar and vocabulary. They also contribute to better text cohesion in students’ written work.

2.4.3 Automated paraphrasing tool usage in relation to language proficiency

Although limited, findings from existing research suggest that language proficiency plays a role in how learners engage with automated paraphrasing tools. Higher-proficiency writers tend to rely less on digital assistance and are more capable of generating accurate paraphrases using their own linguistic resources, whereas lower-proficiency learners often depend on tools like synonym finders in a mechanical way, which can result in awkward or incoherent outputs (Bailey & Withers, 2018). Buschek et al. (2021) investigated how native and non-native English speakers interact with AI-

generated paraphrasing suggestions (ChatGPT) during email writing tasks and found that non-native speakers—presumably lower-proficiency users—accepted more tool-generated suggestions, incorporated more suggested vocabulary, and reported a greater influence on their writing content and phrasing. Similarly, Kim et al. (2024), using think-aloud protocols, demonstrated that lower-proficiency learners (A2) relied heavily on multiple support features of AI writing tools to complete paraphrasing tasks, while higher-proficiency learners (B1 and B2) were more selective, drawing on their linguistic knowledge to evaluate and revise AI outputs more independently.

2.4.4 The use of automated paraphrasing tools by teachers

Few scholars have explored the pedagogical application of automated paraphrasing tools (APTs). Zhao (2022) proposed the use of an APT named Wordtune and suggested that teachers can encourage students to engage critically with the vocabulary, synonyms, clauses, and formal sentence structures they encounter through the rewrites generated by the tool. Another approach to teaching paraphrasing is encouraging students to analyse errors produced by automated tools. Pecorari (2003), studying 17 postgraduate ESL students, found that much of their paraphrasing resembled patchwriting, often seen as plagiarism. However, through interviews and text analysis, she argued this reflected a developmental stage rather than intent, echoing Howard's (1995) view that patchwriting is part of novice academic writing. Following Pecorari, several scholars have supported error demonstration as a teaching strategy (Abasi, 2008; Chen et al., 2013; Li & Flowerdew, 2007; Niño, 2009; Rogerson & McCarthy, 2017; Tan & Carnegie, 2022; Ansorge et al., 2021; Roe & Perkins, 2022). For instance, Niño (2009) recommended openly showing students the errors generated by online tools, while Rogerson and McCarthy (2017) highlighted issues like “word salads” and poor expression. Ansorge et al. (2021) recommended warning students about risks such as missing data or confusing meanings, and Chen et al. (2013) called for clear instructional plans distinguishing pedagogical use from assessed work. Other studies proposed encouraging students to practice with paraphrasing (Hiverla & Du, 2013) or comparing student writing with tool-generated alternatives to promote self-directed learning (Barrot, 2023).

2.4.5 The use of paraphrasing tools by students

Overall, little empirical work has examined how students actually use APTs in their writing. In one of the few studies, Sulistyaningrum (2021) surveyed 30 English majors and found that 83% used free tools such as Paraphrasing Tool and Spinbot to rewrite literature review sources, mainly to select appropriate vocabulary and parts of speech. Students may also combine APTs with machine translation in “back translation” cycles (Jones, 2009; Dinneen, 2021), translating a source into another language and then back into English before sending it through an APT again. This can produce “word salad” (Rogerson & McCarthy, 2017) that writers must then revise, and such practices can constitute academic dishonesty and plagiarism (Roe & Perkins, 2022), even if often unintentional due to confusion between patchwriting and acceptable paraphrasing (Rogerson & McCarthy, 2017; Keck, 2010; Bowen & Nanni, 2021). In controlled settings, however, teachers may tolerate these strategies and use them as opportunities for vocabulary learning and error noticing (Pecorari, 2003; Niño, 2009; Ansorge et al., 2021). Similar combinations of APTs and machine translation in ESL writing are also reported by Prentice and Kinden (2018).

2.4.6 Perceptions of teachers and students towards automated paraphrasing tools

2.4.6.1 Perceptions of teachers

A small but growing body of research has examined how Automated Paraphrasing Tools (APT) are used by research scholars and EFL teachers, highlighting both their pedagogical benefits and their ethical and practical limitations. Singh et al. (2023) surveyed 200 research scholars at Babasaheb Bhimrao Ambedkar University and the Central University of South Bihar and found high awareness and frequent use of paraphrasing tools, particularly Quillbot and Duplichecker. While respondents reported using these tools to rephrase text, manage their workload, and, in their view, “avoid plagiarism” by rewording sentences, the authors also call for training on their “ethical” use, underscoring that such tools can just as easily obscure plagiarism through surface-level rewriting rather than genuinely prevent it.

Alzubi et al. (2024) studied EFL teachers at Najran University, Saudi Arabia, who rated Quillbot highly for enhancing paraphrasing, vocabulary, grammar, and overall writing.

Teachers valued its role in teaching grammar and supporting shy students' participation but noted drawbacks such as over-reliance, uneven effectiveness across proficiency levels, and the need for revision of outputs.

Marzuki et al. (2023) interviewed four teachers about Quillbot, WordTune, ChatGPT, and Jenni, with Quillbot emerging as particularly useful for paraphrasing and sentence structure. Teachers highlighted benefits such as improving clarity, fluency, and reducing repetition, but also cautioned that advanced vocabulary suggestions could lead to awkward writing and that excessive reliance might hinder critical thinking.

In summary, the three studies consistently acknowledge the effectiveness of APTs in enhancing paraphrasing, improving writing, and helping users avoid plagiarism by generating alternative sentence structures. QuillBot, in particular, is recognised for supporting vocabulary development and saving time (Alzubi et al., 2024; Marzuki et al., 2023). Nonetheless, challenges remain: Singh et al. (2023) reported issues such as subscription costs, word limits, and loss of meaning, while Alzubi et al. (2024) and Marzuki et al. (2023) cautioned against over-reliance, which may reduce originality, and stressed the importance of revising tool outputs for accuracy. Despite these insights, research on teacher usage of APTs remains limited, and none of the existing studies explicitly identify which stage of the writing process these tools are most suitable for.

2.4.6.2 Perceptions of students

The three reviewed studies reveal both positive and negative perceptions of automated paraphrasing tools (APTs), particularly QuillBot, among ESL/EFL students. Overall, students expressed predominantly positive attitudes, frequently citing improvements in grammar, vocabulary, sentence structure, and overall writing quality. For instance, in Kurniati and Fithriani's (2022) study of 20 Indonesian postgraduate students, 75% reported that QuillBot boosted their confidence, improved the quality of their academic writing, and reduced writing anxiety. Similarly, 95% believed it enhanced their vocabulary, and 80% noted improvements in grammar. Mohammad et al. (2023) also found that 31 preparatory-year EFL students at Najran University perceived QuillBot as helpful in addressing paraphrasing challenges, particularly in synonym selection, sentence variety, and word choice, while also increasing their

enthusiasm for writing and reducing anxiety. In Alammar and Amin's (2023) study, Saudi students appreciated QuillBot and other APTs for improving grammar and facilitating paraphrasing, with many highlighting their usefulness for saving time and avoiding plagiarism. These consistent findings demonstrate the widespread belief among learners that APTs, when used strategically, can support multiple aspects of L2 academic writing.

To sum up, studies exploring student perceptions (Kurniati & Fithriani, 2022; Alammar & Amin, 2023; Mohammad et al., 2023) and those focused on teachers' perceptions (Singh et al., 2023; Marzuki et al., 2023; Alzubi et al., 2024) demonstrate a consensus regarding APTs' positive role in developing English writing skills. Students value the tool for improving paraphrasing, especially through more effective use of synonyms, sentence structures, and word choice and appreciate its capacity to boost confidence, alleviate anxiety, and assist with complex grammatical forms. Likewise, teachers emphasise QuillBot's contribution to students' paraphrasing and its enhancement of sentence fluency and clarity.

In terms of negative perceptions, a similar concern was found between students and teachers, which is the risks of overreliance on QuillBot (Alammar & Amin, 2023), aligning with teachers' apprehensions in Marzuki et al. (2023) that excessive use of the tool could impair students' critical thinking and writing independence. However, differences in perception also emerged. While some students viewed the tool negatively for inaccuracies in word choice and contextual errors produced by QuillBot (Alammar & Amin, 2023), they did not mention the concern that the tool's output quality largely depends on their own English proficiency—a factor that teachers associate with a low quality of writing (Marzuki et al., 2023). Teachers also emphasised the differences in proficiency levels when it comes to revising Quillbot's output (Alzubi et al., 2024), yet the students' English proficiency level in the three studies by Kurniati & Fithriani (2022); Mohammad et al. (2023); and Alammar & Amin (2023) was not measured.

2.4.7 Summary of literature review on automated paraphrasing tools

The literature on automated paraphrasing tools (APTs), particularly QuillBot, highlights their increasing role in second language (L2) writing. APTs use algorithms to rephrase

input while preserving meaning (Rogerson & McCarthy, 2017), with QuillBot recognised for its accessibility and AI-enhanced features (Zhao, 2022). Studies have shown that APTs can improve paraphrasing quality, grammar, vocabulary, and writing coherence, especially for less proficient but motivated learners (Truong, 2023; Chen et al., 2013). However, limitations remain, including incoherent output (Callison-Burch & Flournoy, 2001), meaning distortion (Prentice & Kinden, 2018), and concerns about academic integrity (Bailey, 2018). Tool effectiveness often varies by proficiency: more advanced learners tend to engage with APTs more critically (Kim et al., 2024; Bailey & Withers, 2018). Pedagogically, APTs have been proposed as useful tools for developing vocabulary and fluency when paired with explicit instruction and critical feedback (Li & Flowerdew, 2007; Rogerson & McCarthy, 2017). Despite these potentials, research on actual student usage remains limited. Available evidence suggests students frequently use APTs during composing, often alongside back-translation, which may lead to patchwriting and unintentional plagiarism (Keck, 2010; Roe & Perkins, 2022). Perception studies report increased confidence and writing support (Alammar & Amin, 2023), though both students and teachers caution against overreliance, which may hinder deeper language learning (Alzubi et al., 2024; Singh et al., 2023). Overall, while APTs show pedagogical promise, more research is needed on their real-world use and impact across proficiency levels.

2.5. Empirical studies on automated writing evaluation tools

2.5.1 Introduction to automated writing evaluation tools

An automated writing evaluation tool (AWET), also referred to as automated written corrective feedback (AWCF) or automated essay scoring (AES), is a technology tool developed to give writers instant feedback in the form of writing suggestions and automated scores, enabling writers to assess and reflect on their writing skill (Grimes & Warschauer, 2010). The first AWET was called Page Essay Grade (PEG) and was developed by Ellis Page. Page used “multiple regression to associate surface text features in a targeted essay to those in a corpus of essays on the same topic that English teachers had scored” (Page, 1967, cited in Grimes & Warschauer, 2010, p.6). Following PEG, many algorithms underlying AWETs were developed based on artificial intelligence machine learning models, including natural language processing (NLP) (Vantage Learning, 2001; cited in Wang & Brown, 2007; Landauer et al., 2003).

Some widely merchandised AWETs listed in Stevenson and Phatiki (2014) are Project Essay Grader (PEG), e-rater, Intelligent Essay Assessor, IntelliMetric, Criterion, MYAccess!, and Write To Learn, which are mostly used by schools and institutions and are not free of charge.

2.5.2 Grammarly

Grammarly is a free-access automated writing evaluation tool. It can locate grammar, spelling, punctuation, and other errors in writing. According to Cavaleri and Dianati (2016), Grammarly operates based on an interconnected system that integrates various principles, innovations, and techniques of AI, such as machine learning, deep learning, and natural language processing. Particularly, Grammarly relies on a text corpus that includes a vast collection of sentences organised and labelled by human experts in a way that AI algorithms can interpret. Since 2009, Grammarly has been marketed as a writing assistance application. Free-access users can receive immediate feedback on spelling, punctuation, grammar structures, and overall score when using Grammarly's free version to check their writing. In 2020, Grammarly was widely used with over 30 million daily active users, according to a report conducted by Lytvyn (2022). Figure 5 presents the interface of Grammarly and its examples of suggestions.

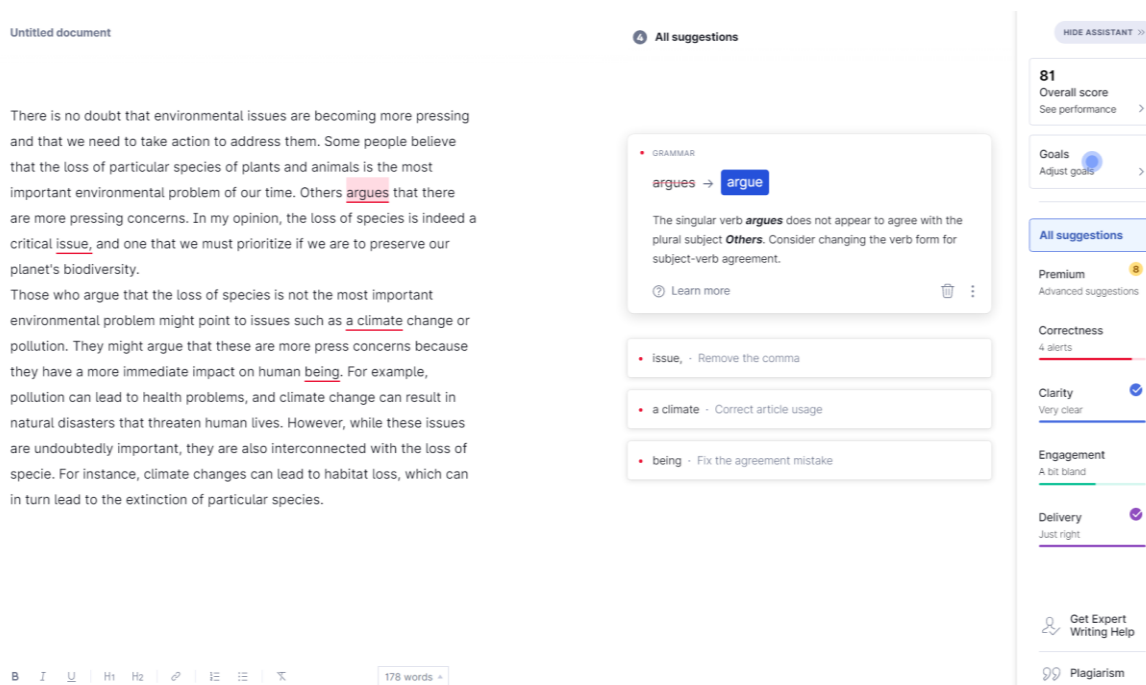


Figure 5. The interface of Grammarly feedback is presented in the free version of Grammarly (2023)

This study utilises Grammarly as an example of an AWET for three reasons. First, Grammarly is a well-known and easily accessible tool compared to other extensively-researched AWE tools like Criterion, e-rater, My Access!, and Intellimetric, which are commercialised to institutions and not freely available to students (Dikli, 2006; Strobl et al., 2019). To illustrate its popularity, Gain, Rao and Bhat (2019) delivered a questionnaire to 134 students and found that 88.6 % of their respondents are using Grammarly for writing research papers, and 96% of participants recommended Grammarly to others. The questionnaire also asked participants whether they needed training to know and use Grammarly's platform, and 63% responded that they did not need training to use this tool. Some studies demonstrated Grammarly's reliability and accuracy in detecting errors (Kotsyuk, 2015; Dembsey, 2017; Gain et al., 2019; O'Neill and Russell, 2019; Barrot & Gabinete, 2021; Koltovskaia, 2020; Bailey & Lee, 2020; Barrot, 2021; Guo et al., 2021). For instance, the study by Kotsyuk (2015) employed Grammarly as a research methodology to identify Ukrainian learners' errors in English. Second, despite its popularity, Grammarly is still an under-researched tool compared to copious research on the above-mentioned AWET tools (Barrot, 2021). Third, Grammarly offers many features that could potentially enhance L2 learners' writing accuracy (Guo et al., 2021; Koltovskaia, 2020) and has similar characteristics to other

AWETs, such as instant corrective feedback, error identification, and suggestions for corrections.

2.5.3. Merits and shortcomings of using automated writing evaluation tools for L2 writing

A large body of research highlights the value of Automated Writing Evaluation Tools (AWETs) in improving English writing. Studies consistently show that AWET feedback enhances accuracy, reduces errors, and improves overall writing quality (Al-Othman, 2024; Bai & Hu, 2016; Barrot, 2021; Dikli & Bleyle, 2014; Ghuftron & Rosyida, 2018; Liao, 2016; Palermo & Wilson, 2020; Pitukwong & Saraiwang, 2024; Ranalli et al., 2017; Sanosi, 2022; Wang et al., 2012; Wei et al., 2023). Tools such as Criterion, e-rater, Grammarly, and MI Write have been shown to improve grammar, vocabulary, and coherence—dimensions aligned with IELTS criteria. For instance, Wang et al. (2012), Liao (2016a, 2016b), and Pitukwong and Saraiwang (2024) reported consistent gains in accuracy and cohesion, while multiple studies on Grammarly found notable improvements in grammar, vocabulary, and clarity (Barrot, 2021; Ghuftron & Rosyida, 2018; Sanosi, 2022; Wei et al., 2023). Although some research noted limited effects on holistic scores (Huang & Wilson, 2021; Palermo & Wilson, 2020), AWETs were still found to reduce errors and enhance writing processes. Concerns remain over their ability to detect complex errors and the absence of human interaction (Bai & Hu, 2016; Dikli & Bleyle, 2014; Hyland, 1998), yet advances in AI have improved performance. Overall, recent evidence supports AWETs as effective tools for improving writing proficiency when used critically and strategically.

On the other hand, Automated Writing Evaluation (AWE) technology has faced considerable criticism from scholars in the field of English language teaching. In its early stage in the 1990s, scholars like Leki (1990) and Reid (1994) argued that computer-generated feedback cannot replace teacher interaction, as ESL students require explicit instruction, modelling, and practice in various writing perspectives. Similarly, Drechsel (1999) and Herrington and Moran (2001) emphasised the need for human feedback, cautioning that students may lose sight of writing as a communicative act if they begin to write primarily for machines. Truscott (1996) went further, contending that grammar correction should be abandoned altogether, as it promotes superficial learning and may negatively impact writing quality. Hyland (1998)

raised concerns that excessive reliance on automated scoring might suppress students' individual voice, while Ferris and Hedgcock (2005) criticised AWE software for lacking the capacity to negotiate meaning in real-world contexts. They also warned that overdependence on AWE tools might discourage students from engaging in essential language-learning strategies, such as consulting dictionaries or seeking teacher support, ultimately hindering long-term language development.

A broader concern among scholars is that AWETs may encourage ESL learners to produce formulaic writing or prioritise length and complexity over clarity and conciseness, as these characteristics are often more highly rewarded by automated scoring systems. As noted by Chung and O'Neil (1997) and Kukich (2000), students may focus more on surface-level features like word order or essay length at the expense of content depth and lexical appropriateness. Similarly, McGee (2006, cited in McNamara et al., 2015) highlighted how AWETs may overlook higher-order issues such as coherence, which require deeper linguistic processing.

Another key limitation is the inconsistency and inaccuracy of AWE feedback. Studies show that error detection varies widely across categories (Bai & Hu, 2017; Dikli & Bleyle, 2014; Ranalli et al., 2017). For example, Dikli and Bleyle (2014) found that Criterion missed most sentence structure errors (100%), unclear messages (75%), and run-ons (70.6%), all of which human instructors identified. Ranalli et al. (2017), analysing 370 drafts, reported accuracy rates of 60–88.6%, but below 70% for critical errors such as articles, prepositions, run-ons, and confused words. Similarly, Bai and Hu (2017) found that Pigai achieved high precision in mechanics (98%) but weaker results in grammar (58%) and collocations (22%). These findings highlight AWE's difficulty in handling complex linguistic features despite strengths in simpler error categories.

In summary, although AWETs have attracted criticism, they offer clear benefits for L2 writing. Studies show that tools such as Criterion, Grammarly, and ProWritingAid can improve grammatical accuracy, reduce errors, increase lexical variety, and support more efficient revision, while also enhancing engagement, autonomy, and confidence through timely, accessible feedback across multiple drafts (e.g. Barrot, 2021; Li et al., 2015; Palermo & Wilson, 2020; Wei et al., 2023). At the same time, scholars have long warned that AWETs may lack meaningful human interaction, encourage mechanical

or formulaic writing, emphasise surface-level corrections over deeper content, and struggle with nuanced issues such as cohesion and complex sentence structure (e.g. Bai & Hu, 2016; Drechsel, 1999; Ranalli et al., 2017). Many of these problems, however, relate to earlier generations of AWE systems. Recent advances in AI and NLP have improved accuracy and feedback quality, and there is little empirical evidence that classroom use of AWETs harms writing outcomes, suggesting they remain valuable pedagogical aids when paired with teacher guidance and critical student use.

2.5.4. Automated writing evaluation tools usage in relation to language proficiency

The effectiveness of Automated Writing Evaluation (AWE) tools appears closely tied to learners' language proficiency levels, yet the results are inconclusive. For low-proficiency learners, AWE can offer considerable benefits, particularly in relation to error correction and writing confidence. Studies have shown that low-proficient learners tend to respond more actively to automated corrective feedback (Chou et al., 2016; Shen et al., 2023; Xu & Zhang, 2022). Moreover, low-proficiency students often exhibit higher writing self-efficacy, which may lead to greater effort in writing tasks and more gains in specific skill areas (Wei et al., 2023). Instant feedback from AWE tools can therefore serve as a motivational and instructional aid, helping low-proficient learners to reduce errors and develop confidence (Saricaoglu & Bilki, 2021). However, comprehension remains a challenge—some studies report that students with limited language knowledge struggle to understand AWE explanations, which hinders their ability to benefit fully from the feedback (Shang, 2019; Gao, 2021; Yousofi, 2022).

By contrast, other research has shown that high-proficiency learners benefit more from AWETs, particularly when they possess the metacognitive strategies necessary to evaluate and refine the feedback. Liao (2016a) argued that advanced learners are more capable of leveraging automated feedback to support long-term grammatical development. Similarly, Cavaleri and Dianati (2016) noted that proficient writers are better equipped to filter out inaccurate or irrelevant suggestions, whereas learners with limited language proficiency may find such feedback overwhelming. Zhang and Hyland's (2018) longitudinal case study of two ESL students with differing proficiency levels further illustrates this divergence. Flora, the high-proficiency student (IELTS

7.5), showed strong emotional and motivational engagement with AWET feedback and actively revised her work, while David, the lower-proficiency student (IELTS 5.5), disengaged from both teacher and tool feedback, reporting embarrassment and demotivation. A similar pattern emerged in Koltovskaia's (2020) study: the advanced L2 writer relied on linguistic knowledge to critically assess and reject Grammarly's suggestions, while the low-intermediate student neither engaged deeply with the tool's feedback nor consulted external resources for revision.

2.5.5 The use of automated writing evaluation tools by teachers

Research shows that teachers integrate AWE tools into instruction in various ways to offset their limitations. One approach is using automated scores as benchmarks before providing feedback. Chen and Cheng (2008) and Li et al. (2014, 2015) found that low Criterion or MYAccess! scores alerted teachers to problematic writing, prompting closer grading and targeted feedback. Teachers also saw these scores as useful for students, helping them identify strengths and weaknesses. Some teachers required students to reach a minimum automated score before peer review, revisions, or final submission (Chen & Cheng, 2008; Li et al., 2015), effectively using AWE as formative assessment. Beyond formative use, AWETs have been employed as raters in timed writing practice. For example, Chen and Cheng (2008) used MYAccess! to score final exams under test conditions, while Link et al. (2014) observed Criterion being used to administer a midterm writing exam.

To sum up, scant attention has been paid to the ways teachers can optimise AWETs for English writing instructions. Among the few studies on this research topic, there are two approaches for integrating AWETs in the classroom. The first one is to use this tool as continuous assistance throughout a student's writing process, with a certain benchmark before the teacher's final feedback (Chen & Cheng, 2008; Li et al., 2014, 2020). The second one is to employ this tool as an official evaluator for in-class writing assessments, particularly in the post-writing step (Link et al., 2014).

2.5.6 The use of automated writing evaluation tools by students

Research on students' use of Automated Writing Evaluation Tools (AWETs) remains limited, but existing studies point to emerging patterns, especially in composing and revising. Quantitative work shows varied engagement with AWET feedback:

Chodorow et al. (2010) found that students either ignored, read, verified, or accepted Criterion suggestions, while adoption rates in Lavolette et al. (2015) ranged from 16% to 96% (M = 73%). Chapelle et al. (2015) similarly noted that learners often ignored over half the feedback and mostly made word- or phrase-level revisions, suggesting a focus on surface editing.

Qualitative studies reveal more nuanced use. Roscoe et al. (2017) found that ESL undergraduates revised mainly through additions, substitutions, and reorganisations. Zhang (2017) documented a shift from surface edits to strategic evaluation, with learners using dictionaries and peer input to check AWET accuracy. Koltovskaia (2020) observed that proficiency shaped engagement: advanced learners critically verified Grammarly feedback, while lower-proficiency learners relied more passively on it.

AWETs are also used during composing. McCarthy et al. (2022) showed students applying W-Pal feedback not only to correct errors but also to expand ideas and restructure arguments, suggesting recursive movement between drafting and revision. Koh (2017) reported that continuous, real-time feedback improved grammar more than post-draft support, highlighting learners' preference for immediate assistance.

In summary, prior research suggested that AWETs can be used flexibly across different writing stages. More proficient learners use them for idea refinement and critical evaluation, while others apply them mainly for surface-level edits. However, engagement with free-access tools such as Grammarly remains underexplored, and much existing work draws on proprietary systems (Warschauer & Ware, 2006), leaving gaps in understanding how different learners interact with AWETs in authentic contexts.

2.5.7 Perceptions of ESL teachers and students towards AWE tools

2.5.7.1 Perceptions of teachers

Several studies have reported favourable attitudes among teachers regarding the use of Automated Writing Evaluation Tools (AWETs) in the classroom. Shermis et al. (2004, as cited in Warschauer & Grimes, 2008) found that ESL teachers valued AWE tools for their time-saving benefits. These teachers noted that automatic grading

reduced their workload outside of class and allowed for more efficient in-class management by promoting student independence, thereby enabling teachers to offer more personalised support. Similarly, Grimes and Warschauer (2008), in a study across four middle schools and one high school in the USA using MYAccess!, found that six out of seven teachers believed AWE feedback was beneficial for student writing. They also reported a shift in their instructional role—from evaluator to facilitator—as they were required to explain mismatches between automated scores and human judgment when students questioned AWE-generated feedback. Li et al. (2015) interviewed four teachers using Criterion and found that AWETs helped raise students' awareness of grammatical errors. Teachers indicated that automated feedback could effectively highlight recurring issues, reducing the need for repetitive grammar instruction and allowing them to focus on higher-order aspects of writing. In a broader study involving Criterion and MYAccess!, Warschauer and Grimes (2008) concluded that these tools were a moderate but useful addition to teachers' instructional strategies. AWETs encouraged more student revision and helped streamline the writing process, although they did not significantly enhance content or organisation. Wilson and Czik (2016) also observed how AWETs influenced feedback practices. In their study with eighth-grade English students, those who received feedback via PEG Writing were compared with peers who only received teacher feedback through Google Docs. Results showed that AWET integration did not reduce the quantity of teacher feedback. Instead, it shifted focus: teachers using PEG Writing offered more feedback on sentence structure and idea development, whereas those using Google Docs focused more on formatting and capitalisation. This suggests that AWETs may support more substantive, higher-level feedback from teachers.

Despite these advantages, some teachers expressed concerns about the limitations of AWETs. Li et al. (2015) also reported that teachers were wary of the tools' occasional inaccuracies—such as missed errors or inappropriate correction suggestions—which they feared could hinder long-term improvements in language accuracy. Teachers also voiced concerns that organisational feedback generated by Criterion resembled a rigid “robotic template,” potentially constraining students' creativity and expression. Further concerns were raised by Link et al. (2014), who interviewed seven ESL university teachers. They identified two primary issues: technical limitations and pedagogical drawbacks. Teachers found the Criterion

interface restrictive, reporting that they could not carry out all desired instructional functions within the system. Additionally, some questioned the reliability of AWE-generated scores. Mismatches between teacher-assigned grades and AWE scores led to discomfort, particularly when students challenged their grades based on conflicting automated feedback. These concerns underscore a broader apprehension about over-reliance on AWETs and the potential for such tools to undermine teacher judgment and student trust in the writing assessment process.

2.5.7.2 Perceptions of students

A substantial body of research reports positive student perceptions of Automated Writing Evaluation (AWE) tools, especially for improving writing, building confidence, and supporting ease of use. Students often view them as useful for enhancing grammar, encouraging revision, and motivating self-directed learning (Barrot, 2021; Cavaleri & Dianati, 2016; Chen & Cheng, 2008; Dikli, 2006; Grimes & Warschauer, 2010; Hoon, 2006; Koltovskaia, 2020; Li et al., 2015; Luo & Liu, 2017; Moore & MacArthur, 2016; O'Neill & Russell, 2019; Palermo & Wilson, 2020; Suci & Syafryadin, 2024; Wang et al., 2012; Warschauer & Grimes, 2008). For example, Cavaleri and Dianati (2016) found that 78% of 37 ESL students surveyed felt Grammarly boosted their confidence, and half believed it improved their academic writing scores.

Positive views of AWETs are reported across several studies. Hakiki (2021) found that most of the 44 EFL students in their study credited Grammarly with improving grammar awareness, confidence, and academic writing. O'Neill and Russell (2019) showed that students receiving both Grammarly and teacher feedback were more satisfied than those with teacher feedback alone, while Suci and Syafryadin (2024) highlighted postgraduate students' appreciation of Grammarly's ease of use and real-time support. Students also valued its ability to address surface-level errors and encourage frequent revision, which fostered autonomy (Li et al., 2015; Palermo & Wilson, 2020).

Despite these benefits, concerns persist about feedback quality and accuracy. Cavaleri and Dianati (2016) reported that some students found Grammarly's suggestions unclear or overly prescriptive, while Barrot (2021) noted frustrations with overcorrections and limited explanations. Hakiki (2021) likewise documented neutral

or negative responses, with doubts about reliability, dissatisfaction with the free version, and difficulties with the interface.

Perceptions also differ by proficiency level. Lower-proficiency learners often respond more positively, finding surface-level corrections manageable (Chen & Cheng, 2008; Koltovskaia, 2020; O'Neill & Russell, 2019). By contrast, higher-proficiency students report less satisfaction, since feedback focuses on grammar rather than argumentation or rhetorical structure (Chen & Cheng, 2008; Xu & Zhang, 2022). Chen and Cheng (2008) noted that advanced learners viewed MYAccess! as overly form-focused, while Fan (2023) found both low- and high-proficiency students saw limited value in Grammarly for deeper revisions.

Taken together, student perceptions of AWETs are largely positive when it comes to improving grammatical accuracy, increasing writing confidence, and supporting revision. However, recurring concerns about feedback quality, tool limitations, and the mismatch between feedback type and writing proficiency highlight the need for critical use and pedagogical support. While AWETs are seen as valuable supplementary tools, students continue to recognise the irreplaceable role of human feedback in developing higher-level writing competence.

2.5.8 Summary of literature review on the use of automated writing evaluation tools

A growing body of research has examined the use of Automated Writing Evaluation Tools (AWETs) such as Criterion, e-rater, MYAccess!, and Grammarly, showing their potential to improve grammar, vocabulary, and writing coherence among ESL learners (Bai & Hu, 2016; Wang et al., 2012). Early critiques warned of reduced human interaction, formulaic writing, and shallow learning (Hyland, 1998; Truscott, 1996). More recent studies suggest that AWET effectiveness is influenced by learner proficiency: lower-proficiency students often benefit from basic feedback but may struggle to interpret it, while higher-proficiency learners engage more critically and strategically (Koltovskaia, 2020; Zhang & Hyland, 2018). Teachers commonly use AWETs for formative assessment and instructional support (Chen & Cheng, 2008), and students generally report increased confidence and improved revision practices (Barrot, 2021). Nonetheless, concerns remain about feedback accuracy and the

potential for overreliance on automation, especially for advanced writing development (Xu & Zhang, 2022).

2.6 Generative AI tools and L2 writing

While much research on teacher and student perceptions has focused on established tools such as MTTs, APTs, AWETs, and corpus tools, recent studies increasingly examine views of Generative Artificial Intelligence (Generative AI) in writing. Generative AI refers to machine learning models that create new content (text, images, audio) from patterns in large datasets. Since tools like ChatGPT became publicly available in late 2022, they have rapidly advanced and are now widely used for answering questions, summarising texts, generating essays, and translating language (Dwivedi et al., 2023; Kasneci et al., 2023; Ray, 2023). In language learning, Generative AI tools such as ChatGPT, Bing Chat, and Google Gemini can provide immediate, personalised feedback on grammar, coherence, and content, helping learners overcome linguistic challenges and improve fluency (Duong & Chen, 2025; Seo, 2024). At the same time, they raise concerns about how to distinguish human from AI-generated text and have revived debates over human authorship in academic writing (Deng et al., 2025; Dwivedi et al., 2023). Although Generative AI systems such as Gemini and ChatGPT are highly relevant to broader debates on digital writing support and are rapidly reshaping how learners and teachers approach academic writing, they do not form the primary focus of the present study. Instead, this thesis concentrates on more established, non-generative tools that were already widely used and institutionally accepted at the time of data collection. The reasons for not including Generative AI tools within the core analysis, are discussed in greater detail in the section on research gaps (Chapter 2, Section 2.7.3).

2.6.1 Perceptions of teachers and students towards Generative AI

Teachers hold mixed views on integrating Generative AI systems, such as Gemini or ChatGPT into education. They acknowledge its value for lesson planning, instructional materials, prompts, differentiated support, and modelling writing (Jeon & Lee, 2021), but express concerns about accuracy, misuse, and lack of pedagogical guidance (Yang & Chen, 2023), as well as its role in assessment and feedback (ElSayary, 2023; Deng et al., 2025). A recurring issue is plagiarism and the difficulty of verifying

authentic authorship (Cotton et al., 2023; Derakhshan & Ghiasvand, 2024; Dimitrov, 2023; Mabuan, 2024; Rahma & Fithriani, 2024; Yang & Chen, 2023). Others warn that GenAI may undermine critical thinking, reduce motivation for original work, and encourage surface-level learning (Cotton et al., 2023; Jeon & Lee, 2021; Kim & McGill, 2024; Perkins, 2023; Rukiati et al., 2023).

By contrast, students generally report more positive perceptions. While teachers worry about reduced engagement, students see AI as a support for critical thinking, research, and conceptual evaluation (Darwin et al., 2024; Kohnke et al., 2023). Teachers prefer to use AI for idea generation and routine tasks (Adams et al., 2023; Barrett & Pack, 2023; Dai et al., 2023; Yang & Chen, 2023). Many also view tools like ChatGPT as helpful in reducing grammatical and lexical errors (Duah & McGivern, 2024; Masoudi, 2024; Mun, 2024).

A recent study by Duong and Chen (2025) observed that lower-proficiency English learners primarily used the chatbot during the Planning stage of writing, with somewhat less activity during the Translating stage. In contrast, higher-proficiency students exhibited a more balanced usage pattern, engaging with ChatGPT across all stages, with the greatest emphasis on the Translating stage. These findings suggest that students' engagement with ChatGPT may have a connection with their proficiency levels.

In summary, teachers hold mixed views on Generative AI, appreciating its instructional benefits while raising concerns about accuracy, plagiarism, and its impact on critical thinking and assessment. In contrast, students generally view these tools positively, using them for idea generation, support, and error correction. Although usage appears to vary by proficiency level, little research has examined how proficiency shapes students' perceptions, revealing a key gap in the literature.

2.7. Research gaps and research questions

2.7.1 Research gaps related to the four tools' impact on IELTS writing scores and language development

Although prior studies have shown that the four types of digital writing tools—corpus tools, machine translation tools (MTTs), automated paraphrasing tools (APTs), and

automated writing evaluation tools (AWETs)—can enhance specific aspects of L2 writing such as grammatical accuracy and lexical diversity (e.g., Lee, 2020; Kol et al., 2018; Tsai, 2019; Wang et al., 2012), few studies have examined their direct impact on criteria of an English standardised tests like the IELTS. In particular, no study to date has directly measured which components of the IELTS scoring rubric (i.e., Task Achievement, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy) are most affected by the combined use of the four types of tool. Therefore, RQ1a addresses a critical gap by asking: “Does the use of the writing tools result in an increase in the overall writing scores of students? If so, what aspects of IELTS writing assessment are most affected by students’ use of the tools?”

Another underexplored area in existing research is the relationship between students’ English proficiency and their engagement with digital writing tools. Although several studies have found that higher-proficiency learners tend to use these tools more strategically and critically (Ahn & Chung, 2020; Chung & Ahn, 2021; Crosthwaite et al., 2019; Fauziah et al., 2024; Kim et al., 2024; Lee & Briggs, 2021; Lin, 2021; Pérez-Paredes et al., 2012; Zhao, 2025), no study to date has directly compared how high- and low-scoring students differ in their behavioural and cognitive engagement with these tools during and after their writing tasks. This gap leads to RQ1b, which examines how students with different writing score results (after tool usage) vary in their ability to engage with tool outputs, encompassing both behavioural and cognitive dimensions, and use them to support their language development.

Additionally, while several studies have noted potential negative effects of digital writing tool use—such as overreliance, superficial revision, and increased cognitive load among less proficient learners (Gao, 2021; Kim et al., 2024; Shang, 2019; Yousofi, 2022; Zhang & Hyland, 2018; Zhao, 2025)—there remains limited empirical evidence explaining the specific factors that may contribute to declines in writing performance. To address this gap, RQ1c asks: “What factors contribute to any observed decrease in students’ writing scores when using these tools?”

Taken together, these gaps establish the basis for the main Research Question 1 (RQ1), which investigates the impact of the four digital writing tools on students’ IELTS essay writing scores. The corresponding sub-questions (RQ1a, RQ1b, and RQ1c) are designed to offer more detailed insights into score outcomes, patterns of learner

engagement, and factors that may contribute to potential negative effects, as detailed below:

RQ1: What is the impact of the four writing tools on students' IELTS essay writing scores?

RQ1a: Does the use of the writing tools result in an increase in the overall writing scores of students? If so, what aspects of IELTS writing assessment (i.e., Task Achievement, Coherence and Cohesion, Lexical Resource, Grammatical Range and Accuracy) are most affected by students' use of the tools?

RQ1b: How do high-scoring and low-scoring students differ in their ability to engage with the tools and use them to support their language development?

RQ1c: What factors contribute to any observed decrease in students' writing scores when using these tools?

2.7.2 Research gaps related to tool usage strategy alignment between teachers and students

Although numerous studies have examined the use of digital writing tools in second language (L2) instruction, limited research has directly compared teachers' recommended strategies with students' actual usage patterns in test preparation contexts such as IELTS. Existing research shows that teachers often adopt structured, pedagogical approaches—for example, using corpus tools during planning and revision (Bunting, 2013; Ma et al., 2022a), applying machine translation tools to raise error awareness (Niño, 2004; Shei, 2002), or encouraging critical analysis of paraphrasing outputs (Rogerson & McCarthy, 2017)—though these practices are usually studied in general academic or grammar instruction rather than IELTS-specific settings. In addition, most studies examine individual tools in isolation, rather than the combined use of all four tool types. These gaps motivate Research Question 2a (RQ2a): “What specific strategies do teachers recommend for using each of the four writing tools in IELTS essay preparation?”

In contrast, students tend to use tools in a task-oriented manner. Corpus tools are commonly consulted during composing and revision to address lexical and

grammatical concerns (Bridle, 2019; Conroy, 2010; Lai & Chen, 2015; Yoon, 2008, 2016), while machine translation tools (MTTs) are widely used—especially during composing—for vocabulary support and across all writing stages for grammar and clarity, often supplemented by dictionaries or tools like Grammarly and QuillBot (Briggs, 2018; Chandra & Yuyun, 2018; Clifford et al., 2013; Dahmash, 2020; Farzi, 2016; Fountain & Fountain, 2009; Larson-Guenette, 2013; Lewis-Kraus, 2016; Luton, 2003; Musk, 2022; Pham et al., 2022; Steding, 2009; Tsai, 2019; Williams, 2006). Research on AWETs suggests that higher-proficiency learners engage more critically—verifying and revising—while lower-proficiency learners rely more passively on surface-level edits; yet, usage patterns across proficiency levels and widely used tools like Grammarly remain underexplored (Chapelle et al., 2015; Chodorow et al., 2010; Koltovskaia, 2020; Koh, 2017; Lavolette et al., 2015; McCarthy et al., 2022; Roscoe et al., 2017; Warschauer & Ware, 2006; Zhang, 2016). Despite these findings, few studies have systematically examined students’ tool strategies in IELTS preparation, prompting Research Question 2b (RQ2b): “What strategies do students employ when using the four writing tools?”

More importantly, no prior research has empirically investigated whether students’ tool use aligns with teachers’ recommended strategies, and how such alignment may reflect in students’ writing performance. Although some studies hint at mismatches between teacher guidance and learner behaviour (Bridle, 2019; Ma et al., 2022a), the implications for learning outcomes remain unclear. Therefore, RQ2c aims to address this gap: “Did students whose tool usage aligned with teachers’ strategies achieve higher Essay 2 scores compared to those who used different approaches?”

Taken together, these gaps establish the basis for the main Research Question 2 (RQ2), which investigates the extent to which teacher-recommended strategies align with students’ actual use of four digital writing tools in IELTS essay preparation. The corresponding sub-questions (RQ2a, RQ2b, and RQ2c) are designed to provide deeper insights into instructional practices, student strategy use, and how these alignments or divergences relate to writing performance, as detailed below:

RQ2: What similarities and differences exist between teachers’ recommended strategies for using the four writing tools in IELTS essay preparation and

students' actual usage patterns, and how are these similarities and differences reflected in students' essay writing performance?

RQ2a: What specific strategies do teachers recommend for using each of the four writing tools in IELTS essay preparation?

RQ2b: What strategies do students employ when using the four writing tools?

RQ2c: Did students whose tool usage shared similarities with teachers' strategies achieve higher Essay 2 scores compared to those whose tool usage differed?

2.7.3 Research gaps related to perceptions of tools between teachers and students

While a substantial body of literature has explored perceptions of digital writing tools, most studies have examined either teachers' or students' views in isolation, without offering a comparative analysis across all four tool types. For example, teachers have expressed both positive feedback and concerns regarding corpus tools and machine translation tools, citing their pedagogical potential alongside challenges such as data complexity, instructional control, and curriculum alignment (Breyer, 2009; Lin, 2019; Ma et al., 2022a). Students, in contrast, often report positive attitudes toward tools like Quillbot and Google Translate, particularly for vocabulary retrieval and writing correction (Alm & Watanabe, 2021; Chandra & Yuyun, 2018), while also acknowledging issues like mistranslations, incoherent paraphrasing, and ambiguous feedback (Nguyen, 2022; Prentice & Kinden, 2018). Despite these insights, few studies have systematically compared both groups' perceptions across all four tool types within the same writing context.

More importantly, these perceptions reflect an often overlooked yet significant dimension of emotional engagement—which includes learners' affective responses such as interest, confidence, anxiety, or frustration when interacting with digital tools (Fredricks et al., 2004; Hiver et al., 2021a). A growing body of research highlights the strong link between emotional engagement and academic outcomes, such as persistence, motivation, and performance (Chan et al., 2024; Christenson et al., 2012;

Lei et al., 2018; Luo et al., 2023; Schrader & Kalyuga, 2020; Skinner et al., 2009; Wu et al., 2020). Building on this, the current study treats emotional engagement as central to understanding how learners respond to and evaluate the tools they use. RQ3a (“What are the strengths and limitations of each tool, as evaluated by teachers and students?”) is designed to explore how emotional engagement is reflected in both groups’ perceptions of the four tools under study.

Although students’ positive perceptions of tools are well documented, there is little empirical evidence linking these attitudes to actual writing performance. While many studies report that learners feel more confident and satisfied when using tools (Alhaisoni & Alhaysony, 2017; Alm & Watanabe, 2021; Chung & Ahn, 2022; Conroy, 2010; Garcia & Pena, 2011; Li et al., 2015; Nguyen, 2022; Palermo & Wilson, 2020; Truong, 2023), few have tested whether these perceptions align with measurable improvements in writing outcomes. RQ3b addresses this gap by asking: “How do students’ evaluations relate to their writing performance?”. Furthermore, while both teachers and students occasionally suggest changes to improve tool usability, most studies report these suggestions in scattered, tool-specific terms. Students frequently express a desire for clearer feedback and simpler interfaces (Charles, 2014; Mueller & Jacobsen, 2015), whereas teachers often call for better integration with pedagogical goals and greater instructional control (Breyer, 2009; Stapleton & Kin, 2019). However, a broader, comparative understanding of what both groups expect from digital writing tools remains missing. RQ3c responds to this by asking: “What additional features do teachers and students expect beyond the tools’ current capabilities?”.

Although Generative AI (GenAI) systems such as Gemini and ChatGPT are relevant to broader discussions of digital writing support, they are not the focus of the present study for three main reasons. First, as this thesis was conducted in 2023, the very recent emergence of GenAI means that its algorithms and empirical evidence on their use are still rapidly evolving, making its integration into the current research design difficult. Second, the present study is explicitly concerned with four established, non-generative tools—Sketch Engine, Google Translate, QuillBot, and Grammarly—that were already widely accessible and institutionally accepted at the time of data

collection (*¹). Additionally, because each tool in this study is built around one separate and clearly defined function—collocation and concordance searching, translating, paraphrasing, or error correction—it enables ESL/EFL students to engage purposefully with each type of support during writing and to reflect on how they used these functions afterwards. This clear functional separation also allows researchers and teachers to trace the specific contribution of each tool to learners' language development. If, as seems likely, tools such as Quillbot and Grammarly increasingly merge these functions into seamless GenAI-driven suites, this kind of unique pedagogical and research insight may become much harder to achieve, which can be a loss for English language teaching, where understanding the process of how ESL learners use particular forms of support is as important as the quality of the final product.

Nonetheless, the growing influence of Generative AI is still acknowledged in this thesis as an important avenue for future research in L2 academic writing. Therefore, even though this study focuses on four distinctive tools, the findings have broader implications for how we understand the evolving relationship between user perceptions, engagement, and tool effectiveness. Insights from RQ3 and its sub-questions may help clarify not only how current tools are used and evaluated, but also what users are increasingly expecting from new forms of digital writing support, such as Generative AI.

Taken together, these research gaps establish the basis for the main Research Question 3 (RQ3): "What additional features do teachers and students expect beyond the tools' current capabilities?", which investigates how teachers and students perceive the effectiveness and limitations of four digital writing tools—corpus tools, machine translation tools (MTTs), automated paraphrasing tools (APTs), and automated writing evaluation tools (AWETs). The corresponding sub-questions (RQ3a, RQ3b, and RQ3c) are designed to provide more detailed insights into how student evaluations relate to writing performance, the perceived strengths and weaknesses of each tool, and users' expectations for additional features beyond current capabilities, as detailed below.

¹ This may be changing as Generative AI technologies are being integrated into tools such as Grammarly and Quillbot.

RQ3: How do teachers and students perceive the effectiveness and limitations of the four writing tools

RQ3a: What are the strengths and limitations of each tool, as evaluated by teachers and students?

RQ3b: How do students' evaluations relate to their writing performance?

RQ3c: What additional features do teachers and students expect beyond the tools' current capabilities?

2.7.4 Research gaps in methodology and context

Several methodological gaps have been identified in the literature on corpus tools, machine translation tools (MTTs), automated paraphrasing tools (APTs), and automated writing evaluation tools (AWETs). Existing research tends to examine only one tool at a time, and few studies have explored how students interact with multiple tools during a time-restricted writing task. Although think-aloud protocols and interviews are occasionally used (e.g. Kim et al., 2024; Yoon, 2016), direct observation of learners' engagement during writing tasks is also uncommon and few studies triangulate data sources—such as tool usage logs, writing scores, and learner reflections—to provide a comprehensive understanding of how students navigate tool use in authentic writing contexts. Additionally, research examining the accuracy and effectiveness of MTTs for the Vietnamese-English language pair remains limited, as well as the overall English language teaching context in Vietnam. Together, these research gaps highlight the need for context-sensitive, methodologically robust investigations focusing on Vietnamese learners of English, with particular attention to how behavioural, cognitive, and emotional engagement with the four tool types relates to English proficiency level, writing performance outcomes, and language development.

2.8 Chapter summary

This chapter reviews key theoretical literature on second language acquisition and empirical studies on four digital writing tools commonly used in writing: corpus tools, machine translation tools (MTTs), automated paraphrasing tools (APTs), and automated writing evaluation tools (AWETs).

The theoretical literature on second language acquisition provides the theoretical foundation for this study. The Output Hypothesis (Swain, 1995) and the Noticing Hypothesis (Schmidt, 1990, 2001) explain the cognitive engagement that underpins language learning through active language production (writing) and attention to linguistic forms. In parallel, the multidimensional concept of learner engagement suggests that interactions with digital writing tools involve not only cognitive engagement—including noticing language gaps and reflecting on form-meaning relationships—but also behavioural engagement (e.g., frequency and purpose of tool use) and emotional engagement (e.g., confidence, motivation, or frustration) (Fredricks et al., 2004; Hiver et al., 2021a; Svalberg, 2009). Collectively, these theoretical frameworks provide an integrated lens for analysing how learners engage with digital tools during L2 writing tasks, and whether such engagement leads to language development. Moreover, theoretical literature on learner engagement highlights the absence of a standardised framework for defining and measuring its dimensions, and calls for more context-sensitive research grounded in actual learner behaviours and observations rather than self-report data (Fredricks et al., 2004; Hiver et al., 2021a; Svalberg, 2017).

Empirical studies reveal several insights on the impact of the four writing tools under study on students' writing outcomes and language improvements, yet they also present specific shortcomings. For instance, corpus tools enhance learners' lexical, grammatical, and collocational awareness (Gaskell & Cobb, 2004; Huang, 2014; Hong, 2010; Lin, 2021), but learners might not use them effectively due to data complexity (Crosthwaite et al., 2019; Fang et al., 2021; Pérez-Paredes et al., 2012). Machine translation tools (MTTs), such as Google Translate, can help vocabulary acquisition, syntactic fluency, and text expansion; however, their utility is limited by inconsistent translation accuracy across language pairs (Groves & Mundt, 2015; Nguyen & Ngo, 2021) and concerns about overreliance and academic integrity

(Correa, 2011; Jones, 2009; Prentice & Kinden, 2018; Roe & Perkins, 2022). Automated paraphrasing tools (APTs) like QuillBot can improve vocabulary diversity through rephrasing functions and synonyms; however, they can produce superficial rewrites (Prentice & Kinden, 2018; Rogerson & McCarthy, 2017). Automated writing evaluation tools (AWETs), such as Grammarly, offer feedback on grammar, spelling, and style, promoting revision and autonomy. Nonetheless, their feedback is often surface-level and algorithm-driven, with limited capacity to support higher-order writing skills unless accompanied by strategic engagement (Ranalli et al., 2021). Across all four tool types, learner proficiency remains a critical moderating factor of the tool's effectiveness, as higher-proficiency students tend to engage more strategically and critically, reflecting deeper behavioural and cognitive engagement in the writing process (Bailey & Withers, 2018; Kim et al., 2024; Nguyen, 2022).

Empirical research also reveals distinct differences in how teachers and students use and perceive digital writing tools. In terms of tool usage, teachers typically incorporate these tools during the planning and revision stages, using corpus tools to design instructional materials or guide grammar-focused feedback (Bunting, 2013; Ma et al., 2022a; Quinn, 2015; Todd, 2001). In contrast, students predominantly engage with tools during composing and editing, often relying on MTTs for translation, APTs for rephrasing, and AWETs for surface-level error correction (Bridle, 2019; Conroy, 2010; Lai & Chen, 2015). In terms of perceptions, teachers generally acknowledge the pedagogical value of digital writing tools, especially for promoting grammar awareness, learner autonomy, and inductive learning (Breyer, 2009; Lin, 2019). However, they also express persistent concerns about data complexity, diminished instructional control, and difficulties integrating these tools into existing curricula (Bunting, 2013; Ma et al., 2022a). Such concerns are particularly pronounced with machine translation and automated paraphrasing tools, which are often viewed as undermining academic integrity and discouraging active language production (Somers et al., 2006; Stapleton & Kin, 2019). In contrast, students tend to report more favourable attitudes, citing the tools' immediacy, vocabulary support, and assistance with error correction as helpful features in their writing processes (Chandra & Yuyun, 2018; Pham et al., 2022). Nonetheless, learners also acknowledge limitations, including mistranslations, overwhelming interfaces, and uncertainties about what constitutes ethical use, particularly when paraphrasing or translating academic texts

(Rogerson & McCarthy, 2017). Moreover, recent studies have increasingly highlighted growing interest among teachers and students in the use of Generative Artificial Intelligence (Generative AI) in writing. Tools such as ChatGPT, Google Gemini, and Bing Chat are being widely adopted in educational settings for their ability to support various stages of the writing process (Dwivedi et al., 2023; Kasneci et al., 2023; Ray, 2023). While these tools retain functions found in earlier technologies—such as grammar correction and vocabulary enhancement—they also introduce more advanced features, including real-time content generation and sophisticated writing feedback, which open up new possibilities for L2 writing instruction (Dwivedi et al., 2023). At the same time, teachers and educators also raise important concerns regarding authorship, academic integrity, and the depth of learners' cognitive engagement (Deng et al., 2025; Seo, 2024).

Lastly, the existing empirical studies reveal several methodological limitations in studies on digital writing tools. These include a scarcity of observational research that captures students' simultaneous use of multiple tools and a lack of data triangulation to validate findings (Conroy, 2010; Ma et al., 2022a; Todd, 2001). Moreover, there is a notable absence of research situated in Vietnamese-English learning contexts, which limits the generalisability of current findings (Nguyen, 2022; Pham et al., 2022). These gaps underscore the need for context-sensitive, methodologically rigorous research that links students' engagement with the four types of writing tool under study, especially to their language proficiency, writing performance after tool usage, and broader language development. In response to these theoretical and empirical gaps, the present study examines the impact of four digital writing tools on IELTS writing outcomes, their relationship to learner proficiency and engagement, the alignment between teacher-recommended strategies and students' actual tool use, and both students' and teachers' perceptions. The following chapter outlines the research methodology, including the study's design, participant selection, data collection instruments, and analytic procedures.

CHAPTER 3: METHODOLOGY

Overview of Chapter 3 sections

Chapter 3 outlines the mixed-methods design used to investigate the research questions. Section 3.1 describes the IELTS Academic Writing Task 2 test format and the writing rubric (Task Achievement, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy), which are used to assess students' essays. Section 3.2 introduces the 20 Vietnamese university participants and explains how the two essay topics were selected and administered (Essay 1 without tools and Essay 2 with Sketch Engine, Google Translate, QuillBot, and Grammarly). Section 3.3 presents the three raters' profiles and the marking procedures, as well as the inter-rater reliability. Section 3.4 details the quantitative analyses used to compare students' scores for Essay 1 and Essay 2, focusing on changes in overall scores and in each IELTS criterion. Section 3.5 outlines the qualitative methodology to analyse students' tool usage strategy and interview responses. Next, Section 3.6 explains the process of teacher participants' recruitment, interview questions, and lesson plan analysis. Section 3.7 describes the analysis procedures used to compare teachers' and students' tool usage across three writing stages, evaluation and recommendation. Finally, Section 3.8 presents the ethical considerations and overall research design, confirming informed consent and adherence to relevant ethical guidelines.

RESEARCH QUESTIONS

RQ1: What is the impact of the four writing tools on students' IELTS essay writing scores?

RQ1a: Does the use of the writing tools result in an increase in the overall writing scores of students? If so, what aspects of IELTS writing assessment (i.e., Task Achievement, Coherence and Cohesion, Lexical Resource, Grammatical Range and Accuracy) are most affected by students' use of the tools?

RQ1b: How do high-scoring and low-scoring students differ in their ability to engage with the tools and use them to support their language development?

RQ1c: What factors contribute to any observed decrease in students' writing scores when using these tools?

RQ2: What similarities and differences exist between teachers' recommended strategies for using the four writing tools in IELTS essay preparation and students' actual usage patterns, and how are these similarities and differences reflected in students' essay writing performance?

RQ2a: What specific strategies do teachers recommend for using each of the four writing tools in IELTS essay preparation?

RQ2b: What strategies do students employ when using the four writing tools?

RQ2c: Did students whose tool usage shared similarities with teachers' strategies achieve higher Essay 2 scores compared to those whose tool usage differed?

RQ3: How do teachers and students perceive the effectiveness and limitations of the four writing tools

RQ3a: What are the strengths and limitations of each tool, as evaluated by teachers and by students?

RQ3b: How do students' evaluations relate to their writing performance?

RQ3c: What additional features do teachers and students expect beyond the tools' current capabilities?

To address the research questions, the researcher has conducted separate research procedures for students and teachers. The researcher has then analysed the data from each group independently, employing both quantitative and qualitative approaches, before comparing and synthesising the findings between the two groups. The two flowcharts below provide an overview of the research and data processing sequences for each participant group: students (Figure 6) and teachers (Figure 7).

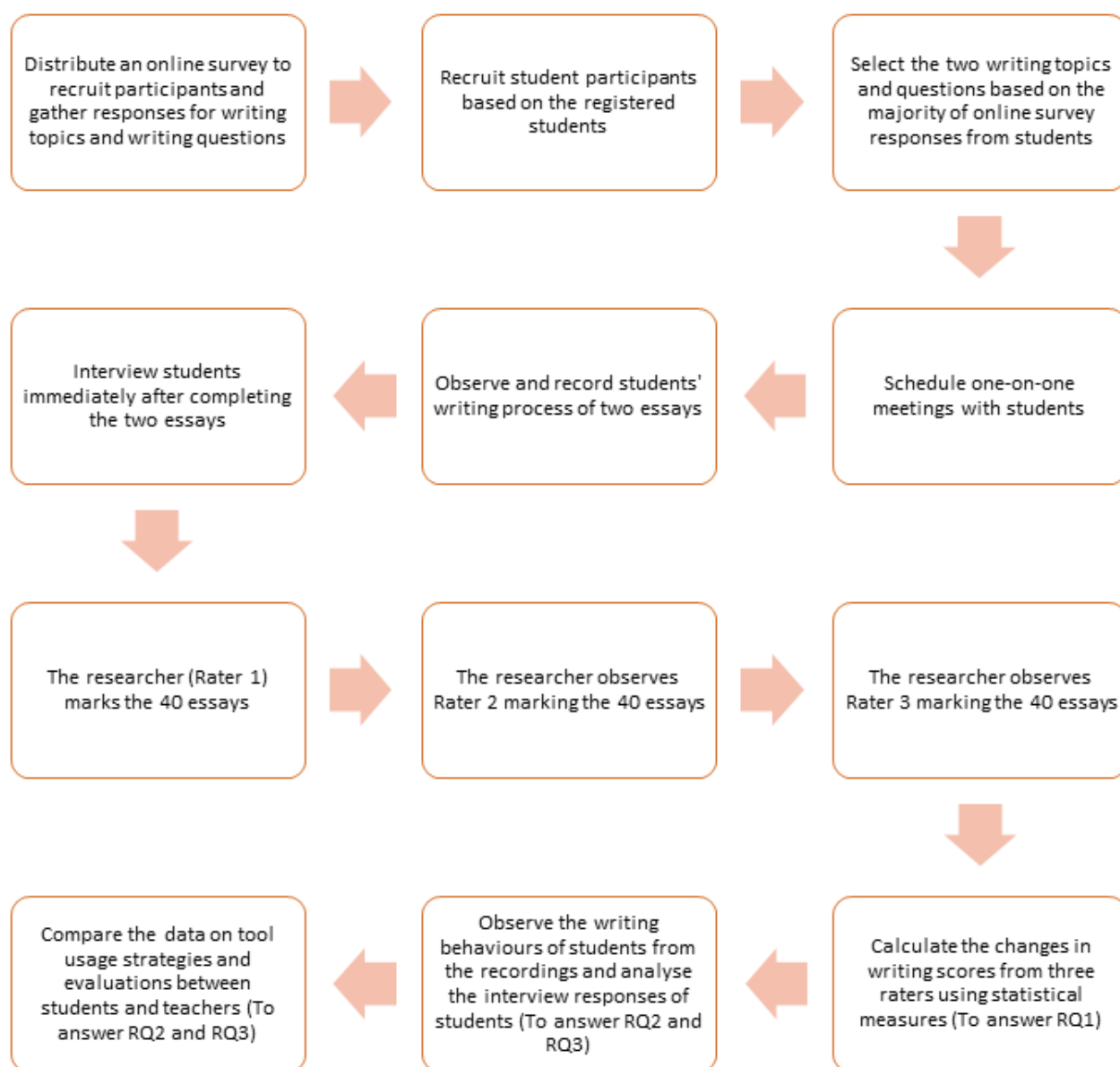


Figure 6. The research procedure and data processing sequence for students

The research procedure for students (Figure 6) involved 12 detailed steps. Initially, the researcher distributed an online survey to several Facebook groups about teaching English and the university email system to identify potential participants who had completed the IELTS test and expressed an interest in using technology tools for essay composition. Once participants were selected, the researcher arranged individual meetings with each participant, observed, and recorded their writing processes in the composition of two essays. Following these sessions, the researcher conducted interviews to gather student perspectives on the tools' usefulness and reflections on the writing tasks. The researcher then marked the 40 essays and

observed the evaluation process of two additional raters (Rater 2 and Rater 3). Using parametric statistical methods (as detailed in Section 4), the researcher analysed score variations among the three raters to address RQ1. Finally, the researcher examined and analysed the recordings of student sessions to address RQ2 and RQ3.

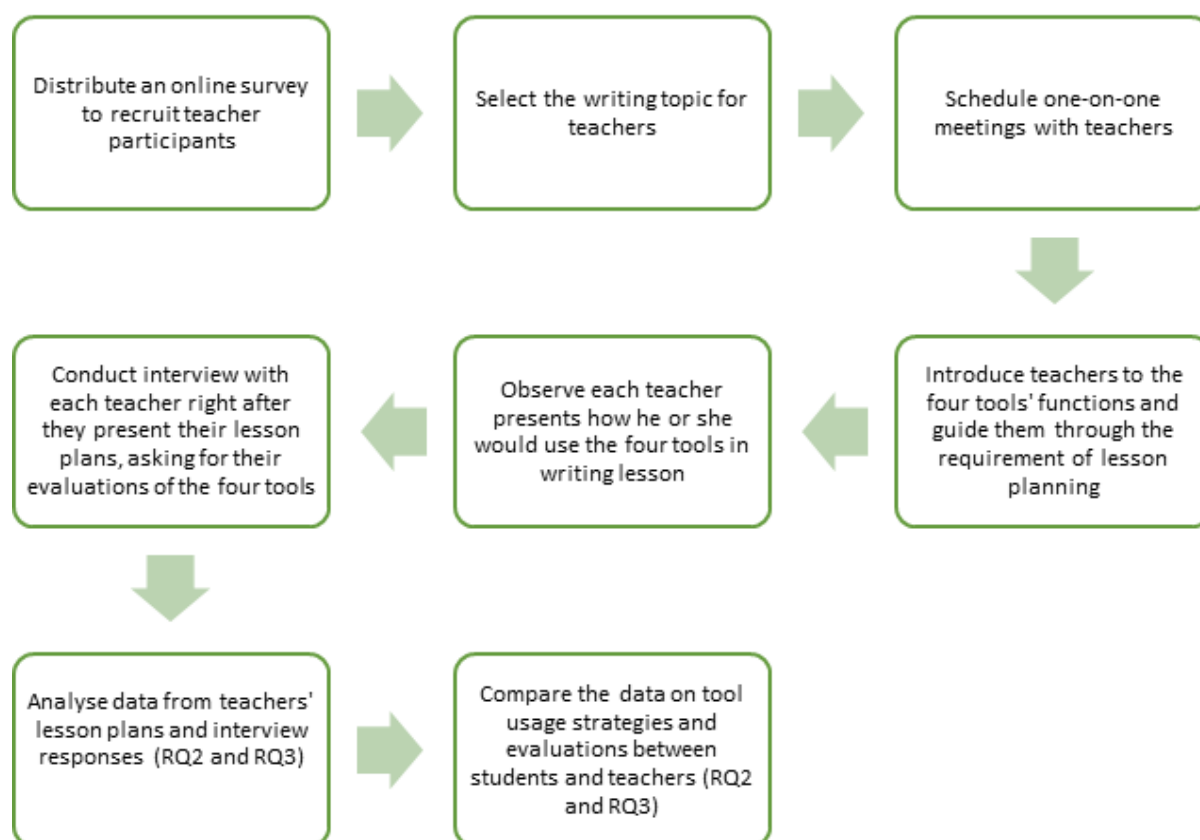


Figure 7. The research procedure and data processing sequence for teachers

The research procedure for teachers (Figure 7) was conducted separately and comprised 9 steps. The researcher started with a similar process of distributing an online survey to find potential teacher participants who expressed interest in using technology tools to teach writing and had experience teaching the IELTS test. Next, the researcher selected the writing topics for teachers and scheduled one-on-one meetings with them. After guiding the teachers in the four tools' functions and lesson planning, the researcher observed how teachers would use the four writing tools in their lessons. Then, the researcher interviewed the teachers regarding their experience and evaluation of the four tools. Finally, the researcher analysed interview

recordings from teachers and compared these data to students' data, to address RQ2 and RQ3.

3.1 Rationale for using the International English Language Testing System (IELTS) in this study

This study uses the International English Language Testing System (IELTS) as the assessment format for evaluating students' writing performance. The details of this standardised English test and the validity of the IELTS writing rubric are presented in this section.

3.1.1 Detail and validity of the IELTS

The International English Language Testing System (IELTS) is a long-established standardised test of English proficiency for non-native speakers launched in 1980 by Cambridge English Language Assessment, and run by IDP Education and the British Council (IELTS n.d). The first rationale for selecting the IELTS Writing Task 2 format for this study is because of its ubiquity in the field of English language teaching, especially in academic writing and preparing for university admission in Vietnam and around the world. This includes universities in Australia and the UK, over 3,400 institutions in the US, and others in various countries (IELTS Asia, 2019). In 2017, more than three million IELTS tests were taken, reflecting the growing importance of this test (British Council, 2017). The IELTS test has also become increasingly popular in the English language teaching context in Vietnam over the past decade, since its introduction in 1994 (Tran-Thanh, 2025). Although the exact number of IELTS tests administered annually in Vietnam is not publicly disclosed, various indicators suggest that the figure is substantial. For one thing, there are departments of both IDP Education and the British Council in Vietnam. Additionally, the frequency of test dates offered by these providers, with multiple options available each month, points to a consistent year-round demand for IELTS testing in the country. This is further supported by the Vietnamese government's goals to enhance the English language proficiency of its citizens, as more and more high schools and universities in Vietnam require students to provide an IELTS certificate (Nguyen Hoai, 2024).

The IELTS test includes four skills: reading, listening, speaking, and writing. The IELTS writing test consists of two writing tasks and has a duration of 60 minutes

(IELTS, n.d). Writing Task 1 requires the candidate to describe, summarise, or explain some visual information, such as graphs, charts, or tables, in their own words. Writing Task 2 involves writing a personal argumentative essay in response to a point of view, discussing both sides of a statement, or pointing out causes and providing solutions. The topics can be quite diverse and cover a range of fields and subjects. Some common topic areas that may appear are education, technology, environment, health, language, business, tourism, music, arts, transportation, and crime (IELTS Writing Info, n.d).

The question types of the IELTS Writing Task 2 tend to include opinion essays, where test-takers express agreement or disagreement with a given statement, and discussion essays, which involve exploring both sides of an issue and providing a personal viewpoint. Other common formats are advantage/disadvantage essays, which assess the pros and cons of a topic, and cause/solution essays, where candidates identify a problem, analyse its causes, and suggest possible solutions (IDP IELTS, 2024; IELTS Advantage, 2024; IELTS, 2024; IELTS Liz, 2024) . Since the IELTS test is designed to assess language proficiency, the essay questions are intended to allow students to demonstrate their English language skills and general knowledge through free responses. As a result, there is no requirement for academic citations on any topic. To answer the essay question, IELTS test-takers would have to rely on their personal opinions and personal knowledge.

The second rationale for choosing this test format for the study is the transparency, thorough revision, and public accessibility of the IELTS Test marking rubric. Figure 8 shows the old version of the writing rubric, which used to be published on the website of the British Council.

WRITING TASK 2: Band Descriptors (public version)

Band	Task response	Coherence and cohesion	Lexical resource	Grammatical range and accuracy
9	- fully addresses all parts of the task - presents a fully developed position in answer to the question with relevant, fully extended and well supported ideas	- uses cohesion in such a way that it attracts no attention - skilfully manages paragraphing	uses a wide range of vocabulary with very natural and sophisticated control of lexical features; rare minor errors occur only as 'slips'	uses a wide range of structures with full flexibility and accuracy; rare minor errors occur only as 'slips'
8	- sufficiently addresses all parts of the task - presents a well-developed response to the question with relevant, extended and supported ideas	- sequences information and ideas logically - manages all aspects of cohesion well - uses paragraphing sufficiently and appropriately	- uses a wide range of vocabulary fluently and flexibly to convey precise meanings - skilfully uses uncommon lexical items but there may be occasional inaccuracies in word choice and collocation - produces rare errors in spelling and/or word formation	- uses a wide range of structures - the majority of sentences are error-free - makes only very occasional errors or inappropriacies
7	- addresses all parts of the task - presents a clear position throughout the response - presents, extends and supports main ideas, but there may be a tendency to over-generalise and/or supporting ideas may lack focus	- logically organises information and ideas; there is clear progression throughout - uses a range of cohesive devices appropriately although there may be some under-use - presents a clear central topic within each paragraph	- uses a sufficient range of vocabulary to allow some flexibility and precision - uses less common lexical items with some awareness of style and collocation - may produce occasional errors in word choice, spelling and/or word formation	- uses a variety of complex structures - produces frequent error-free sentences - has good control of grammar and punctuation but may make a few errors
6	- addresses all parts of the task although some parts may be more fully covered than others - presents a relevant position although the conclusions may become unclear or repetitive - presents relevant main ideas but some may be inadequately developed/unclear	- arranges information and ideas coherently and there is a clear overall progression - uses cohesive devices effectively, but cohesion within and/or between sentences may be faulty or mechanical - may not always use referencing clearly or appropriately - uses paragraphing, but not always logically	- uses an adequate range of vocabulary for the task - attempts to use less common vocabulary but with some inaccuracy - makes some errors in spelling and/or word formation, but they do not impede communication	- uses a mix of simple and complex sentence forms - makes some errors in grammar and punctuation but they rarely reduce communication
5	- addresses the task only partially; the format may be inappropriate in places - expresses a position but the development is not always clear and there may be no conclusions drawn - presents some main ideas but these are limited and not sufficiently developed; there may be irrelevant detail	- presents information with some organisation but there may be a lack of overall progression - makes inadequate, inaccurate or over-use of cohesive devices - may be repetitive because of lack of referencing and substitution - may not write in paragraphs, or paragraphing may be inadequate	- uses a limited range of vocabulary, but this is minimally adequate for the task - may make noticeable errors in spelling and/or word formation that may cause some difficulty for the reader	- uses only a limited range of structures - attempts complex sentences but these tend to be less accurate than simple sentences - may make frequent grammatical errors and punctuation may be faulty; errors can cause some difficulty for the reader
4	- responds to the task only in a minimal way or the answer is tangential; the format may be inappropriate - presents a position but this is unclear - presents some main ideas but these are difficult to identify and may be repetitive, irrelevant or not well supported	- presents information and ideas but these are not arranged coherently and there is no clear progression in the response - uses some basic cohesive devices but these may be inaccurate or repetitive - may not write in paragraphs or their use may be confusing	- uses only basic vocabulary which may be used repetitively or which may be inappropriate for the task - has limited control of word formation and/or spelling; errors may cause strain for the reader	- uses only a very limited range of structures with only rare use of subordinate clauses - some structures are accurate but errors predominate, and punctuation is often faulty
3	- does not adequately address any part of the task - does not express a clear position - presents few ideas, which are largely undeveloped or irrelevant	- does not organise ideas logically - may use a very limited range of cohesive devices, and those used may not indicate a logical relationship between ideas	- uses only a very limited range of words and expressions with very limited control of word formation and/or spelling - errors may severely distort the message	attempts sentence forms but errors in grammar and punctuation predominate and distort the meaning
2	- barely responds to the task - does not express a position - may attempt to present one or two ideas but there is no development	has very little control of organisational features	uses an extremely limited range of vocabulary; essentially no control of word formation and/or spelling	cannot use sentence forms except in memorised phrases
1	answer is completely unrelated to the task	fails to communicate any message	can only use a few isolated words	cannot use sentence forms at all
0	- does not attend - does not attempt the task in any way - writes a totally memorised response			

Figure 8. The old version of the public scoring criteria for Academic IELTS Writing Task 2 by the British Council, retrieved from Uysal (2009).

In 2023, a study was conducted by the IELTS Research Group to review the old criteria of the IELTS Writing test. The study involved analysing feedback from over 400 examiners across 17 countries and identifying areas for clarity enhancement in scoring scales and descriptors (Clark et al., 2023), and to form the new IELTS rubric in May 2023, as presented in Figure 9.

IELTS Writing Task 2 Band Descriptors

Scoring criteria for Academic and General Training tests

Updated May 2023

Please visit ielts.org for updates

Page 3

A script must fully fit the positive features of the descriptor at a particular level. **Bolded text** indicates negative features that will limit a rating.

Band Score	Task Response	Coherence & Cohesion	Lexical Resource	Grammatical Range & Accuracy
4	The prompt is tackled in a minimal way, or the answer is tangential, possibly due to some misunderstanding of the prompt. The format may be inappropriate. A position is discernible, but the reader has to read carefully to find it. Main ideas are difficult to identify and such ideas that are identifiable may lack relevance, clarity and/or support. Large parts of the response may be repetitive.	Information and ideas are evident but not arranged coherently and there is no clear progression within the response. Relationships between ideas can be unclear and/or inadequately marked. There is some use of basic cohesive devices, which may be inaccurate or repetitive. There is inaccurate use or a lack of substitution or referencing. There may be no paragraphing and/or no clear main topic within paragraphs.	The resource is limited and inadequate for or unrelated to the task. Vocabulary is basic and may be used repetitively. There may be inappropriate use of lexical chunks (e.g. memorised phrases, formulaic language and/or language from the input material). Inappropriate word choice and/or errors in word formation and/or in spelling may impede meaning.	A very limited range of structures is used. Subordinate clauses are rare and simple sentences predominate. Some structures are produced accurately but grammatical errors are frequent and may impede meaning. Punctuation is often faulty or inadequate.
3	No part of the prompt is adequately addressed, or the prompt has been misunderstood. No relevant position can be identified, and/or there is little direct response to the question/s. There are few ideas, and these may be irrelevant or insufficiently developed.	There is no apparent logical organisation. Ideas are discernible but difficult to relate to each other. There is minimal use of sequencers or cohesive devices. Those used do not necessarily indicate a logical relationship between ideas. There is difficulty in identifying referencing. Any attempts at paragraphing are unhelpful.	The resource is inadequate (which may be due to the response being significantly underlength). Possible over-dependence on input material or memorised language. Control of word choice and/or spelling is very limited, and errors predominate. These errors may severely impede meaning.	Sentence forms are attempted, but errors in grammar and punctuation predominate (except in memorised phrases or those taken from the input material). This prevents most meaning from coming through. Length may be insufficient to provide evidence of control of sentence forms.
2	The content is barely related to the prompt. No position can be identified. There may be glimpses of one or two ideas without development. Responses of 20 words or fewer are rated at Band 1.	There is little relevant message, or the entire response may be off-topic. There is little evidence of control of organisational features.	The resource is extremely limited with few recognisable strings, apart from memorised phrases. There is no apparent control of word formation and/or spelling. Responses of 20 words or fewer are rated at Band 1.	There is little or no evidence of sentence forms (except in memorised phrases).
1	The content is wholly unrelated to the prompt. Any copied rubric must be discounted.	Responses of 20 words or fewer are rated at Band 1. The writing fails to communicate any message and appears to be by a virtual non-writer.	Responses of 20 words or fewer are rated at Band 1. No resource is apparent, except for a few isolated words.	Responses of 20 words or fewer are rated at Band 1. No rateable language is evident.
0	Should only be used where a candidate did not attend or attempt the question in any way, used a language other than English throughout, or where there is proof that a candidate's answer has been totally memorised.			

Figure 9. The updated public scoring criteria for Academic IELTS Writing Task 2 by the British Council (2023).

According to the new rubric (Figure 9), the assessment for essay writing (Writing Task 2) is based on four key criteria: Task Response (TR), Coherence and Cohesion (CC), Lexical Resource (LR), and Grammatical Range and Accuracy (GA). Each criterion in the new rubric is still assessed with a level of band score, ranging from 0 to 9 with descriptions of each criterion, as well as the requirements and differences between each band score.

For Task Response, the emphasis in the new IELTS rubric (2023) has shifted toward addressing all parts of the task fully and logically while managing complex ideas effectively. Coherence and Cohesion now focus on logical sequencing, effective paragraphing, and the quality of cohesive devices rather than their quantity. Lexical Resource offers clearer expectations for vocabulary range and appropriateness, including avoiding overused or repetitive expressions. Grammatical Range and Accuracy highlights control over complex structures and the effective use of varied grammatical forms. Overall, these changes also aim to enhance practical usability for examiners by reducing variability in application and incorporating feedback from stakeholders.

The first criterion, Task Response (TR), assesses the extent to which the test-taker can address all aspects of the task, and maintain a clear position of the argument throughout their response. Spelling errors or grammatical errors are usually bypassed when marking this criterion.

The second criterion, Coherence and Cohesion (CC), looks at the organisation of ideas and information, the effective usage of cohesive devices, and overall coherence in the essay. When marking the Coherence and Cohesion (CC) criterion, the rater would assess the student's ability to organise ideas logically, use paragraphing effectively, and maintain a logical flow of ideas in the essay through the use of cohesive devices (linking words and phrases).

In terms of Lexical Resource (LR), according to the rubric, the rater would assess the accuracy in spelling, and the range of vocabulary, particularly in choosing words appropriately and also paraphrasing skills. However, the measurement for assessing vocabulary range and accuracy within one essay is not clearly defined in the rubric, leaving much to the individual rater's training and experience.

Regarding the Grammatical Range and Accuracy (GA) criterion, the writing rubric assesses how students can use a variety of sentence structures with accuracy. Based on the latest IELTS rubric (Figure 9), to achieve a band 6 and above in the Grammatical Range and Accuracy (GA) criterion, students are expected to write complex sentences or compound-complex sentences in their essays. However, the extent to which these structures are measured and evaluated in an essay is not explicitly mentioned in the rubric and depends largely on each rater's opinion and experience. Therefore, the IELTS examiners usually assess the overall range of sentence structures in English, which typically includes four main types: simple, compound, complex, and compound-complex sentences (Celce-Murcia and Larsen-Freeman, 1983, p.19).

The third rationale for selecting this IELTS test lies in its validity for research purposes and the widespread adoption of its analytic scoring rubric for evaluating English writing proficiency in existing studies. This ensures that the assessment of this study aligns with recognised standards and facilitates comparability across research contexts.

In previous decades, the assessment of writing has been viewed as challenging, partly due to the participation of multiple raters, which can involve variability and subjectivity in the evaluation process (Godshalk et al., 1966; Huot, 1990, Kondo-Brown 2002; Schaefer 2008; Fahim & Bijani, 2011). To minimise variability and subjectivity in the assessment of second language (L2) writing, scholars recommend establishing clear and standardised rating scales, such as rubrics (Diederich et al., 1961; Hamp-Lyons 2007; Crossley et al., 2014; Dawson, 2015). In terms of designing rubrics to assess specific types of writing assignments, Huot (1990) outlined three methods to assess writing quality for writing rubrics: primary trait, holistic, and analytic scoring. Primary trait assessment, as described by Lloyd-Jones (1977), involves creating specific criteria for each writing assignment, and it is recommended for small samples due to its time and cost (Huot, 1990). Holistic scoring assessment generally reflects a rater's overall impression of a piece of writing (White 1985; Brown & Hudson 2002). Analytic scoring, developed by Diederich (1974), is based on several identifiable aspects of a text, with raters scoring individual traits. This analytic scoring method has been regarded as the most reliable (Veal & Hudson, 1983, Scherer, 1985; Kuo 2007; Shaw & Weir 2007; Hamp-Lyons 2007; Knoch 2009, 2011; Becker 2011; Coombe et al. 2012).

Some studies were conducted to demonstrate the role of analytic scoring rubrics in assessing writing performance in terms of reliability and accuracy in scoring (Jacobs et al., 1981; Bauer, 1981; Veal & Hudson, 1983; White, 1985; Brossell, 1986; Hamp-Lyons, 1990; Kroll 1998; Bacha, 2001; Brown & Hudson, 2002; Kondo-Brown, 2002; Diab & Balaa 2011; Lu & Zhang, 2013, Ghalib & Hattami, 2015). Jacobs et al. (1981) developed and introduced an early example of an analytic scoring rubric in ESL composition, which comprised five key aspects: content, organization, vocabulary, language use, and mechanics, which was named as “Scoring Profile” (Figure 10).

ESL COMPOSITION PROFILE				
STUDENT		DATE	TOPIC	
SCORE	LEVEL	CRITERIA	COMMENTS	
CONTENT	30-27	EXCELLENT TO VERY GOOD: knowledgeable • substantive • thorough development of thesis • relevant to assigned topic		
	26-22	GOOD TO AVERAGE: some knowledge of subject • adequate range • limited development of thesis • mostly relevant to topic, but lacks detail		
	21-17	FAIR TO POOR: limited knowledge of subject • little substance • inadequate development of topic		
	16-13	VERY POOR: does not show knowledge of subject • non-substantive • not pertinent • OR not enough to evaluate		
ORGANIZATION	20-18	EXCELLENT TO VERY GOOD: fluent expression • ideas clearly stated/ supported • succinct • well-organized • logical sequencing • cohesive		
	17-14	GOOD TO AVERAGE: somewhat choppy • loosely organized but main ideas stand out • limited support • logical but incomplete sequencing		
	13-10	FAIR TO POOR: non-fluent • ideas confused or disconnected • lacks logical sequencing and development		
	9-7	VERY POOR: does not communicate • no organization • OR not enough to evaluate		
VOCABULARY	20-18	EXCELLENT TO VERY GOOD: sophisticated range • effective word/idiom choice and usage • word form mastery • appropriate register		
	17-14	GOOD TO AVERAGE: adequate range • occasional errors of word/idiom form, choice, usage <i>but meaning not obscured</i>		
	13-10	FAIR TO POOR: limited range • frequent errors of word/idiom form, choice, usage • <i>meaning confused or obscured</i>		
	9-7	VERY POOR: essentially translation • little knowledge of English vocabulary, idioms, word form • OR not enough to evaluate		
LANGUAGE USE	25-22	EXCELLENT TO VERY GOOD: effective complex constructions • few errors of agreement, tense, number, word order/function, articles, pronouns, prepositions		
	21-18	GOOD TO AVERAGE: effective but simple constructions • minor problems in complex constructions • several errors of agreement, tense, number, word order/function, articles, pronouns, prepositions <i>but meaning seldom obscured</i>		
	17-11	FAIR TO POOR: major problems in simple/complex constructions • frequent errors of negation, agreement, tense, number, word order/function, articles, pronouns, prepositions and/or fragments, run-ons, deletions • <i>meaning confused or obscured</i>		
	10-5	VERY POOR: virtually no mastery of sentence construction rules • dominated by errors • does not communicate • OR not enough to evaluate		
MECHANICS	5	EXCELLENT TO VERY GOOD: demonstrates mastery of conventions • few errors of spelling, punctuation, capitalization, paragraphing		
	4	GOOD TO AVERAGE: occasional errors of spelling, punctuation, capitalization, paragraphing <i>but meaning not obscured</i>		
	3	FAIR TO POOR: frequent errors of spelling, punctuation, capitalization, paragraphing • poor handwriting • <i>meaning confused or obscured</i>		
	2	VERY POOR: no mastery of conventions • dominated by errors of spelling, punctuation, capitalization, paragraphing • handwriting illegible • OR not enough to evaluate		
TOTAL SCORE		READER	COMMENTS	

Figure 10. Scoring Profile (Jacob et al.,1981)

At present, several writing rubrics from various recognised English proficiency tests, including the IELTS test, follow the format of analytic scoring. They share common criteria, such as task achievement, organisation, vocabulary, grammatical precision, and mechanic error types (i.e. spelling, punctuation, capitalisation) (Shabani & Panahi, 2020). These criteria, though labelled differently across various rubrics, are designed

to assess the same aspects of writing. For instance, what the IELTS writing rubric refers to as "Task Response" is called "Content" in the Cambridge C1 Advanced English (CAE) rubric (as shown in Figure 11) and in the Cambridge C2 Proficiency in English (CPE) rubric (as shown in Figure 12), as well as in the Graduate Record Examinations (GRE).

Writing Assessment subscales for C1 Advanced

As with other Cambridge English exams, candidates' writing for C1 Advanced is assessed in terms of Content, Communicative Achievement, Organisation and Language. The detailed band descriptors are as follows:

C1	Content	Communicative Achievement	Organisation	Language
5	All content is relevant to the task. Target reader is fully informed.	Uses the conventions of the communicative task with sufficient flexibility to communicate complex ideas in an effective way, holding the target reader's attention with ease, fulfilling all communicative purposes.	Text is a well-organised, coherent whole, using a variety of cohesive devices and organisational patterns with flexibility.	Uses a range of vocabulary, including less common lexis, effectively and precisely. Uses a wide range of simple and complex grammatical forms with full control, flexibility and sophistication. Errors, if present, are related to less common words and structures, or occur as slips.
4	<i>Performance shares features of Bands 3 and 5.</i>			
3	Minor irrelevances and/or omissions may be present. Target reader is on the whole informed.	Uses the conventions of the communicative task effectively to hold the target reader's attention and communicate straightforward and complex ideas, as appropriate.	Text is well organised and coherent, using a variety of cohesive devices and organisational patterns to generally good effect.	Uses a range of vocabulary, including less common lexis, appropriately. Uses a range of simple and complex grammatical forms with control and flexibility. Occasional errors may be present but do not impede communication.
2	<i>Performance shares features of Bands 1 and 3.</i>			
1	Irrelevances and misinterpretation of task may be present. Target reader is minimally informed.	Uses the conventions of the communicative task to hold the target reader's attention and communicate straightforward ideas.	Text is generally well organised and coherent, using a variety of linking words and cohesive devices.	Uses a range of everyday vocabulary appropriately, with occasional inappropriate use of less common lexis. Uses a range of simple and some complex grammatical forms with a good degree of control. Errors do not impede communication.
0	Content is totally irrelevant. Target reader is not informed.	<i>Performance below Band 1.</i>		

Figure 11. The Writing Rubric of Cambridge C1 Advanced English (CAE). Retrieved from Cambridge English (2020)

C2	CONTENT	COMMUNICATIVE ACHIEVEMENT	ORGANISATION	LANGUAGE
5	All content is relevant to the task. Target reader is fully informed.	Demonstrates complete command of the conventions of the communicative task. Communicates complex ideas in an effective and convincing way, holding the target reader's attention with ease, fulfilling all communicative purposes.	Text is organised impressively and coherently using a wide range of cohesive devices and organisational patterns with complete flexibility.	Uses a wide range of vocabulary, including less common lexis, with fluency, precision, sophistication and style. Use of grammar is sophisticated, fully controlled and completely natural. Any inaccuracies occur only as slips.
4	Performance shares features of Bands 3 and 5.			
3	Minor irrelevances and/or omissions may be present. Target reader is on the whole informed.	Uses the conventions of the communicative task with sufficient flexibility to communicate complex ideas in an effective way, holding the target reader's attention with ease, fulfilling all communicative purposes.	Text is a well-organised, coherent whole, using a variety of cohesive devices and organisational patterns with flexibility.	Uses a range of vocabulary, including less common lexis, effectively and precisely. Uses a wide range of simple and complex grammatical forms with full control, flexibility and sophistication. Errors, if present, are related to less common words and structures, or occur as slips.
2	Performance shares features of Bands 1 and 3.			
1	Irrelevances and misinterpretation of task may be present. Target reader is minimally informed.	Uses the conventions of the communicative task effectively to hold the target reader's attention and communicate straightforward and complex ideas, as appropriate.	Text is well organised and coherent, using a variety of cohesive devices and organisational patterns to generally good effect.	Uses a range of vocabulary, including less common lexis, appropriately. Uses a range of simple and complex grammatical forms with control and flexibility. Occasional errors may be present but do not impede communication.
0	Content is totally irrelevant. Target reader is not informed.	Performance below Band 1.		

Figure 12. The Writing Rubric of Cambridge C2 Proficiency in English (CPE). Retrieved from Cambridge English (2023)

The description for "Task Response" is similarly referred to as "Relevant and well-elaborated explanations, exemplifications, and/or details" (Band 5) in the rubric of the TOEFL Internet-based test (TOEFL iBT), as shown in Figure 13.

SCORE	DESCRIPTION
5	A fully successful response The response is a relevant and very clearly expressed contribution to the online discussion, and it demonstrates consistent facility in the use of language. A typical response displays the following: • Relevant and well-elaborated explanations, exemplifications, and/or details • Effective use of a variety of syntactic structures and precise, idiomatic word choice • Almost no lexical or grammatical errors other than those expected from a competent writer writing under timed conditions (e.g., common typos or common misspellings or substitutions like there/their)
4	A generally successful response The response is a relevant contribution to the online discussion, and facility in the use of language allows the writer's ideas to be easily understood. A typical response displays the following: • Relevant and adequately elaborated explanations, exemplifications, and/or details • A variety of syntactic structures and appropriate word choice • Few lexical or grammatical errors
3	A partially successful response The response is a mostly relevant and mostly understandable contribution to the online discussion, and there is some facility in the use of language. A typical response displays the following: • Elaboration in which part of an explanation, example, or detail may be missing, unclear, or irrelevant • Some variety in syntactic structures and a range of vocabulary • Some noticeable lexical and grammatical errors in sentence structure, word form, or use of idiomatic language
2	A mostly unsuccessful response The response reflects an attempt to contribute to the online discussion, but limitations in the use of language may make ideas hard to follow. A typical response displays the following: • Ideas that may be poorly elaborated or only partially relevant • A limited range of syntactic structures and vocabulary • An accumulation of errors in sentence structure, word forms, or use
1	An unsuccessful response The response reflects an ineffective attempt to contribute to the online discussion, and limitations in the use of language may prevent the expression of ideas. A typical response may display the following: • Words and phrases that indicate an attempt to address the task but with few or no coherent ideas • Severely limited range of syntactic structures and vocabulary • Serious and frequent errors in the use of language • Minimal original language; any coherent language is mostly borrowed from the stimulus
0	The response is blank, rejects the topic, is not in English, is entirely copied from the prompt, is entirely unconnected to the prompt, or consists of arbitrary keystrokes.

Figure 13. Rubric for Writing an academic discussion in TOEFL iBT. Retrieved from Educational Testing Service (n.d).

Similarly, the assessment of text organisation, vocabulary and grammar (syntactic structures) are all included in the five rubrics of the abovementioned tests. This indicates that the IELTS writing rubric follows the general analytic scoring method proposed by scholars in recent decades, aligning it with the writing rubrics of several other widely recognised English standardised tests.

The study by Shabani and Panahi (2020) demonstrates that the original IELTS rubric (Figure 9) is akin to other well-known tests. Particularly, Shabani and Panahi examined

the level of agreement among the rubrics used for assessing writing tasks from four internationally recognised tests of English language proficiency, and they used IELTS as standard rubrics for the study. The two researchers randomly selected 200 IELTS essays that were taken between 2015 and 2016 and asked one rater (Rater 2) to mark these essays based on the IELTS rubric. After that, they asked three more raters to mark these essays using four other sets of rubrics from tests created by the Educational Testing Service (ETS) and Cambridge English Language Assessment. The assessment and evaluation of writing tasks in these five standard examinations employed analytic rubrics with varying point scales: a nine-point scale for the IELTS Task 2, a five-point scale for the Graduate Record Examinations (GRE) and the Test of English as a Foreign Language internet-based (TOEFL iBT), and a six-point scale for the Cambridge Advanced English (CAE) and Cambridge Proficiency English (CPE). Generally, these rubrics evaluate key elements of essay writing, such as the range and use of vocabulary and grammar, cohesion and organisation, and the use of cohesive devices. To analyse the consistency of test ratings and the agreement among four raters, Cronbach's alpha was initially calculated collectively for the five sets of scores (i.e. IELTS, CAE, CPE, TOEFL iBT, and GRE). With acceptable inter-rater correlation rates (.67 to .80), and generally acceptable intercorrelations between CAE and GRE test ratings and IELTS scores, the study of Shabani and Panahi (2020) indicates that the IELTS writing rubric is comparable with other popular English writing rubrics. Other linguistic scholars have also utilised or adapted the IELTS writing rubric for writing assessment in their studies (Bagheri & Zare, 2009; Bostič Bishop, 2011; Nguyen & Le, 2022; Shabani & Panahi, 2020; Tieu & Baker, 2022), further affirming its validity for research purposes.

In summary, the first rationale for selecting the IELTS Writing Task 2 format for this study is its popularity in the field of English language teaching, especially in academic writing and preparing for university admission in Vietnam and around the world (Barnes, 2010; Tran, 2015; British Council, 2017; IELTS Asia, 2019). The second rationale for choosing this test format for the study is the transparency, thorough revision (Clark et al., 2023), and public accessibility of the IELTS Test marking rubric (British Council, 2023). The third rationale for selecting this IELTS test lies in its validity for research purposes and the widespread adoption of its analytic scoring rubric for evaluating English writing proficiency in existing studies (Abe, 2009; Bagheri & Zare,

2009; Bostič Bishop, 2011; Shabani & Panahi, 2020; Nguyen & Le, 2022; Tieu & Baker, 2022).

3.2. The recruitment of student participants and the writing topic selection

This section presents a pre-testing questionnaire used both to select participants and to shape the design of the main study. In the main study, students are asked to complete Writing Task 2 of IELTS, with and without the use of writing tools. In accordance with the IELTS writing test format, student participants were asked to write two essays with the same length and time constraints: an essay of at least 250 words within a 45-minute time limit. For the first essay (Essay 1), students wrote independently, without the use of support tools, and the scores from this essay served as pre-test results to evaluate their baseline writing abilities. In the second essay (Essay 2), students were allowed to use four tools: a corpus tool (Sketch Engine), a machine translation tool (Google Translate), a paraphrasing tool (Quillbot), and an automated writing evaluation tool (Grammarly). Each participant was provided with two writing questions to choose from: one to be completed without any tools and the other with the support of the four tools. The study involved 20 Vietnamese university students, all of whom use English as a second language (ESL), representing a spectrum of English proficiency levels.

To select student participants for this study, the researcher used the method of stratified sampling (Cohen et al., 2017a). First, an online questionnaire was delivered to approximately 300 Vietnamese students in the university and English language institute where the researcher works. The researcher also shared the survey with a Facebook group called “People of TESOL/TEFL/Linguistics”, which, in 2023, had a total number of over 19,000 members, including ESL teachers and students. The survey had a total of 16 questions. The full online questionnaire is presented in Appendix A in the Appendices.

In Questions 1, 2 and 3, participants must confirm that they have already taken the IELTS test and were required to provide their latest IELTS writing score and the test year. These questions indicate the English proficiency level of students and verify that the students already knew the time restriction and the format of the IELTS test, reducing the training time and facilitating the participant selection process for the main

study. Questions 4 through 9 assess students' awareness and experience with the four types of tools under investigation, utilizing a variety of question formats to gather diverse insights. These include a Likert scale (Question 4), an open-ended question asking, "What are the AI writing tools you have used or heard about?" (Question 5), and multiple-choice questions (Questions 6, 7, 8, and 9). This mixed-methods approach ensures a comprehensive understanding of students' familiarity with the tools and their usage patterns, providing both quantitative data from structured questions and qualitative insights from open-ended responses.

Questions 10 and 11 ask for steps that they generally feel are difficult in the IELTS Task 2 writing, which indicates that they would need additional support from the tools. In Question 12 of the questionnaire, participants were asked: "What topics of the IELTS Task 2 essay do you feel most confident with?" They were instructed to select their familiar topics from a list of 12 themes that frequently appear in the IELTS Writing test. This question aimed to help the researcher identify appropriate topics for the main study. Based on the participants' responses, the two most popular topics selected were "Education" (chosen by 55 out of 60 participants) and "Environment" (chosen by 47 out of 60 participants). This response informed the design of the main study, as Education and Environment questions were chosen as the writing topics for the research session.

In Question 13, participants were asked: "What types of IELTS Task 2 questions do you feel most confident with?" They were instructed to select familiar question types from this list of question types:

- "To what extent do you agree or disagree?"
- "Discuss both sides and give your opinion"
- "What are the advantages and disadvantages?"
- "What are the causes and problems?"
- "What are the causes and solutions?"
- "What are the problems and solutions?"

Based on the responses, the researcher identified the two most popular question types to include in the study. These were “What are the advantages and disadvantages?”, selected by 54 out of 60 participants, and “To what extent do you agree or disagree?”, chosen by 38 out of 60 participants. Accordingly, the two writing questions that the researcher used in the study are presented in Figure 14.

Topic 1

Distance learning is becoming more and more popular. Do the advantages of it outweigh the disadvantages?

Give reasons for your answer and include any relevant examples from your own knowledge or experience.

Write at least 250 words.

Topic 2

Some people think that renewable energy should replace fossil fuels. To what extent do you agree or disagree?

Give reasons for your answer and include any relevant examples from your own knowledge or experience.

Write at least 250 words.

Figure 14. The essay questions in the main study

While these two writing topics reflect the majority preference indicated in the survey, it is important to note that they may not align with the preferences of every participant. Consequently, some students might prefer other topics and find the selected ones less familiar or engaging, potentially impacting their ability to perform at their best. Questions 14, 15, and 16 aim to confirm students' interest in participating in the main study and collect their personal information, such as names and email addresses. This information enables the researcher to contact participants and arrange one-on-one meetings for the subsequent research phase.

A total of 23 students registered for the study via an online questionnaire, reporting their most recent IELTS writing scores in the range of 5.0 to 8.0. Because the study required in-depth, individual data collection (two writing tasks with close observation

and subsequent one-to-one interviews), the researcher set the participant number at 20 to keep the workload feasible while maintaining sufficient variation in proficiency for comparison. To achieve this, stratified sampling was applied. Registered students were first divided into two proficiency strata using Band 6.5 as the threshold because it aligns with the CEFR B2/C1 boundary and is widely used as a benchmark for UK student-visa and university-entry language requirements (CEFR B2 is commonly mapped to IELTS 5.5–6.5) (Council of Europe, 2020; UK Home Office, 2024; British Council, 2021; IELTS Asia, n.d.). The resulting pool comprised 8 higher-proficiency students (Bands 7.0–8.0) and 12 lower-proficiency students (Bands 5.0–6.5), yielding a final sample of 20 participants. This ensured that both proficiency groups were represented in a balanced and analytically useful way, while keeping the sample size manageable for the study’s qualitative components. It should also be noted that some participants had taken IELTS some time ago, meaning their self-reported scores might not accurately reflect current writing ability. To obtain a more reliable and up-to-date proficiency indicator, the researcher therefore used performance on the first in-study essay task (Essay 1) as the baseline measure of participants’ writing proficiency. This provided a consistent, current assessment for subsequent analysis and reduced the risk of bias from outdated test results.

Following recruitment, one-to-one interviews were scheduled for each participant. Students were assigned ID numbers according to the order of interview scheduling. Table 3 presents participant IDs alongside their most recent self-reported IELTS Writing scores and the test year.

Table 3. The participants with their coded names, their latest IELTS writing scores and their latest test year.

Students	Latest IELTS writing score	Test year
Student 1	8.0	2022
Student 2	6.0	2022
Student 3	7.5	2021

Student 4	6.5	2022
Student 5	8.0	2020
Student 6	6.5	2012
Student 7	6.5	2015
Student 8	6.0	2018
Student 9	6.0	2022
Student 10	8.0	2022
Student 11	7.0	2016
Student 12	7.0	2018
Student 13	6.5	2021
Student 14	5.5	2016
Student 15	6.5	2022
Student 16	5.0	2017
Student 17	7.0	2020
Student 18	6.5	2018
Student 19	6.0	2016
Student 20	7.0	2022

The main research sessions were conducted individually and online via Zoom, with participants sharing their screens as they completed two writing tasks. The Zoom's screen-capture feature was utilised to record their writing processes, enabling the

researcher to observe tool usage and analyse the generated outputs. Students were required to compose Essay 1 with no tool usage and were observed directly by the researcher. Before starting the second essay, participants were given a 10-minute introduction to the four selected writing tools, ensuring they understood the four functionalities. Following this, participants had 45 minutes to complete the second task. Afterwards, the researcher conducted a 20-minute interview with each participant, focusing on their experiences with the tools, including perceived strengths and limitations. In the interview, students also self-evaluated both essays, reflecting on any writing improvements and tool assistance they had experienced in the second task.

The 20 essays composed without tools (Essay 1) are presented in Appendix C, and the 20 essays composed with tools (Essay 2) are presented in Appendix D. Details on how students' writing processes were tracked and recorded, along with the interview questions regarding tool usage strategy, writing reflections, and tool evaluations, are described in Section 3.5.

3.3 The profiles of raters and the essay marking procedure

After collecting the 40 essays from 20 student participants, the essays were marked by three raters using the latest IELTS writing rubric (2023): the researcher (Rater 1), a British IELTS teacher-examiner (Rater 2), and an experienced Vietnamese IELTS teacher. This section provides information about the raters and the reasons for their selection. Even when employing well-developed rubrics, there can be variation between markers in their writing assessment. Factors that may influence writing assessment include the raters' professional training experience, familiarity with the rating scale, and the scoring procedure (Cumming, 1990; Lumley, 2002; Erdosy, 2004; Fahim & Bijani, 2011; Humphry & Heldsinger, 2018; Isbell, 2017; Wind, 2020). Therefore, the researcher chose to select English teachers who also have experience in marking IELTS to become the raters for this study.

Three raters, including the researcher, participated in this study to mark a total of 40 essays. The first rater is the researcher, who had 5 years of English teaching experience, obtained a score of 8.5 on the IELTS test and had 3 years of teaching and grading IELTS tests and essays. The second rater is a British native speaker. He has

more than 10 years of experience teaching English to Asian students. He worked for 8 years as an official examiner in the official IELTS test of the British Council, so he has considerable experience and expertise in marking IELTS essays. The third rater, a Vietnamese teacher, has obtained a score of 8.5 on the IELTS test and has over 5 years of experience teaching and marking IELTS essays. The marking procedure is summarised in Table 4.

Table 4. Summary of the marking procedure.

Phase	Marking procedure
Phase 1	The researcher, who is also Rater 1, independently marked 40 essays to gain an understanding of the overall writing quality, noting down comments on a document.
Phase 2	The researcher conducted three sessions one-to-one with Rater 2 to observe Rater 2 marking 40 essays on paper in a private meeting room. Each session lasted approximately 1 hour and was recorded.
Phase 3	The researcher conducted four online one-to-one sessions to observe Rater 3 marking 40 essays via Zoom. Rater 3 commented on shared documents via Google Docs. Each session lasted approximately 1 hour and was recorded.

As shown in Table 4, in Phase 1, the researcher (Rater 1) independently marked 40 essays, writing down the comments and highlighting the errors and good phrases. By doing this initially, the researcher can obtain a comprehensive understanding of each essay before the follow-up marking sessions with Rater 2 and Rater 3 in Phase 2 and Phase 3. This individual marking procedure of Rater 1 was not recorded. In Phase 2, the researcher set up direct, face-to-face meetings with Rater 2 and recorded the process of marking and assessing these essays to ascertain a transparent marking process. The researcher observed Rater 2 marked the essay and highlighted the

errors directly on the paper. Each meeting lasted approximately one hour. In Phase 3, the researcher conducted an online meeting with Rater 3 through Zoom. The researcher shared the essay files on Google Docs. Rater 3 shared his screen, marked the essay, and wrote comments on the essay on these shared online documents. This process was recorded using Zoom's recording function.

During the time observing raters' decision-making processes, the researcher encouraged Rater 2 and Rater 3 to engage in concurrent verbal protocols, also known as think-aloud protocols, when they assessed a random sequence of 40 essays. For example, the raters are encouraged to read a sentence and verbalise their comments such as: "The introduction is clear", "The meaning of this sentence is unclear", or "This word does not fit in this context". These verbal comments are also used as research data to evaluate the essays, in case of the raters did not write them down. According to Bowles (2010, p.111), concurrent verbal protocols are verbal reports completed during a task and have been widely used as a research methodology to investigate many aspects of first and second language acquisition, including the decision-making process of raters (Cohen, 1986; Cohen & Cavalcanti, 1987; Hosenfeld, 1984; Pressley & Afflerbach, 1995; Qi & Lapkin, 2001; Cumming et al., 2001, 2002; Lumley, 2002; Erdosy, 2004; Sachs & Polio, 2007; Eckes, 2008; Barkaoui, 2010). After assessing the whole essay, each rater independently gave four scores on four criteria (Task Response, Coherence and Cohesion, Lexical Resource, Grammar range and Accuracy) using the rubric of the IELTS Writing Band descriptor for Task 2, which is provided publicly by the British Council (2023), with each score being a round number, ranging from band 1 to band 9, as presented in Figure 9 (Chapter 3, Section 3.1.1).

The researcher also asked the raters to highlight good examples of written language, as well as any lexical or grammatical errors in each essay. This approach is intended to help the raters concentrate on evaluating language proficiency in the essay, rather than considering too much on the facts of the arguments. Figure 15 provides examples of Rater 2's comments on paper and Rater 3's comments in Google Docs.

Some people think that renewable energy should replace fossil fuels. To what extent do you agree or disagree? Give reasons for your answer and include any relevant examples from your own knowledge or experience.

Write at least 250 words.

Some people believe that fossil fuels should be replaced by alternative energy. Personally, I partly agree with this statement.

On the one hand, there are several reasons why people should use renewable energy instead of fossil fuels. First, it reduces polluted environment. The amount of carbon emission will not be released when using alternative energy so it reduces air pollution. In addition, polluted water might be lessened because there are not any oil spill cases. Second, renewable energy is endless so it can be used for long-term. For example, wind energy can provide electric power for a whole city without afraid of power cutting.

However, I still believe that alternative energy cannot replace fossil fuels due to these reasons. Firstly, installation fee of renewable energy is quite expensive which can create a heavy financial burden on the government and companies. For example, the average cost of installing a wind turbine might be up to 4 million dollars and require a wind farm with hundreds of turbines to supply power. Secondly, some certain machine still need fossil fuels to run, such as vehicles. Therefore, it is hard to get rid of fossil fuels.

In brief, although many people think that renewable energy should replace fossil fuels because of its advantages, I believe that it is difficult for countries to replacing fossil fuels.

5565

Some people think that renewable energy should replace fossil fuels. To what extent do you agree or disagree? Give reasons for your answer and include any relevant examples from your own knowledge or experience. Write at least 250 words.

Some people believe that fossil fuels should be replaced by alternative energy. Personally, I **partly agree** with this statement. **[not clear enough]**

On the one hand, there are several reasons why people should use renewable energy instead of fossil fuels. First, it reduces **polluted environment**. The amount of **carbon** emission **will** **[claim]** not be **decrease** when using alternative energy so **it reduces** **[claim]** air pollution. In addition, polluted water might be lessened because there are not any oil spill cases. Second, renewable energy is endless so it can be used **for long term**. For example, wind energy can provide **electricity** **web** for a whole city **without afraid of** power cutting.

However, I still believe that alternative energy cannot replace fossil fuels due to these reasons. Firstly, installation fee of renewable energy is quite expensive which can create a **heavy financial burden** on the government and companies. For example, the **average cost of installing a wind turbine** might be up to 4 million dollars and require a wind farm with hundreds of turbines to supply power. Secondly, some certain machine still need fossil fuels to run such as vehicles. Therefore, it is hard to **get rid of fossil fuels**.

In brief, although many people think that renewable energy should replace fossil fuels because of its advantages, I believe that it is difficult for countries to replacing fossil fuels.

TR: 7 CC 7 GRA 6 LR 6

Figure 15. Examples of Rater 2's comments on paper and Rater 3' comments on Google Docs.

Using those highlighted examples of words and phrases that the marker recognises as demonstrating that the student meets a particular band score to assess the contribution of the tools to students scores, the researcher was able to identify which of these examples were suggested by the tools and which were produced independently by the students. This method allows the researcher to collect qualitative information regarding the impact of the four tools, alongside an emphasis on quantitative data such as scores and statistical calculations.

3.3.1 Summary of each rater's assessment

To assess each essay, each rater follows the writing rubric published by the British Council (2023). The detailed rubric is provided in Figure 9 of Chapter 3, Section 3.1.1.

Each essay is assessed with a band score on a scale from 0 to 9 for the four criteria: Task Response (TR), Coherence and Cohesion (CC), Lexical Resource (LR), and Grammatical Range and Accuracy (GA). The Task Response (TR) criterion evaluates the content of the essay by examining how thoroughly the essay addresses the question, maintains a consistent position of argument, and effectively develops and supports arguments with relevant examples. The Coherence and Cohesion (CC) criterion assesses the structural organisation of information and the accurate use of cohesive devices between sentences. Lexical Resource (LR) focuses on the range of vocabulary, the accuracy of vocabulary usage, the appropriate academic writing style and the ability to paraphrase the writing topic. Lastly, Grammatical Range and Accuracy (GA) evaluates the variety and precision of grammatical structures used throughout the essay. The overall score for each essay was calculated by using the average of these four scores.

3.3.2 Summary of scores given by each rater

Tables 5, 6, and 7 below present the average and range of scores provided by each rater for the four criteria in two sets of 20 essays: those completed before the intervention of the four tools (Essay 1) and those completed after the intervention (Essay 2). Detailed data for each rater are available in Table 5a, Table 6a, and Table 7a in the Appendices. These summaries include the minimum scores, mean, standard deviations, and maximum scores. For the 20 Essay 1 sets, the four criteria are represented as TR1 (Task Response), CC1 (Coherence and Cohesion), LR1 (Lexical Resource), and GA1 (Grammatical Accuracy). For the 20 Essay 2 sets completed using the four tools, the four criteria are represented as TR2 (Task Response), CC2 (Coherence and Cohesion), LR2 (Lexical Resource), and GA2 (Grammatical Accuracy).

Table 5. Descriptive statistics of scores by Rater 1

Scores by Rater 1	TR1	CC1	LR1	GA1	TR2	CC2	LR2	GA2
Minimum	4	5	5	5	5	6	6	5
Mean	6.7	6.4	6.65	6	6.9	6.95	7.15	6.65
Standard Deviation	1.23	0.86	0.96	0.84	1.09	0.74	0.85	0.73
Maximum	9	8	8	7	9	8	9	8

Table 6. Descriptive statistics of scores by Rater 2

Scores by Rater 2	TR1	CC1	LR1	GA1	TR2	CC2	LR2	GA2
Minimum	4	4	5	5	5	5	6	6
Mean	6.65	6.6	6.5	6.3	6.85	6.7	6.95	6.95
Standard Deviation	1.01	0.97	0.81	0.56	0.79	0.78	0.67	0.5
Maximum	8	8	8	7	8	8	8	8

Table 7. Descriptive statistics of scores by Rater 3

Scores by Rater 3	TR1	CC1	LR1	GA1	TR2	CC2	LR2	GA2
Minimum	5	5	5	5	5	5	6	6
Mean	6.65	6.6	6.45	6.6	6.55	6.7	7	7
Standard Deviation	0.91	0.80	0.86	0.86	0.92	0.78	0.71	0.71
Maximum	8	8	8	9	8	8	8	8

3.3.3. The inter-rater reliability rates

Using data in Tables 5, 6a and 7a, the researcher employed the Intraclass Correlation Coefficient (ICC) to evaluate inter-rater reliability among three raters. Specifically, the ICC(C,3) model, a two-way random-effects model, was applied. This model assumes that both the subjects (essays) and the raters were randomly selected. It emphasises the consistency of ratings rather than absolute agreement, assessing how reliably the raters rank the essays, even if their individual scores vary slightly. Table 8 below presents the ICC values for the first set of 20 essays (Essay 1), reflecting the consistency of scores assigned by three raters across Task Response (TR), Coherence and Cohesion (CC), Lexical Resource (LR), and Grammatical Accuracy (GA).

Table 8. Inter-rater reliability for Essay 1 across four criteria

ICC(C,3)	TR1	CC1	LR1	GA1
	0.779	0.796	0.892	0.807

According to Table 8, there was a generally high level of agreement among three raters for the first 20 essays, with ICC(C,3) values of 0.779 for TR1, 0.796 for CC1, 0.892 for LR1, and 0.807 for GA1. Overall, the results suggest good reliability across four criteria in Essay 1. Koo and Li (2016) provide a guideline for interpreting Intraclass Correlation Coefficient (ICC) values in reliability research. According to their classification, ICC values less than 0.5 indicate poor reliability, values between 0.5 and 0.75 suggest moderate reliability, values between 0.75 and 0.9 represent good reliability, and values greater than 0.9 indicate excellent reliability (Koo & Li, 2016, p.161).

In the same way, the inter-rater reliability rates of the second set of 20 essays (Essay 2) are computed and presented in Table 9.

Table 9. Inter-rater reliability for Essay 2 across four criteria

ICC(C,3)	TR2	CC2	LR2	GA2
	0.787	0.842	0.78	0.802

Regarding Table 9, the ICC analysis also indicated a high level of agreement among the three raters in four criteria in the second set of 20 essays, with ICC(C,3) values of 0.787 for TR2, 0.842 for CC2, 0.78 for LR2, and 0.802 for GA2. The agreement rate of GA remained mostly unchanged, whereas the rates of TR and CC slightly increased. These findings also indicate good reliability across the four criteria in Essay 2, as the values sit in the good reliability range (between 0.75 and 0.9), and the changes in inter-rater reliability between Essay 1 and Essay 2 were minimal.

In summary, the high ICC values from both essay sets demonstrate the reliability of the standardised IELTS writing rubric used in this study as the methodology. They also highlight the advantages of employing experienced human raters to ensure consistent assessments across the four writing criteria. The following section will analyse the data using more statistical methods to identify and interpret changes in students' writing scores.

3.4. Quantitative analysis methodology for writing scores

Section 3.4 details the statistical methods employed by the researcher to analyse the inter-rater reliability and the extent to which students' writing scores change between essay one and essay two. Specifically, this analysis involves statistical calculations at a group level to address two research questions: RQ1a: "Does the use of the writing tools lead to a change in students' overall writing scores?" and RQ1b: "Which aspects of the IELTS writing assessment (i.e., Task Achievement, Coherence and Cohesion, Lexical Resource, Grammatical Range and Accuracy) are most influenced by the students' use of the tools?".

For each student essay, the researcher computed the mean scores of those given by the three raters for each criterion, including Task Response (TR), Coherence and Cohesion (CC), Lexical Resource (LR), and Grammatical Range and Accuracy (GA). Following this, the overall scores were derived from the average scores of these four criteria as rated by the three raters. Subsequently, the researcher employed three statistical measures—Intraclass Correlation Coefficient (ICC), Paired t-test, and Pearson's Correlation Coefficient—to analyse the data. All calculations were performed using IBM's Statistical Package for the Social Sciences (SPSS).

3.4.1 Intraclass Correlation Coefficient (ICC)

Purpose: The Intraclass Correlation Coefficient (ICC) analysis is used to measure the inter-rater reliability (also known as the inter-rater agreement) of the three raters in this study.

Model: The researcher used the model ICC(C,3): a two-way random effects model, in which both the subjects and raters are randomly selected, focusing on the consistency of ratings rather than absolute agreement. The concept of "consistency of ratings" in this context of reliability analysis using the Intraclass Correlation Coefficient refers to how consistent the raters are in their ranking of the participants, despite there being some variation in the exact scores they assign (Shrout & Fleiss, 1979; McGraw & Wong, 1996; Landers, 2015).

Interpretation: The ICC value ranges from 0 to 1, with values closer to 1 indicating higher reliability or agreement.

3.4.2 Paired T-tests

Purpose: The researcher used paired t-tests to compare the means of two groups of essays' scores to see if they are statistically different from each other. A total of eight paired t-tests were used to calculate the mean score changes in each criterion, including Task Response (TR), Coherence and Cohesion (CC), Lexical Resource (LR), Grammatical range and Accuracy (GA).

Type: Paired t-test across two groups (i.e. all students' scores in Essay 1 and Essay 2)

Interpretation: If the p-value is less than or equal to a threshold (commonly lower than 0.05), it is typically interpreted as the groups being significantly different. If the p-value is higher than 0.05, it suggests that the observed data are likely under the null hypothesis, and it is interpreted as insufficient evidence to conclude that there is a significant difference between the groups (Urdan, 2022).

3.4.3 Pearson's Correlation Coefficient

Purpose: Pearson's correlation coefficient (r) was used for two analyses: (1) to examine the relationship between overall scores in Essay 1 (written without tools) and Essay 2 (written with tools) in Chapter 4, Section 4.1.3, Table 20; and (2) to test whether Essay 2 overall scores were associated with use of each tool (coded 1 = used, 0 = not used; see Section 4.2.1, Table 23).

Interpretation: The ICC value ranges from -1 to +1. Values close to 1 indicate a strong relationship, while values close to -1 indicate a weak relationship (Field, 2009; Koo & Li, 2016).

3.4.4 Calculating overall scores and individual score changes

For each essay, the overall score was the average of the four IELTS criteria: Task Response, Coherence and Cohesion, Lexical Resource, and Grammar. To find each student's change in overall performance, the researcher subtracted their overall score in Essay 1 from their overall score in Essay 2; this shows whether they improved or declined.

3.4.5 Quantifying students' tool usage instances

To examine a student's tool use in Essay 2, the researcher watched the screen recordings (and checked interview answers) to see whether each tool—Sketch Engine, Google Translate, Quillbot, and Grammarly—was used by a student at any point while writing. Each tool usage instance was coded as “1” if it was used and “0” if it was not. This constitutes a quantitative, dichotomous coding of screen-recording data (0 = not used; 1 = used) for the record of tool usage instance in Essay 2, not how many times or how long a tool was used (see Chapter 4, Section 4.2.5 for details).

3.5. Qualitative analysis of students' tool usage strategy and interview responses

3.5.1 Interview questions for students

This section details the interview questions for students. The purpose of the interview is to identify how students perceive the usefulness of the tools and their reflections on their tool usage. The interview consists of eight questions, as shown in Table 10 below:

Table 10. Interview questions for students

1/ In general, how do you feel about the usage of the four AI writing tools for writing this IELTS essay? Do you find them helpful?

2/ Overall, do you think you produced a higher quality essay with the tools compared to the first essay that you wrote without the tools? Which part or sentence do you think you wrote better with the tools?

3/ In the brainstorming and planning step (before writing), did you find the tools useful for you in this step? Which tool helped you the most? How did it help you?

4/ In the composing step, did you find the tools useful? Which tool helped you the most? How did it help you?

5/ In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?

6/ Can you give me some examples of challenges that the tools could not help you in this writing task?

7/ What do you want more from these four tools: SketchEngine, Google Translate, QuillBot, and Grammarly?

8/ Do you recommend or would you use any other tools for this IELTS essay writing practice? Why?

The interview questions were designed to thoroughly examine the usefulness and limitations of the four AI writing tools (SketchEngine, Google Translate, QuillBot, and Grammarly) across different stages of the IELTS essay writing process. Question 1 aims to examine students' general impressions of the tools. Question 2 asks students to compare their tool-assisted essay with a previous, tool-free essay to determine if the tools led to noticeable quality improvements and, if so, to identify specific parts or sentences that benefited. Questions 3, 4, and 5 delve into the four tools' usefulness at different stages of writing: brainstorming and planning (Q3), composing (Q4), and revising and editing (Q5), asking students to reflect on which tools were most effective at each phase and how they helped the students. Question 6 encourages students to share specific challenges where the tools did not provide adequate support, identifying

any limitations in functionality. Question 7 asks students to suggest improvements, giving insight into additional features or enhancements they would find beneficial. Finally, Question 8 invites students to recommend other tools they consider useful for IELTS essay writing, providing a broader context for evaluating these AI tools within the landscape of technology for language learning. Questions 7 and 8 contribute to RQ3: “How do teachers and students perceive the effectiveness of the four writing tools?”.

3.5.2 Qualitative analysis of student tool engagement and language learning

3.5.2.1 Classification of language proficiency levels for comparative engagement analysis

To compare engagement patterns among high-achieving and low-achieving students, the 20 student participants were grouped using a criterion-referenced IELTS/CEFR cut-off. Each student’s Essay 2 score was the unrounded mean of three raters; <6.5 was classified as low and ≥ 6.5 as high. The 6.5 threshold aligns with the CEFR B2/C1 boundary and is widely used for UK student-visa proficiency (CEFR B2 $\approx$ IELTS 5.5–6.5) and university entry minima, providing a meaningful standard rather than a sample-dependent split (Council of Europe, 2020; UK Home Office, 2024; British Council, 2021; IELTS Asia, n.d.). This classification served two purposes. First, it provided a clear frame for comparative qualitative analysis, allowing us to examine how students at different performance levels engaged with the tools (e.g., the extent of checking, revising, or rejecting tool output). Second, it enabled descriptive contrasts on the quantitative indicators (e.g., presence/absence of tool use) and guided the selection of illustrative extracts in the thematic analysis. The focus on Essay 2 ensures that grouping reflects performance under tool-available conditions, which is the context in which engagement with the four tools was observed.

3.5.2.2 Coding for thematic analysis of students’ interview responses

To answer RQ1b: “How do high-scoring and low-scoring students differ in their ability to engage with the tools and use them to support their language development?” the researcher conducted a thematic analysis of the interview transcripts following Braun and Clarke’s (2006) six steps. Drawing on Chapter 2’s theoretical framework (Section 2.1.4), including the multidimensional view of engagement, the Output Hypothesis,

and the Noticing Hypothesis, the researcher coded students' responses into three themes: Cognitive Engagement (CE), Vocabulary Acquisition (VA), and Grammar Improvement (GI). Each theme instance is the number of times in the transcript that the student discusses their usage of a particular tool.

Cognitive Engagement (CE) is coded when a learner demonstrates both behavioural and cognitive engagement. Behavioural engagement is indicated by students' tool usage behaviours such as exploring different tool functions, cross-checking the output of one tool with another tool, editing the draft, or actively accepting or rejecting a suggestion. Cognitive engagement is indicated when the student reports noticing language gaps, comparing alternatives, and explaining why one option is more appropriate.

Vocabulary Acquisition (VA) is coded when a student reports noticing new lexis (including synonyms and collocations) through tool use in their interview. Drawing on prior empirical findings that the four tools can support lexical development, this code examines whether comparable gains occur in the present study.

Grammar Improvement (GI) is coded when a student reports noticing a grammatical problem and uses a tool to correct or refine it, leading to clearer and more accurate writing. Drawing on findings from earlier empirical studies that report grammar benefits from corpus- and MT-supported revision as well as paraphrasing and AWE tools, this code assesses whether and how such grammar improvements appear in these data.

3.5.2.3. Analysis of case study

To address RQ1c ("What factors contribute to any observed decrease in students' writing scores when using these tools?"), the researcher conducted a multiple-case analysis of four students whose overall scores declined from Essay 1 (without tools) to Essay 2 (with access to the four tools). The analysis of these case studies was based on four main data sources: students' overall score changes (Section 4); raters' comments on errors and language use (as mentioned in Section 3); video recordings of each student's writing process; and recordings of student interviews.

Figure 16. below outlines the process that is used to assess the tools' impact and identify additional factors that may have contributed to the decrease in each case.

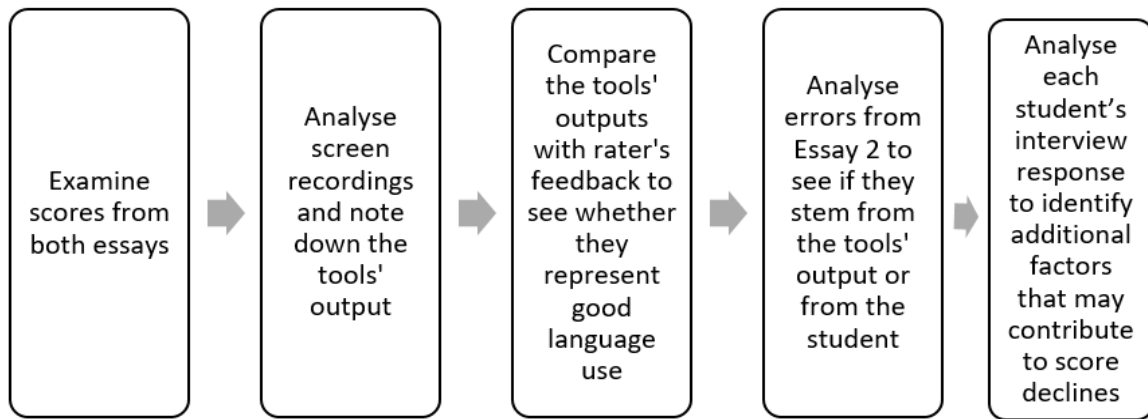


Figure 16. Summary of the process for analysing the impact of the four tools and additional factors contributing to score declines.

For each case, the researcher first examined criterion scores of both essays—Task Response, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy—marked by three raters. The researcher then analysed the process recordings to document the sequence of tool use and the outputs produced by each tool, and compared those outputs with rater feedback and the resulting text. When a tool output corresponded to an error flagged by raters or introduced new errors, it was treated as a negative tool effect. By contrast, outputs reflecting effective language use indicated no direct tool-related score decrease, making student factors the more plausible contributors. Subsequently, the researcher analysed each student's interview response to identify additional factors that may contribute to declined scores. This approach enables within-participant comparisons of writing performance before and after tool use, tracing score changes in each criterion and the role of specific tool-mediated revisions. It also reveals challenges that each student encountered during Essay 2 and assesses whether the tools mitigated or exacerbated those issues, insights that may be obscured in quantitative results or broad thematic analyses. Accordingly, findings from case study analysis can identify the factors that contribute to the students' observed score declines.

3.6 Teacher recruitment, interview questions, and lesson plan analysis

This section details the recruitment process, lesson plan requirements, and interview questions for the 10 teacher participants to specifically address RQ2 and RQ3 and their sub-questions:

RQ2: What similarities and differences exist between teachers' recommended strategies for using the four writing tools in IELTS essay preparation and students' actual usage patterns, and how are these similarities and differences reflected in students' essay writing performance?

RQ2a: What specific strategies do teachers recommend for using each of the four writing tools in IELTS essay preparation?

RQ2b: What strategies do students employ when using the four writing tools?

RQ2c: Did students whose tool usage shared similarities with teachers' strategies achieve higher Essay 2 scores compared to those whose tool usage differed?

RQ3: How do teachers and students perceive the effectiveness and limitations of the four writing tools

RQ3a: What are the strengths and limitations of each tool, as evaluated by teachers and students?

RQ3b: How do students' evaluations relate to their writing performance?

RQ3c: What additional features do teachers and students expect beyond the tools' current capabilities?

The researcher compared the teachers' data with the students' tool usage strategy data and interview responses.

3.6.1 Recruitment process of teacher participants

To explore the general attitudes of teachers regarding the integration of four specific writing tools in their lessons, an online questionnaire was disseminated to

approximately 200 ESL teachers through the internal email network of the University of Economics Ho Chi Minh City and the English language centre where the researcher was working. The primary objectives of this survey were to assess teachers' awareness of popular writing tools and to recruit potential participants for the study. This approach enabled access to a wider range of practising IELTS instructors beyond a single institution and supported voluntary participation. Moreover, the fact that these teachers did not teach the student participants can avoid any teacher–student power relationship or conflict of interest. To qualify for participation, teachers needed to meet the following criteria: (1) familiarity with the four writing tools and the four functions under study, (2) willingness to incorporate these tools into an IELTS writing lesson, and (3) a minimum of three years of experience teaching IELTS writing.

The questionnaire included 18 questions in a variety of formats. For example, open-ended questions such as Q2: “Have you ever used AI tools in your writing lessons/curriculum?” and Q3: “Please list some AI tools for improving English writing that you have used or heard about” were used to gauge teachers' familiarity with AI writing tools. Additionally, 5-point Likert scale questions, like Q4: “In general, what is your attitude towards using technology and AI tools in teaching writing?” and 3-point Likert scale questions (Questions 10-13), as well as multiple-choice questions (Questions 5-9), aimed to confirm the teachers' awareness of the tools' specific functions, including collocation search, translation, paraphrasing, and grammar correction. Three questions focused on the teachers' experience in teaching IELTS writing, such as Q17: “How long have you been teaching the IELTS Task 2 Writing skill?” and Q18: “What are the outcome levels of IELTS writing classes that you have taught?” Additionally, two questions collected teachers' names and email addresses to identify potential participants for follow-up lesson-planning interviews and confirm their willingness to be re-contacted. The full questionnaire for teachers is presented in Appendix B in the Appendices.

Out of the survey respondents, 14 teachers registered for the study. Two participants were excluded for having less than one year of teaching experience, one did not attend the initial session, and another withdrew after the first session. Consequently, the final participant pool consisted of 10 teachers. These participants were Vietnamese instructors with at least five years of general English teaching experience and a

minimum of three years specialising in IELTS preparation. They demonstrated a positive attitude towards using writing tools and had experience with specific tools such as Grammarly (an automated writing evaluation tool) and Quillbot (an automated paraphrasing tool) for their writing tasks. Table 6 below provides a detailed overview of the teaching experience of the 10 teacher participants.

Table 11. The teaching experience of 10 teacher participants

Teacher	Years of teaching experience in IELTS
Teacher 1	3 years
Teacher 2	4 years
Teacher 3	3 years
Teacher 4	5 years
Teacher 5	5 years
Teacher 6	4 years
Teacher 7	3 years
Teacher 8	5 years
Teacher 9	3 years
Teacher 10	4 years

Subsequently, each of the teachers attended two sessions (one-on-one) with the researcher. In the first session, which lasted for about 30 minutes, the researcher introduced teachers to the four selected writing tools: Sketch Engine, Google Translate, Quillbot, and Grammarly, and demonstrated their main features. After that, teachers were required to design a lesson plan that instructs students to use a combination of the four writing tools above to assist their writing task: an argumentative essay following the IELTS Task 2 essay format (i.e., a 250-word essay) with the topic:

“Spoken communication is more powerful than written communication. To what extent do you agree or disagree?”. The next section explains the rationale for selecting this essay topic.

3.6.2 Essay topic selection for teachers

For the writing topic of teachers, the researcher opted for the topic of comparing spoken and written communication (Figure 17).

IELTS Writing Task 2

You should spend about 40 minutes on this task

Spoken communication is more powerful than written communication. To what extent do you agree or disagree?

Give reasons for your answer and include any relevant examples from your own knowledge or experience

Write at least 250 words

Figure 17. The essay topic for teachers.

This topic avoids niche knowledge areas, requires minimal specialised knowledge to address, and offers plenty of scope for argument, which makes it a fitting choice for the assistance of writing tools. It can also give space for the data-driven learning approach of Johns (1991). For example, to strengthen the argument between spoken communication and written communication, the teacher can search for two words on the topic on Sketch Engine, such as “speaking” and “writing”, and compare their collocations, as illustrated in Figure 18.

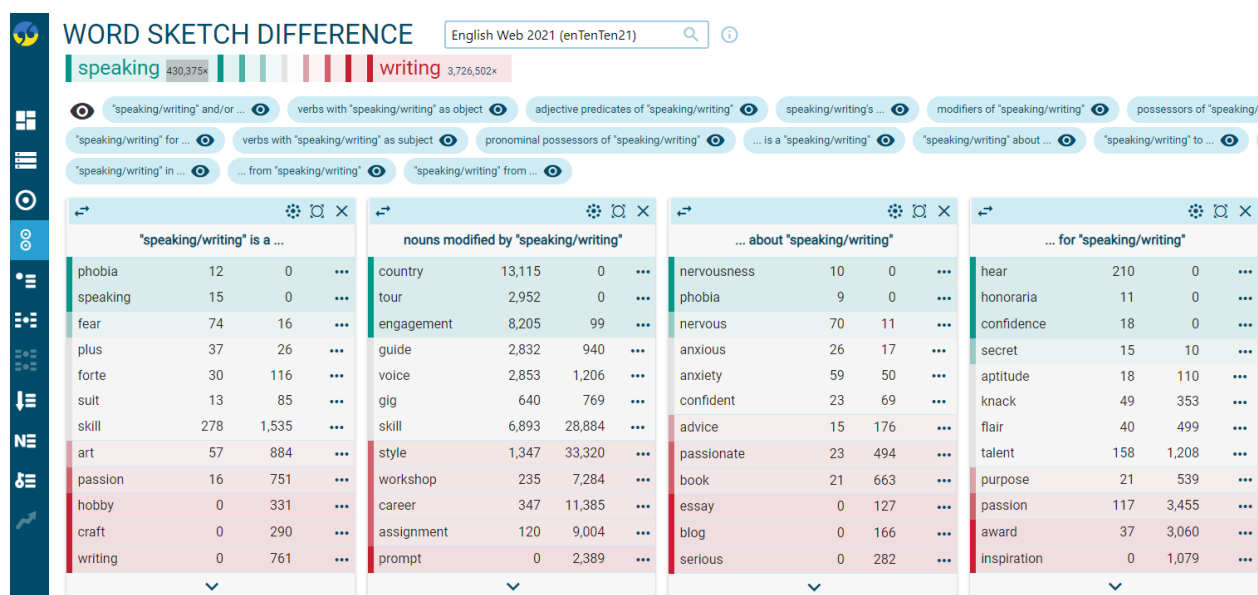


Figure 18. An example of how teachers can use Sketch Engine for lesson planning on the chosen topic.

Accordingly, the teacher can form an argument that written communication is more powerful than spoken communication, because the word “speaking” usually goes with negative words, such as “phobia”, “nervousness” or “fear”, while more positive words usually appear with “writing”, such as “art” or “passion”. Drawing on collocation results from Sketch Engine, teachers might instruct students to form full sentences describing these ideas in a writing lesson, and that writing craft can help students work on and develop a complete essay.

3.6.3 Interview questions for teachers

In the second session, the participating teachers were required to elaborate on their lesson plans, specifically highlighting the optimal utilisation of the four identified writing tools within each phase of a writing lesson by sharing their screens and presenting their lesson plans. Finally, immediately after their presentation, the researcher elicited more on teachers’ evaluations of the four tools in the interview session. This interview consists of 19 specific questions regarding different aspects of using writing tools in teaching IELTS essay writing. Questions 1, 5, 9, and 13 ask for participants’ experiences and perceptions of using Sketch Engine, Google Translate, Quillbot, and Grammarly, respectively, for this purpose. Questions 2, 6, 10, and 14 delve into teaching methodologies or instructional steps for effectively integrating these tools into

IELTS Task 2 essay instruction. Concerns related to each tool are addressed in questions 3, 7, 11, and 15, while questions 4, 8, 12, and 16 ask for suggestions or additional features or improvements to these four writing tools to better support teachers with writing lessons. General reflections on the lesson plan activity are gathered through question 17. Questions 18 and 19 focus on whether teachers would recommend other tools or functions that could be beneficial for teaching or for students to practice IELTS Task 2 essays at home. The full list of 19 interview questions for teachers is presented in Table 7 below.

Table 12. The 19 interview questions for teachers

- 1/ Based on your experience of participating in this study, how do you feel about using Sketch Engine for teaching IELTS essay writing?
- 2/ If you must use Sketch Engine to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology do you think would be the most suitable for Sketch Engine?
- 3/ Do you have any concerns about using Sketch Engine as a writing tool? What are they?
- 4/ Can you think of anything you would want more from Sketch Engine to help teachers with writing lessons?
- 5/ Based on your experience of participating in this study, how do you feel about using Google Translate for teaching IELTS essay writing?
- 6/ If you must use Google Translate to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology would be the most suitable?
- 7/ What are your concerns about using Google Translate as a writing tool?
- 8/ Can you think of anything you would want more from Google Translate to help teachers with writing lessons?
- 9/ Based on your experience of participating in this study, how do you feel about using Quillbot for teaching IELTS essay writing?
- 10/ If you must use Quillbot to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology would be the most suitable?
- 11/ What are your concerns about using Quillbot as a writing tool?
- 12/ Can you think of anything you would want more from Quillbot to help teachers with writing lessons?
- 13/ Based on your experience of participating in this study, how do you feel about

using Grammarly for teaching IELTS essay writing?

14/ If you must use Grammarly to teach an IELTS Task 2 essay, what step of instructions or teaching methodology would be the most suitable for this tool?

15/ What are your concerns about Grammarly?

16/ Can you think of anything you would want more from Grammarly to help teachers with writing lessons?

17/ In general, do you have any other reflections on the lesson plan activity?

18/ Do you recommend any other tools for teaching the IELTS Task 2 essay?

19/ Do you recommend any other tools for students to practice the IELTS Task 2 essay at home?

3.6.4 Lesson plan analysis

To analyse how the four tools were integrated into teaching the essay, each teacher completed a three-stage process-writing template (Planning, Composing, Revising). The template is anchored in established models of the writing process which, despite terminological variation, converge on a planning → composing → revising sequence: Flower and Hayes (1981) differentiate planning, translating (i.e., composing), and reviewing (which encompasses revising); Seow (2002) specifies planning, drafting (i.e., composing), revising, and editing; and Hyland (2003) organises instruction around preparation, composition, and revision. Guided by this convergence, lesson planning and subsequent analysis were operationalised around three core stages. For each stage, teachers documented a corresponding activity and specified which of the four tools was employed. The analysis then sought to identify the particular functions through which these tools supported students at different points in the writing process. An example of a lesson plan from Teacher 1 is provided in Table 8 below.

Table 13. An example of the lesson plan of Teacher 1.

Teacher	Tool	Stage	Activity
Teacher 1	Sketch Engine	Planning	The teacher selects some keywords: “speak” vs “write”, “nonverbal” and “verbal”, “communication”, “and powerful”.

			From the results of Sketch Engine, suggest students choose relevant nouns that go with “speak”: dialect, accent, tone, voice.
	Google Translate		The teacher suggest that students write separate sentences as their initial thoughts about this topic and use Google Translate to assist in translating words. Example: “Spoken communication is less powerful than writing because it can be affected by dialect and accent and it is difficult to reach many people.”
	No tool	Composing	Students work individually for 40 minutes and write a full essay.
	Quillbot	Revising	This is a session for individual revision. The teacher suggests that students use Quillbot to get suggestions for their writing. Students will then selectively choose the sentences they want to paraphrase. This will help improve the quality and variety of their writing. Students will selectively choose the sentences to paraphrase
	Grammarly		Students receive their own essays back after the teacher has seen them. The teacher allows students to use Grammarly to run the final check.

3.7. Comparison of teachers’ and students’ tool usage, evaluation and recommendation

3.7.1 Comparison of quantified data from teachers' and students' tool usage

To address RQ2c: “Did students whose tool usage shared similarities with teachers’ strategies achieve higher Essay 2 scores compared to those whose tool usage differed?”, the researcher first quantified alignment between teachers’ recommended tool-use strategies and students’ reported practices. Alignment was defined as cases where a student used the same tool, for the same function, and at the same stage of writing as recommended by at least one teacher. This was calculated by comparing the percentage of teachers who endorsed a specific tool–function–stage combination with the percentage of students who adopted that same combination. For example, if 20% of teachers (2 out of 10) recommended using Sketch Engine to check phrase concordances during revision, and 20% of students (4 out of 20) reported doing so, this was recorded as a point of alignment.

After identifying alignment patterns for each writing stage, the researcher identified whether aligned students belonged to the high- or low-scoring groups based on their Essay 2 scores (as mentioned in Section 3.5.2.1). This classification enabled an examination of whether students whose tool usage closely mirrored teacher-recommended strategies were more frequently represented in the high-scoring group, thereby providing insight into the potential relationship between alignment in tool use and writing performance. Table 9 below provides an example of quantified data of tool usage comparison between teachers and students in the Planning stage.

Table 14. An example of quantified data of tool usage comparison between teachers and students in the planning stage

Stage	Tool’s name	Tool’s function	Percentage of teachers	Percentage of students
Planning	Sketch Engine	Looking for collocations of a word or a phrase to form ideas	70%	10%
	Google	Translating words or	90%	0%

	Translate	sentences from Vietnamese to English		
	Quillbot	Replacing a word, rewriting a phrase or a sentence	10%	0%
	Grammarly	Checking spelling and grammatical errors of students' self-written sentences	20%	0%

3.7.2. Comparison of teachers' and students' tool evaluation

To address RQ3a— “What are the strengths and limitations of each tool, as evaluated by teachers and students?”—this study adopts the premise that perceptions of digital tools often reflect an important dimension of emotional engagement, encompassing affective responses such as interest, confidence, anxiety, or frustration (Fredricks et al., 2004; Hiver et al., 2021a). Accordingly, the method was designed not only to document the perceived strengths and limitations of each tool but also to capture the emotional undercurrents that shape these evaluations. Emotional engagement has been widely recognised as a significant predictor of academic outcomes, influencing learners' persistence, motivation, and overall performance (Chan et al., 2024; Christenson et al., 2012; Lei et al., 2018; Luo et al., 2023; Schrader & Kalyuga, 2020; Skinner et al., 2009; Wu et al., 2020). Building on that, teachers' and students' responses were categorised into “positive feedback”, representing positive emotional expressions of usefulness (e.g., satisfaction with a tool's accuracy, accessibility, or convenience), and “negative feedback”, representing expressions of dissatisfaction (e.g., frustration with errors, anxiety over misleading suggestions, or disappointment with limited functionality). This categorisation facilitated the interpretation of participants' affective responses embedded in their evaluations. An example of teachers' and students' evaluations of Sketch Engine is presented in Table 10, with detailed findings discussed in Chapter 5.

Table 15. Example comparison of teachers' and students' evaluations of Sketch Engine

Sketch Engine		
	% Teacher	% Student
Positive feedback:		
Useful for planning ideas	4 out of 10 40% (Teachers 3, 5, 7, 8).	2 out of 20 10% (Students 6, 8)

3.7.3 Comparison of students' evaluations and writing performance

To address RQ3b— “How do students’ evaluations relate to their writing performance?”—the researcher employed a triangulation approach, integrating qualitative and quantitative data. Students’ evaluations of the four writing tools, gathered through semi-structured interviews, were cross-referenced with their Essay 2 mean overall scores (see Chapter 4, Section 4.2.4, Table 21) and their classification as high-scoring or low-scoring, as mentioned in Chapter 3, Section 5.2.1. This comparison seeks to reveal patterns connecting evaluative trends—such as satisfaction, confidence, or confusion—with two performance groups, taking into account both emotional engagement (e.g., satisfaction, frustration, motivation) and cognitive engagement (e.g., critical assessment, problem-solving, reflection on tool feedback). By examining these emotional and cognitive engagement of students alongside their writing scores, the analysis aims to clarify whether distinct engagement patterns, such as highly positive evaluation or criticisms on the tools’ shortcomings, were associated with higher scores, and whether disengagement corresponded with lower scores.

3.7.4 Comparison of teachers' and students' recommendations for additional tool features

To address RQ3c—"What additional features do teachers and students expect beyond the tools' current capabilities?"—the researcher analysed responses to Interview Question 8: "Do you recommend or would you use any other tools for this IELTS essay writing practice? Why?" Each response was coded by (a) the tool or feature mentioned, (b) its intended purpose (e.g., vocabulary expansion, grammar checking, idea generation, concordance search), and (c) participant ID. This coding enabled a direct comparison of teachers' and students' expectations, moving beyond isolated, tool-specific suggestions to identify broader trends in desired functionality. An example of tool recommendation from teachers and students, along with its purpose, is presented in Table 16 below.

Table 16. An example of teachers' and students' recommendations of other tools and their usage purposes

Tool's name	Purpose	Teacher %	Teacher IDs
ChatGPT	Brainstorming ideas	4 out of 10 40%	1, 2, 7, 8

Tool's name	Purpose	Student %	Student IDs
ChatGPT/ Google Bard/ Bing AI	Brainstorming ideas and giving feedback	11 out of 20 55%	3, 4, 5, 6, 7, 8, 9, 10, 14, 15, 20

3.8 Ethical considerations

This research project adheres strictly to ethical standards, ensuring the protection and well-being of all participants throughout all phases of the study. For the initial online

survey, participants, including both students and teachers, were required to provide their consent by checking a box before proceeding. This step served as an essential procedure for obtaining informed consent.

Upon registration for the main study, participants were given a comprehensive information sheet detailing the study's objectives, methodologies, and anticipated duration. Before participating in the primary research sessions, all participants signed and returned a formal consent form to the researcher. Students and teachers were informed of their right to withdraw from the study and retract their data within the project's first three months.

Data collected is stored in a pseudo-anonymised format on an encrypted computer to ensure security. A separate list of participants' identities and corresponding ID codes is securely stored in a locked cabinet, accessible only to the researcher. To preserve confidentiality, participant identities are anonymised in this thesis and subsequent publications. All research outputs will maintain participant anonymity, ensuring that data are used exclusively for research purposes. This study has also received ethical clearance from an independent ethics committee, which thoroughly reviewed all elements of the research design to confirm adherence to the highest ethical standards.

CHAPTER 4: FINDINGS IN SCORE IMPACT OF THE FOUR TOOLS

Chapter 4 presents the findings addressing Research Question 1 (RQ1) and its three sub-questions:

RQ1: What is the impact of the four writing tools on students' IELTS essay writing scores?

RQ1a: Does the use of the writing tools result in an increase in the overall writing scores of students? If so, what aspects of IELTS writing assessment (i.e., Task Achievement, Coherence and Cohesion, Lexical Resource, Grammatical Range and Accuracy) are most affected by students' use of the tools?

RQ1b: How do high-scoring and low-scoring students differ in their ability to engage with the tools and use them to support their language development?

RQ1c: What factors contribute to any observed decrease in students' writing scores when using these tools?

Overview of Chapter 4 sections

Section 4.1 reports group-level statistical analyses, including paired t-tests and correlation measures, to determine whether score changes across criteria are significant and to examine relationships between tool usage and performance outcomes. Building on these quantitative findings, Section 4.2 compares individual score changes, highlighting cases of both improvement and decline, before turning to qualitative analyses of students' tool engagement. Section 4.3 focuses on thematic patterns in how students interacted with Sketch Engine, Google Translate, Quillbot, and Grammarly, distinguishing differences between high- and low-scoring groups. Section 4.4 details case studies of four students whose scores declined after using the tools, providing further insight into how tool-mediated revisions, learner decisions, and technical issues contributed to reduced performance. A summary of all findings to RQ1 and its sub-questions is presented in Section 4.5.

4.1. Statistical analysis of writing scores and the relationship between tool usage and score changes

4.1.1 Summary of mean scores from three raters and the overall changes

To address Research Question 1a: “Does the use of the writing tools result in an increase in the overall writing scores of students? If so, what aspects of IELTS writing assessment (i.e., Task Achievement, Coherence and Cohesion, Lexical Resource, Grammatical Range and Accuracy) are most affected by students’ use of the tools?”, the researcher calculated the mean scores assigned by three raters for each criterion across both sets of 20 essays. Table 17 provides a summary of mean scores by three raters, including the minimum, mean, standard deviation, and maximum scores across four criteria. Full results are available in Table 17a in the Appendices. The average scores from three raters for the first set of 20 essays (Essay 1) across four criteria are presented as TR1 (Task Response), CC1 (Coherence and Cohesion), LR1 (Lexical Resource), and GA1 (Grammatical Accuracy). For the second set of 20 essays (Essay 2), which was composed after students had access to the four tools, the average scores from three raters across four criteria are presented as TR2, CC2, LR2, and GA2.

Table 17. Summary of mean scores among the four criteria from three raters

	Essay 1				Essay 2			
Mean scores by three raters	TR1	CC1	LR1	GA1	TR2	CC2	LR2	GA2
Minimum	4.6	4.7	5	5.3	5.3	5.3	6	5.7
Mean	6.52	6.54	6.53	6.31	6.77	6.78	7.03	6.87
Standard Deviation	0.88	0.74	0.80	0.65	0.79	0.67	0.62	0.55
Maximum	8	8	8	7.7	8.3	8	8.3	8

According to Table 12, for the Task Response (TR) criterion of Essay 1, mean scores in this criterion (TR1) ranged from a minimum of 4.6 to a maximum of 8, while mean scores of Essay 2 (TR2) ranged from 5.3 to 8.3. The mean scores of Coherence and Cohesion criterion in Essay 1 (CC1) spanned from 4.7 to 8, while those of Essay 2 (CC2) spanned from 5.3 to 8.0. In Lexical Resource (LR), mean scores of this criterion in Essay 1 (LR1) spanned from 5.0 to 8, whereas those of Essay 2 (LR2) showed a shorter range from 6.0 to 8.3. The mean scores of the Grammatical Range and Accuracy criterion in Essay 1 (GA1) ranged from 5.3 to 7.7, while Essay 2's mean scores (GA2) were between 5.7 and 8. All four criteria showed an increase in their mean values, with TR increasing by 0.25, CC by 0.24, LR by 0.5, and GA experiencing the highest increase of 0.56. These findings indicate that the intervention of the four writing tools generally had a positive impact on all evaluated criteria, as shown in the increased mean scores, increased minimum scores for TR2, CC2, LR2, and GA2, as well as higher maximum scores for TR2 and LR2. Furthermore, the data in Table 10 reveal that the use of these four tools was insufficient to raise any criterion to the highest possible Band 9 score.

The mean overall score for Essay 1 for each student was calculated by averaging the scores assigned by three raters for Task Response (TR1), Coherence and Cohesion (CC1), Lexical Resource (LR1), and Grammatical Accuracy (GA1), using data in Table 10a (Appendices). Similarly, the mean overall score for Essay 2 was calculated by averaging the corresponding scores for TR2, CC2, LR2, and GA2 from the same three raters. Table 13 provides a statistical summary of the mean overall score changes of all 20 students between Essay 1 and Essay 2 from Table 10a (Appendices), including minimum, mean, standard deviation, and maximum values.

Table 18. Descriptive statistics of mean overall score changes between Essay 1 and Essay 2

	Mean overall score Essay 1	Mean overall score Essay 2
Minimum	4.9	5.9
Mean	6.48	6.865
Standard Deviation	0.733	0.634
Maximum	7.8	8.2

According to Table 13, students demonstrated an overall improvement in their scores, with the mean rising from 6.48 to 6.865. The increase in the minimum score from 4.9 in Essay 1 to 5.9 in Essay 2, along with the rise in the maximum score from 7.8 to 8.2, also indicates an overall improvement in students' performance. The reduction in standard deviation from 0.733 in Essay 1 to 0.634 in Essay 2 indicates that the scores became slightly more consistent in Essay 2. This could indicate that the tools put students on a more even level of performance.

4.1.2 T-test analysis of mean score changes across four criteria and overall scores

In this section, the researcher conducted five paired t-test analyses to assess changes in mean scores across the four evaluation criteria and the overall scores for both sets of all 20 essays. The results are presented in Table 19 below.

Table 19. Paired t-tests on the mean score change from each criterion of three raters

	TR1 vs TR2	CC1 vs CC2	LR1 vs LR2	GA1 vs GA2	Mean overall scores of Essay 1 vs Essay 2
t-statistic	-1.993	-2.165	-3.405	-4.453	-3.46
p-value	0.061	0.043	0.003	0.0002	0.0026
Significant ($\alpha=0.05$)	No	Yes	Yes	Yes	Yes
Significant ($\alpha=0.01$)	No	No	Yes	Yes	Yes

As shown in Table 19, the paired-samples t-test results indicate that the comparison between TR1 and TR2 yielded a p-value of 0.061, which slightly exceeds the conventional significance threshold of 0.05, suggesting no statistically significant difference. In contrast, the comparisons for CC1 vs. CC2 ($p = 0.043$), LR1 vs. LR2 ($p = 0.003$), and GA1 vs. GA2 ($p = 0.0002$) all fall below the 0.05 level, indicating statistically significant improvements across these criteria. Among these, the most substantial change was observed in Grammatical Range and Accuracy (GA), which recorded the lowest p-value. Furthermore, the analysis of mean overall scores between Essay 1 and Essay 2 also revealed a statistically significant difference ($p = 0.0026$).

However, to account for the increased likelihood of Type I error due to multiple comparisons, a Bonferroni correction was applied. By dividing the standard alpha level of 0.05 by the number of tests conducted ($n = 5$), a more stringent adjusted significance threshold of 0.01 was obtained. Under this corrected criterion, the previously significant result for CC1 vs. CC2 ($p = 0.043$) no longer reached statistical significance, as its p-value exceeded the adjusted alpha level. The TR1 vs. TR2 comparison remained non-significant under both the original and adjusted alpha levels. The differences for LR1 vs. LR2 ($p = 0.003$), GA1 vs. GA2 ($p = 0.0002$), and the mean overall scores ($p = 0.0026$) remained statistically significant, as all values were well below the corrected threshold.

4.1.3 Correlation analysis between the mean overall scores of Essay 1 and Essay 2

The researcher employed the Pearson correlation coefficient to examine the relationship between the mean overall scores of Essay 1 and Essay 2, using data from Table 10a (Appendices). This statistical method is widely used to assess the strength and direction of a linear relationship between two continuous variables (Cohen et al., 2017a). The Pearson correlation coefficient value ranges from -1 to +1, where values closer to +1 indicate a strong positive correlation, values near -1 suggest a strong negative correlation, and values around 0 denote little to no linear relationship (Cohen et al., 2017b). By employing this test, the study aims to determine whether changes in students' writing scores across the two essays are systematically related. Results are presented in Table 15 below.

Table 20. Pearson correlation coefficient between the mean overall scores of Essay 1 and Essay 2

Correlations				
			MEAN E1	MEAN E2
MEAN E1	Pearson Correlation		1	.759**
	Sig. (2-tailed)			<.001
	N		20	20
MEAN E2	Pearson Correlation		.759**	1
	Sig. (2-tailed)		<.001	
	N		20	20
**. Correlation is significant at the 0.01 level (2-tailed).				

The value of Pearson correlation coefficient between the mean overall scores of Essay 1 and Essay 2, as shown in Table 15, is 0.759, indicating a moderately strong positive relationship between the scores of the two essays. With a p-value <0.001, this correlation is statistically significant. This result suggests that students who performed well in the first essay tended to achieve similarly high scores in the second essay, while those with lower scores in the first essay generally maintained lower scores in the second. This shows that while the tools may improve scores, they are not

substantially changing the assessment of the students as high and low-performing, in relation to one another.

4.2 The individual changes in mean scores

The mean overall score for Essay 1 for each student was calculated by averaging the scores assigned by three raters for Task Response (TR1), Coherence and Cohesion (CC1), Lexical Resource (LR1), and Grammatical Accuracy (GA1), using data in Table 10a from the Appendices. Similarly, the mean overall score for Essay 2 was calculated by averaging the corresponding scores for TR2, CC2, LR2, and GA2 from the same three raters. To observe changes in each student's performance, the mean score of E1 was subtracted from that of E2, as presented in Table 21 below.

Table 21. The mean overall scores of each essay and the mean overall score change of each student

Student ID	Mean overall score Essay 1	Mean overall score Essay 2	Mean overall score changes
1	7.8	7.7	-0.1
2	6.1	6.8	0.8
3	7.3	7.0	-0.3
4	6.3	6.4	0.2
5	7.8	8.2	0.4
6	6.3	6.6	0.3
7	7.0	7.1	0.1
8	6.1	6.8	0.7
9	6.5	7.4	0.9
10	7.3	8.2	0.8
11	6.9	7.3	0.3

12	7.1	6.4	-0.7
13	6.2	6.4	0.3
14	5.3	6.3	1.1
15	6.5	6.8	0.3
16	4.8	5.9	1.1
17	6.5	5.9	-0.6
18	6.3	6.8	0.5
19	5.9	6.3	0.4
20	6.5	6.9	0.4

According to Table 21, the mean overall score changes of 20 students ranged from -0.7 to 1.1, indicating that while some students improved by up to one band, others experienced modest declines of approximately half a band. Four students demonstrated a decrease in their mean overall scores (IDs 1, 3, 12, 17). In contrast, 16 students showed an increase in their mean overall scores (IDs 2, 4, 5, 6, 8, 9, 10, 11, 13, 14, 15, 16, 18, 19, and 20). These results indicate a predominantly positive impact of the four writing tools, as the number of students who achieved higher scores exceeded those who experienced declines.

4.2.1 Relationship between the instances of tool usage and Essay 2 scores

For Essay 2, the 20 student participants were allowed to use up to four tools, though not all made use of every tool. To examine each student's instances of tool usage, the researcher observed screen recordings, examined interview responses and documented instances of usage for each of the four writing tools. All instances of a specific tool used were coded as "1", and non-usage was coded as "0". Table 22 below presents pre-usage essay scores (Essay 1) and post-usage essay scores (Essay 2) from Table 21 as well as the tool usage data for 20 students, including individual and total tool usage.

Table 22. Essay scores, score changes, individual and total tool usage instances of each of 20 students.

Student ID	Essay 1 score	Essay 2 score	Score change	Sketch Engine	Google Translate	Quillbot	Grammarly	Total tool usage
1	7.8	7.7	-0.1	1	0	0	1	2
2	6.1	6.8	0.7	0	1	0	1	2
3	7.3	7	-0.3	1	1	1	1	4
4	6.3	6.4	0.1	0	0	0	1	1
5	7.8	8.2	0.4	1	0	1	1	3
6	6.3	6.6	0.3	1	1	1	0	3
7	7	7.1	0.1	0	1	1	1	3
8	6.1	6.8	0.7	1	1	1	1	4
9	6.5	7.4	0.9	1	1	1	0	3
10	7.3	8.2	0.9	1	1	1	1	4
11	6.9	7.3	0.4	1	1	1	1	4
12	7.1	6.4	-0.7	1	0	0	1	2
13	6.2	6.4	0.2	1	1	1	1	4
14	5.3	6.3	1	0	1	1	1	3
15	6.5	6.8	0.3	1	1	1	1	4
16	4.8	5.9	1.1	0	1	0	1	2
17	6.5	5.9	-0.6	0	1	1	1	3
18	6.3	6.8	0.5	0	1	1	1	3
19	5.9	6.3	0.4	0	1	1	1	3
20	6.5	6.9	0.4	0	1	1	1	3

Next, using data from Table 22, the researcher employed the Pearson correlation coefficient to calculate the correlation between tool usage instances and Essay 2 scores. Results are presented in Table 23.

Table 23. Pearson correlation coefficient between tool usage instances and Essay 2 scores.

Tool	Pearson Correlation coefficient with Essay 2 scores	P-value
Sketch Engine	0.529	0.016
Google Translate	-0.248	0.291
Quillbot	0.200	0.397
Grammarly	-0.074	0.758
Total tools used	0.273	0.244

The correlation analysis between Essay 2 scores and tool usage instances revealed that Sketch Engine had a moderate positive correlation ($r = 0.529$, $p = 0.016$), suggesting a statistically significant relationship with higher scores in Essay 2. This implies that there is a moderate tendency for usage of Sketch Engine to be associated with higher Essay 2 scores. In other words, students who scored higher in Essay 2 tend to use Sketch Engine more than students who scored lower. Google Translate showed a weak negative correlation ($r = -0.248$, $p = 0.291$), suggesting a weak negative relationship with Essay 2 scores, though the relationship was not statistically significant. This implies that there is a slight tendency for higher usage of Google Translate to be associated with lower Essay 2 scores, but the relationship is weak and not statistically significant. In other words, students who scored lower in Essay 2 tend to use Google Translate more than students who scored higher. Quillbot ($r = 0.200$, $p = 0.397$) showed a weak positive correlation with Essay 2 scores, suggesting that its usage had a minimal relationship with Essay 2 scores. Grammarly ($r = -0.074$, $p = 0.758$) exhibited a very weak negative correlation with Essay 2 scores. This implies that there is a slight tendency for higher usage of Grammarly to be associated with lower Essay 2 scores. Finally, the total number of tool usage instances had a weak positive correlation, but was not statistically significant ($r = 0.273$, $p = 0.244$). This finding implies that while there is a slight trend suggesting that students who use more writing tools may achieve higher Essay 2 scores, the weak and non-significant

correlation indicates that tool usage alone is not a decisive factor in writing improvement. Building on this finding, the qualitative differences in how students use these tools will be explored in Section 4.3.

4.2.2 Summary of findings in score changes after the four tools' intervention

In response to the first part of RQ1a: “Does the use of the writing tools result in an increase in the overall writing scores of students?”, the group-level analysis revealed that the intervention of the four writing tools generally led to an improvement in students' mean writing scores for the second set of essays (Essay 2). Specifically, 15 out of 20 essays showed improved overall scores, four experienced declines, and one remained unchanged, as detailed in Table 21. Moreover, the statistical analysis of overall score changes yielded a p-value of 0.0026 (Table 19), confirming that this improvement is statistically significant. The observed changes ranged from a minimum decrease of 0.6 points to a maximum increase of 1.1 points. Pearson correlation coefficient of 0.759 between the mean overall scores of Essay 1 and Essay 2, as detailed in Table 15, reflects a moderately strong positive relationship. The statistical significance of this correlation ($p < 0.001$) highlights that the observed relationship is consistent, suggesting that students who performed well in Essay 1 generally continued to do well in Essay 2, indicating that the assessment of students as relatively high or low-performing does not change even if their scores increase after the intervention of the tools.

In response to the second part of RQ1a, “What aspects of IELTS writing assessment (Task Response, Coherence and Cohesion, Lexical Resource, Grammatical Range and Accuracy) are most affected by students' use of the tools?”, the findings presented in Table 21 indicate that the use of the four writing tools had a generally positive impact across all evaluated criteria, as reflected in the increased minimum scores for TR2, CC2, LR2, and GA2, along with higher maximum scores for TR2 and LR2, and the unchanged maximum score for CC2. However, the maximum scores presented in Table 21 in all four criteria reveal that the intervention of tools was not sufficient to achieve the highest possible Band 9 score in any criterion. After the Bonferroni correction was applied, the paired t-test results of the average scores among three raters (Table 19) reveal significant improvements in two of the four writing criteria. Grammatical range and accuracy (GA) exhibited the most significant improvement,

with a p-value of 0.0002, highlighting the tools' pronounced impact on this criterion of writing. Lexical Resource (LR) also showed meaningful increases, supported by a p-value of 0.003. Coherence and Cohesion (CC) and Task Response (TR) did not show a statistically significant change. These findings emphasise the tools' stronger efficacy on vocabulary and grammar while having less impact on the content and cohesion.

Furthermore, the Pearson correlation coefficient analysis for Sketch Engine, Quillbot, and Grammarly (Table 23) showed that use of Sketch Engine had a moderate positive correlation with Essay 2 scores ($r = 0.529$), indicating a positive relationship of this corpus tool to higher scores in Essay 2, and it is statistically significant ($p = 0.016$). By contrast, Google Translate showed a slight negative relationship with scores, indicating that lower-scoring students tended to use it more, though this was not statistically significant. Quillbot and Grammarly had minimal relations with Essay 2 scores. This shows that students' improvement in IELTS scores through tool usage is not automatic, but is dependent on how the student utilises these tools, which will be explored in Section 3.

The total number of tool usage instances showed a weak positive correlation with Essay 2 scores, though the relationship was not statistically significant ($r = 0.273$, $p = 0.244$). Overall, statistical findings in this section suggests that the integration of the four tools in the set of 20 essays (Essay 2) generally have an overall positive impact on writing scores as a group-level and these increases were notable in the criteria of Lexical Resource (LR) and Grammar range and accuracy (GA). However, the assessment of students as relatively high or low-performing does not change even if their scores increase after the intervention of the tools. Among the four tools analysed, only Sketch Engine demonstrated a statistically significant positive correlation with higher Essay 2 scores, suggesting a potential relationship between its use and improved writing performance. However, Quillbot and Grammarly showed minimal relationships with Essay 2 scores, and Google Translate exhibited a slight negative correlation. Additionally, the total number of tool usage instances had a weak and non-significant positive correlation with Essay 2 scores, indicating that simply using more tools does not necessarily lead to higher scores. Since the quantitative results suggest that overall tool usage does not directly and automatically determine score improvements, it is crucial to examine how these 20 student participants, whose

different levels of English proficiency, as indicated by their official IELTS scores and their Essay 2 scores, engage with these tools. This cognitive engagement with tools will be explored further in Section 4.3.

4.3 Analysis of tool engagement

Since the quantitative findings in Section 4.2 indicated that the instances of tool usage did not directly correlate with writing score improvements—suggesting that score gains in Essay 2 were not simply a result of using the available tools—the researcher further examined how students engaged with the four writing tools, with particular attention to potential differences between high-scoring and low-scoring students.

This section presents an exploratory qualitative analysis of student responses from the interview transcripts of 20 students, aiming to address RQ1b: “How do high-scoring and low-scoring students differ in their ability to engage with the tools and use them to support their language development?”. To facilitate this analysis, students were classified into high-scoring and low-scoring groups based on their Essay 2 scores (Section 4.3.1). Subsequently, thematic analysis was applied to interview transcripts from all 20 participants, focusing on how each student critically assessed tool-generated outputs and described their language development, including new vocabulary acquisition and grammatical accuracy improvements after using the tools to compose Essay 2 (Section 4.3.2). After establishing the thematic findings, the researcher compared responses from high- and low-scoring groups to identify key differences and examine potential links between Essay 2 scores and students’ cognitive engagement with the four tools (Section 4.3.3). A summary of these exploratory qualitative findings is presented in Section 4.3.4.

4.3.1 Classification of language proficiency levels

To compare high-scoring and low-scoring students, the researcher classified student proficiency levels using a criterion-referenced approach aligned with the IELTS proficiency scale. According to the UK Home Office (2024), international students applying for a UK Student Visa must demonstrate English proficiency at the CEFR B2 level, equivalent to IELTS scores ranging from 5.5 to 6.5 (Council of Europe, 2020; UK Home Office, 2024). Additionally, an IELTS score of 6.5 is commonly recognised as the minimum requirement for entry into undergraduate and postgraduate programs at

English-speaking institutions (British Council, 2021; IELTS Asia, n.d.). Based on these proficiency guidelines and the scores of their Essay 2 (as presented in Table 16), students in this study were classified into two groups: the low-scoring group (Essay 2 scores below 6.5; n=7, IDs: 4, 12, 13, 14, 16, 17, 19) and the high-scoring group (Essay 2 scores from 6.5 and above; n=13, IDs: 1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 15, 18, 20).

4.3.2 Thematic analysis of interview data

To gain deeper insights into how high-scoring and low-scoring students (in relation to their Essay 2 scores) interact with the four tools during the writing process and their impact on writing improvement, the researcher conducted an exploratory thematic analysis of interview transcripts from 20 student participants. The thematic analysis follows Braun and Clarke's (2006) six-phase approach: (1) familiarisation with the data, (2) generating initial codes, (3) searching for themes, (4) reviewing themes, (5) defining and naming themes, and (6) producing the report. The analysis begins with a thorough review of all transcripts to identify key themes. Each transcript was read multiple times to ensure a comprehensive understanding of the data. Open coding was then employed to systematically annotate meaningful excerpts that reflected a student's behavioural and cognitive engagement with the text, as well as any particular phrases that students discuss their language learning, such as learning new vocabulary or recognising grammar errors. These codes were subsequently organised into three broader themes based on similar patterns: Cognitive Engagement (CE), Vocabulary Acquisition (VA), and Grammar Improvement (GI). The researcher then recorded the specific tools referenced within each response under the identified themes, facilitating a detailed comparative analysis of the four tools across both high- and low-scoring students. Each theme instance is the number of times in the transcript that the student discusses their usage of a particular tool.

Cognitive Engagement (CE) theme refers to instances where students actively read and assess the tool-generated language or suggestions before incorporating them into their writing. For example, when Student 1 described their experience when using Sketch Engine to search for more words, he stated: "I also used Sketch Engine to help me find more words related to 'top-notch' and then among the collocations, I chose 'world-class' to complete the phrase 'world-class education'. I think this word sounds

more suitable compared to other high-frequency words like 'matchless', 'triple-A', or 'first-rate'." This response demonstrates how this student critically considered different lexical options before selecting "world-class" based on contextual appropriateness, thus aligning with the Cognitive Engagement (CE) theme. Sketch Engine is the tool mentioned in Student 1's response, which the student also demonstrated the ability to critically assess.

Vocabulary Acquisition (VA) refers to responses demonstrating students' acquisition of new vocabulary after using a tool. For example, when Student 9 answered: "I was struggling to remember the English word for 'ban hành luật', and Google Translate helped me solve this problem, now I know it is 'promulgate'." (Student 9), the response shows that this student learned a new word "promulgate", and the tool that was mentioned by this student for this particular theme is Google Translate.

Grammar Improvement (GI) refers to instances where students identify grammatical errors and apply tool-generated suggestions to enhance the accuracy and clarity of their writing. For example, Student 1 stated: "Grammarly helped me correct errors in spelling, prepositions, and articles, such as changing 'with affordable prices' to 'at affordable prices' and 'choose the suitable time' to 'choose a suitable time'". This response highlights how Student 1 used Grammarly, then recognised grammatical errors and accepted corrections. Table 24 below presents the three themes, their corresponding coding colours, an example response, and an example of tools used in each theme.

Table 24. The coding by colour and theme for the 20 interview transcripts and examples of responses.

Code themes	Example	Tools used in each theme
Cognitive Engagement (CE)	<i>I also used Sketch Engine to help me find more words related to 'top-notch,' and then among the collocations, I chose 'world-class' to complete the phrase 'world-class education.' I think this word sounds more suitable compared to other high-frequency words like 'matchless,' 'triple-A,' or 'first-rate.'"</i>	Sketch Engine
Vocabulary Acquisition (VA)	<i>"I was struggling to remember the English word for 'ban hành luật', and Google Translate helped me solve this problem, now I know it is 'promulgate'."</i>	Google Translate
Grammar Improvement (GI)	<i>"Grammarly helped me correct errors in spelling, prepositions, and articles, such as changing 'with affordable prices' to 'at affordable prices' and 'choose the suitable time' to 'choose a suitable time.'"</i>	Grammarly

Full thematic coding for the interview transcripts of all 20 students is presented in Table 24a in the Appendices. Table 25 presents an example of the thematic coding from the interview transcript of Student 1:

Table 25. Example of the thematic findings from Student 1's interview transcript

STUDENT 1

1/ In general, how do you feel about the usage of the four AI writing tools for writing this IELTS essay? Do you find them helpful?

I find Grammarly and Google Translate helpful for checking synonyms, grammar, and spelling.

2/ Overall, do you think you produced a higher quality essay with the tools compared to the first essay that you wrote without the tools? Which part or sentence do you think you wrote better with the tools?

Yes, I believe I produced a higher quality essay with the tools, particularly in terms of grammar and spelling, which were improved using Grammarly. Additionally, I replaced synonyms using SketchEngine and Google Translate.

3/ In the brainstorming and planning step (before writing), did you find the tools useful for you in this step? Which tool helped you the most? How did it help you?

I did not need any tools during the brainstorming and planning step because the topic was quite common and simple. I can write in English with simple words first. Then I need to replace the words that are too simple.

4/ In the composing step, did you find the tools useful? Which tool helped you the most? How did it help you?

Quillbot was not working during my composing time so I had to find synonyms suggested by Google Translate. I can write in English with simple words first. Then I need to replace the words that are too simple. For example, I didn't feel that the phrase "increased relaxation time" was correct, so I checked the meaning in Vietnamese, and I found more synonyms of this word suggested by Google Translate. Among them, I picked the word "extend" and modified my phrase to "extended relaxation time". I also used Sketch Engine to help me find more words related to "top-notch", and then among the collocations, I chose "world-class" to complete the phrase "world-class education". I think this word sounds more suitable compared to other high-frequency words like "matchless", "triple-A", or "first-rate".

5/ In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?

In the revising and editing step, Grammarly was particularly useful for checking overall spelling and grammar. Additionally, I used Sketch Engine to affirm the usage of terms like "real-life interactions". Another word was the word "re-interact" from the phrase "re-interacting with their peers", which was my term, but I am not confident in using it. The Vietnamese meaning of this word seemed accurate when it was translated by Google Translate, so I used this word, and also because there were no other suggestions from the tool for this word. Grammarly is also helpful in correcting my errors in spelling, prepositions, and articles, such as "with affordable prices" to "at affordable prices", and "choose the suitable time" should be "choose a suitable time".

6/ Can you give me some examples of challenges that the tools could not help you in this writing task?

The tools could not help me find results for "re-interact" in Sketch Engine, Google Translate, and QuillBot. I wanted to paraphrase the sentence, "Moreover, online learning offers students flexibility as they can choose a suitable time for their education, boosting their productivity and learning effectiveness" using a different, more formal structure but I did not succeed. And as I said, Quillbot was not working during my composing time.

7/ What do you want more from these four tools: SketchEngine, Google Translate, QuillBot, and Grammarly?

I would like Grammarly to evaluate coherence and cohesion. I wish Google Translate could find synonyms for phrases. Sketch Engine is a bit difficult to use and does not always generate correct synonyms. Some collocations suggested by the tools did not fit the context. Additionally, I wanted to evaluate the range of vocabulary (level of vocabulary based on CEFR) and suggest some collocations that fit the context and topic of education.

8/ Do you recommend or would you use any other tools for this IELTS essay writing practice? Why?

I recommend using ChatGPT to brainstorm more arguments and an AI-based IELTS essay scoring tool for evaluating cohesion and coherence.

According to Table 25, the response to Interview Question 5 reveals two instances of Cognitive Engagement (CE). In particular, when using Sketch Engine, Student 1 stated: “I can write in English with simple words first. Then I need to replace the words that are too simple. For example, I didn’t feel that the phrase ‘increased relaxation time’ was correct, so I checked the meaning in Vietnamese, and I found more synonyms of this word suggested by Google Translate. Among them, I picked the word “extend” and modified my phrase to “extended relaxation time” (Interview Question 4), showing that Student 1 actively assessed the output of Google Translate, and modified the phrase to “extended relaxation time”, instead of passively copying the output. This reflects clear evidence of behavioural engagement by Student 1 because he used Google Translate as a tool for lexical enhancement rather than for direct translation. Moreover, when using Sketch Engine, Student 1 demonstrated behavioural engagement by exploring multiple collocations for the term “top-notch” before selecting “world-class”, as he answered Interview question 4: “I also used Sketch Engine to help me find more words related to ‘top-notch,’ and among the collocations I chose ‘world-class’ to complete the phrase ‘world-class education.’ I think this word sounds more suitable compared to other high-frequency words like ‘matchless’, ‘triple-A’, or ‘first-rate’.” This demonstrates the student’s cognitive engagement of lexical choices rather than passively or randomly accepting other tool-generated suggestions such as “matchless”, “triple-A”, or “first-rate”. Another instance was observed in Interview Question 5, as Student 1 stated:

[1] I used Sketch Engine to affirm the usage of terms like ‘real-life interactions.’ Another word was ‘re-interact’ from the phrase ‘re-interacting with their peers,’ which was my term, but I was not confident in using it. The Vietnamese meaning of this word seemed accurate when it was translated by Google Translate, so I used this word.

The excerpt of response [1] demonstrates an example of cognitive engagement, as the student used Sketch Engine to confirm the meaning and contextual appropriateness of the phrase “real-life interactions”, rather than relying on the tool for content generation. Notably, when uncertain about the self-created term “re-interact”,

the student cross-verified its meaning using another tool—Google Translate—and made a context-based decision regarding its usage.

The instance of grammar improvement (GI) was also noted when Student 1 acknowledged the effectiveness of Grammarly in improving grammar, particularly in correcting spelling, prepositions, and articles. Moreover, Student 1 specifically highlighted how Grammarly helped refine phrases such as "with affordable prices" to "at affordable prices" and "choose the suitable time" to "choose a suitable time", demonstrating their recognition of errors in prepositions, article usage, and overall grammatical accuracy. However, there is no clear evidence that they gained new vocabulary after using the four tools. Throughout the interview, he described their use of Google Translate and Sketch Engine primarily for synonym replacement and word refinement rather than learning new words. Therefore, Student 1's responses suggest a focus on refining his existing linguistic knowledge rather than acquiring new words. Overall, the data indicates that Student 1, a high-proficiency student in the high-scoring group, actively engaged with and critically assessed the outputs of Sketch Engine, Google Translate, and Grammarly for lexical refinement and grammar accuracy improvement, but did not learn new vocabulary from these tools. The thematic analysis instances for Student 1 are noted as "1" for Cognitive Engagement (CE), "0" for vocabulary acquisition (VA) and "1" for grammar improvement (GI).

4.3.3 Comparison of thematic findings from the high-scoring and low-scoring

After conducting thematic analysis on interview transcripts from 20 students using the established framework, the researcher summarised and compared the number of these theme instances between the high-scoring group (n=13) and the low-scoring group (n=7). Table 26 below provides a summary of the total instances of each theme, the specific tools mentioned in each theme, and the number of students mentioning each tool. A theme instance refers to the number of times in the transcript where a student discusses using a specific tool. Cognitive Engagement (CE) indicates the number of times a student shows both behaviour and thinking—e.g., exploring functions, cross-checking outputs, accepting/rejecting suggestions, assessing outputs, noticing gaps, comparing options, and justifying choices of a particular tool. Vocabulary acquisition (VA) indicates students' reports of learning new lexis (synonyms, collocations) after using a particular tool. Grammar improvement (GI)

indicates students' reports of noticing a grammatical issue and using a tool to resolve it. This quantitative comparison serves only as an exploratory overview due to the different group sizes.

Table 26. A summary of the number of theme instances and the specific tool mentioned in each theme between the low-scoring and high-scoring groups.

Coding theme	High-scoring group (scored ≥ 6.5 in Essay 2) (n=13)	Low-scoring group (scored < 6.5 in Essay 2) (n=7)
Cognitive Engagement (CE)	Total instances: 11 Google Translate (6), Quillbot (7), and Sketch Engine (7)	Total instances: 4 Quillbot (3) and Sketch Engine (1)
Vocabulary Acquisition (VA)	Total instances: 9 Google Translate (5) and Quillbot (4)	Total instances: 4 Google Translate (4)
Grammar Improvement (GI)	Total instances: 10 Refined grammar accuracy by using Grammarly (9) and Quillbot (1)	Total instances: 6 Refined grammar by using Grammarly (6)

Using data from Table 26, the researcher compares the thematic analysis of responses from each group across the four established themes: Cognitive Engagement (CE), Vocabulary Acquisition (VA), and Grammar improvement (GI).

Cognitive Engagement (CE)

The high-scoring group recorded 11 instances of the Cognitive Engagement (CE) theme from 11 out of 13 students (IDs 1, 3, 5, 6, 7, 8, 9, 10, 11, 18, and 20), whereas two students (IDs 2 and 15) did not self-report. The three tools that were mentioned by students in this group are Google Translate (6), Quillbot (7), and Sketch Engine (7).

The responses from 11 high-scoring students show that they carefully assessed each output and considered the contextual appropriateness of the language in their Essay 2. For example, the interview extract [2] below showed Student 1's indications of both behavioural and cognitive engagement:

[2] For example, I didn't feel that the phrase "increased relaxation time" was correct, so I checked the meaning in Vietnamese, and I found more synonyms of this word suggested by Google Translate. Among them, I picked the word "extend" and modified my phrase to "extended relaxation time". I also used Sketch Engine to help me find more words related to "top-notch", and then among the collocations, I chose "world-class" to complete the phrase "world-class education". I think this word sounds more suitable compared to other high-frequency words like "matchless", "triple-A", or "first-rate" (Student 1)

Student 1 initially questioned the phrase "increased relaxation time" and used Google Translate to explore its Vietnamese meaning, reviewed its synonyms, and then selected "extended relaxation time" as the most suitable option. The student also used Sketch Engine to refine lexical choices, choosing "world-class" over options like "matchless," "triple-A," and "first-rate" for the phrase "world-class education" based on contextual suitability. When Student 1 compared multiple outputs from Google Translate to identify the most contextually appropriate option for his essay, this indicates both behavioural and cognitive engagement. Taken together, his interview response demonstrates behavioural engagement (querying, cross-checking, revising) alongside cognitive engagement (evaluating contextual fit, weighing alternatives, and justifying selections) with Google Translate and Sketch Engine.

Likewise, Students 6, 7, 9, 11, and 18 reflected in their interview response (excerpts [3], [4], [5], [6], [7]) on how they assessed and considered the tools' outputs for contextual appropriateness before incorporating them into their Essay 2, rather than accepting all suggestions instantly:

[3] At first, I used Quillbot to paraphrase the phrase "distance learning" from the topic and I got the synonym phrase "remote learning". Then I put both "distance learning" and "remote learning" in Sketch Engine to search for relevant concordances. From that, I can access other sentences about distance learning. I picked out some other good phrases that I can use, which are "social distancing", "faculty and staff", and "academic facilities". (Student 6)

[4] I composed most of my writing without any tools. I only used Google Translate to find the English equivalent of the Vietnamese word "ủng hộ", then

I got the word “support” and “favor”, and I chose “support” because it fits more with the context of expressing an opinion. (Student 7)

[5] In the revising and editing step, Quillbot was the most helpful tool for checking my work. [...] After a few trials, I find that some phrases are even better than my original words, such as “mitigate”, “alleviate the negative environmental implications” and “pose ecological challenges”, which seem to be more formal than my original phrase “reduce the adverse environmental impacts” and “have potential ecological risks”. (Student 9)

[6] I used Quillbot to paraphrase the introduction, as it provided useful synonyms to enhance my writing. For example, it suggested changing “people’s lives” to “human’s life”, “many people” to “many individuals”, and “people can take advantage of it” to “it may be used indefinitely by humans”, and I agree with these suggestions, I think they are better than my own writing, they sound more formal and suit with the context of the environment. (Student 11)

[7] Quillbot is also quite useful. For example, it helps me to paraphrase from “Countries with less resources are put in the lower hand in negotiation and slower in development” to “Countries with fewer resources are given a weaker negotiating position and develop more slowly.”. [...]. Also, the phrase “develop more slowly” sounds more formal than my original phrase “slower in development”. (Student 18)

Moreover, examples of responses from the high-scoring group also show that high-scoring students used a combination of tools to verify their own word choices or the output of another tool, for example, using Google Translate to review suggestions from Quillbot or Sketch Engine. This combined use of tools is clearly reflected in the responses of Students 1, 6, 9, and 10, who described how they used Quillbot, Google Translate, and Sketch Engine in combination in their writing process:

[8] Additionally, I used Sketch Engine to affirm the usage of terms like “real-life interactions”. Another word was the word “re-interact” from the phrase “re-interacting with their peers”, which was my term, but I am not confident in using it. The Vietnamese meaning of this word seemed accurate when it was translated by Google Translate, so I used this word. (Student 1)

[9] In the brainstorming and planning step, I found the “Concordance function” of Sketch Engine and Quillbot quite useful. At first, I used Quillbot to paraphrase the phrase “distance learning” from the topic and I got the synonym phrase “remote learning”. Then I put both “distance learning” and “remote learning” in Sketch Engine to search for relevant concordances. From that, I can access other sentences about distance learning. I picked out some other good phrases that I can use, which are “social distancing”, “faculty and staff”, and “academic facilities”. (Student 6)

[10] I usually have a difficult time rewriting my own sentences, but Quillbot outputs helped me to do it faster and explore more synonyms. After a few trials, I find that some phrases are even better than my original words, such as “mitigate”, “alleviate the negative environmental implications” and “pose ecological challenges”, which seem to be more formal than my original phrase “reduce the adverse environmental impacts” and “have potential ecological risks”. I also tried the word “mitigate” in Sketch Engine to check for its collocations. After checking with Sketch Engine, I found the noun “risk” was the first collocation that can combine with the verb “mitigate”, so I chose the phrase “mitigate risk”. (Student 9)

[11] I tried Sketch Engine to find collocations and it suggested a collocation “tacitly” for “agree”. But after checking it with Google Translate, I found that this word doesn’t fit in my sentence, as I want to have a strong adverb with “agree”, such as “wholeheartedly” rather than “tacitly” so I skipped it and did not use this tool anymore. (Student 10)

High-scoring students (IDs 3, 5, 10, and 20) further displayed both behavioural and cognitive engagement by discarding tool-generated language—particularly from Google Translate, Quillbot, and Sketch Engine—when they were not contextually suitable:

[11] Quillbot was not very useful for me when I used it to look for synonyms, as the phrase “harmful by-products” that I tried with Quillbot did not result in better words. (Student 3)

[12] Additionally, I used Sketch Engine's Word Sketch to find more verbs related

to “attention” but I did not find a relevant verb combination among its suggested, such as “turn”, “shift”, “focus” or “flatter”, so I composed my own phrase “hold the attention”. (Student 5)

[13] I also adopted Quillbot’s version of the first sentence: “Some may object to this shift, citing the enormous expenses required to develop the necessary infrastructure and technology to implement it globally.”. I retained the following sentence as it was. Additionally, I changed “unavoidable” to “inevitable” and “making it essential” to “necessitating”. However, instead of using Quillbot’s suggestion of “immediate action”, I kept my own phrase, “prompt action”. (Student 10)

[14] I think the texts generated by Quillbot sound more formal to me in general, compared to those of Google Translate. I accepted all of its output, except for the last sentence of the conclusion: “Overall, online learning is an inexorable advancement of technology and time, therefore we must learn to accept it and find solutions to the remaining issues. It still has a lot of strengths and disadvantages that need to be analyzed.”. The last sentence: “It still has a lot of strengths and disadvantages that need to be analyzed.” I think is a bit redundant, so I did not include it. (Student 20)

Compared with the high-scoring group (11 out of 13), the low-scoring group reported fewer instances of critical assessment (4 out of 7), from Students 12, 13, 17, and 19, mentioning the use of Quillbot (3) and Sketch Engine (1). This pattern indicates weaker evaluation of tool-generated output and different usage patterns among lower-scoring students.

While high-scoring students (e.g., IDs 1, 6, 7, 9, 11, and 18) showed consideration of whether a tool’s suggestions were contextually appropriate for Essay 2, the four low-scoring students, who showed cognitive engagement (IDs 12, 13, 17 and 19), did so in a more surface-level manner. Their discussion is also shorter and less detailed. For example, Student 12 demonstrated a more surface-level cognitive engagement with the output of Sketch Engine:

[15] I tried using Sketch Engine to check whether the phrase “exhausted fumes” is used in a corpus, and it provided me with some results, so I kept this phrase.
(Student 12)

According to this interview extract, Student 12 relied solely on the “exhausted fumes” phrase from Sketch Engine, which appeared in the concordance results, and accepted it without further verification, showing a more superficial consideration than those responses of the high-scoring group.

The critical assessment theme identified in the response of Student 17 also reveals a surface-level comparison:

[16] I tried using Quillbot one time. It helped me paraphrase the sentence from “People on earth develop once they are healthy” to “On earth, once a person is healthy, they develop”, I think it sounded better than my previous writing.
(Student 17)

When stating “I tried using Quillbot one time...I think it sounded better...”, this response indicates that the Student 17’s focus is primarily on the instant refinement of sentence, without deeper exploration of vocabulary or semantic context.

Similarly, Student 19 also concentrated on instant refinement of a paragraph:

[17] For Quillbot, I used it to paraphrase the last paragraph, but I only changed these sentences: “On the other hand, renewable energy is still too new for people to get used to, to be completely honest. Despite the incorporation of renewable energy into daily life, fossil fuels still have a significant influence on many of our activities.” because I think this output contains better word choice and phrases than my original sentences, particularly the phrase “the incorporation of renewable energy into daily life” is better than my previous sentence. (Student 19)

The response of Student 19 mentioning phrases like “the incorporation of renewable energy into daily life” as preferable to their original wording, indicates that the assessment remained limited to surface-level improvements. In other words, these students talk about one phrase “sounding better” than another, without really

explaining why.

However, there was a notable exception in Student 13, who assessed more critically with the outputs from Sketch Engine to consider the context appropriateness of her language input:

[18] I used Sketch Engine to check if my phrases “CO2 emissions” and “green energy” were correct. I saw that both phrases appeared in concordance examples with other technical words like “generating power” and “greenhouse emissions”. That showed me they fit the topic, so I decided to use both in my essay. (Student 13)

Unlike most students in the low-scoring group, she actively assessed the contextual appropriateness of her lexical choices using Sketch Engine. Specifically, she used the tool’s concordance function to verify whether phrases like “CO2 emissions” and “green energy” were suitable for her topic. By observing that these terms frequently co-occurred with other technical expressions such as “generating power” and “greenhouse emissions” in authentic usage examples, she concluded that her chosen phrases aligned well with the academic context of her essay. This example illustrates a more reflective and context-aware approach to tool usage, which stood out among responses from the low-scoring group.

Second, high-scoring students demonstrated more critical assessment by using multiple tools in combination—for example, verifying Quillbot’s suggestions with Google Translate or cross-checking word usage through Sketch Engine. Students such as 1, 6, 9, and 10 described how they integrated Quillbot, Google Translate, and Sketch Engine to refine their language choices, while others (e.g., Students 3, 5, 10, and 20) actively rejected tool-generated suggestions they deemed contextually inappropriate. In contrast, low-scoring students tended to rely on a single tool at a time and accepted its output without additional verification or cross-referencing.

[19] I tried using Sketch Engine to check whether the phrase “exhausted fumes” is used in a corpus, and it provided me with some results, so I kept this phrase. (Student 12)

[20] In the composing step, I found Quillbot useful for paraphrasing the

"renewable sources" to "alternative energy" and I chose its suggestions.
(Student 13)

[21] These tools are so useful for fast writing within 45 minutes. I can write in Vietnamese, and then use Google Translate, check with Grammarly, and use Quillbot to produce an essay. But I didn't have time to use Quillbot. With Google Translate helping me with my whole essay I don't even have to think in English and I think its translation quality is better than my English level [...] Google Translate and Grammarly helped me to produce an essay that is faster and has no grammar or spelling errors. (Student 16)

[22] Among these four tools, I only need Google Translate the most because I need a dictionary to search for words that I do not know in English, such as "tsunami", "compete", "mineral" and "instability". (Student 17)

These qualitative findings on differences in critical assessment between high- and low-scoring students provide deeper insight into the quantitative results presented in Section 4.2.1 regarding the usage of Google Translate. The Pearson correlation analysis (Table 23) of Google Translate showed a weak negative correlation ($r = -0.248$, $p = 0.291$), suggesting a weak negative relationship with Essay 2 scores, implying that there is a slight tendency for higher usage of Google Translate to be associated with lower Essay 2 scores. The qualitative data provides more insights to this trend by revealing that this use was often uncritical, with limited attention to the contextual appropriateness of the tool's suggestions. In contrast, high-scoring students engaged more reflectively, often verifying outputs across multiple tools and rejecting inappropriate suggestions.

Another notable difference lies in how a high-scoring student (ID 6) and a low-scoring student (ID 12) articulated their use of Sketch Engine:

[23] In the brainstorming and planning step, I found the "Concordance function" of Sketch Engine and Quillbot quite useful. At first, I used Quillbot to paraphrase the phrase "distance learning" from the topic and I got the synonym phrase "remote learning". Then I put both "distance learning" and "remote learning" in Sketch Engine to search for relevant concordances. From that, I can access other sentences about distance learning. I picked out some other good phrases

that I can use, which are “social distancing”, “faculty and staff”, and “academic facilities”. (Student 6)

[24] I tried using Sketch Engine to check whether the phrase “exhausted fumes” is used in a corpus, and it provided me with some results, so I kept this phrase. (Student 12)

Student 6, a high-scoring participant, demonstrated strategic use of multiple tools during the brainstorming stage, using Quillbot to paraphrase “distance learning” into “remote learning,” and then employing Sketch Engine’s Concordance function to explore both terms in context. This allowed the student to extract relevant academic phrases that appear in texts about distance learning, such as “social distancing”, “faculty and staff” and “academic facilities”. These critical assessments reflect the student’s cognitive engagement. In contrast, Student 12, a low-scoring participant, used Sketch Engine in a limited way, merely checking if the phrase “exhausted fumes” appeared in the corpus and accepting it without further analysis. This response lacks the depth of critical assessment.

This difference in how high- and low-scoring students engaged with Sketch Engine further explains the quantitative results calculated in Section 4.2.1. The Pearson correlation analysis (Table 23) showed a moderate and statistically significant positive correlation between Sketch Engine usage and Essay 2 scores ($r = 0.529$, $p = 0.016$), suggesting that students who used this corpus tool tended to achieve higher writing scores. The in-depth, strategic engagement demonstrated by high-scoring students—such as Student 6, who combined Quillbot and Sketch Engine during the brainstorming stage to explore synonyms, collocations, and contextually appropriate academic phrases—illustrates the way of using Sketch Engine that likely contributed to Student 6’s high score outcome. In contrast, the more limited use observed in the low-scoring student—such as Student 12’s reliance on Sketch Engine merely to verify the presence of a phrase—suggests that limited engagement would result in lower score outcomes. These contrasting approaches support the statistical finding by highlighting that it is not merely the instance of using Sketch Engine that correlates with higher performance, but rather the quality of engagement with the tool.

In summary, the exploratory thematic analysis on Cognitive Engagement (CE)

highlights differences between high- and low-scoring students in their assessment of the four writing tools. The responses from 11 high-scoring students (IDs 1, 3, 5, 6, 7, 8, 9, 10, 11, 18, and 20) reflected the theme of cognitive engagement, as they carefully evaluated the contextual appropriateness of tool-generated language (e.g., Students 1, 6, 7, 9, 11, and 18), actively rejected unsuitable suggestions (e.g., Students 3, 5, 10, and 20)—and used multiple tools in combination to cross-verify both their original language input and outputs generated by the tools (e.g., Students 1, 6, 9, and 10). In contrast, although four out of seven low-scoring students (IDs 12, 13, 17, and 19) also demonstrated a theme of cognitive engagement in their response, their language assessment was less critical as they did not consider the contextual appropriateness of the output and instantly accepted outputs without rejecting or using other tools to verify them.

Vocabulary Acquisition (VA)

In the high-scoring group, 9 out of 13 students discussed aspects of their usage that can be analysed as attending to the theme of vocabulary acquisition. Among the tools used, Google Translate facilitated vocabulary learning in five instances, while Quillbot was mentioned in four instances. The following excerpts demonstrate how students in this group engaged with these tools to improve their vocabulary:

[25] I know one of the words to describe “cận thị” is “being short-sighted”, but from Google Translate I learned another word that is “myopic”. (Student 2)

[26] For composing, I mostly needed Google Translate to look for some phrases in English, such as the phrase “petrol-powered car”, “coal”, and “burning coal releases CO2” and I learned these new words. (Student 3)

[27] Yes, in the composing step, I found the tools useful. I used Quillbot to find synonyms of “distance learning”, and I learned that there are more words that can replace my phrase, those are “online learning”, “remote learning” and “distance education”. (Student 5)

[28] Quillbot also helped me paraphrase lots of my sentences, for example, from “I totally agree with this standpoint for the following reasons.” to “I completely concur with this viewpoint for the reasons listed below.”, and I

learned that “concur” is a synonym of “agree”. Quillbot also helped me rewrite the order and structure of this sentence: “First, because it does not emit as much hazardous exhaust as fossil fuels do, recyclable energy can help to prevent the environment from becoming increasingly polluted.”. I also learn more words such as: “entirely”, “conventional energy sources”. (Student 7)

[29] Additionally, for the sentence: "For example, young learners find it hard to concentrate fully on the lesson because they can stop themselves from using social media or playing games while studying at a distance", Quillbot helped me learn that "feel it hard" can be paraphrased as "find it hard". (Student 8)

[30] I used Google Translate and Quillbot. Google Translate helped me to remember words in this context, such as “ecological risk”, “wind propeller”, “indestructible”, and “the most optimal energy source”. (Student 9)

[31] Google Translate also helped me learn two new words. From using it, I learned that the translation of the phrase “vô tận” is “inexhaustibility”, “trữ lượng” is “reserves”. (Student 11)

[32] I also used Quillbot to improve my body paragraphs after completing my essay. It allowed me to modify words I thought were not suitable and see more options. For example, I learned that “successfully implement the usage of these energies into real life” can be paraphrased as “properly execute the use of these energies in real life” and I learned the new phrase “properly execute” I also discovered that “alter” is a synonym for “change” and that “much cheaper” can be rewritten as “significantly cheaper”. (Student 15)

[33] I remember the word “landbound” while I was composing, but I wasn’t sure if it was the correct meaning, so I needed Google Translate. It also helped me realise that my word “hydro power” was wrong, and it was “hydroelectric energy”. (Student 18)

These examples of responses contain phrases such as “I learned” (Students 2, 3, 5, 11, 15), “helped me learn” (Student 11), “I also discovered that” (Student 15), and “It also helped me realise that...” (Student 18), which suggest that the high-scoring students’ use of tools went beyond task completion and contributed to their language

development—particularly in vocabulary acquisition. In addition, what appears to distinguish this group is their tendency to use tools like Sketch Engine, Google Translate and Quillbot not only to identify unknown English words but also to acquire synonyms of their self-written words or phrases, alternative expressions, and more appropriate collocations. For example, Student 2 learned the more formal synonym “myopic” as an alternative to his phrase “being short-sighted” after using Google Translate. Student 7 learned the term “concur” as a synonym for “agree” and also learned more synonyms from her original sentence input after using Quillbot. Similarly, Student 5 discovered synonymous phrases for “distance learning”, including “online learning”, “remote learning” and “distance education” after using Quillbot. Student 7 explicitly mentioned using Quillbot to restructure her self-written sentence while also acquiring new vocabulary such as “entirely” and “conventional energy sources”. Student 8 acquired new synonyms from Sketch Engine, such as “home study” and “open learning courses”, as well as improving her collocation from “feel it hard” to “find it hard” after using Quillbot. Student 15 acquired new vocabulary from Quillbot’s suggestions, such as learning that her self-written phrase “successfully implement” could be paraphrased into the more formal and precise expression “properly execute.”

Within the low-scoring group, four instances of vocabulary acquisition were identified, each from a different student (IDs 13, 14, 17, and 19), all of whom referenced the use of Google Translate. Three other students in the low-scoring group (IDs 4, 12, and 16) did not articulate any explicit instances of vocabulary learning during their interview, nor did they use reflective expressions such as “I learned”, “I found out”, or “I discovered”. The four instances of vocabulary acquisition in the low-scoring group are presented in the following responses:

[34] I also used Google Translate to find words that I don't know in English. For example, I needed to translate the word “bệnh hô hấp” and I learned that it is “respiratory disease”. (Student 13)

[35] For this step, I only used Google Translate to check for English words and my spelling. I translated the word “thúc đẩy” and I learned that it is “promote”, and “chểnh mạnh” is “negligence”. (Student 14)

[36] Google Translate was the most helpful tool during the composing step.

Among these four tools, I only need Google Translate the most because I need a dictionary to search for words that I do not know in English, such as “tsunami”, “compete”, “mineral” and “instability”. (Student 17)

[37] I was struggling to find the English word for “ban hành luật”, and Google Translate helped me solve this problem, now I know it is “promulgate”. (Student 19)

Accordingly, expressions from these low-scoring group’s responses, such as “I learned that it is...” (Students 13 and 14) and “Google Translate helped me solve this problem, now I know it is...” (Student 19), are similar to the high-scoring group, suggesting that these four low-scoring students also acquired new vocabulary through their interactions with the tool. However, their approach to vocabulary learning appears different from that of the high-scoring group. In particular, some high-scoring students (IDs 1, 5, 7 8, and 15) often used the tools to find more synonyms from their own English words:

[38] I know one of the words to describe “cận thị” is “being short-sighted”, but from Google Translate, I learned another word that is “myopic” (Student 1)

[39] Yes, in the composing step, I found the tools useful. I used Quillbot to find synonyms of “distance learning”, and I learned that there are more words that can replace my phrase, those are “online learning”, “remote learning” and “distance education” (Student 5)

[40] Quillbot also helped me paraphrase lots of my sentences, for example, from “I totally agree with this standpoint for the following reasons.” to “I completely concur with this viewpoint for the reasons listed below.”, and I learned that “concur” is a synonym of “agree” (Student 7)

[41] Additionally, for the sentence: “For example, young learners find it hard to concentrate fully on the lesson because they can stop themselves from using social media or playing games while studying at a distance”, Quillbot helped me learn that “feel it hard” can be paraphrased as “find it hard” (Student 8)

[42] For example, I learned that “successfully implement the usage of these

energies into real life” can be paraphrased as “properly execute the use of these energies in real life” and I learned the new phrase “properly execute” I also discovered that “alter” is a synonym for “change” and that “much cheaper” can be rewritten as “significantly cheaper” (Student 15)

In contrast to those abovementioned responses in the high-scoring group, the four low-scoring students (Students 13, 14, 17, and 19) engaged with vocabulary in a more limited and superficial manner, relying almost entirely on Google Translate as a dictionary to translate unfamiliar Vietnamese words into English. Their responses indicate a primary focus on word-for-word translation rather than lexical expansion or refinement. For instance, Student 13 stated, “I also used Google Translate to find words that I don’t know in English. For example, I needed to translate the word: ‘bệnh hô hấp’ and I learned that it is ‘respiratory disease’” (Student 13). Similarly, Student 14 explained, “For this step, I only used Google Translate to check for English words and my spelling. I translated the word ‘thúc đẩy’ and I learned that it is ‘promote,’ and ‘chểnh mảng’ is ‘negligence’” (Student 14). Student 17 emphasised their reliance on the tool, noting, “Google Translate was the most helpful tool during the composing step. Among these four tools, I only need Google Translate the most because I need a dictionary to search for words that I do not know in English, such as ‘tsunami,’ ‘compete,’ ‘mineral,’ and ‘instability’” (Student 17). Likewise, Student 19 stated, “I was struggling to find the English word for ‘ban hành luật,’ and Google Translate helped me solve this problem; now I know it is ‘promulgate’” (Student 19). These responses collectively suggest that low-scoring students primarily used Google Translate to address immediate lexical gaps through direct translation, rather than to explore synonyms, alternative phrases or collocations, which was demonstrated in the group of high-scoring students (e.g. Students 5, 7, 8, and 15). These findings add more insights to the quantitative results in Section 4.2 (Table 21), which indicated a slight tendency for higher usage of Google Translate to be associated with lower Essay 2 scores—that is, low-scoring students tended to use Google Translate more frequently than their high-scoring peers. The thematic analysis reveals that this usage was typically more passive, as low-scoring students primarily used the tool for instant translations of unknown Vietnamese words into English, rather than as a means to explore synonyms and develop broader English lexical resources.

In summary, both high- and low-scoring students demonstrated vocabulary acquisition through the use of Google Translate. In both groups, students described learning new words using phrases like “I learned that...” or “now I know it is...”, showing that tool usage supported their lexical development during the writing process. However, notable differences can be observed between the responses of the two groups. High-scoring students engaged with tools more strategically, often combining Quillbot with Google Translate to explore synonyms, alternative phrases or collocations (e.g. Students 5, 7, 8, and 15), thus their tool usage reflected a more active approach to language learning. In contrast, low-scoring students (e.g. Students 13, 14, 17 and 19) relied almost solely on Google Translate for instant translations to learn new vocabulary, rather than exploring alternatives to expand their vocabulary. Collectively, these findings between the high-scoring and low-scoring groups suggest that writing score gains and vocabulary development are influenced more by the level of tool engagement rather than by tool access or frequency of usage.

Grammar Improvement (GI)

The high-scoring group exhibited 10 instances related to the theme of grammar improvement, with Grammarly mentioned in nine cases and Quillbot in one. Among the high-scoring group, 9 out of 13 students (IDs 1, 2, 3, 5, 7, 8, 15, 18, and 20) mentioned using Grammarly during the editing phase to improve their grammar accuracy. The following interview excerpts illustrate these students’ responses within this theme:

[43] Grammarly is also helpful in correcting my errors in spelling, prepositions, and articles, such as “with affordable prices” to “at affordable prices”, and “choose the suitable time” should be “choose a suitable time”. (Student 1)

[44] I used Grammarly one time to check for spelling and grammar. Grammarly was particularly helpful in fixing articles (the, a), plurals, singulars, verbs, and my spelling errors. For example, Grammarly helped me learned that I need to add “the” before “problems” in the phrase “one of the many problems”, add “s” to make plurals with “students”, “classmates”, add “are” to the phrase “applications that are available”, “distracted by” rather than “distracted into”. I

also learned that the past tense of “set up” is “set up” and not “setted up”. (Student 2)

[45] For editing, Grammarly is useful in correcting errors. Thanks to it, I can identify some article errors that I made, such as missing “a” before “electric car” and “petrol-powered car”. However, it did not provide good suggestions for vocabulary and writing style. (Student 3)

[46] After completing my essay, I used Grammarly to check for errors. The tool helped me identify several grammatical mistakes, such as changing “reduce the exhaust” to “reduce exhaust”, correcting “Another factor causes” to “Another factor that causes” modifying “the earth surface” to “the earth’s surface” and adjusting “clean energy can be produce” to “clean energy can be produced” and “clean energy can be use” to “clean energy can be used”. Additionally, it corrected the spelling of “sea water” to “seawater”. (Student 7)

[47] Grammarly is helpful for correcting articles and prepositions, such as from “from a distance” to “at a distance,” changed “on tuition fees” to “for tuition fees” and adjusted “participate on” to “participate in” (Student 8)

[48] Grammarly helped me the most during the editing step. It helped me recognise my mistakes with plural nouns, such as correcting “hydros power” to “hydro power” and “this ideas” to “this idea”. It also reminded me to add a comma after “For that reason” and corrected my preposition mistake by changing “replace fossil fuels to other energy” to “replace fossil fuels with other energy”. (Student 15)

[49] Quillbot is also quite useful. For example, it helped me paraphrase “Countries with less resources are put in the lower hand in negotiation and slower in development” to “Countries with fewer resources are given a weaker negotiating position and develop more slowly”. This tool also helped me realise that “resources” should be written with “fewer” rather than “less”. Grammarly was the most helpful tool during the revision and editing process. For instance, it corrected my phrase “into daily life” to “in life” and helped refine my use of prepositions, articles, and spelling. (Student 18)

[50] Grammarly was the most helpful tool in the revising and editing step, especially for checking punctuation and articles. For instance, it helped me recognise punctuation errors, one spelling error (Google meet should be GoogleMeet), remove the preposition “out” in “list out”, and turning “times” to “time”. (Student 20)

Generally, the tool Grammarly supported these high-scoring students to improve certain points of grammar, spelling and punctuation errors. For example, Student 1 detailed how Grammarly corrected their preposition and article errors, such as changing “with affordable prices” to “at affordable prices” and “choose the suitable time” to “choose a suitable time”, showing awareness of grammar rules. Student 2 discussed how Grammarly helped with articles, plurals, verb forms, and spelling, and explained specific grammar knowledge gained, such as learning that the verb “set up” does not change in the past tense. Similarly, Student 3 acknowledged Grammarly's usefulness in correcting articles. Student 7 provided examples of how Grammarly corrected verb forms and noun phrases (e.g., “clean energy can be produced” instead of “can be produce”). Student 8 highlighted corrections related to prepositions and article usage. Student 15 recognised Grammarly's role in correcting plural forms and punctuation. Notably, Student 18 also used Quillbot alongside Grammarly to paraphrase and correct grammatical structures, such as replacing “Countries with less resources” with “Countries with fewer resources,” thereby showing both lexical and grammatical refinement. Student 20 reflected on how Grammarly helped with punctuation, article usage, and preposition removal.

Among the 13 students in the high-scoring group, four responses (from Students 6, 9, 10, and 11) did not include any references to grammar improvement or Grammarly usage. For Students 6 and 9, the main barrier was time constraints, as both stated they were unable to incorporate Grammarly into their writing process: “I did not have enough time to use Grammarly” (Student 6) and “But I didn't have time to use Grammarly” (Student 9). Student 11 similarly limited his engagement with the tool, citing both access issues—namely, the need for a paid subscription—and a perceived

lack of utility from this tool ⁽²⁾, as she answered: “Grammarly required me to buy a subscription and it did not give me any useful correction.” (Student 11). Meanwhile, Student 10, who had previously demonstrated Cognitive Engagement with other tools by rejecting Sketch Engine’s inappropriate synonym suggestion (‘tacitly’ for ‘agree’) after verifying with Google Translate, explicitly dismissed Grammarly’s usefulness. He preferred to rely on his own editing skills and criticised the tool’s feedback as not useful:

[51] Grammarly was not useful for revising my essay because I can identify most errors myself. There was one sentence: ‘In conclusion, I completely agree with the suggestion that fossil fuels be replaced by their renewable counterparts’, which was flagged as erroneous but it did not offer any better suggestion. So I don’t think Grammarly can help me at all. (Student 10)

Moreover, Students 6, 10 and 11 reported limited support from Grammarly:

[52] However, I did not need much help with language use. (Student 6)

[53] So I don’t think Grammarly can help me at all. (Student 10)

[54] I used Grammarly for checking, but since I did not have access to the premium version, it did not provide any other useful corrections for my essay. (Student 11)

Collectively, these responses indicate that higher-proficiency students may be less inclined to use Grammarly because it offers limited language support for them. In this subset (IDs 6, 9, 10, 11), students either encountered constraints (e.g., time, lack of premium features) or intentionally relied on their own grammar. Despite not using Grammarly in revision, all four attained high scores, suggesting that grammatical accuracy was more closely associated with their existing grammar competence than with access to the tool.

Regarding the low-scoring group, six out of seven students (Students 4, 12, 13, 14,

² This issue is part of the methodological limitation (see Chapter 7, Section 7.4) because participants used only the free versions of the tools, which offer reduced functionality compared to paid subscriptions.

17, and 19) mentioned using Grammarly for grammar corrections, but their responses suggest a more passive engagement with this tool. The following interview excerpts present the six low-scoring students' responses within this theme:

[55] Grammarly was helpful, for example, it helped me correct from "the mesmerizing alternatives of other fuels" to "mesmerizing alternative to other fuels", adding "the" before "equipment". (Student 4)

[56] I still think that Grammarly was useful in the revising and editing step. For example, I edited the sentence from "They can create" to "They can be created by unlimited sources in nature such as the wind heat or sunlight" based on its highlighted suggestions and the "Passive voice misuse". (Student 12)

[57] In the revising and editing step, Grammarly was the most helpful tool for checking grammar and spelling. It helped me notice mistakes, especially with plural nouns. For example, I wrote "I partly agree with this ideas" but Grammarly corrected it to "I partly agree with this idea" It also fixed "too much traditional sources" to "too many traditional sources" when talking about fossil fuels. I also made some mistakes with verbs. For example, I wrote "it completely disappear" but Grammarly changed it to "it completely disappears". Another mistake was with prepositions. I wrote "put a financial burden for some nations" but it corrected it to "put a financial burden on some nations". These corrections helped me improve my writing a lot. (Student 13)

[58] Grammarly was the most useful tool for revising and editing. It helped me fix mistakes with plural nouns, like changing "so many course" to "so many courses", "online course" to "online courses", and "material" to "materials". It also reminded me to add "the" before "internet". Another mistake I made was with "is" and "are". I wrote "there are no waste", but Grammarly corrected it to "there is no waste". It also helped me with verb tenses. I wrote "should followed", but Grammarly pointed out that it should be "should follow". Another error was "talent pay", and Grammarly corrected it to "talent pays" (Student 14)

[59] Grammarly were the most helpful tools during the revising and editing step. Grammarly helped with corrections like "The Earth" and pointing out errors that I need to write in plural form, such as from "human" to "humans", correct article

usage from "the lack of water" to "a lack of water", and I learned that I made a spelling mistake with "breathe" (Student 17)

According to these responses, the tool Grammarly generally supported these six low-scoring students to improve certain points of grammar. For instance, Student 4 described Grammarly's correction of phrases like "the mesmerizing alternatives of other fuels" to "mesmerizing alternative to other fuels". Student 12 mentioned accepting Grammarly's suggestions related to the passive voice without elaboration. Student 13 listed various grammar corrections, including pluralisation and verb tense changes (e.g., "it completely disappear" to "it completely disappears"). Student 14 described recognising several types of grammatical errors, such as changing "so many course" to "so many courses," and "online course" to "online courses.", errors in article usage (e.g., adding "the" before "internet"), errors in subject-verb agreement ("there are no waste" to "there is no waste"), and verb form ("should followed" to "should follow"). Student 17 was aware of corrections in plural forms (e.g., "human" to "humans"), article use, and spelling errors such as "breathe" after using Grammarly. Collectively, these responses indicate some degree of improvement following the use of Grammarly, as the students demonstrated an awareness of their errors. However, in comparison to the high-scoring group, these responses from the low-scoring group are less critical towards the tool's output. For instance, the phrases used by these lower-scoring students in their responses —such as "Grammarly was the most helpful tool for checking grammar and spelling," "Grammarly changed it to 'it completely disappears'" (Student 13), "Grammarly was the most useful tool for revising and editing" (Student 14), "the most helpful tool" (Student 17) —suggest that it was the tool, rather than the students themselves, that took primary responsibility for making grammatical corrections. Moreover, none of the six responses reflected any assessment of Grammarly's feedback or showed evidence of a conscious decision to reject or avoid the tool—practices that were demonstrated among high-scoring students (for example, Students 3 and 10). Particularly, Student 3 remarked, 'However, it did not provide good suggestions for vocabulary and writing style,' and Student 10 stated, 'There was one sentence: "In conclusion, I completely agree with the suggestion that fossil fuels be replaced by their renewable counterparts," which was flagged as erroneous but it did not offer any better suggestion. So I don't think Grammarly can help me at all.'

Notably, Student 16 in the low-scoring group showed no instances of cognitive engagement, vocabulary learning and grammar improvement after using Google Translate and Grammarly:

[60] These tools are so useful for fast writing within 45 minutes. I can write in Vietnamese, and then use Google Translate, check with Grammarly, and use Quillbot to produce an essay. But I didn't have time to use Quillbot. With Google Translate helping me with my whole essay I don't even have to think in English and I think its translation quality is better than my English level [...] Yes. Google Translate and Grammarly helped me to produce an essay that is faster and has no grammar or spelling errors. (Student 16)

According to interview excerpt [60], Student 16 described using the tools for the composing process—writing in Vietnamese, translating with Google Translate, and then checking with Grammarly—without evaluating the accuracy or quality of the output. Her comment, “I don’t even have to think in English,” indicates passive reliance on translation, with no verification or reflection. Moreover, Student 16 did not articulate learning any new words or grammar rules, and references to Grammarly lacked specific examples, suggesting this student applied automated corrections without understanding the underlying grammar errors. Overall, Student 16’s response exemplifies a minimal level of engagement with the tools, indicating that language improvement is hindered by overreliance and a lack of behavioural and cognitive engagement.

To sum up, while both high- and low-scoring students used Grammarly to identify errors and enhance their grammar accuracy for their essays, the key distinction lies in their existing level of grammar knowledge and their critical assessment of the tool’s output. High-scoring students (e.g., Students 2, 3, 7, 15, 18, and 20) actively assessed and compared Grammarly’s suggestions in relation to their original writing, demonstrating a reflective understanding of grammar rules. Notably, some high-scoring students—such as Students 6, 9, 10, and 11—did not use Grammarly during the revision process, suggesting that their high writing score outcomes were grounded in their existing grammar competence rather than this tool’s corrective function. It also suggests that highly proficient students may reject Grammarly because it does not help them to develop their language skills beyond the current level.

In the low-scoring group, four students (e.g., Students 4, 12, 13, and 14) also showed an awareness of their grammatical errors and reported learning from comparing Grammarly's feedback with their drafts. However, compared to the high-scoring group, these four students did not question or reject any of the tool's suggestions, indicating a more passive approach to revision. These qualitative findings contribute to a more comprehensive understanding of the correlation reported in Table 20, which indicated a slight tendency for higher Grammarly usage to be associated with lower Essay 2 scores. The interview data helps explain this pattern by showing that low-scoring students tended to rely more heavily on Grammarly's corrections. In contrast, high-scoring students demonstrated independence and cognitive engagement with the tool's suggestions. Furthermore, the interview responses highlight different reflections on grammar improvement between the two groups, suggesting that even when using the same tool, their depth of language learning differed and can be hindered by overreliance and a lack of behavioural and cognitive engagement with the use of Grammarly.

4.3.4 Summary of thematic findings

To answer RQ1b: "How do high-scoring and low-scoring students differ in their ability to engage with the tools and use them to support their language development?", the exploratory thematic analysis in Section 3 reveals notably differences between high- and low-scoring students in their use and assessment of the four writing tools, particularly in terms of Cognitive Engagement (CE), Vocabulary Acquisition (VA), and Grammar Improvement (GI).

In the theme of Cognitive Engagement (CE), high-scoring students (IDs 1, 3, 5, 6, 7, 8, 9, 10, 11, 18, and 20) demonstrated a deeper level of critical assessment of language and tool usage compared to low-scoring students. In particular, students in the high-scoring group considered contextual appropriateness (e.g., Students 1, 6, 7, 9, 11, and 18), used a combination of tools to verify another tool's output (e.g., Students 1, 6, 9, and 10), and rejected unsuitable suggestions from the tools (e.g., Students 3, 5, 10, and 20). In contrast, although four of the seven low-scoring students (IDs 12, 13, 17, and 19) demonstrated some signs of critical assessment, their evaluations were generally limited, focusing primarily on immediate corrections without

considering the contextual appropriateness of the tools' outputs. Moreover, these students did not cross-check suggestions across multiple tools or reject any outputs.

Some notable differences emerged in the theme of Vocabulary Acquisition (VA). While both high- and low-scoring students reported learning new words after using Google Translate, the depth of engagement with the tool's outputs differed. High-scoring students used multiple tools, such as combining Google Translate with Quillbot to explore synonyms, collocations, or alternative phrasing (e.g., Students 5, 7, 8, and 15). This indicates a more strategic and reflective approach to lexical development among high-scoring students. In contrast, low-scoring students tended to use only Google Translate more passively, mainly for quick translations of Vietnamese words into English, without the purpose of exploring alternative lexical choices (e.g., Students 13, 14, 17, and 19). Notably, some low-scoring students (e.g., Students 4, 12, and 16) did not articulate any vocabulary gains, indicating limited engagement with the tools and limited impact of the tools on vocabulary development within this group.

Within the theme of Grammar Improvement (GI), Grammarly was used by students in both groups; however, its impact on writing scores and its effectiveness in supporting grammar development varied according to the students' existing level of grammatical competence. Some high-scoring students (e.g., Students 2, 3, 7, 15, 18, and 20) actively engaged with Grammarly's feedback by comparing its corrections with their original writing. This reflective process demonstrated their grammatical awareness and suggested language improvement. Meanwhile, other high-scoring students (e.g., Students 6, 9, 10, and 11) either did not use Grammarly during the revision process or explicitly rejected its suggestions, indicating that high writing outcomes can be achieved independently of access to this tool. In comparison, some low-scoring students (e.g., Students 4, 12, 13, and 14) also reported using Grammarly and improving their grammar by comparing its corrections with their drafts, indicating some learning. However, the main differences lie in the fact that none of the students in the low-scoring group questioned its suggestions, suggesting a more passive approach and reliance on Grammarly in this group.

Overall, these exploratory qualitative findings elaborate on the quantitative results regarding tool usage instances presented in Section 4.2.5 (Table 17), underscoring that gains in writing scores and language development—particularly in vocabulary and

grammar—are more closely linked to students’ ability to critically assess tool-generated output rather than to tool access or tool usage instances.

4.4 Factors contributing to decreased writing scores

This section addresses RQ1c: “What factors contribute to any observed decrease in students’ writing scores when using these tools?” To explore this question, the researcher conducts a qualitative analysis of four students (IDs 1, 3, 12, and 17), whose Essay 2 scores (composed with tools) declined according to the results in Table 16 (Section 4.2.4). See the process for analysing the impact of the four tools and additional factors contributing to score declines. In 33Figure 16 (Chapter 3, Section 3.5.2.3)

For each case, the researcher began by presenting the criterion scores for both essays—Task Response, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy—based on evaluations from three raters. Next, the process recordings were analysed to document the sequence of tool use, the outputs generated by each tool, and their correspondence to rater feedback and the final text. Tool outputs that aligned with rater-identified errors or introduced new errors were classified as negative tool effects. Conversely, outputs demonstrating effective language use, without contributing to score decreases, were attributed to student-related factors. Subsequently, the researcher analysed each student’s interview response to identify additional factors that may contribute to declined scores. This procedure enabled within-participant comparisons of writing performance before and after tool use, mapping score changes for each criterion and identifying the influence of tool-mediated revisions. It also highlighted challenges encountered during Essay 2 and evaluated whether the tools alleviated or intensified these issues—patterns that might be overlooked in purely quantitative results or broad thematic analyses. Consequently, the case study findings provided insights into the factors underlying students’ observed score declines.

Section 5 of Chapter 4 is organised as follows: Section 4.4.1 details the case of Student 1, Section 4.4.2 details the case of Student 3, Section 4.4.3 details the case of Student 12, and Section 4.4.4 details the case Student 17. Finally, Sections 4.4.5

presents a summary and discussion of all identified factors contributing to score declines.

4.4.1 The Case of Student 1

The Essay 2 of Student 1 is shown in Figure 19.

Remote learning has become ubiquitous in modern society. Despite drawbacks in terms of decreased real-life interactions and increased distractions, the financial and social benefits of such a mode of learning are far more significant.

Due to its inherent convenience and lack of real-life interactions, online learning may distract undisciplined students and worsen their communication skills. Remote learners may be tempted by numerous distractions on the internet that would be prohibited under conventional classroom conditions. As a consequence, students who lack self-control may squander their time and not fully focus on online lessons, resulting in declining academic results. In addition, since online learners may suffer from insufficient social interactions, their communication skills may be exacerbated. After the Covid-19 pandemic in Vietnam, for example, a number of students have reported difficulties when re-interacting with their peers in real life.

Despite the aforementioned downsides, remote learning offers much greater merits for learners worldwide. Thanks to technological advances, students from all over the world can now have access to prestigious educational institutions regardless of geographic distance. Off-campus learning has empowered them to pursue a world-class education at affordable prices, especially for underprivileged pupils. In addition, distance learning saves precious time for learners. Not having to spend time commuting, students can enjoy extended relaxation time which can enhance their mental health and academic performance. Moreover, online learning offers students with flexibility as they can choose a suitable time for their education, boosting their productivity and learning effectiveness.

In conclusion, although distance learning may be distracting and detrimental to the communication skills of some individuals, its economic and social upsides to learners are much more considerable.

Figure 19. Essay 2 of Student 1

Analysis of tool usage

In composing their second essay, Student 1 first composed a draft in English without any intervention of the tools during the planning stage. After that, he employed Sketch Engine to check the concordance of “real-life interactions” (Figure 20).

CONCORDANCE English Web 2020 (enTenTen20) SUBSCRIBE 30 days left

simple **real-life interactions** • 255
0.01 per million tokens • 5.9e-7%

Left context KWIC Right context

1	softpanorama.in...	And then they would say, 'No, no, no. </s><s>I value my real-life interactions </s><s>I value the moment. </s><s>I value presence.</s><s>
2	jonmower.com	ournalist, traveling through the United States recording real-life interactions with Americans.</s><s>It is the second film built aroun
3	fslhome	
4	kten.cor	
5	oce-ont	...
6	justmag	come up to me and they would say, 'I'm not on social media.' </s><s> And I would say, 'OK. </s><s> You know, you will be.' </s><s>
7	anarchis	And then they would say, 'No, no, no. </s><s> I value my real-life interactions . </s><s> I value the moment. </s><s> I value
8	boom.o	presence. </s><s> I value intimacy.' </s><s> And I would say, 'We'll get you eventually.' </s><s> " </s> <s> "I don't know if I
9	eatingdi	really understood the consequences of what I was saying, because [of]
10	ed.ac.uk	...
11	nealrc.o	
12	idea.int	
13	dragonf	

Figure 20. Student 1 used Sketch Engine to review the concordance of “real-life interaction”

After seeing the result of this phrase, Student 1 kept it in the essay. Student 1 was looking for words and phrases associated with “remote learning” with Sketch Engine and chose the word “off-campus” (Figure 21)

WORD SKETCH English Web 2020 (enTenTen20) SUBSCRIBE 30 days left

remote learning 10,672x ...

remote_{adjective} > learning_{noun}

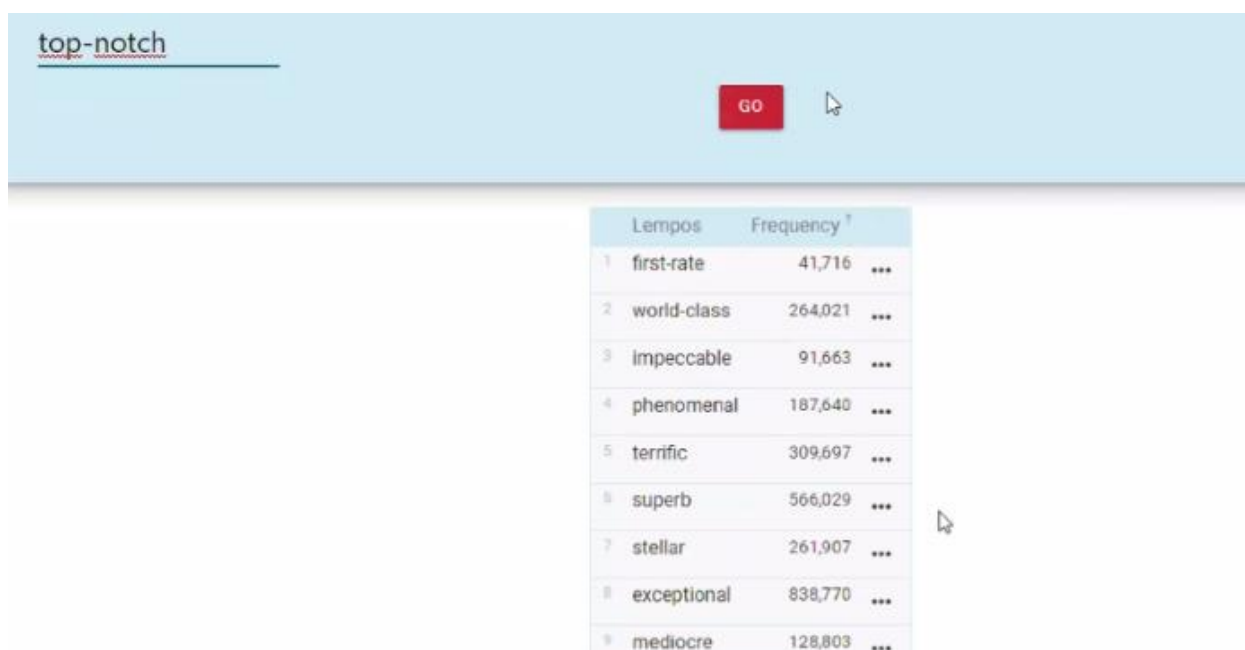
"remote" and/or ...		"learning" and/or ...	
"remote"		"learning"	
in-class	...	in-person	...
synchronous	...	teleworking	...
pandemic	...	telehealth	...
full-time	...	in-school	...
asynchronous	...	work-from-home	...
face-to-face	...	at-home	...
hybrid	...	on-campus	...
experiential	...	telemedicine	...
continued	...	telework	...

modifiers of "remote learning"
districtwide
at-home
off-campus
synchronous
telehealth
distancing
full-time
asynchronous
pandemic
experiential

verbs with "remote learning" as object
juggle
navigate
necessitate
facilitate
continue

Figure 21. Student 1 used Sketch Engine to search for the word “remote learning” and chose the phrase “off-campus”.

Then, the student appears to be looking for a synonym. He typed in the word “top-notch” and replaced it with the word “world-class”, as presented by this tool (Figure 22).



The screenshot shows the top-notch tool interface. At the top, there is a light blue header with the 'top-notch' logo on the left and a red 'GO' button on the right. Below the header, a table displays a list of synonyms for the word 'top-notch'. The table has two columns: 'Lempos' and 'Frequency'. The words are ranked from 1 to 9. A mouse cursor is visible over the 'world-class' entry.

	Lempos	Frequency	
1	first-rate	41,716	...
2	world-class	264,021	...
3	impeccable	91,663	...
4	phenomenal	187,640	...
5	terrific	309,697	...
6	superb	566,029	...
7	stellar	261,907	...
8	exceptional	838,770	...
9	mediocre	128,803	...

Figure 22. Student 1 typed in the word “top-notch” in Sketch Engine and replaced this word with the word “world-class”.

Next, Student 1 used Google Translate for checking and revising his English words, as illustrated in Figures 23, 24 and 25.

Figure 25. Student 1 used Google Translate to check the meaning of the word “re-interact” in Vietnamese

After using Google Translate, Student 1 did not make any changes to his essay. Subsequently, Student 1 attempted to use Quillbot; however, the screen recording showed that he had encountered a technological problem while composing, which prevented the successful use of Quillbot, as shown in Figure 26:

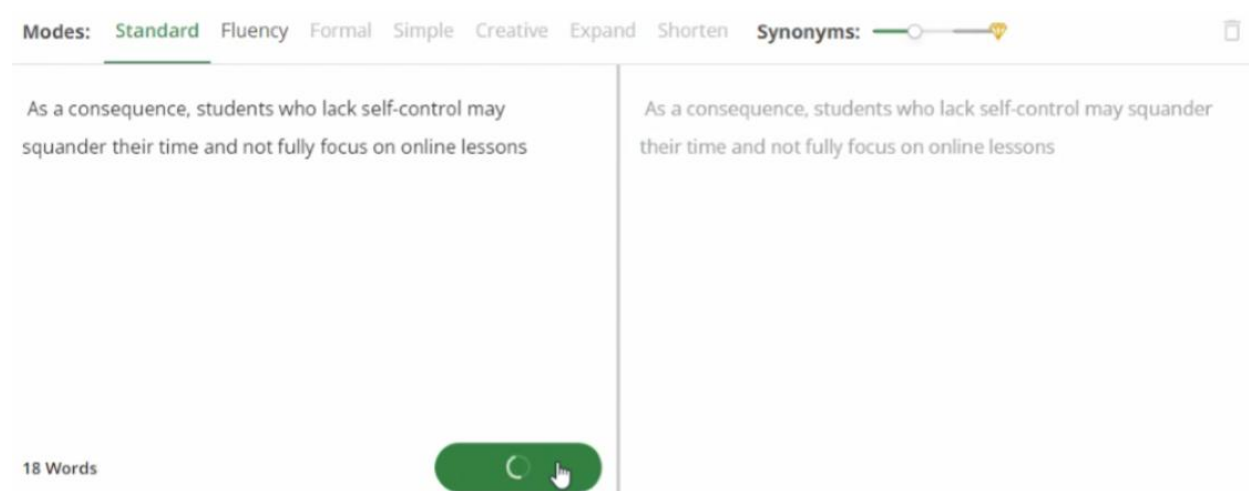


Figure 26. Student 1 experiencing a technology issue with Quillbot.

After the unsuccessful attempt with Quillbot, Student 1 switched to Grammarly for revision (Figure 27).

conventional classroom conditions. As a consequence, students who lack self-control may squander their time and not fully focus on online lessons, resulting in declining academic results. In addition, since online learners may suffer from insufficient social interactions, their communication skills may be exacerbated. After the Covid-19 pandemic in Vietnam, for example, a number of students have reported difficulties when re-interacting with their peers in real life.

Despite the aforementioned downsides, remote learning offers much greater merits for learners worldwide. Thanks to technological advances, students from all over the world can now access educational institutions regardless of geographic distance. This has empowered them to pursue world-class education with a more flexible schedule. In addition, remote learning saves time by eliminating the need for commuting, allowing students to enjoy their time more effectively. Furthermore, remote learning can enhance their mental health and academic performance by offering students with flexibility as they can choose the suitable time for their education, boosting their productivity and learning effectiveness.

In conclusion, although distance learning may be distracting and detrimental to the communication skills of some individuals, its economic and social upsides to learners are much more considerable.

Figure 27. Screenshot of Student 1 checking the suggestion from Grammarly to correct article usage

In summary, Student 1 used three tools during the writing process: Sketch Engine, Google Translate, and Grammarly. Sketch Engine was used to confirm the English phrase “real-life interaction”, to replace the phrase “top-notch” with “world-class”, and to compose the phrase “off-campus” (Figures 21 and 22). Google Translate assisted him in checking the meanings of the English words “fully” and “re-interact” in Vietnamese and Student 1 did not make any changes, so there was no output from this tool. Lastly, Grammarly was employed for spelling and article corrections, which will be examined to determine whether these corrections are accurate or faulty, or if there are any extra errors in this essay that Grammarly could not detect. Notably, the student encountered a technical issue with Quillbot, which resulted in the paraphrasing tool being unusable during the task.

Analysis of score changes and writing issues marked by the three raters

In this section, the researcher analysed the score changes of Student 1, as marked by the three raters (Table 27). Subsequently, the researcher summarised the comments and errors identified by the three raters in Essay 2 and compared these errors to

determine whether they were caused by the tools used or were originally from the student's writing.

Table 27. The comparison of Essay 1 and Essay 2 scores of Student 1

Student 1	TR1	CC1	LR1	GA1
Rater 1 (the researcher)	8	8	8	7
Rater 2	8	8	8	7
Rater 3	8	8	8	7

Student 1	TR2	CC2	LR2	GA2
Rater 1 (the researcher)	8	8	8	7
Rater 2	8	8	8	7
Rater 3	8	8	7	7

According to Table 27, most of Student 1's scores remained stable, but the Lexical Resource score decreased by one band, as noted by Rater 3. He highlighted three word choice problems in the sentence: "students who lack self-control may squander their time... resulting in declining academic results... their communication skills may be exacerbated." He judged *squander* too strong for this context and suggested "misuse" or "waste" instead; considered "declining academic results" vague and preferred "declining grades"; and noted that "exacerbated" cannot logically collocate with "communication skills," recommending alternatives such as "weakened" or "hindered." Rater 2 also identified the misuse of "exacerbated", and Rater 1 agreed with all three word choice issues. Rater 1 further observed that these errors appeared

in the student's original draft and were not introduced by any of the four tools. The corpus-derived phrases "off-campus" and "world-class" from Sketch Engine were neither penalised nor explicitly praised, suggesting they had no clear impact on the LR score. Overall, the findings indicate that (1) the lexical problems stemmed from the learner's own writing rather than tool output, and (2) although Grammarly successfully corrected most other errors, it did not detect or correct these three problematic lexical choices.

Analysis of interview excerpts

Table 28 below provides the complete transcript of Student 1's interview responses.

Table 28. The interview transcript of Student 1

STUDENT 1

1/ In general, how do you feel about the usage of the four AI writing tools for writing this IELTS essay? Do you find them helpful?

I find Grammarly and Google Translate helpful for checking synonyms, grammar, and spelling.

2/ Overall, do you think you produced a higher quality essay with the tools compared to the first essay that you wrote without the tools? Which part or sentence do you think you wrote better with the tools?

Yes, I believe I produced a higher quality essay with the tools, particularly in terms of grammar and spelling, which were improved using Grammarly. Additionally, I replaced synonyms using Sketch Engine and Google Translate.

3/ In the brainstorming and planning step (before writing), did you find the tools useful for you in this step? Which tool helped you the most? How did it help you?

I did not need any tools during the brainstorming and planning step because the topic was quite common and simple. I can write in English with simple words first. Then I need to replace the words that are too simple.

4/ In the composing step, did you find the tools useful? Which tool helped you the most? How did it help you?

Quillbot was not working during my composing time so I had to find synonyms suggested by Google Translate. I can write in English with simple words first. Then I need to replace the words that are too simple. For example, I didn't feel that the phrase "increased relaxation time" was correct, so I checked the meaning in Vietnamese, and I found more synonyms of this word suggested by Google Translate. Among them, I picked the word "extend" and modified my phrase to "extended relaxation time". I also used Sketch Engine to help me find more words related to "top-notch", and then among the collocations I chose "world-class" to complete the phrase "world-class education". I think this word sounds more suitable compared to other high-frequency words like "matchless", "triple-A", or "first-rate".

5/ In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?

In the revising and editing step, Grammarly was particularly useful for checking overall spelling and grammar. Additionally, I used Sketch Engine to affirm the usage of terms like “real-life interactions”. Another word was the word “re-interact” from the phrase “re-interacting with their peers”, which was my term, but I am not confident in using it. The Vietnamese meaning of this word seemed accurate when it was translated by Google Translate, so I used this word, and also because there were no other suggestions from the tool for this word. Grammarly is also helpful in correcting my errors in spelling, prepositions, and articles, such as “with affordable prices” to “at affordable prices”, and “choose the suitable time” should be “choose a suitable time”.

6/ Can you give me some examples of challenges that the tools could not help you in this writing task?

The tools could not help me find results for “re-interact” in Sketch Engine, Google Translate, and QuillBot. I wanted to paraphrase the sentence, “Moreover, online learning offers students flexibility as they can choose a suitable time for their education, boosting their productivity and learning effectiveness” using a different, more formal structure but I did not succeed. And as I said, Quillbot was not working during my composing time.

7/ What do you want more from these four tools: Sketch Engine, Google Translate, QuillBot, and Grammarly?

I would like Grammarly to evaluate coherence and cohesion. I wish Google Translate could find synonyms for phrases. Sketch Engine is a bit difficult to use and does not always generate correct synonyms. Some collocations suggested by the tools did not fit the context. Additionally, I wanted to evaluate the range of vocabulary (level of vocabulary based on CEFR) and suggest some collocations that fit the context and topic of education.

8/ Do you recommend or would use any other tools for this IELTS essay writing practice? Why?

I recommend using ChatGPT to brainstorm more arguments and an AI-based IELTS essay scoring tool for evaluating cohesion and coherence.

The responses regarding Interview Question 4: “In the composing step, did you find the tools useful? Which tool helped you the most? How did it help you?” reveal some of the challenges that Student 1 faced after composing Essay 2:

[61] Quillbot was not working during my composing time so I had to find synonyms suggested by Google Translate.

According to Student 1’s response [61]: “Quillbot was not working during my composing time, so I had to find synonyms suggested by Google Translate”, the technical issue stemmed from Quillbot was a factor that disrupts his writing process, prevented the student from utilising Quillbot, depriving him of a tool for synonym selection and paraphrasing, which in turn disrupted his writing process.

Furthermore, regarding Interview Question 5: “In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?”, Student 1 stated: “In the revising and editing step, Grammarly was particularly useful for checking overall spelling and grammar.” Accordingly, Grammarly was useful for this student in revising, particularly in correcting errors in prepositions and articles. However, his final version of Essay 2 still contained three lexical errors— “squander their time,” “declining academic results,” and “exacerbated”—as identified by the raters. These inaccuracies were less a result of Grammarly’s shortcomings and more a consequence of the student’s limited error awareness and uncritical acceptance of certain word choices, reflecting a reliance on the tool without independent evaluation. While Grammarly did not detect these lexical issues, its role in reducing the score was indirect. The decisive factor was the student’s limited capacity to recognise and address them without assistance.

4.4.2 The Case of Student 3

The Essay 2 of Student 3 is shown in Figure 28.

The use of fossil fuels as an energy resource has created many problems for the natural environment. Therefore, there has been an increasing demand for an alternative source, namely renewable energy, to mitigate fossil fuels' dangerous outcomes. From my perspective, I totally agree with this solution because of the benefits that these clean energies bring about.

First and foremost, the amount of fossil fuels is limited, whereas the amount of renewable energy is infinite. The more we use fossil fuels, the more they become depleted, which results in their continually rising cost to prevent people from using them up. Gradually, people can no longer afford to buy fossil fuels in order to generate energy for their daily needs. By contrast, natural energy resources like solar or wind power can be harnessed from endless force. Therefore, people do not need to worry about facing the risk of their depletion. Some may argue that the cost of installation of a system to collect this type of energy is higher than the current cost of fossil fuels. However, in the long run, it will be much more cost-effective as the energy from nature is automatically converted into electricity without any further expense for by-processes.

Furthermore, it is obvious that the use of fossil fuels leaves detrimental impacts on the environment while renewable energy is totally clean and produces no harmful by-products. The burning of coal to generate energy, for instance, releases harmful emissions to the environment, which increases the carbon footprint and contributes to global warming. Another typical example of the adverse effect of fossil fuels is that the exhaust fumes from those means of transport that use petrol are among the factors leading to air pollution. Meanwhile, solar power generates energy directly from the heat and light of the sun without any by-products; and an electric car is far more environmentally friendly than a petrol-powered car for no exhaust fumes being produced.

In conclusion, due to their limitless amount and clean generating process, renewable resources deserve to be practically applied in daily life and mass production for the sustainable development of our globe. The government may need to fund the beginning stage of installation but this can bring about positive effects in the long term.

Figure 28. The Essay 2 of Student 3

Analysis of tool usage

To complete their second essay, Student 3 first composed a draft in English without tools, and then used the tools to edit this draft. In particular, Student 3 used Sketch Engine as she typed in the keyword “energy” and found the word “harness” among the collocation results (Figure 29).

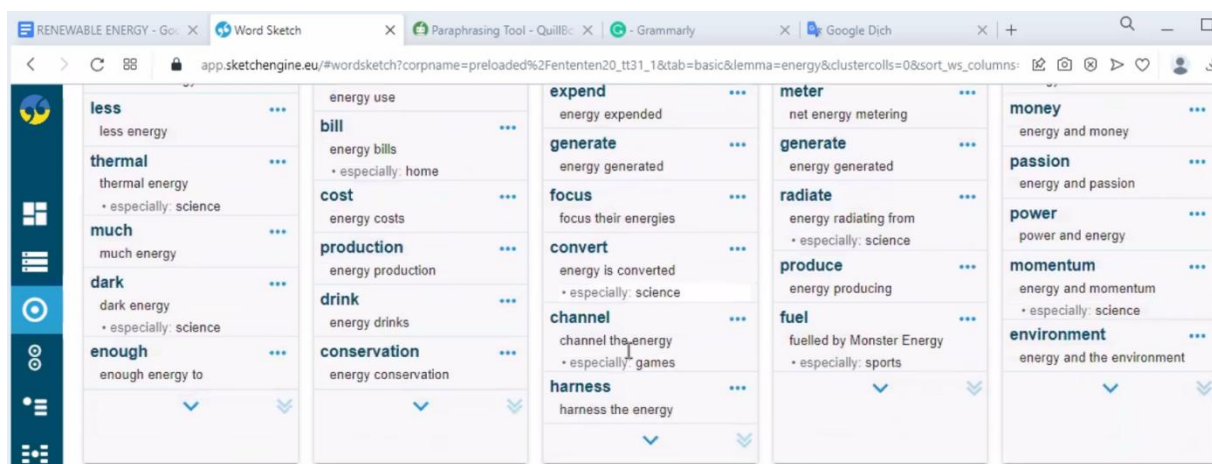


Figure 29. Student 3 used Sketch Engine to the collocation of “energy” and found the word “harness”

With this result, Student 3 went back to her draft and wrote the sentence: “By contrast, natural energy resources like solar or wind power can be harnessed from endless force”. After that, Student 3 inserted Vietnamese words to Google Translate to identify the English meanings.

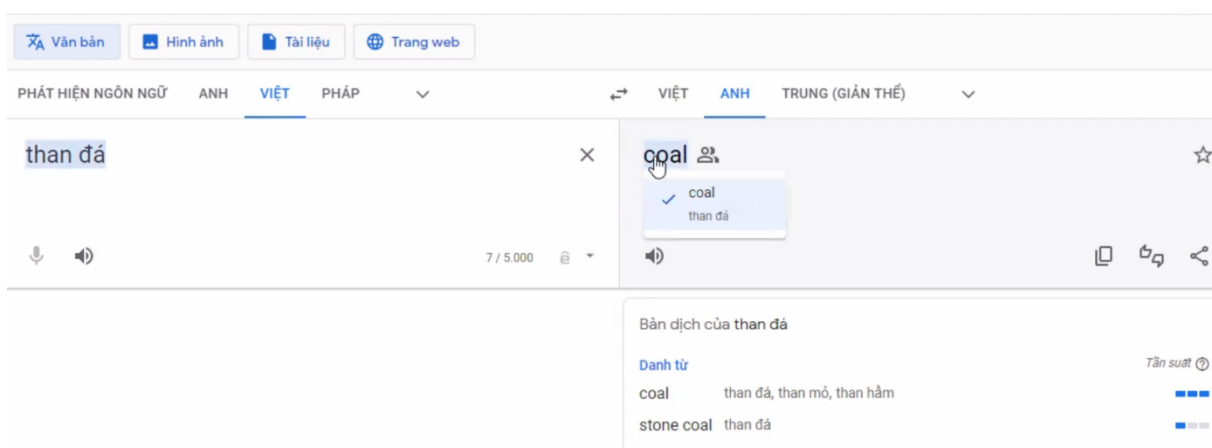


Figure 30. Student 3 used Google Translate to find the word “coal” in English.

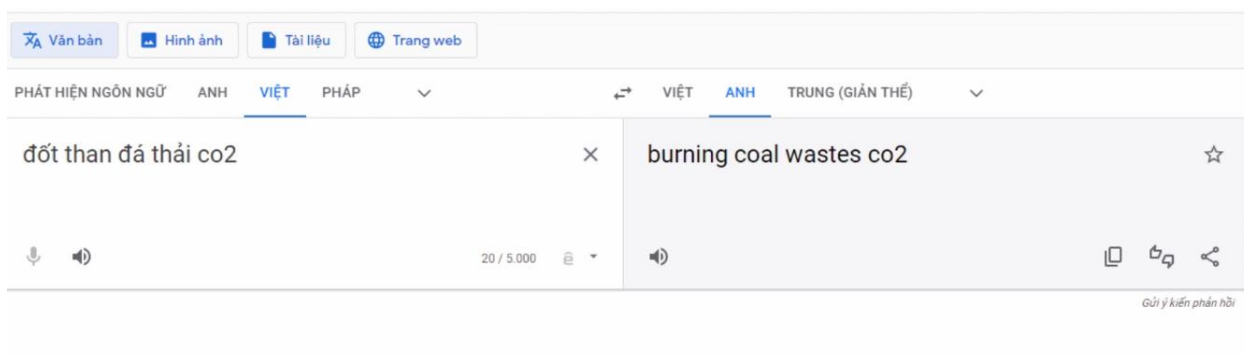


Figure 31. Student 3 used Google Translate to find the phrase “burning coal wastes CO2”, but she did not use this result.

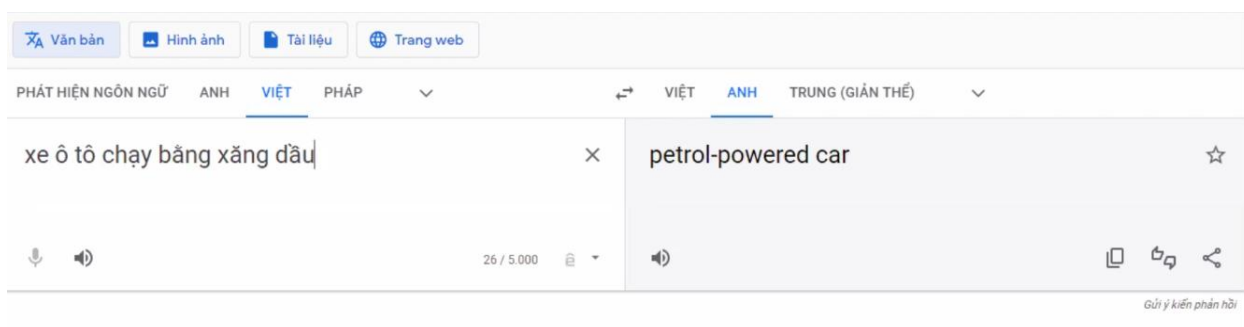


Figure 32. Student 3 used Google Translate to find the phrase “petrol-powered car” in English

In Figures 30, 31 and 32, Student 3 attempted to translate the phrase but chose not to use the generated output. Instead, she independently wrote her own sentence: "The burning of coal to generate energy, for instance, releases harmful emissions into the environment, which increases the carbon footprint and contributes to global warming.". This demonstrates her thoughtful consideration and critical assessment of the tool's output.

Subsequently, she used Quillbot to paraphrase the phrase “harmful by-products”, then reviewed suggestions, but in the end, retained her original wording (Figure 33).

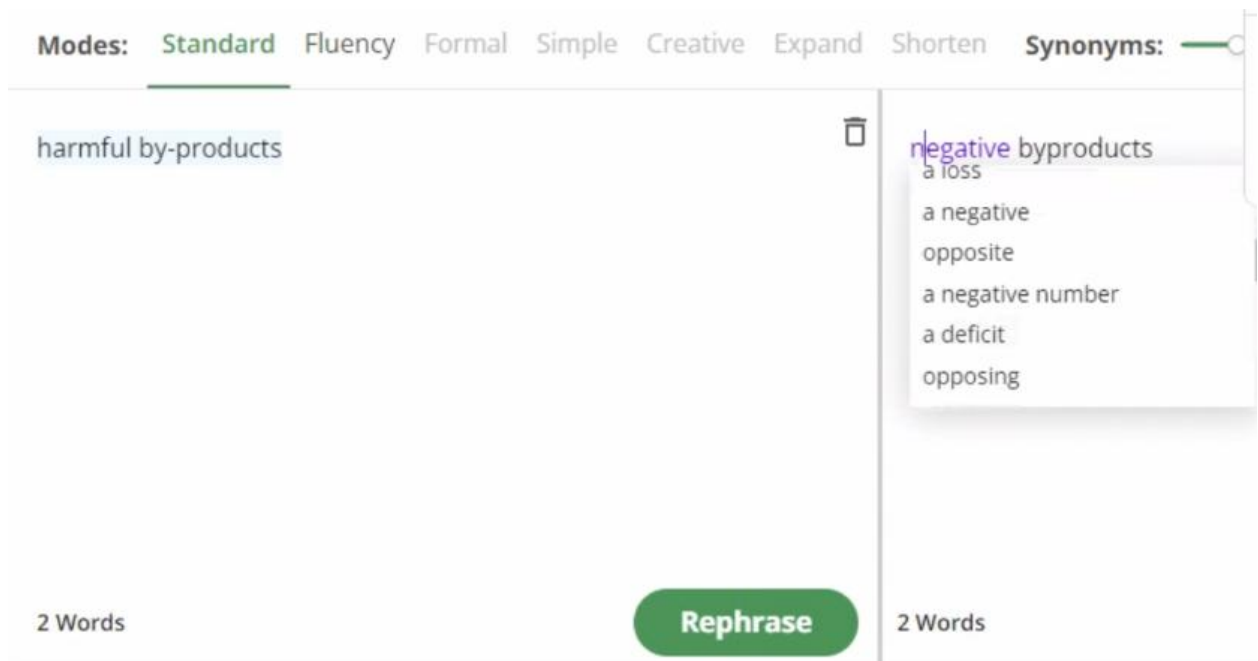


Figure 33. Student 3 was looking for synonyms of “harmful by-product” in Quillbot

These actions also indicate that Student 3 demonstrated a critical evaluation of Quillbot's output, as she reviewed all the suggested changes and decided to retain her original wording. After that, Student 3 used Grammarly and also demonstrated a cognitive engagement usage to Grammarly's output by carefully reviewing all the suggested changes, as presented in Figure 34, before accepting all changes.

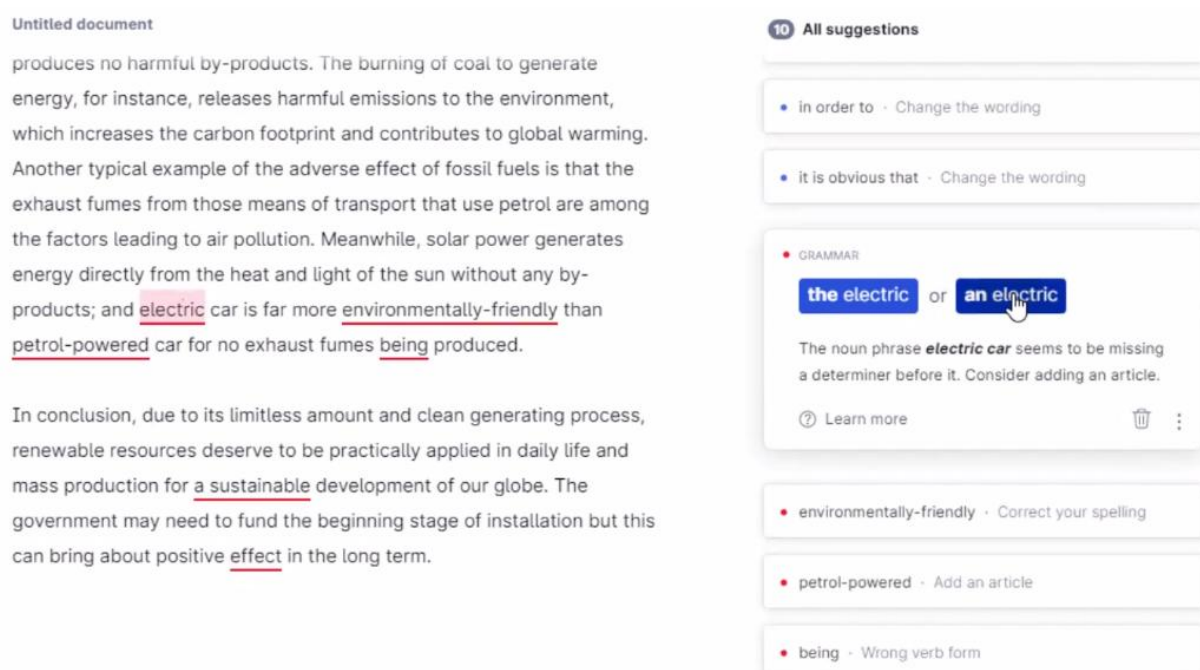


Figure 34. Student 3 reviewed suggestions from Grammarly before accepting all suggestions

In summary, the screen recording of Student 3 showed that this student did not use any tools for planning; she drafted the essay in English and used four tools for composing and revising. She used Sketch Engine one time to find collocations for the word "energy", critically evaluated the output, and then selected the word "harness" to write a complete sentence. Consequently, the word "harness", as generated by Sketch Engine will be examined in the next section. When searching for English vocabulary, she used Google Translate but opted not to accept its translated phrases, instead composing her own sentence with new phrases. She chose not to incorporate the output from Quillbot. For Grammarly, Student 3 used this tool for revision and accepted all suggestions.

Analysis of score changes

The scores from both essays of Student 3, as marked by the three raters, is presented in Table 29.

Table 29. The comparison of Essay 1 and Essay 2 scores of Student 3

Student 3	TR1	CC1	LR1	GA1
Rater 1 (the researcher)	8	7	8	7
Rater 2	7	7	8	7
Rater 3	7	7	7	7

Student 3	TR2	CC2	LR2	GA2
Rater 1 (the researcher)	7	7	7	7
Rater 2	7	7	8	7
Rater 3	6	7	7	7

According to Table 29, Student 3's Essay 2 scores fell from 8 to 7 in Task Response and Lexical Resource (Rater 1) and from 7 to 6 in Task Response (Rater 3). The three raters identified three word choice errors: "endless force", "collect this type of energy", and "by-processes". Rater 2 noted that "endless force" is inappropriate with energy and should be "inexhaustible energy" or "renewable energy"; "collect this type of energy" should be "harness" or "generate"; and "by-processes" should be "by-products". These issues lowered the Lexical Resource score. However, tool-use logs show that these errors did not come from Sketch Engine, Google Translate, or Quillbot but from the student's own language choices, suggesting the tools did not directly reduce writing quality and that Grammarly failed to flag these lexical errors. By contrast, the phrase "solar or wind power can be harnessed", taken from Sketch Engine, was praised by all three raters. Overall, as with Student 1, the main cause of

the score drop appears to be Student 3's unawareness of these lexical problems and reliance on automated correction, alongside Grammarly's limitation in detecting such errors.

Analysis of interview excerpts

Table 30 below presents the complete transcript of Student 3's interview responses.

Table 30. The interview transcript of Student 3

STUDENT 3

1/ In general, how do you feel about the usage of the four AI writing tools for writing this IELTS essay? Do you find them helpful?

I think these tools are useful for improving my writing in terms of language.

2/ Overall, do you think you produced a higher quality essay with the tools compared to the first essay that you wrote without the tools? Which part or sentence do you think you wrote better with the tools?

Yes, I believe I produced a higher quality essay in terms of lexical resource and grammar, especially because topic 2 was less familiar to me than topic 1. For instance, I used Sketch Engine to find keywords and phrases like "harness energy". I used Google Translate to translate parts of this sentence: "Meanwhile, solar power generates energy directly from the heat and light of the sun without any by-products; and an electric car is far more environmentally friendly than a petrol-powered car due to the lack of exhaust fumes.", and I was satisfied with its output. Grammarly also helped me with the usage of articles (a/an/the). However, these four tools cannot help me with more ideas, and the "Environment" topic is more difficult for me compared to the first topic about "Distance learning".

3/ In the brainstorming and planning step (before writing), did you find the tools useful for you in this step? Which tool helped you the most? How did it help you?

For brainstorming, I like Sketch Engine because it helped me to find the verb "harness" in "harness energy" when I searched for the word "energy" to look for more ideas. I did not use any tools during the brainstorming and planning step.

4/ In the composing step, did you find the tools useful? Which tool helped you the most? How did it help you?

Yes, in the composing step, I found the tools useful. For composing, I mostly needed Google Translate to look for some phrases in English, such as the phrase "petrol-powered car", "coal", and "burning coal releases CO₂" and I learned these new words. Quillbot was not very useful for me when I used it to look for synonyms, as the phrase

"harmful by-products" that I tried with Quillbot did not result in better words.

5/ In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?

For editing, Grammarly is useful in correcting errors. Thanks to it, I can identify some article errors that I made, such as missing "a" before "electric car" and "petrol-powered car". However, it did not provide good suggestions for vocabulary and writing style.

6/ Can you give me some examples of challenges that the tools could not help you in this writing task?

Quillbot did not provide good suggestions for the phrase "harmful by-products". Additionally, I needed vocabulary that fit the context, but both Quillbot and Google Translate did not help with this.

7/ What do you want more from these four tools: Sketch Engine, Google Translate, QuillBot, and Grammarly?

I would like Sketch Engine to suggest collocations classified by topics.

8/ Do you recommend or would use any other tools for this IELTS essay writing practice? Why?

I recommend brainstorming ideas with ChatGPT. Additionally, tools that can organise the arguments in a paragraph with cohesion and coherence would be beneficial.

The response regarding Interview Question 2: "Overall, do you think you produced a higher quality essay with the tools compared to the first essay that you wrote without the tools? Which part or sentence do you think you wrote better with the tools?" reveals some of the challenges that Student 3 faced after composing Essay 2, as she stated: "Grammarly also helped me with the usage of articles (a/an/the). However, these four tools cannot help me with more ideas, and the "Environment" topic is more difficult for me compared to the first topic about distance learning." Based on this response, the decrease in her writing scores can be attributed to the student's unfamiliarity of the writing topic in the second essay. This difficulty relates to the decrease in her Task Response (TR) criteria dropping from 8/7/7 to 7/7/6. Moreover, while Student 3 valued

Sketch Engine for finding collocations to compose the phrase “harness energy”, she admitted that: “They cannot help me with more ideas”. This answer also highlights the ineffectiveness of the four tools in supporting Student 3 with idea generation.

Regarding Interview Question 5: “In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?”, Student 3 responded:

[62] For editing, Grammarly is useful in correcting errors. Thanks to it, I can identify some article errors that I made, such as missing “a” before “electric car” and “petrol-powered car”. However, it did not provide good suggestions for vocabulary and writing style.

Excerpt [62] indicates that while Grammarly helped Student 3 identify article errors, it did not assist in recognising three word choice errors—“endless force,” “collect this type of energy,” and “by-processes”—which were identified by the raters. These inaccuracies contributed to a decline in her Lexical Resource (LR) score, which dropped from 8/8/7 to 7/8/7. Moreover, her sole usage on Grammarly for revision, despite its inability to provide nuanced vocabulary refinements or suitable collocation suggestions, suggests a reliance on this tool and also reveals the student’s limited ability for self-assessing lexical accuracy. This issue aligns with the case of Student 1, highlighting a recurring pattern: students depend on Grammarly for revision, yet the tool falls short in offering advanced lexical corrections. This finding emphasises both students’ challenges with language self-assessment and Grammarly’s limitations in refining vocabulary beyond grammatical accuracy.

4.4.3 The Case of Student 12

The Essay 2 of Student 12 is shown in Figure 35 below.

When people find different sources of alternative energies such as wind power, and thermal energy, many people think that these power sources should replace fossil fuels. This paper will prove how this proposition should be partially agreed on.

The search and use of alternative sources of energy should be promoted for several reasons. First, new sources of energy such as wind power, thermal power, or solar power are renewable energies. They can be created by unlimited sources in nature such as the wind, heat, or sunlight. That is the reason why more and more facilities are created and improved to set up the infrastructure for the widespread production, transmission and use of these energy sources. In Vietnam, for instance, there is an increasing amount of wind turbines established along the provinces in the Central and sea areas. In addition, wind or solar power are proven to release less exhausted fumes than fossil fuels such as coal or oil and gas. That motivates not only the governments but also the producers, vendors and users to prefer these sources of energy to petroleum or coal.

However, regardless of the economic and environmental benefits of renewable energies, they should never totally replace traditional energy sources such as coal or petroleum. Concerning the availability of underground coal mines, and oil wells offshore, it would be a terrible waste of natural resources if they are left unused. In addition, taking all the infrastructures and labor workforce living on the execution of exploitation, production, and transmission of these energy sources into consideration, it could hardly believe that why we can stop all these activities immediately without a transition period. Finally, the provision of power from alternative sources needs time and effort; therefore, we still need to partially rely on traditional power sources for a long time.

In conclusion, even though renewable sources of energy can bring many benefits, it should not totally replace fossil fuels.

Figure 35. The Essay 2 of Student 12

Analysis of tool usage

Student 12 started by translating three phrases from Vietnamese to English, as shown in Figure 36.

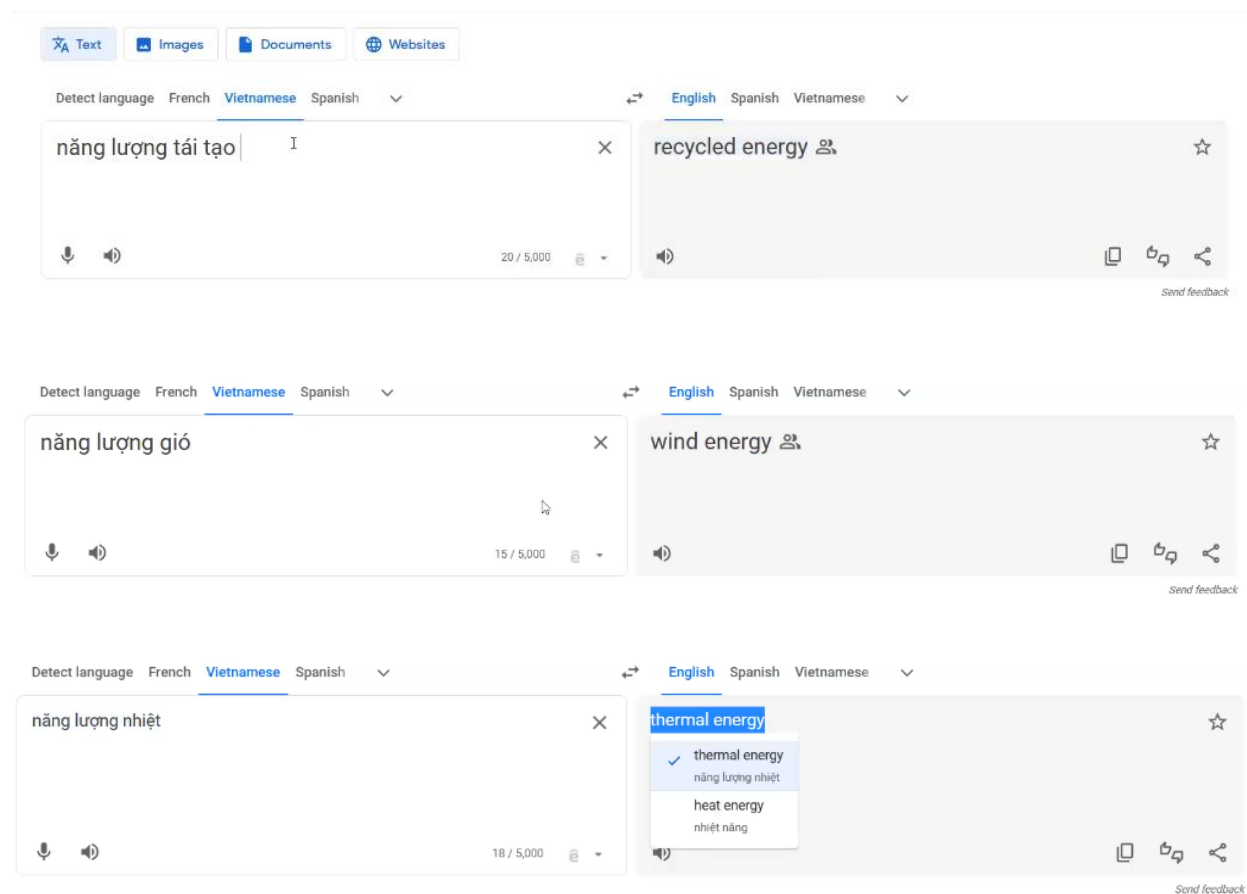


Figure 36. Student 12 used Google Translate to translate three phrases into English during the planning stage

recycled energy

wind energy

thermal energy

energy transition

Coal, petroleum, gas : black gold

Agree: replace the amount / proportion: economical and environmental benefits

Disagree: not replace all: infrastructures, the petroleum wells, coal mine

When people find different sources of alternative energies such as wind power, and thermal energy, many people think that these power sources should replace fossil fuels. This paper will prove how this proposition should be partially agreed on.

I
First, new sources of energies such as wind power, or solar power are renewable energies. They can be created by unlimited sources of the nature such as wind, he

Figure 37. Student 12 included the three phrases in her draft during the planning stage

Student 12 included these three phrases— “recycled energy,” “wind energy,” and “thermal energy”—generated by Google Translate into her draft and they will be examined in Section 4.3.2. After planning, screenshots revealed that Student 12 composed a complete English essay without using tools. After completing her essay, she used Sketch Engine to analyse the concordance of the phrase “exhausted fumes”, as illustrated in Figure 38.

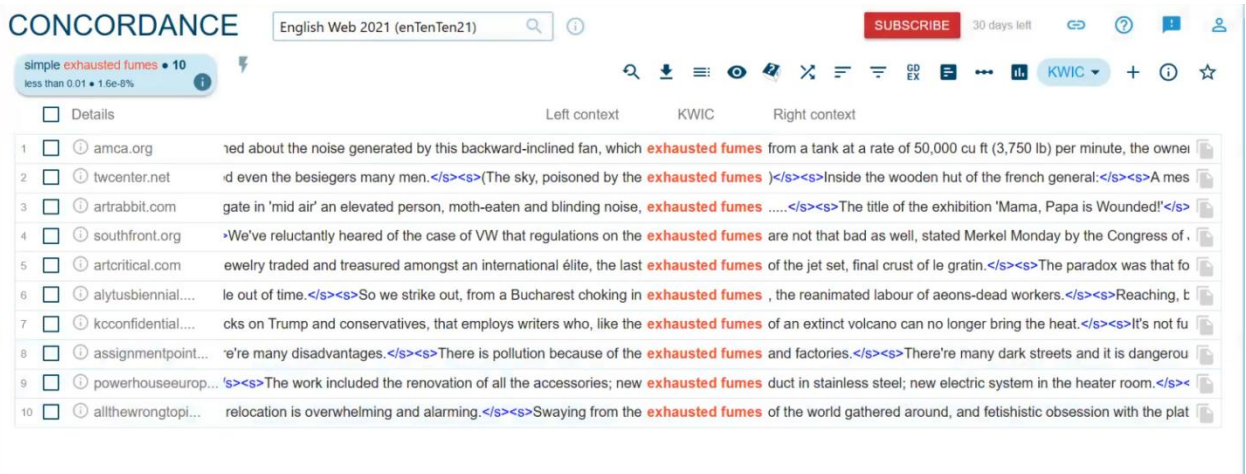


Figure 38. Student 12 used Sketch Engine to check the concordance of the phrase “exhausted fumes”

After seeing the result, Student 12 did not make any changes to her essay. Subsequently, Student 12 used Grammarly to check for errors in her essay. However, as shown in Figure 39, Grammarly appeared with a requirement for premium access, therefore error corrections were hidden from her screen, which is considered as a technology issue.

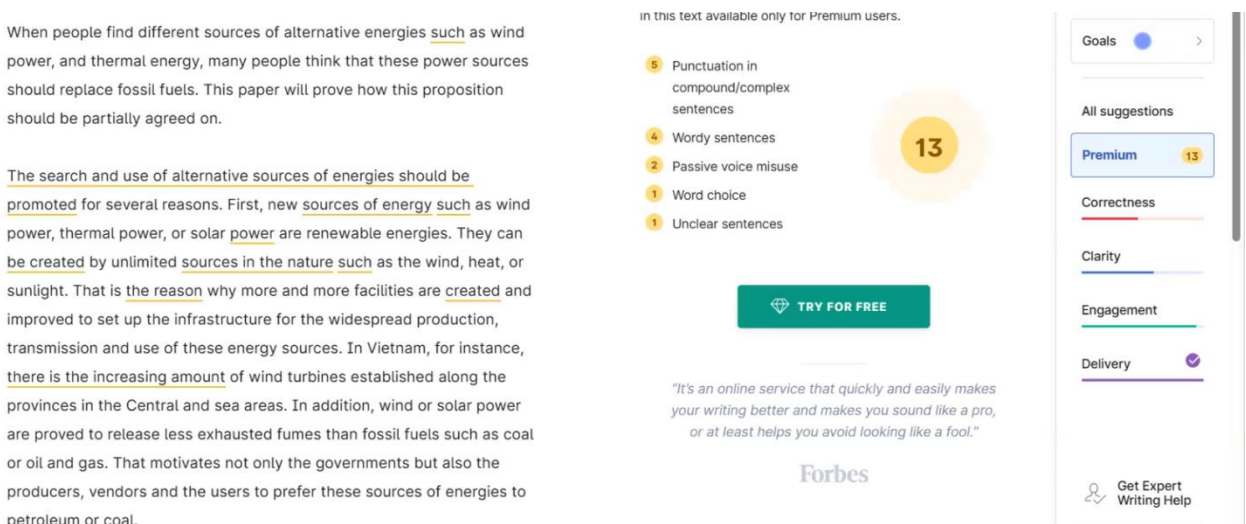


Figure 39. Student 12 used Grammarly to correct her second essay but did not make any changes.

After not being able to make corrections with Grammarly, Student 12 corrected three errors independently. In particular, she replaced the phrase “This is the reason why”

with “That is the reason why”, changed the word order from “coal mines underground” to “underground coal mines”, and changed the phrase “a huge waste” to “a terrible waste”. This instance also highlights a technological limitation in Grammarly that impeded the student's ability to fully utilise the tool for error correction.

Analysis of score changes

The scores from both essays of Student 12, as marked by the three raters, are presented in Table 31. Subsequently, the researcher summarised the errors and comments provided by the raters on Essay 2 and compared these errors to determine whether they stemmed from the outputs of the four tools or were inherent to the student's writing.

Table 31. Essay 1 and Essay 2 scores of Student 12

Student 12	TR1	CC1	LR1	GA1
Rater 1 (the researcher)	8	7	8	7
Rater 2	8	7	7	7
Rater 3	6	6	7	7

Student 12	TR2	CC2	LR2	GA2
Rater 1 (the researcher)	6	6	6	6
Rater 2	6	6	6	6
Rater 3	6	6	6	6

According to Table 31, Student 12's Essay 2 scores decreased on all four criteria. Task Response dropped by 2 bands (Raters 1 and 2); Coherence and Cohesion by 1 band (Raters 1 and 2); Lexical Resource by 2 bands (Rater 1) and 1 band (Raters 2 and 3); and Grammar Range and Accuracy by 1 band (Raters 1 and 3). The three raters identified five lexical errors ("find", "search", "created", "unlimited", and "exhausted fume(s)") under LR. None were produced by Google Translate, so this tool did not reduce writing quality. However, "exhausted fumes" came directly from Sketch Engine's English Web 2021 corpus (Figure 38), and Student 12 adopted it in her essay. All raters flagged this as incorrect, noting that the appropriate term is "exhaust fumes". Because the English Web 2021 corpus includes nonstandard web language, this is a rare case where engagement with a corpus tool led to an inappropriate lexical choice and a negative impact on writing quality.

Moreover, the three raters marked one error in verb tense in the sentence "Concerning the availability of underground coal mines, and oil wells offshore, it would be a terrible waste of natural resources if they are left unused.". The grammatical issue in the sentence lies in the tense inconsistency between "would be" (conditional) and "are left" (present tense). To maintain consistency, "are left unused" should be changed to "were left unused" to match the phrase "would be", as it is the second conditional structure. Additionally, all three raters pointed out that the following sentence is problematic: "In addition, taking all the infrastructures and labour workforce living on the execution of exploitation, production, and transmission of these energy sources into consideration, it could hardly believe that why we can stop all these activities immediately without a transition period.". The sentence contains an issue in pronoun reference in the phrase "it could hardly believe that why we can stop all these activities immediately without a transition period.". The subject "it" is unclear as it does not align with any logical subject in the sentence. Additionally, the word "why" is redundant and unnecessary in this context. The correct sentence would be: "It is hard to believe that we can stop all these activities immediately without a transition period.". These issues affect both the LR and GA criteria.

Analysis of interview excerpts

Table 32 below provides the complete transcript of Student 12's interview responses.

Table 32. The interview transcript of Student 12

STUDENT 12
1/ In general, how do you feel about the usage of the four AI writing tools for writing this IELTS essay? Do you find them helpful? I find the AI writing tools quite helpful. My order of preference is: 1. Google Translate (for checking word meaning of some phrases) 2. Quillbot 3. Sketch Engine 4. Grammarly 2/ Overall, do you think you produced a higher quality essay with the tools compared to the first essay that you wrote without the tools? Which part or sentence do you think you wrote better with the tools? No, I did not produce a higher-quality essay with the tools due to time constraints. I don't think these tools helped me write any better in 45 minutes. I am not good with ideas for the environment topic and I truly need something that can help me search for more ideas. In terms of language, I used Sketch Engine to check whether the phrase "exhausted fumes" is used in a corpus, and it provided me with some results so I kept this phrase. I didn't use Google Translate because I did not have any problem with writing my ideas in English. I did not have time to use Quillbot as well. For Grammarly, I did not offer any corrections because it required me to pay for premium access. Overall, I feel like I didn't do as well on this second essay compared to the first one. I am not familiar with the topic of environment, so I cannot think of good examples. I need a tool that helps me with more ideas. 3/ In the brainstorming and planning step (before writing), did you find the tools useful for you in this step? Which tool helped you the most? How did it help you?

I generated ideas by myself and I used Google Translate and Sketch Engine to check words.

4/ In the composing step, did you find the tools useful? Which tool helped you the most? How did it help you?

In the composing step, Quillbot was useful. I also used Sketch Engine (concordance) to check words in context, such as "exhausted fumes".

5/ In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?

As I mentioned earlier, I tried using Sketch Engine to check whether the phrase "exhausted fumes" is used in a corpus, and it provided me with some results so I kept this phrase. For the remaining parts, I used Grammarly. However, it did not offer specific corrections because it required me to pay for premium access and I only see some highlighted phrases. I still think that Grammarly was useful in the revising and editing step. For example, I edited the sentence from "They can create" to "They can be created by unlimited sources in nature such as the wind heat or sunlight" based on its highlighted suggestions and the "Passive voice misuse".

6/ Can you give me some examples of challenges that the tools could not help you in this writing task?

Quillbot could not generate synonyms for "fossil fuels", so I didn't find this tool useful. From my previous experience, Grammarly usually identifies the use of passive voice as incorrect, yet I was taught that the passive voice is important for an IELTS essay.

7/ What do you want more from these four tools: Sketch Engine, Google Translate, QuillBot, and Grammarly?

I think Quillbot should offer more input data. Sketch Engine should provide more accurate synonyms and antonyms or allow imports from Google. Grammarly should offer better support for academic passive voice.

8/ Do you recommend or would use any other tools for this IELTS essay writing

practice? Why?

I recommend using a collocation dictionary (OZdict), thesaurus, Ctrl + F7 (Word), and searching for sample readings to generate ideas.

The responses regarding Interview Question 2 reveal some insights in Student 12's challenges after composing Essay 2, as she stated:

[63] No, I did not produce a higher-quality essay with the tools due to time constraints. I don't think these tools helped me write any better in 45 minutes. I am not good with ideas for the environment topic and I truly need something that can help me search for more ideas. In terms of language, I used Sketch Engine to check whether the phrase "exhausted fumes" is used in a corpus, and it provided me with some results so I kept this phrase. I didn't use Google Translate because I did not have any problem with writing my ideas in English. I did not have time to use Quillbot as well. For Grammarly, I did not offer any corrections because it required me to pay for premium access.

Accordingly, Student 12 stated that the 45-minute time constraint limited her ability to effectively utilise the four AI writing tools; she felt more confident with her first essay even though she spent a similar amount of time on both essays. Therefore, the time constraint was not the direct cause of her reduced writing performance; rather, it was the time spent on using tools that proved ineffective.

Additionally, her response to Interview Question 2: "I am not good with ideas for the environment topic, and I truly need something that can help me search for more ideas." reveals her struggle to generate ideas for the given topic likely contributed to her lower performance in Essay 2. This struggle can be referred to the decline in Task Response scores, which dropped from 8/8/6 in Essay 1 to 6/6/6 in Essay 2, as well as in her Coherence and Cohesion scores, which dropped from 7/7/6 to 6/6/6. This difficulty with idea generation is similar to the case of Student 3, resulting from an unfamiliarity with the topic of "renewable energy".

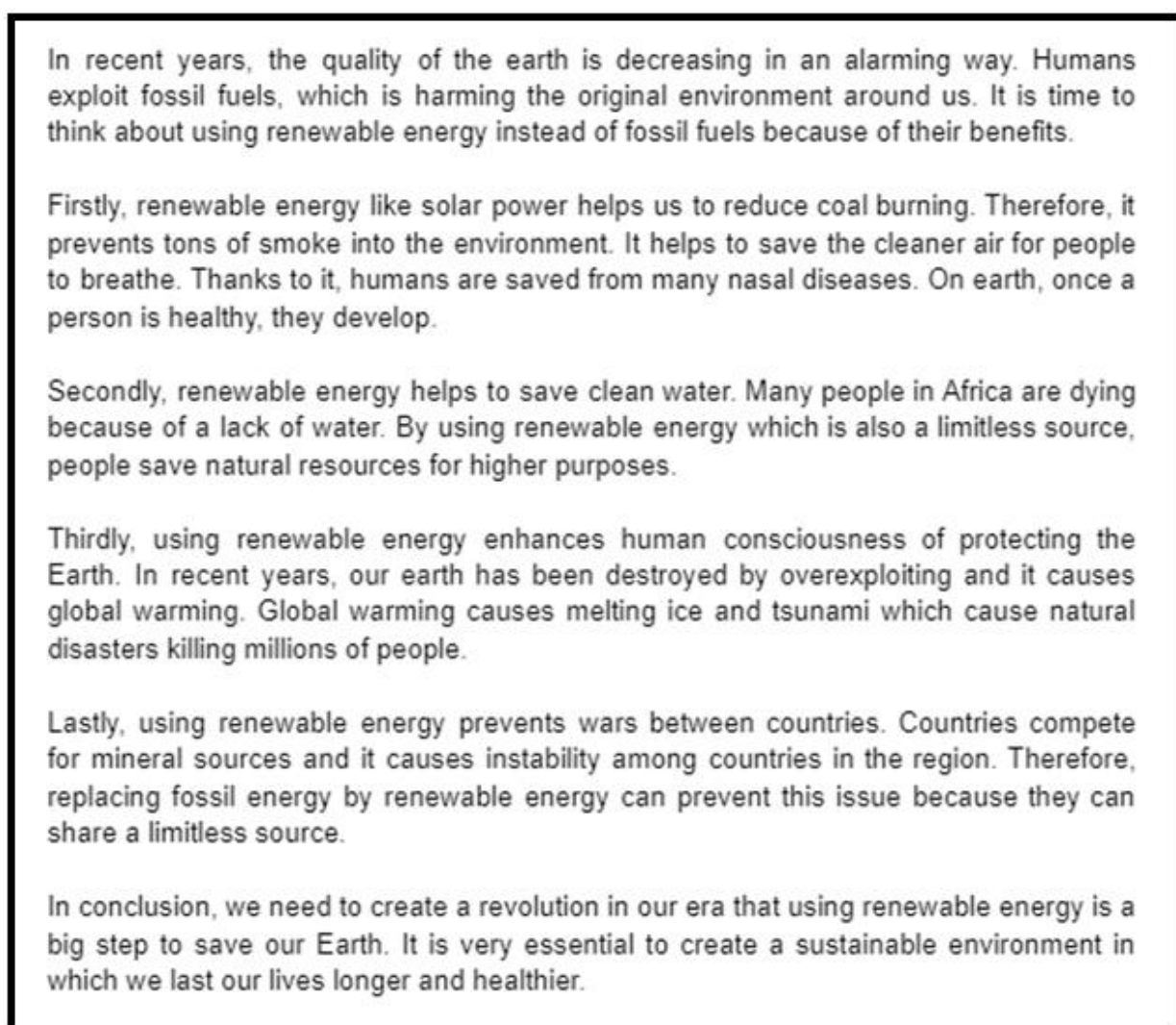
In response to Interview Question 5: “In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?”, Student 12 reported using Sketch Engine to check whether the phrase “exhausted fumes” appeared in a corpus before deciding to retain it:

[64] As I mentioned earlier, I tried using Sketch Engine to check whether the phrase “exhausted fumes” is used in a corpus, and it provided me with some results so I kept this phrase. For the remaining parts, I used Grammarly, However, it did not offer specific corrections because it required me to pay for premium access and I only see some highlighted phrases. I still think that Grammarly was useful in the revising and editing step. For example, I edited the sentence from “They can create” to “They can be created by unlimited sources in nature such as the wind heat or sunlight” based on its highlighted suggestions and the “Passive voice misuse”.

Moreover, after using Grammarly, Student 12 only edited one sentence: “They can be created by unlimited sources in nature such as the wind heat or sunlight” and did not find any other errors.”. However, there were five instances of imprecise word usage, as identified by three raters (Table 26), remained in her Essay 2: “find”, “search”, “created”, “unlimited”, and “exhausted fumes”. More grammatical errors in verb tense, conditional sentences, and unclear pronoun references were also noted. These errors indicate that the student had limited ability to critically assess her word choices and grammar errors. These challenges mirror those observed in Students 1 and 3, highlighting a recurring issue among the three participants in accurately assessing vocabulary and grammar accuracy in their work.

4.4.4 The Case of Student 17

The Essay 2 of Student 17 is shown in Figure 40.



In recent years, the quality of the earth is decreasing in an alarming way. Humans exploit fossil fuels, which is harming the original environment around us. It is time to think about using renewable energy instead of fossil fuels because of their benefits.

Firstly, renewable energy like solar power helps us to reduce coal burning. Therefore, it prevents tons of smoke into the environment. It helps to save the cleaner air for people to breathe. Thanks to it, humans are saved from many nasal diseases. On earth, once a person is healthy, they develop.

Secondly, renewable energy helps to save clean water. Many people in Africa are dying because of a lack of water. By using renewable energy which is also a limitless source, people save natural resources for higher purposes.

Thirdly, using renewable energy enhances human consciousness of protecting the Earth. In recent years, our earth has been destroyed by overexploiting and it causes global warming. Global warming causes melting ice and tsunami which cause natural disasters killing millions of people.

Lastly, using renewable energy prevents wars between countries. Countries compete for mineral sources and it causes instability among countries in the region. Therefore, replacing fossil energy by renewable energy can prevent this issue because they can share a limitless source.

In conclusion, we need to create a revolution in our era that using renewable energy is a big step to save our Earth. It is very essential to create a sustainable environment in which we last our lives longer and healthier.

Figure 40. The Essay 2 of Student 17

Analysis of tool usage

Student 17 began by drafting the initial version of the essay in English with no tools, relying on her English language skills. Her initial draft is shown in Figure 41.

In recent years, the quality of earth is decreasing in an alarming way. Humans exploit fossil fuels, which is harming the original environment around us. It is time to think about using renewable energy instead of fossil fuels because of their benefits.

Firstly, renewable energy like solar power helps us to reduce coal burning. Therefore, it prevents tons of smoke into the environment. It helps to save the cleaner air for people to breathe. Thanks to it, humans are saved from many nasal diseases

Secondly, renewable energy helps to save clean water. Many people in Africa are dying because of a lack of water. By using renewable energy which is also limitless, people save natural resources for higher purposes.

Thirdly, using renewable energy enhances human consciousness of protecting the Earth. In recent years, our earth has been destroyed by overexploiting and it causes global warming.

In conclusion, we need to create a revolution in our era that using renewable energy is a big step to save our Earth. It is very essential to create a sustainable environment in which we last our lives longer and healthier.

Figure 41. Student 17 composed the first draft of Essay 2 independently in English.

After composing the first draft, Student 17 translated a clause from Vietnamese to English, including: “melting ice, tsunami, destroying people because of natural disasters”, as shown in Figure 42



Figure 42. Student 17 translated a clause from Vietnamese to English

After that, Student 17 went back to Paragraph 4 and used some of the translated phrases, composing this sentence (Sentence 14): “Global warming causes melting ice and tsunami which causes natural disasters killing millions of people.” (Figure 43).

Thirdly, using renewable energy enhances human consciousness of protecting the Earth. In recent years, our earth has been destroyed by overexploiting and it causes global warming. Global warming causes melting ice and tsunami which causes natural disasters killing millions of people.]

Figure 43. Student 17 used some translated phrases and completed Sentence 14 independently

Following that, Student 17 returned to the second paragraph for editing, where she edited the word “causes” to “cause” (Figure 44).

Thirdly, using renewable energy enhances human consciousness of protecting the Earth. In recent years, our earth has been destroyed by overexploiting and it causes global warming. Global warming causes melting ice and tsunami which cause|natural disasters killing millions of people.

Figure 44. Student 17 revised the word “causes” to “cause” in Sentence 16 “

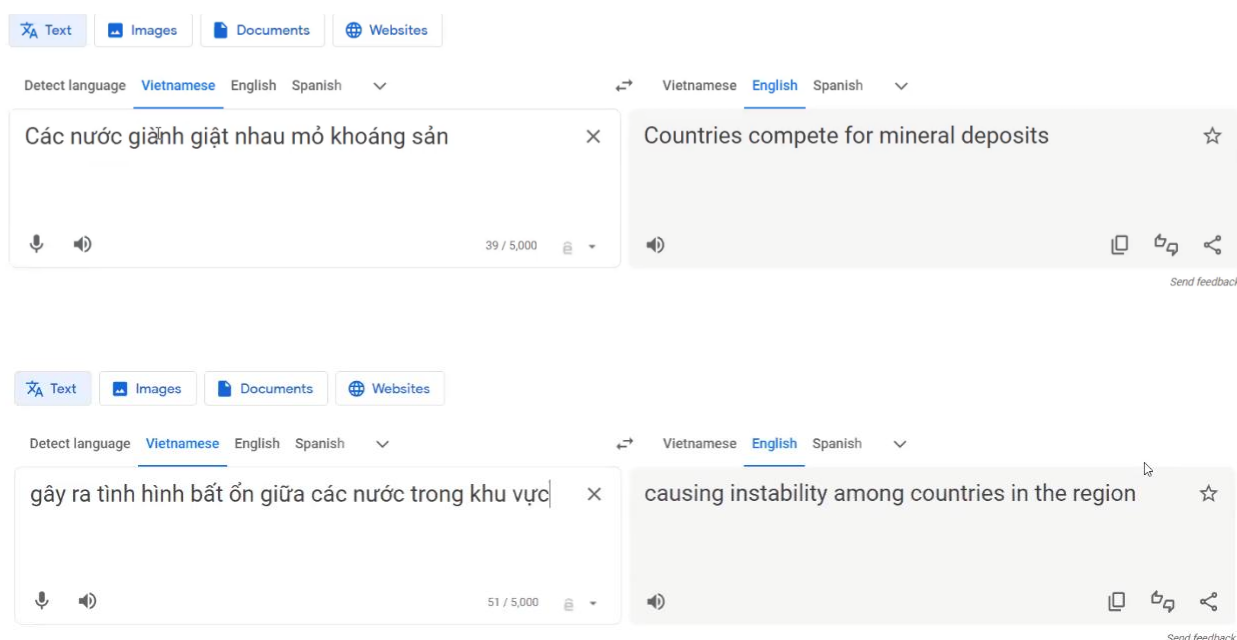


Figure 45. Student 17 translated two clauses of Sentence 16

After writing Sentence 16, Student 17 went back and composed a new paragraph with sentences 15, 16, and 17, before the conclusion paragraph. She composed Sentence 15— “Lastly, using renewable energy enhances human consciousness of protecting the Earth”—without any tools. She then used Google Translate to translate two separate clauses from Vietnamese into English: “Countries compete for mineral deposits” and “causing instability among countries in the region” (Figure 45), then combined these translated phrases into a single sentence without adjusting for grammatical accuracy, resulting in a clause fragment (“causing instability...”) that was not properly integrated into the sentence. This indicates that, although the tool assisted her in translating accurate phrases, she did not critically evaluate its grammatical correctness or restructure the output to form a full sentence, showing a reliance on Google Translate for direct clause translation. This reliance reflects limited behavioural and cognitive engagement.

Following that, Student 17 modified the phrase “mineral deposits” suggested by Google Translate to “mineral sources”, and composed the complete Sentence 16: “Countries compete for mineral sources and it causes instability among countries in the region.” (Figure 46).

Lastly, using renewable energy prevents wars between countries. Countries compete for mineral sources and it causes instability among countries in the region.

Figure 46. Student 17 revised the phrase “mineral deposits” – a phrase suggested by Google Translate – to “mineral sources” in order to compose Sentence 16

Subsequently, Student 17 returned to the second paragraph and inserted Sentence 8: “People on earth develop once they are healthy.” She then employed Quillbot to rephrase this sentence, as shown in Figure 47 and adopted the paraphrased version directly in her final draft without any modifications.

People on earth develop once they are healthy.

On earth, once a person is healthy, they develop. I

Figure 47. Student 17 used Quillbot to paraphrase Sentence 8

In the final step, Student 17 employed Grammarly to check for spelling and grammatical errors (Figure 48).

In recent years, the quality of earth is decreasing in an alarming way. Humans exploit fossil fuels, which is harming the original environment around us. It is time to think about using renewable energy instead of fossil fuels because of their benefits.

Firstly, renewable energy like solar power helps us to reduce coal burning. Therefore, it prevents tons of smoke into the environment. It helps to save the cleaner air for people to breathe. Thanks to it, humans are saved from many nasal diseases. On earth, once a person is healthy, they develop.

Secondly, renewable energy helps to save clean water. Many people in Africa are dying because of a lack of water. By using renewable energy which is also a limitless source, people save natural resources for higher purposes.

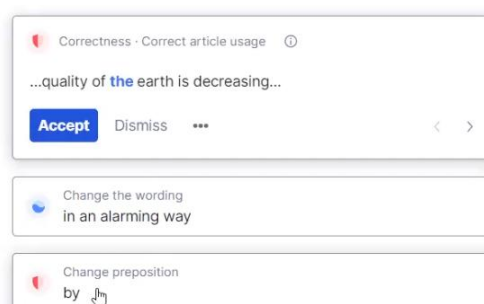


Figure 48. Grammarly identified an error with article usage in Sentence 1 and the preposition error in Sentence 17, but Student 17 only accepted the first change.

According to Figure 48, Grammarly found no spelling errors and proposed three suggestions for this essay: adding “the” before “earth”, changing the wording of the phrase “in an alarming way”, and changing the preposition in Sentence 17, from “replacing fossil energy by renewable energy” to “replacing fossil energy with renewable energy”. Student 17 only accepted the first suggestion, changing from “the quality of earth” to “the quality of the earth”. Therefore, she did not benefit from this suggestion of Grammarly.

To sum up, the screenshots from Student 17's work indicate that her composing process was not sequential; instead, she composed and revised her work nearly simultaneously with three tools: Google Translate, Quillbot, and Grammarly. Among these three tools, Google Translate was the primary one she needed to assist with her English writing. The screenshots also reveal that throughout the process, Student 17 actively engaged in decision-making, writing, and editing sentences based on her own English language skills. She critically assessed the outputs generated by these tools rather than passively accepting them. The following outputs— “melting ice, tsunami, destroying people because of natural disasters”, “Countries compete for mineral

deposits”, and “causing instability among countries in the region”—produced by Google Translate, as well as the phrase “On Earth, once a person is healthy, they develop” generated by Quillbot, will be analyzed in the next section.

Analysis of score changes

The score from both essays of Student 17, as marked by the three raters, is presented in Table 33.

Table 33. Essay 1 and Essay 2 scores of Student 17

Student 17 (IELTS 7.0)	TR1	CC1	LR1	GA1
Rater 1 (the researcher)	7	7	6	6
Rater 2	7	7	7	6
Rater 3	6	7	6	6

Student 17	TR2	CC2	LR2	GA2
Rater 1 (the researcher)	6	6	6	6
Rater 2	6	6	6	5
Rater 3	6	6	6	6

Compared with Essay 1, Student 17’s Essay 2 received lower scores from both Raters 1 and 2 in Task Response (–1 band) and from all three raters in CC (7/7/7 to 6/6/6). In LR and GA, only Rater 2 gave a lower score, while the other raters’ scores remained unchanged.

All three raters found 11 lexical errors in Essay 2: seven word-choice problems, one informal expression, one unclear pronoun, one preposition error, and one redundancy. Lexical precision was weak in items such as “source” instead of “resource,” “consciousness” instead of “awareness,” vague phrases like “the quality of the earth” and “original environment,” and misuse of connectors (e.g., “once” instead of “if,” “very essential” as redundant). They also noted four grammatical errors involving subject–verb agreement, missing words, article use, punctuation, and word form (e.g., “which is harming” instead of “which are harming,” missing elements in Sentence 5, misuse of “the,” missing commas before relative clauses, faulty noun modification, and “overexploiting” instead of “overexploitation”). Notably, none of these errors were generated by Google Translate or Quillbot; for example, the connector “once” came from the student’s own input, so Quillbot did not harm the text. Grammarly corrected only 2 of 6 grammatical errors and missed problems with agreement, omissions, relative clauses, and punctuation. One helpful suggestion (“replace by” → “replace with”) was not accepted. This limited uptake, combined with weak grammatical awareness, likely contributed to the lower LR and GA scores.

Analysis of interview excerpts

Table 34 below provides the complete transcript of Student 17’s interview responses.

Table 34. The interview transcript of Student 17

STUDENT 17
1/ In general, how do you feel about the usage of the four AI writing tools for writing this IELTS essay? Do you find them helpful? I find Google Translate the most helpful, followed by Grammarly and Quillbot. 2/ Overall, do you think you produced a higher quality essay with the tools compared to the first essay that you wrote without the tools? Which part or sentence do you think you wrote better with the tools? No, I did not produce a higher-quality essay with the tools. Among these four tools, I

only need Google Translate the most because I need a dictionary to search for words that I do not know in English, such as “tsunami”, “compete”, “mineral” and “instability”. Quillbot is also good at searching more synonyms but I did not have time to try out other paraphrase options with Quillbot or try Sketch Engine to check collocations. But these tools cannot help me when I run out of ideas. I am not familiar with any environmental issues, so it is more difficult for me to write the second essay. I think my vocabulary for this topic is not good enough compared to the first topic of distance learning.

3/ In the brainstorming and planning step (before writing), did you find the tools useful for you in this step? Which tool helped you the most? How did it help you?

I did not use any tools during the brainstorming and planning step.

4/ In the composing step, did you find the tools useful? Which tool helped you the most? How did it help you?

Google Translate was the most helpful tool during the composing step. Among these four tools, I only need Google Translate the most because I need a dictionary to search for words that I do not know in English, such as “tsunami”, “compete”, “mineral” and “instability”.

5/ In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?

I tried using Quillbot one time. It helped me paraphrase the sentence from “People on earth develop once they are healthy” to “On earth, once a person is healthy, they develop”, I think it sounded better than my previous writing, but I didn’t learn any synonyms.

Grammarly were the most helpful tools during the revising and editing step. Grammarly helped with corrections like “The Earth” and pointing out errors that I need to write in plural form, such as from “human” to “humans”, correct article usage from “the lack of water” to “a lack of water”, and I learned that I made a spelling mistake with “breathe”.

6/ Can you give me some examples of challenges that the tools could not help you in this writing task?

The tools did not help me generate more ideas.

7/ What do you want more from these four tools: Sketch Engine, Google Translate, QuillBot, and Grammarly?

I ran out of time for the second essay and I did not have enough time to fully exploit the potential of Sketch Engine and Quillbot.

8/ Do you recommend or would use any other tools for this IELTS essay writing practice? Why?

I recommend using Google Search for more ideas and searching for more sample essays to learn from them.

The responses regarding Interview Question 2: “Overall, do you think you produced a higher quality essay with the tools compared to the first essay that you wrote without the tools? Which part or sentence do you think you wrote better with the tools?” reveal some insights in Student 17’s challenges after composing Essay 2, as she answered:

[65] [...] But these tools cannot help me when I run out of ideas. I am not familiar with any environmental issues, so it is more difficult for me to write the second essay. I think my vocabulary for this topic is not good enough compared to the first topic of distance learning.

Based on interview excerpt [65], Student 17 struggled with the unfamiliar topic of environmental issues and found that the tools did not help her develop new ideas. This struggle can be referred to the decline in Student 17’s Task Response scores, which dropped from 7/7/6 in Essay 1 to 6/6/6 in Essay 2

This challenge was similar to the case of Student 3 and Student 12. This struggle can be referred to the decline in TR scores, which dropped from 8/8/6 in Essay 1 to 6/6/6

in Essay 2, as well as in her CC scores, which dropped from 7/7/6 to 6/6/6. Her difficulty with generating ideas is similar to the challenges experienced by Student 3.

Regarding Interview Question 5: "In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?", Student 17 answered:

[66] I tried using Quillbot one time. It helped me paraphrase the sentence from "People on earth develop once they are healthy" to "On earth, once a person is healthy, they develop", but I didn't learn any synonyms. [...] Grammarly were the most helpful tools during the revising and editing step. Grammarly helped with corrections like "The Earth" and pointing out errors that I need to write in plural form, such as from "human" to "humans", correct article usage from "the lack of water" to "a lack of water", and I learned that I made a spelling mistake with "breathe".

Based on the excerpt [66], Quillbot and Grammarly were evaluated highly by this student in helping her recognise errors in plural nouns, article usage, and spelling. However, despite using these tools, the student still made 11 lexical errors, including seven word choice errors (in sentences 1, 2, 8, 11, 12, 16, and 17), one instance of an informal writing style (sentence 18), one unclear pronoun reference (sentence 16), one preposition error (sentence 17), and one case of word redundancy (sentence 19), as identified by three raters. These errors suggest that the student had difficulty critically evaluating her own word choices and grammatical accuracy. This challenge mirrors those observed in Students 1, 3, and 12, highlighting a recurring issue among these four participants in accurately assessing vocabulary and grammar in their writing. Moreover, Grammarly corrected only 2 out of 6 grammatical errors identified in Essay 2 of Student 17, and it did not identify the remaining 4 errors, particularly those related to subject-verb agreement in a sentence with relative clauses (Sentence 2 and Sentence 18), missing words (Sentence 5), and punctuation (Sentence 14). This limitation in the automated writing evaluation tool was also observed in the cases of Students 1, 3, and 12, highlighting its inconsistent effectiveness in addressing both lexical errors and the aforementioned grammatical issues.

Regarding Interview Question 7: "What do you want more from these four tools: Sketch Engine, Google Translate, QuillBot, and Grammarly?", Student 17 responded:

[67] *“I ran out of time for the second essay, and I did not have enough time to fully exploit the potential of Sketch Engine and Quillbot.”*

The excerpt [67] indicates that, despite having the same time allocation, Student 17 felt she did not perform well on the second essay. It suggests that her use of the four AI tools was ineffective under time constraints, as they did not help her complete the essay within the given timeframe. This suggests that perhaps the tools took time away from essay composition. This difficulty of Student 17 is similar to Student 12, as discussed in Section 4.4.3, highlighting a recurring issue from the two student participants in using the four writing tools and accessing their output effectively under a restricted time.

4.4.5 Factors contributing to score decreases

First, interview data identified the first factor, an external technical issue for Student 1: a Quillbot's malfunction during composing limited the student's ability to use all four tools and may have hindered Essay 2 performance. As Student 1 noted, “Quillbot was not working during my composing time, so I had to find synonyms suggested by Google Translate”. While this external technical factor limited Student 1's access to all four tools during composing, it appears isolated, as no comparable failures were reported by other participants. More broadly across the four cases, the second internal factor emerges that is more consistently linked to score decreases: internal limitations in behavioural and cognitive engagement with tool outputs. In several instances, these four students accepted suggestions without checking contextual fit or accuracy, allowing inaccuracies to persist despite access to multiple tools. For example, in the case of Student 12, a highly proficient participant (based on Essay 1 scores) who also achieved a high score in Essay 2, incorporated the phrase “exhausted fumes” from Sketch Engine without recognising it as erroneous. This illustrates how insufficient critical evaluation and filtering of inaccurate tool outputs can lead to the inclusion of incorrect language in the essay, thus lowering writing scores outcomes after tool use. A similar pattern was observed in the case of Student 17, who uncritically combined translated phrases from Google Translate. With Grammarly, this internal factor was evident across all four students—Student 1 (7.7), Student 3 (7.0), Student 12 (6.4), and Student 17 (6.1)—regardless of changes in their Essay 2 scores. Even after using Grammarly, their final submissions still contained multiple errors, suggesting that the

primary cause of score decline lay in their lack of error awareness rather than in the tool itself. However, issues with the tools were not a consistently significant contributor to score decline, as the use of Sketch Engine, Google Translate and Grammarly did not produce comparable errors in the other cases. In contrast, these tools even facilitated vocabulary expansion for some learners, as discussed in the thematic analysis (Section 4.3.3). Thus, although the external factor reduced opportunities for lexical refinement in one case, the internal factor explains the use of inappropriate language and unresolved grammatical errors across these cases more directly.

Moreover, while unfamiliarity with the Essay 2 topic (“renewable energy”) posed difficulties for some students, the observed declines in writing performance were not attributable to the writing tools themselves. Instead, the main challenge lies in the students’ ability to generate ideas, include relevant arguments, and supporting details to address the writing topic within the time limit. This limitation in topic unfamiliarity did not uniformly result in score declines. Several high-scoring students, such as Students 7, 11, and 14, reported similarly restricted knowledge of the topic, yet they still achieved improved Essay 2 scores. Their interview responses reveal different coping strategies that helped offset topic-related challenges. For instance, Student 7 acknowledged, “I was not familiar with the environmental topic. Grammarly helped me with reviewing grammar tense, but most of the errors were identified by myself” (Student 7, high-scoring), indicating strong self-monitoring skills and effective grammar-focused tool use. Another noted, “Not really, because of the time demand and the topic. I only have time to use Quillbot and Google Translate” (Student 11, high-scoring), suggesting selective tool use to maximise efficiency under time pressure. Similarly, Student 14 (low-scoring, increased scores) reflected, “No, the second essay took more time, and I felt a lack of vocabulary, perhaps due to the topic familiarity. I did not have time to fully try Sketch Engine, as it provides too much information, but not exactly what I am looking for. I only used Google Translate and Grammarly. I also did not have time to try Quillbot” (high-scoring), which points to strategic prioritisation of these tools despite topic unfamiliarity. Taken together, these examples from Students 7, 11 and 14, whose Essay 2 scores increased, suggest that topic unfamiliarity and writing time constraint were not a main factor because they did not uniformly lower writing scores.

In summary, findings from the case studies analysis suggest that even highly proficient students (based on Essay 1 scores) and high-scoring students (based on Essay 2 scores) can still experience score declines when their tool-use strategies are constrained by external factors (e.g., technical issues) and internal factors (e.g., insufficient evaluation of outputs; inadequate behavioural and cognitive engagement).

4.5 Summary of findings in Chapter 4

Quantitative findings in 4.1, indicate that the intervention of four writing tools—Sketch Engine, Google Translate, Quillbot, and Grammarly—led to statistically significant improvements in students' writing scores. In Essay 2, 15 out of 20 participants increased their overall scores, yet none of the participants reached the maximum band score of 9. At a group-level analysis (see Table 19), writing score changes were significant in the criteria of Lexical Resource (LR) and Grammatical Range and Accuracy (GA). However, Task Response (TR) and Coherence and Cohesion (CC) showed no significant changes. Among the four tools, only Sketch Engine usage correlated moderately and significantly with higher Essay 2 scores ($r = .529, p = .016$), suggesting a potential relationship between its use and improved writing performance; whereas Google Translate, Quillbot, and Grammarly showed minimal correlations (see Table 20). The total number of tool usage instances showed a weak positive correlation with Essay 2 scores, though the relationship was not statistically significant ($r = 0.273, p = 0.244$), suggesting that simply using more tools does not automatically lead to higher scores. Moreover, a strong correlation was observed between mean overall scores of Essay 1 and Essay 2 ($r = .759, p < .001$); this indicates that while the tools contributed to overall score improvements, they did not substantially alter students' relative performance rankings. In other words, students who initially performed at lower levels in Essay 1 largely remained in the low-scoring group, while high-scoring students maintained their relative standing in Essay 2.

Findings from the exploratory thematic analysis of 20 students' interview responses in Chapter 4, Section 4.4 further indicate that quantitative gains in writing scores were not determined by tool access or usage frequency, but mostly by the students' behavioural and cognitive engagement with the tools (the CE theme), which were predominantly presented among the responses of high-scoring students. High-scoring students (who scored 6.5 and above in Essay 2) demonstrated more indicators of

behavioural and cognitive engagement, such as their ability to assess contextual appropriateness of the outputs, cross-checking outputs across tools, and rejecting unsuitable suggestions. This may explain why some high-scoring students (i.e. Students 7, 11 and 14) who reported having limited topic familiarity were still able to maintain or even improve their scores. By contrast, low-scoring students (who scored below 6.5 in Essay 2) demonstrated in their interview responses that they did not assess the outputs, or did so in a very superficial level, they only used one output at a time, and most notably, they did not reject any outputs, indicating less behavioural and cognitive engagement. The analysis of Vocabulary Acquisition (VA) and Grammar Improvement (GI) coding themes also reveal that high-scoring students in Essay 2 demonstrate more instances of new vocabulary acquisition and grammar improvement, compared to low-scoring students' responses. Taken together, findings from Section 4.4 suggest that students' cognitive engagement with the four tools is positively associated with both their writing scores and language gains.

Findings from case study analysis in Chapter 4, Section 4.4, reveal that writing score decreases in Essay 2 for Students 1, 3, 12, and 17 were associated with both external and internal factors rather than the tools themselves. For the external factor, a technical problem with Quillbot was found to restrict the successful tool usage of some students. For the internal factor, the main cause was limited behavioural and cognitive engagement. All four students revealed certain instances in which they accepted tool outputs without sufficient evaluation and did not reliably detect errors, for example, adopting "exhausted fumes" from Sketch Engine and uncritically combining phrases from Google Translate. As a result, language errors persisted in the essay, leading to declines in their writing scores. Additionally, findings from these four case studies indicate some that topic unfamiliarity and time limit hindered successful tool usage for some students. However, they were not the primary factor, as several other participants facing the same topic unfamiliarity challenge still successfully utilised the tools under time pressure and improved their Essay 2 scores. Furthermore, analysis of tool-use strategies and interview responses for Students 1, 3, and 12, who were high-scoring students, indicates that, although such students typically demonstrated more engagement and benefited more from these tools in terms of language gains (as discussed in Section 4.3.3) compared to low-scoring students, they can still experience score declines when their tool use is constrained by these external factors

(technical issues and writing time limits) and internal factors (insufficient behavioural and cognitive engagement).

CHAPTER 5: FINDINGS IN TOOL USAGE AND EVALUATIONS BETWEEN TEACHERS AND STUDENTS

RESEARCH QUESTIONS

RQ2: What similarities and differences exist between teachers' recommended strategies for using the four writing tools in IELTS essay preparation and students' actual usage patterns, and how are these similarities and differences reflected in students' essay writing performance?

RQ2a: What specific strategies do teachers recommend for using each of the four writing tools in IELTS essay preparation?

RQ2b: What strategies do students employ when using the four writing tools?

RQ2c: Did students whose tool usage shared similarities with teachers' strategies achieve higher Essay 2 scores compared to those whose tool usage differed?

Chapter 5 examines the strategies teachers and students employ when using the four writing tools to answer the main research question RQ2. Section 1 provides a summary and quantification of teachers' strategies to address the sub-question RQ2a: and students' tool usage to answer RQ2b. Additionally, an overall comparison of tool usage between teacher recommendations and student practices are discussed. Subsequently, detailed analyses for each tool are presented in the following sections: Sketch Engine (Section 5.2), Google Translate (Section 5.3), Quillbot (Section 5.4), and Grammarly (Section 5.5). In each tool analysis section, the researcher compared teachers' recommended strategies with students' usage patterns by examining data from teachers' lesson plans, students' screen recordings and interview transcripts, identifying both similarities and differences. Next, drawing on students' Essay 2 writing scores presented in Chapter 4, the researcher examined whether similarities in students' tool usage patterns with teachers' strategies were observed among high-scoring students (Essay 2 scores of 6.5 and above), or low-scoring students (Essay 2 scores below 6.5) to answer the sub-question RQ2c. A summary and discussion of

findings to answer the main RQ2 and its three sub-questions is presented in Section 5.6.

5.1 Comparative analysis of tool usage between teachers and students

To address the sub-research question RQ2a: “What specific strategies do teachers recommend for using each of the four writing tools in IELTS essay preparation?”, the researcher first analysed and summarised the data from ten teachers’ lesson plans, including activity notes from the teachers during their presentation in the meeting. The full lesson plans of ten teacher participants are presented in Table 35a in the Appendices. Table 35 below shows how one teacher’s lesson plans were summarised, giving one example from the larger Table 30a.

Table 35. An example of the lesson plan of Teacher 1.

Tool	Stage	Activity
Sketch Engine	Planning	The teacher selects some keywords: “speak” vs “write”, “nonverbal” and “verbal”, “communication”, “and powerful”. From the results of Sketch Engine, suggest students choose relevant nouns that go with “speak”: dialect, accent, tone, voice.
Google Translate		The teacher suggest that students write separate sentences as their initial thoughts about this topic and use Google Translate to assist in translating words. Example: “Spoken communication is less powerful than writing because it can be affected by dialect and accent and it is difficult to reach many people.”
No tool	Composing	Students work individually for 40 minutes and write a full essay.
Quillbot	Revising	This is a session for individual revision. The teacher suggests that students use Quillbot to get suggestions for their writing. Students will then selectively choose the sentences they want to paraphrase. This will help improve the quality and variety of their writing. Students will selectively choose the sentences to paraphrase

Grammarly		Students receive their own essays back after the teacher has seen them. The teacher allows students to use Grammarly to run the final check.

Next, the researcher provided an analysis draws on the data from across all ten teachers, focusing on the integration of the four writing tools: Sketch Engine, Google Translate, Quillbot, and Grammarly. This analysis involved thematic coding and categorising the usage of these tools according to the three main stages of the writing process: Planning, Composing, and Revising. Specific functions of each tool, as indicated in both lesson plans and teacher interviews, were also documented. To facilitate comparison with student usage, the data was converted to percentages. The percentages show the number or proportion of teachers that recommended a particular usage. However, it is important to note that this conversion is not intended to mask the small sample sizes of participants in this study (10 teachers and 20 students). Table 36 below provides a summary of the recommended usage of the four tools presented at three main stages of the writing process by 10 teachers.

Table 36. A summary of the recommended usage of the four tools presented at three main stages of the writing process by 10 teachers

Stage	Tool's name	Tool's function	% of teachers	Teacher IDs
Planning	Sketch Engine	Looking for collocations of a word or a phrase to form ideas	70%	1, 2, 3, 4, 5, 6, 7
Planning	Google Translate	Translating words or sentences from Vietnamese to English	90%	1, 2, 4, 5, 6, 7, 8, 9, 10
Planning	Quillbot	Replacing a word, rewriting a phrase or a sentence	10%	8
Planning	Grammarly	Checking spelling and grammatical errors of students' self-written sentences	20%	9, 10
Composing	Google Translate	Translating words or phrases from Vietnamese to English	10%	3
Composing	Quillbot	Replacing a word, rewriting a phrase or a sentence	50%	3, 5, 8, 9, 10

Composing	Grammarly	Checking spelling and grammatical errors	10%	8
Revising	Sketch Engine	Checking the concordance of phrases	20%	4, 10
Revising	Quillbot	Replacing a word, rewriting a phrase or a sentence	50%	1, 2, 4, 6, 7
Revising	Grammarly	Checking spelling and grammatical errors	90%	1, 2, 3, 4, 5, 6, 7, 9, 10

Subsequently, the researcher transcribed the interview responses from the 10 teacher participants. Table 37 below presents a sample transcription from Teacher 1. The complete set of transcriptions is available in Table 37a in the Appendices.

Table 37. An example of transcriptions from Teacher 1.

Teacher 1
1/ Based on your experience of participating in this study, how do you feel about using Sketch Engine for teaching IELTS essay writing? The tool can be used in the planning stage, but it is not suitable for other writing steps. 2/ If you were to use Sketch Engine again to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology do you think would be the most suitable for Sketch Engine? The most suitable steps of instruction or teaching methodology for this tool would be planning. 3/ Do you have any concerns about using Sketch Engine as a writing tool? What are they? No, I do not have any concerns about using Sketch Engine as a writing tool.

To address the sub-question RQ2b: “What strategies do students employ when using the four writing tools?”, the researcher first noted the students’ tool usage from their screen recordings and transcribed each student's interview responses. This information is mostly answered in the Interview Question 3: “In the brainstorming and planning step (before writing), did you find the tools useful for you in this step? Which tool helped you the most? How did it help you?”, Question 4: “In the composing step,

did you find the tools useful? Which tool helped you the most? How did it help you?” and Question 5: “In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?”. Table 38 presents an overview of student tool usage across the three primary writing stages: Planning, Composing, and Revising. To facilitate comparison with the teacher data, the number of students using each tool (out of a total of 20) also has been converted to percentages.

Table 38. A summary of the usage of the four tools presented at three main stages of the writing process by 20 students

Stage	Tool	Function	%	Student IDs
Planning	Sketch Engine	Finding concordances of a phrase	10%	6, 8
Composing	Sketch Engine	Looking for collocations	15%	1, 3, 10
	Google Translate	Translating words/phrases (VN → EN)	60%	2, 3, 6, 7, 9, 10, 11, 13, 14, 17, 18, 19
	Google Translate	Translating full essay (VN → EN)	10%	16, 20
	QuillBot	Rewriting/replacing words or sentences	25%	5, 7, 9, 10, 13
Revising	Sketch Engine	Checking concordances	20%	1, 9, 12, 13
	Google Translate	Translating (EN → VN)	10%	18, 19
	QuillBot	Rewriting/replacing words or sentences	45%	6, 7, 9, 11, 15, 17, 18, 19, 20
	Grammarly	Checking spelling & grammar	95%	1–8, 10–20

Subsequently, Table 39 presents a side-by-side comparison of the percentages for both groups, summarising both the recommended and actual use of the four tools at the three key stages of the writing process.

Table 39. A comparative summary of the recommended and actual usage of the four tools presented at three main stages of the writing process by teachers and students.

Stage	Tool	Function	% Teachers	% Students
Planning	Sketch Engine	Looking for collocations to form ideas	70%	10%
	Google Translate	Translating (VN → EN)	90%	0%
	QuillBot	Rewriting/replacing words or sentences	10%	0%
	Grammarly	Checking spelling & grammar	20%	0%
Composing	Sketch Engine	Looking for collocations	0%	15%
	Google Translate	Translating (VN → EN)	10%	60%
	Google Translate	Translating full essay (VN → EN)	0%	10%
	QuillBot	Rewriting/replacing words or sentences	50%	25%
	Grammarly	Checking spelling & grammar	10%	0%
Revising	Sketch Engine	Checking concordances	20%	20%
	Google Translate	Translating (EN → VN)	0%	10%
	QuillBot	Rewriting/replacing words or sentences	50%	45%
	Grammarly	Checking spelling & grammar	90%	95%

According to Table 39, while 70% of teachers recommended the usage of Sketch Engine during the planning stage, only 10% of students employed this corpus tool. At this stage, Google Translate was recommended by 90% of teachers but was not used

by any students. Grammarly was recommended by 20% of teachers, but no student used this tool at this stage.

In the composing stage, while only 10% of teacher recommend using Google Translate, a much higher proportion of students (60%) used Google Translate for word and phrase translation, with an additional 10% translating their entire essays. Quillbot was recommended by 50% of teachers but was utilised by only 25% of students. Additionally, while no teacher suggested using Sketch Engine, 15% of students used this corpus tool in this stage. The usage of Grammarly in this stage was reported by a small usage rate from teachers (10%) and none from students.

In the revising stage, Grammarly saw widespread use from both groups (90% of teachers and 95% of students). Teacher recommendations for Sketch Engine (20%) and Quillbot (50%) were also closely reflected in students' usage patterns, which stood at 20% and 45%, respectively, during this stage. Moreover, there was a small proportion of students (10%) who used Google Translate to do a backward translation from English to Vietnamese for revision, which no teachers advised.

To sum up, the comparative data analysis in percentages from Table 34 reveals both similarities and differences in tool usage between students and teachers. Overall, students' writing practices mostly mirrored with teacher guidance in the revising stage, particularly in the use of Grammarly, Sketch Engine, and Quillbot. A specific point of similarity was observed in the Concordance function of Sketch Engine, where a similar proportion (20%) of both teachers and students utilised this feature for revision purposes. By contrast, major differences were noted between the usage rates of the corpus tool (Sketch Engine) and the machine translation tool (Google Translate) across the planning and composing stages. To further explore these differences, the researcher conducted a more detailed analysis in the following section. This analysis examines the underlying tool usage purposes of teacher and students by drawing on data from screen recordings, lesson plan documents, and interview transcripts.

5.2 Comparative analysis of Sketch Engine

This section analyses similarities and differences between teachers' recommendations and students' usage of Sketch Engine using data from screen

recordings and interview transcripts. The tool usage is separated in three stages according to the participants' responses: Planning, Composing and Revising.

5.2.1 Planning

According to the lesson plans, seven teachers used Sketch Engine at the planning stage to build students' vocabulary and help them turn ideas into sentences. Teacher 1 asked students to enter key words (e.g., "speak," "write") and use Word Sketch to find collocations such as "dialect," "accent," "tone," and "voice," then write sentences with them. Teacher 2 guided students to search for synonyms and related phrases for terms like "spoken communication," "powerful," and "written communication" to expand their lexical range. Teacher 3 used Word Sketch to pre-select items such as "verbal communication," "facilitate communication," and "communication channels," highlighting them as a lexical base for coherent ideas before asking students to write full sentences. Teacher 4 used the Concordance function to show example sentences with "verbal cue" and "potent," asked students to infer meaning, and then had them write sentences such as: "In a conversation, using a verbal cue like saying 'Good job' can be a potent way to encourage someone." Teachers 5 and 6 combined Sketch Engine with Google Translate so students could gather and compare expressions and choose more precise wording. For example, Teacher 6 had students find high-frequency collocations for "communication" and "powerful," then draft outlines using these phrases. Teacher 7 similarly used Word Sketch for "message," "voice," and "emotion" to help students develop arguments. Despite methodological differences, all seven used Sketch Engine primarily as a vocabulary resource that students had to interpret and transform into ideas and sentences.

In comparison, only two high-scoring students (IDs 6 and 8) used Sketch Engine during the planning stage with a relatively similar approach to these seven teachers by extracting keywords from collocations to support idea development, as they stated in the interview:

[68] In the brainstorming and planning step, I found the Concordance function of Sketch Engine and Quillbot quite useful. At first, I used Quillbot to paraphrase the phrase "distance learning" from the topic and I got the synonym phrase "remote learning". Then I put "distance learning" and "remote learning" in

Sketch Engine to search for relevant concordances. From that, I can access other sentences about distance learning. I picked out some other good phrases that I can use, which are “social distancing”, “faculty and staff”, and “academic facilities”. From that, it helped me to think about what to write more regarding the merit of “distance learning” and helped me to write this sentence: “The faculty and staff of academic facilities uploaded the digital content to the cloud services, therefore, students can get an education without the interruption of the social distancing period.” (Student 6)

[69] For brainstorming, I found Sketch Engine useful. I used its concordance function to find more ideas for “distance learning”. I found some terms that appeared in the concordance examples, such as “home study” and “open learning courses”, which helped me with more ideas about its advantages. (Student 8)

Accordingly, Student 6 combined Quillbot and Sketch Engine to locate phrases such as “academic facilities” and “social distancing”, then composed a sentence from these two phrases. Student 8 used concordance lines from Sketch Engine to identify terms like “home study” and “open learning courses” to brainstorm for more ideas:

Overall, data indicates that a majority of teachers (70%) and a minority of students (Students 6 and 8), who are both from the high-scoring group, shared a similar underlying approach to Sketch Engine: extracting contextually relevant vocabulary and collocations from corpus results and using them as building steps for idea in argumentative writing. This suggests that while teachers and students find it useful to use Sketch Engine in the planning stage, and that it may contribute to high essay scores, very few students actually use this tool in planning their essays.

5.2.2 Composing

During the composing phase, teachers did not explicitly instruct Sketch Engine usage. In contrast, three students (IDs 1, 3 and 10), who are high-scoring participants, actively and independently used Sketch Engine during composition:

[70] I also used Sketch Engine to help me find more words related to “top-notch”, and then among the collocations I chose “world-class” to complete the

phrase “world-class education”. I think this word sounds more suitable compared to other high-frequency words like “matchless”, “triple-A”, or “first-rate”. (Student 1)

[71] For instance, I used SketchEngine to find keywords and phrases like “harness energy”. I used Google Translate to translate parts of this sentence: “Meanwhile, solar power generates energy directly from the heat and light of the sun without any by-products; and an electric car is far more environmentally friendly than a petrol-powered car due to the lack of exhaust fumes.”, and I was satisfied with its output. (Student 3)

[72] I tried Sketch Engine to find collocations and it suggested a collocation “tacitly” for “agree”. But after checking it with Google Translate, I found that this word doesn’t fit in my sentence, as I want to have a strong adverb with “agree”, such as “wholeheartedly” rather than “tacitly” so I skipped it and did not use this tool anymore. (Student 10)

These three students’ responses highlight their engagement and outcomes when using Sketch Engine during the composing stage. Student 1 used the tool to refine lexical choices, selecting “world-class” over other synonyms for “top-notch” to form the phrase “world-class education.” Student 3 also demonstrated an effective use of this tool, by identifying collocations like “harness energy” and using Google Translate to support comprehension; the student expressed satisfaction with the combined output. Student 10 demonstrated a not so successful gain with the phrase “tacitly agree” from Sketch Engine, but he still demonstrated his critical assessment of the tool’s output by “checking it with Google Translate”; he rejected the tool’s output because he wants to have “a strong adverb with ‘agree’, such as ‘wholeheartedly’ rather than ‘tacitly’”.

The difference in the number of teachers and students who used Sketch Engine during the composing stage suggests a gap between teaching plans and actual student practice. While teachers did not include Sketch Engine in their lesson plans for this stage, these three high-scoring students (IDs 1, 3, and 10) still demonstrated the ability to critically assess the tool’s output to verify or enhance their word choices during the writing process.

5.2.3 Revising

During the revising stage, only two teachers (Teacher 4 and Teacher 10) integrated Sketch Engine with Quillbot and Grammarly to support vocabulary refinement. Their shared instructional goal was to guide students in verifying the appropriateness of lexical usage with this corpus tool. Teacher 4's lesson plan outlined a structured three-step revision sequence: first using Quillbot for paraphrasing, then applying Sketch Engine to check collocations and concordance lines, and finally conducting peer review. This instructional approach was further elaborated in the interview, as Teacher 4 explained:

[73] In my revision activities, I encouraged students to take a structured approach. First, they paraphrase their original sentences using Quillbot to improve lexical variety. Then, they use Sketch Engine to verify whether the new expressions make sense by checking collocations and concordance lines in authentic contexts. This step helps them evaluate the appropriateness of vocabulary usage. (Teacher 4)

Teacher 10 described a revision strategy involving the combination use of Grammarly, Quillbot, and Sketch Engine:

[74] I believe using the corpus tool as an additional language reference for altering the original words would be the most optimal approach. For example, in terms of preposition errors, after students complete their essays, I can guide them to review the changes made by Grammarly and determine if those changes align with the most frequent collocations. Additionally, I can ask students to compare the synonyms suggested by Quillbot with the highest frequency words based on Sketch Engine, and then decide the final sentence based on the similarity of these two tools. (Teacher 10)

The lesson plans of Teacher 4 and Teacher 10 shared a common approach to revision by integrating other tools with Sketch Engine, including Quillbot, and Grammarly. In their approach, Sketch Engine consistently serves as a final tool to verify the outputs from Grammarly and Quillbot.

In comparison, the use of Sketch Engine during the revising stage was noted in only one high-scoring student (Student 9) and two low-scoring students (Students 12 and 13). Student 9 answered in her interview that she integrated Sketch Engine with Quillbot during revision:

[75] I usually have a difficult time rewriting my own sentences, but Quillbot outputs helped me to do it faster and explore more synonyms. After a few trials, I find that some phrases are even better than my original words, such as “mitigate”, “alleviate the negative environmental implications” and “pose ecological challenges”, which seem to be more formal than my original phrase [76] “reduce the adverse environmental impacts” and “have potential ecological risks”. I also tried the word “mitigate” in Sketch Engine to check for its collocations. After checking with Sketch Engine, I found the noun “risk” was the first collocation that can combine with the verb “mitigate”, so I chose the phrase “mitigate risk”. (Student 9)

Accordingly, Student 9 began by using Quillbot to generate more formal expressions, then turned to Sketch Engine to check if the new word combinations were accurate, such as confirming that “mitigate” commonly collocates with “risk”. This combined use of tools closely reflected the approach of Teacher 4 and Teacher 10 to verify Quillbot’s suggestions by consulting Sketch Engine.

According to the screenshots in Figures 49 and 50 below, Student 13, a low-scoring student, used Sketch Engine twice to verify two phrases, but these are her self-written phrases and not from Quillbot’s output. First, she searched for “green energy” (Figure 49) but decided not to use it. Later, she checked “CO2 emissions” (Figure 50) and incorporated this phrase into her sentence about air pollution.

[78] I used Sketch Engine to check whether the phrase “exhausted fumes” is used in a corpus, and it provided me with some results so I kept this phrase. (Student 12)

As discussed in Chapter 4, Sections 4.3 and 4.4, the case study of Student 12—a low-scoring participant—demonstrated a superficial use of both Sketch Engine and Grammarly, marked by minimal cognitive engagement with the tools’ output in her interview response. Therefore, the analysis of Students 9, 12, and 13 reveals important differences in how Sketch Engine was used during the revising stage among the high- and low-scoring groups of students in comparison with teachers’ recommended strategy. While Student 9, a high-scoring participant, demonstrated a strategic combination of tool use (using Sketch Engine to verify the output of Quillbot), which mirrors teachers’ recommendations (Teachers 4 and 10), the other two low-scoring students showed limited engagement with this tool. Student 13 did not mention Sketch Engine at all in her interview, despite evidence of recorded usage, instead highlighting Grammarly as her primary tool, suggesting a possible lack of engagement with Sketch Engine in her experience or a lack of understanding in how to use the corpus tool effectively. Student 12, though attempting to use Sketch Engine, failed to detect a lexical error in her input (“exhausted fumes”) and failed to assess the outputs from this tool. These findings indicate that effective use of this corpus tool during revision, particularly in the same multiple tool integration approach of teachers, largely depend on students’ ability to critically assess Sketch Engine’s output, a skill that is more evident in the higher-scoring student (Student 9). By contrast, low-scoring students (i.e. Students 12 and 13) were unable to use Sketch Engine effectively as a final verification tool for the outputs generated by Quillbot or Grammarly. Rather than using the corpus to expand their vocabulary or refine their use of contextually appropriate phrases, they employed it merely to confirm whether their chosen words existed in native English usage. As a result, Sketch Engine was not used as a resource to enhance their linguistic knowledge, which does not align with teachers’ aims.

To sum up, across planning, composing, and revising stages, both similarities and differences were observed in how teachers and students used Sketch Engine. In the planning stage, Teachers 1, 2, 3, 4, 5, 6 and 7 guided students to explore collocations, concordances and build ideas from these topic-relevant key words or key phrases

using Sketch Engine. This similar usage pattern, however, was limited to a small number of students, as only two high-scoring students (Students 6 and 8) used the tool with a similar manner: extracting phrases from concordance lines to support idea development. In the composing stage, no teachers incorporated Sketch Engine, possibly viewing vocabulary work as complete after planning. In comparison, three high-scoring students (Students 1, 3, and 10) used it independently to refine word choices. Their critical evaluation of collocations also showed lexical awareness and autonomy, contrasting with the absence of teacher-guided use at this stage. In the revising stage, Teachers 4 and 10 used Sketch Engine with Quillbot and Grammarly in a structured multi-tool process to check collocations and revise vocabulary. Among students, only Student 9 mirrored this approach, while Students 12 and 13 used the tool with limited engagement. These differences suggest that higher-scoring students (e.g., Students 1, 3, 6, 8, and 10) were able to engage with corpus tools more critically, often reflecting tool usage strategies recommended by teachers. In contrast, lower-scoring students (e.g., Students 12 and 13) demonstrated a superficial usage of this tool and were not able to critically assess or apply corpus data effectively in their writing. Furthermore, their usage was limited to checking language accuracy rather than using the tool to expand or develop their ideas. These relatively low percentages of both teachers and students who utilised Sketch Engine during revision suggest that this tool may not have been perceived as useful for refining drafts for IELTS essay. Overall, these findings indicate that students who shared similarities with teachers' strategies in the use of Sketch Engine, especially during planning and revising, achieved higher Essay 2 scores (Students 1, 3, 6, 8, 9 and 10) compared to those whose tool usage differed. However, it cannot be concluded that similarities in Sketch Engine use during the composing stage had the same result, given that teachers did not integrate it for this purpose and very few of the high-scoring students reported using it.

5.3 Comparative analysis of Google Translate

This section analyses similarities and differences between teachers' recommendations and students' usage of Google Translate during Planning, Composing and Revising.

5.3.1 Planning

Lesson plans and interviews show that nine of ten teachers (1, 2, 4, 5, 6, 7, 8, 9, 10) used Google Translate at the planning stage. Teacher 1 used it in planning and composing, asking students to draft in Vietnamese, then translate into separate English sentences to improve lexical accuracy. Teacher 2 had groups brainstorm situations in Vietnamese, translate with Google Translate, and then use the English output for peer sharing and outlining; she described planning as “the most suitable step” because the tool is “easy to use and fast.” Teachers 4, 5, 6, 7, 8, 9, and 10 all followed a similar pattern: students first generated ideas in Vietnamese, then translated them into English to support understanding of the question, idea expansion, and vocabulary building, often in combination with Sketch Engine or Grammarly. For example, Teacher 8 encouraged free brainstorming in Vietnamese, then had students use Google Translate “to explore how those ideas might be expressed in English,” especially for beginners. Teacher 10 likewise asked students to choose a stance, brainstorm in Vietnamese, then translate key points into short English sentences to reduce anxiety at the start of writing. Across these nine teachers, Google Translate was consistently positioned as a planning tool to mediate idea generation between Vietnamese and English, particularly for lower-proficiency learners, helping them gain confidence in expressing ideas before composing the full essay.

In contrast to the teachers' strategies, the students' tool usage was markedly different, as screen recordings revealed that none of the 20 participants employed Google Translate during the planning phase. As presented in Table 33, only two high-scoring students (Student 6 and Student 8) showed an attempt at using a tool for planning, as they used the Concordance function of Sketch Engine to search for keywords and phrases to generate more ideas. The remaining 18 students bypassed the planning stage entirely. Instead, they proceeded directly to the composing stage of their writing task. This difference indicates that students may not fully understand the role of tool usage during planning in improving writing skill.

5.3.2 Composing

In the composing stage, only Teacher 3 included Google Translate in her lesson plan as follows:

[79] In the composing stage, students can use tools like Google Translate to overcome language barriers when expressing their ideas in English. However, it's important to note that the accuracy of translations depends on the clarity of the original ideas. Additionally, students can use Quillbot to paraphrase their thoughts, transforming simple statements into more scholarly language.

She further explained this in her interview:

[80] In the composing stage, I guide students to use Google Translate as a starting point to express their ideas in English, especially when those ideas were originally developed in Vietnamese. Many students struggle to write directly in English, so translating full Vietnamese phrases helps them produce a basic draft and overcome initial language barriers. However, I emphasize that Google Translate is not a shortcut or the final product, because it often produces literal or overly simple sentences. To address this, I encourage students to use Quillbot immediately afterward to paraphrase and improve the sophistication of their writing. For example, a simple translated sentence like “In some situations, speaking is more effective than writing” can be rephrased through Quillbot as “In certain circumstances, verbal communication outperforms written communication.” This process helps students not only express their ideas more clearly but also refine vocabulary, sentence structure, and academic tone. I see this combination of tools as a scaffolded approach that supports reflective learning and enables students to produce clearer, more polished writing. (Teacher 3)

According to her lesson plan and interview, Teacher 3 sees Google Translate as a helpful starting point during the composing stage, allowing students to express ideas initially written in Vietnamese. For example, she suggests students translate a full Vietnamese sentence like “Trong một số tình huống thì giao tiếp bằng lời nói sẽ hiệu quả hơn là bằng văn bản” into English as “In some situations, speaking is more effective than writing.” However, she made it clear that Google Translate is only the first step and should not be used as the final product, as its output can be overly basic or literal. To improve the quality of writing, she instructed students to follow up with Quillbot to paraphrase and enhance sentence structure and vocabulary. This tool combination, in her view, helps students produce more polished, academic writing by

building on an initial draft rather than relying solely on machine translation.

In comparison, 12 students (IDs 2, 3, 6, 7, 9, 10, 11, 13, 14, 17, 18, and 19) reported using Google Translate during the composing stage:

[81] I know one of the words to describe “cận thị” is “being short-sighted”, but from Google Translate I learned another word that is “myopic”. (Student 2)

[82] Yes, in the composing step, I found the tools useful. For composing, I mostly need Google Translate to look for some phrases in English, such as the phrase “petrol-powered car”, “coal”, and “burning coal releases CO2”. (Student 3)

[83] Google Translate played a major role in the composing step. It helped me look for certain words I can’t remember in English. I think I only need this tool at this step. (Student 6)

[84] In the composing step, I used Google Translate to find the English equivalent of the Vietnamese word “ủng hộ”. (Student 7)

[85] I mostly used Google Translate to help me write the two phrases: “water wheel”, “non-degradable” and the two sentences: “Although renewable energy sources help alleviate the negative environmental implications of fossil fuels, they still pose ecological challenges and are not the optimal energy source.”, “Some wind farms are built in the wrong places which threatens the migratory flow of native birds as the turbines are deadly danger to birds if hit”. (Student 9)

[86] Yes, I found the tools useful in the composing step. Quillbot was particularly helpful. I tried Sketch Engine to find collocations and it suggested a collocation “tacitly” for “agree”. But after checking it with Google Translate, I found that this word doesn’t fit in my sentence, as I want to have a strong adverb with “agree”, such as “wholeheartedly” rather than “tacitly” so I skipped it and did not use this tool anymore. (Student 10)

[87] In the composing step, I found Quillbot, and Google Translate useful. I used Google Translate to check the translations of the word “vô tận” (Vietnamese) and decided to use the word “inexhaustibility”. (Student 11)

[88] [...] I also used Google Translate for single words. For example, I needed to check the word: “bệnh hô hấp” and I got “respiratory disease”.. (Student 13)

[89] For this step, I only used Google Translate to check for English words and my spelling. I translated two words from Vietnamese to English: “promote” and “negligence”. (Student 14)

[90] Google Translate was the most helpful tool during the composing step. Among these four tools, I only need Google Translate the most because I need a dictionary to search for words that I do not know in English, such as “tsunami”, “compete”, “mineral” and “instability”. (Student 17)

[91] Google Translate and Quillbot were the most useful in the composing step. I wrote in simple English and then used Google Translate as a dictionary. There were two words: “hydropower” and “landbound” that I was not sure and I needed to confirm their Vietnamese meaning with Google Translate. (Student 18)

[92] During the composing step, I found Google Translate and Quillbot to be the most useful. I wrote in simple English and then used Google Translate as a dictionary. [...] I was struggling to remember the English word for “ban hành luật”, and Google Translate helped me solve this problem, now I know it is “promulgate”. (Student 19)

When compared with the approach advocated by Teacher 3, five high-scoring students (IDs 10, 11, 18, and 19) and one low-scoring student (ID 13) demonstrated similar practices by using Google Translate not only to work with more complex or technical vocabulary but also in combination with other tools during the composing stage. Student 10 used multiple tools but with a different sequence: after trying Sketch Engine for collocations and finding “tacitly agree,” he cross-checked the meaning with Google Translate, realised it didn’t fit his intended idea, and instead opted for a stronger adverb like “wholeheartedly”. Likewise, Student 11 used Google Translate to translate the Vietnamese word “vô tận” and selected the more academic term “inexhaustibility,” then used Quillbot to refine sentence clarity. Student 18 reported writing in simple English and then using Google Translate to verify the meanings of words like “hydropower” and “landbound,” before revising with Quillbot. Similarly, Student 19 used Google Translate to find the correct translation for “ban hành luật” as

“promulgate,” and described how he relied on Quillbot to enhance overall clarity and sentence quality. Student 13 confirmed the translation of “bệnh hô hấp” as “respiratory disease” and complemented this with Quillbot to improve the phrasing of her writing.

In contrast, four high-scoring students (IDs 2, 3, 6, 7) and two low-scoring students (IDs 14 and 17) used Google Translate in a more limited, dictionary-like manner during the composing stage, primarily for instant translation of individual words or short phrases. Their responses suggest a focus on retrieving vocabulary rather than engaging in a broader, reflective writing process. For example, Student 2 reported learning the word “myopic” as another way to express “cận thị,” treating the tool as a quick vocabulary finder. Student 3 used Google Translate to look up specific phrases such as “petrol-powered car” and “burning coal releases CO₂,” without mentioning any subsequent revision or refinement. Similarly, Student 6 stated that she used the tool only to find English words she couldn’t recall, indicating a utilitarian purpose. Student 7 relied on it to find the English equivalent of “ủng hộ,” and Student 14 translated only two words, “promote” and “negligence,” without mentioning any further language development. Student 17 emphasised that Google Translate was the most useful tool because it served as a dictionary for unfamiliar words such as “tsunami,” “mineral,” and “instability.” Unlike the tool combination approach described by Teacher 3, these students did not appear to use follow-up tools such as Quillbot or Grammarly to refine writing, revealing a gap between instructional intention and actual learner practices.

Notably, Student 16 (low-scoring) and Student 20 (high-scoring) composed their entire essays in Vietnamese and translated them with Google Translate—an approach not recommended by teachers (see Figures 51 and 52)—yet both showed score gains in Essay 2.

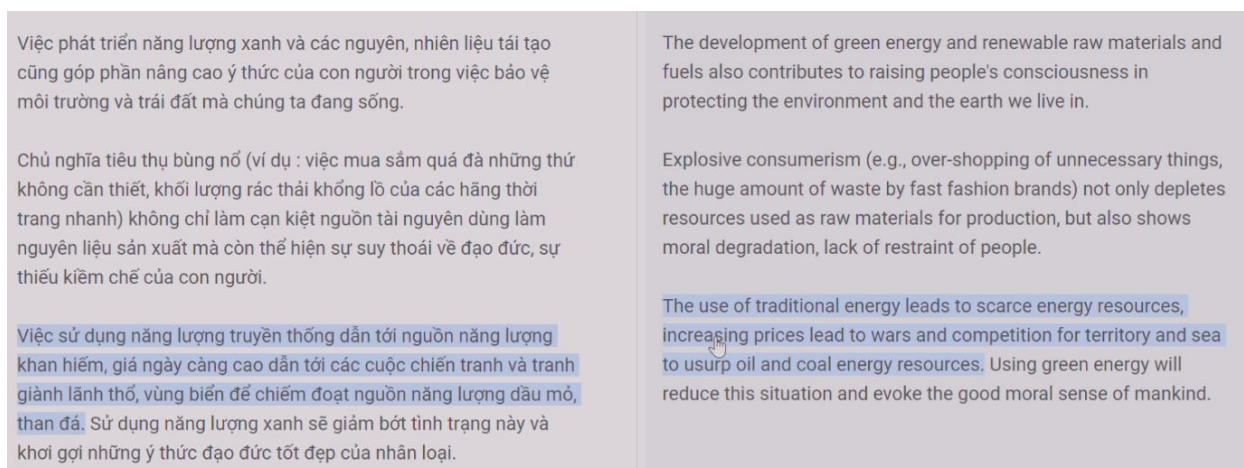


Figure 51. Screenshots from the computer of Student 16 using Google Translate during her second essay composition

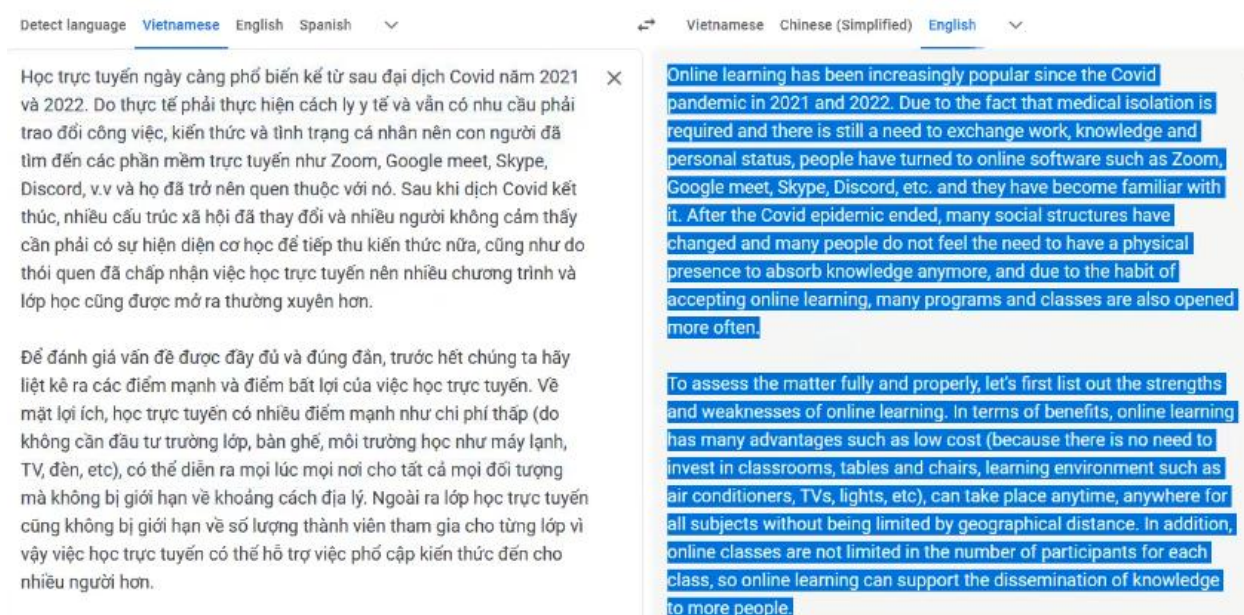


Figure 52. Student 20 composed in Vietnamese and used Google Translate to translate the whole essay

These two students responded to the interview question: “4/ In the composing step, did you find the tools useful? Which tool helped you the most? How did it help you?” as follows:

[93] These tools are so useful for fast writing within 45 minutes. I can write in Vietnamese, and then use Google Translate, check with Grammarly, and use Quillbot to produce an essay. But I didn't have time to use Quillbot. With Google Translate helping me with my whole essay, I don't even have to think in English,

and I think its translation quality is better than my English level. (Student 16)

[94] Google Translate produces great quality English translation for my second essay [...] (Student 20).

The responses of Student 16 and Student 20 indicate that they used Google Translate primarily to quickly produce English texts, rather than to support their language learning. Moreover, their method of translating their entire essays from Vietnamese into English differs from all teachers' instructional recommendations and reflects some teachers' concern, such as "I'm afraid that my students will become too lazy to think in English with Google Translate." (Interview, Teacher 7), and "It's easy to use but I would not recommend it for composing" (Interview, Teacher 9).

As detailed in Table 16 (Section 4.2.4), Student 16's score increased from 4.8 in Essay 1 to 5.9 in Essay 2, while Student 20's rose from 6.5 to 6.9—indicating their respective positions as a lower-proficiency (Student 16) and higher-proficiency (Student 20) student. Despite both using Google Translate to generate full essay drafts, their levels of cognitive engagement with the output diverged significantly, as revealed in (Figure 51) Student 16 relied almost exclusively on Google Translate, made minimal revisions, and did not incorporate support from other tools like Quillbot or Sketch Engine. In this case, the writing score gains that Student 16 achieved demonstrates tool-generated text rather than the student's own work, raising concerns about authorship and assessment validity. Moreover, as revealed from thematic analysis in (Sections 4.3.2 and 4.3.3) there was no indicator of language learning from the way the tool was used by Student 16.

By contrast, Student 20 reviewed and refined the translated text with Grammarly and Quillbot, demonstrating deeper behavioural and cognitive engagement. Student 20 accepted most paraphrases but rejected redundant content, as he noted ([95]):

[95] After checking all errors with Grammarly, I used Quillbot to paraphrase the whole essay. I think the texts generated by Quillbot sound more formal to me in general, compared to those of Google Translate. I accepted all of its output, except for the last sentence of the conclusion: "Overall, online learning is an inexorable advancement of technology and time, therefore we must learn to accept it and find solutions to the remaining issues. It still has a lot of strengths

and disadvantages that need to be analyzed.”. The last sentence: “It still has a lot of strengths and disadvantages that need to be analyzed.” I think is a bit redundant, so I did not include it. I can finish an essay with just Google Translate and Quillbot in 20 minutes and I just need to review. (Student 20)

Taken together, these cases indicate that tool adoption alone does not determine outcomes; rather, writing quality is shaped by the depth of cognitive engagement with tool output. In this respect, only Student 20’s process aligns closely with teachers’ instructional goals that prioritise skill development over surface-level score gains.

5.3.3 Revising

During the revising stage, teachers did not recommend Google Translate. By contrast, screen captures from the computers of two students (IDs 18, and 19) reveal that they used Google Translate for revision, as presented in Figures 53 and 54.

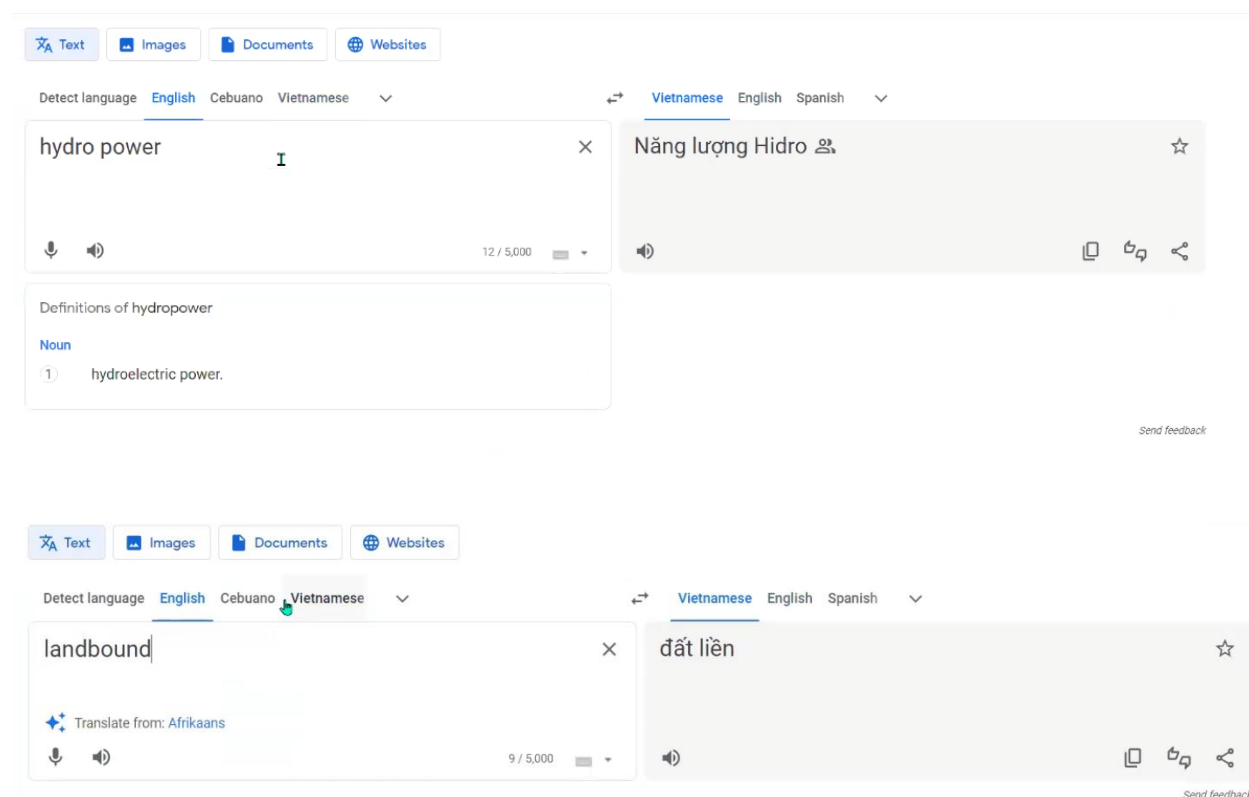


Figure 53. Student 18 used Google Translate from English to Vietnamese to revise her phrases.

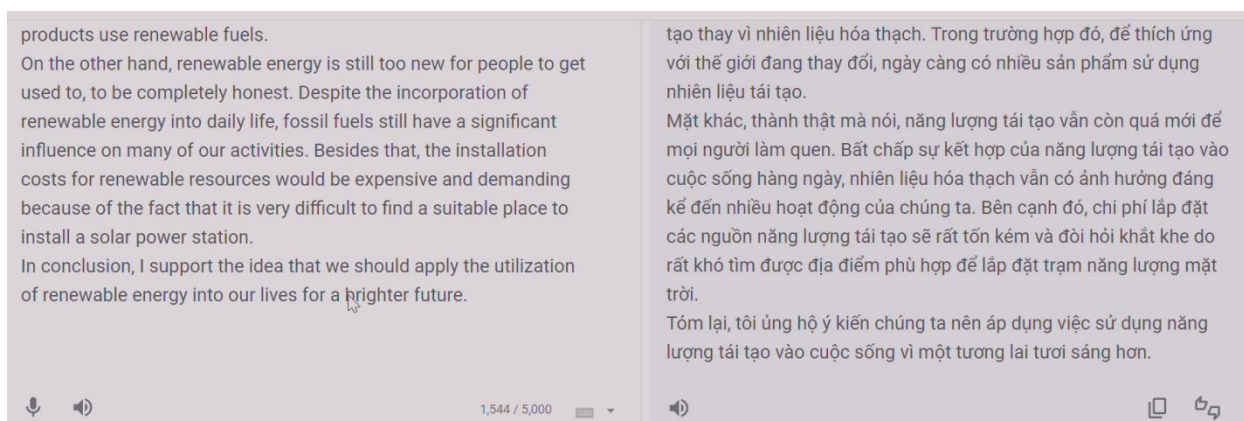


Figure 54. Student 19 used Google Translate from English to Vietnamese to revise her written essay.

Despite clear evidence from screen recordings (Figures 56 and 57), neither student mentioned this practice in their interviews. Instead, their responses to Question 5: “In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?” were as follows:

[95] It is Quillbot. It helps me change my informal writing style, such as “countries with less resources are put in the lower hand in negotiation” to “countries with fewer resources are given a weaker negotiating position”. (Student 18)

[96] Grammarly was the most helpful tool in the revising and editing step. It assisted with prepositions, articles, and spellings. For example, it corrected my phrase “into daily life” to “in daily life” (Student 19).

Accordingly, Students 18 and 19, both of whom showed improvements in their Essay 2 scores, identified Quillbot and Grammarly as the most useful tools during the revision stage, particularly for enhancing sentence structure and correcting grammatical errors. Although neither student explicitly reported using Google Translate at the revision stage—implying they may not have viewed back-translation as a preferred revision strategy—the approach nonetheless appears deliberate and was associated with score gains: Student 18’s score rose from 6.3 to 6.8, and Student 19’s scores rose from 5.9 to 6.3. Moreover, although the use of Google Translate during revision was

not included in any teacher's instructional plan, this student-led strategy reveals a potentially beneficial practice. These findings highlight the importance of providing students with explicit guidance on how to match tools to specific writing stages. As such, teachers may consider incorporating backward translation as an optional but strategic step in the revision process, particularly when paired with cognitive engagement and follow-up editing tools.

To sum up, across the planning, composing, and revising stages, clear differences emerged in how teachers and students approached Google Translate. In the planning stage, nine of ten teachers (Teachers 1, 2, 4, 5, 6, 7, 8, 9, and 10) recommended using Google Translate to translate L1 ideas into English, expand vocabulary, and build confidence—especially for lower-proficiency learners—and often paired it with brainstorming, peer exchanges, or tools such as Sketch Engine and Grammarly. However, screen recordings show that none of the 20 students used Google Translate during planning. As a result, the anticipated planning-phase benefits were not revealed in practice, and the dataset provides no evidence that such a use would improve Essay 2 scores for lower-scoring students, or that students used Google Translate in the way that teachers intended. A plausible explanation is that students—particularly those with lower Essay 2 scores—experienced cognitive load and time pressure at the outset and therefore prioritised producing simple English sentences over an additional translate-while-planning step; unfamiliarity with the topic (e.g., Students 12 and 17) may also have limited idea generation even in L1, rendering a translation tool less useful because it presupposes content to translate. More broadly, most students tend to perceive Google Translate as a vocabulary support—primarily to look up unknown English words—rather than as a planning scaffold, creating a practical and conceptual misalignment with instructional objectives. Therefore, recommending translation at the planning stage is unlikely to be effective unless students receive clear guidance, dedicated time, and explicit reasons for using it. Without these supports, they are unlikely to use the tool in ways that promote learning or improve scores.

During the composing stage, only one teacher (Teacher 3) explicitly guided students to use Google Translate as a first step to express ideas in English, followed by Quillbot to modify to academic writing style. Twelve students (IDs 2, 3, 6, 7, 9, 10, 11, 13, 14, 17, 18, and 19) reported using Google Translate during the composing stage. Among

these 12, five high-scoring students and one low-scoring student adopted a similar tool-combination approach, critically refining their writing with Quillbot. In contrast, other students used Google Translate merely as a dictionary for word-level translation. Notably, two students (Student 16 and Student 20) translated their entire essays from Vietnamese to English—an approach not recommended by any teachers. Student 20 critically revised the translation with Quillbot, while Student 16 relied solely on Google Translate.

During revision, teachers did not recommend using Google Translate. Nevertheless, two students (IDs 18 and 19) used it for back-translation (from English to Vietnamese) to support their revisions, though they did not report this practice in interviews. Neither student explicitly attributed benefits to this strategy, yet both showed gains in Essay 2 scores (Student 18: 6.3 → 6.8; Student 19: 5.9 → 6.3). While based on only two cases, this tool usage suggests that using Google Translate for back-translation could be considered as an optional revision that teachers can adopt in their lessons.

This gap indicates that students did not perceive Google Translate as useful for planning, highlighting the need for clearer, stage-specific guidance from teachers on how to use Google Translate effectively and how this usage approach will benefit students of all proficiency levels in terms of language improvement. Moreover, although some high-scoring students' practices (IDs 10, 11, 18, and 19) partly mirrored one teacher's recommended usage strategy during composing, the widespread divergence in tool usage across the three stages makes it difficult to conclude that students who shared similarities in Google Translate usage with teachers achieved higher writing scores.

5.4 Comparative analysis of Quillbot

5.4.1 Planning

In Teacher 8's lesson plan, Quillbot was integrated into the planning stage:

[97] In the planning stage, the teacher engages students with the topic "Written communication or Spoken communication: Which is more important?" by asking open-ended questions such as: What do you use to communicate? How often do you use language to communicate? Which form do you use more

often, writing or speaking? Which one do you consider more important, written communication or spoken communication? Encourage students to brainstorm ideas using various materials and Google Translate. Then, the teacher allows students to independently brainstorm using Google Translate, and afterwards, they can use Quillbot to paraphrase words like "powerful", "spoken communication" and "written communication". (Lesson plan, Teacher 8)

Accordingly, Teacher 8 paired Quillbot with Google Translate, suggesting a sequential usage order: first translating ideas from Vietnamese to English using Google Translate, then refining those expressions using Quillbot. This approach aimed to enrich students' lexical range and foster deeper engagement with the topic. She further explained this approach in the interview:

[98] Students can first use Google Translate to express their ideas in paragraph form during the composing stage. Once the basic ideas are on screen, they paste that text into Quillbot to experiment with synonyms and more academic phrasing. This back-and-forth helps them flesh out ideas they hadn't considered, expand their vocabulary right from the outset, and walk into the composing stage with richer, better-worded points to develop. (Interview, Teacher 8)

However, despite this structured instructional design, screen recordings indicated that no students used Quillbot during the planning stage. This highlights a clear divergence between teacher intentions and student practice, indicating that students may lack awareness of the tool's planning-stage benefits, may not perceive a need for paraphrasing at that point, or may face difficulties in using multiple tools early in the writing process alongside time constraints.

5.4.2 Composing

In their lesson plans and interview responses, Teachers 3, 5, 8, 9, and 10 instructed students to use Quillbot for the composing stage as follows:

[100] In the composing stage, students can use tools like Google Translate to overcome language barriers when expressing their ideas in English. For instance, they can translate Vietnamese phrases such as "Trong một số tình

huống thì giao tiếp bằng lời nói sẽ hiệu quả hơn là bằng văn bản" to "In some situations, verbal communication is more effective than written communication." However, it's important to note that the accuracy of translations depends on the clarity of the original ideas. Additionally, students can use Quillbot to paraphrase their thoughts, transforming simple statements into more scholarly language. For example, "In some situations, verbal communication is more effective than written communication" can be rephrased as "In certain circumstances, vocal communication outperforms written communication." These two tools can serve as valuable resources for students to enhance their English writing skills and academic expression. (Lesson Plan, Teacher 3)

[101] Students use Quillbot to paraphrase sentences with different structures, phrases, and words, to discover more expressions. For example, instead of saying, "It is said that spoken communication has more power than written one," they might say, "I agree with the statement saying.../I am in favour of the opinion...." In groups of four, students then use Quillbot to modify and customise their sentences, focusing on the two body paragraphs of their essays. The teacher observes and assists the students in effectively employing these tools when necessary (Lesson Plan, Teacher 5)

[102] In the composing process, students can use Google Translate to help express their opinions in paragraph form. To enhance the quality of their writing, students employ Quillbot to paraphrase the paragraph. This is to improve writing style and clarity. After that, students should use Grammarly to review and correct any grammatical errors, ensuring their work is polished and error-free. I hope after Grammarly, they can notice their errors and remember the corrections while composing. (Lesson Plan, Teacher 8)

[103] Use Google Translate as a dictionary when writing. Next, to avoid repetition and improve their writing style, students were guided to use Quillbot to rephrase these words in new ways, such as changing "write" to "compose" or "draft," and "speak" to "discuss" or "articulate" (Lesson Plan, Teacher 9)

[104] Students will continue to compose their essays until the time limit of 45 minutes. Afterwards, they should use Quillbot to paraphrase sections of their writing that have repeated or informal words. (Lesson plan, Teacher 10)

These five teachers (Teachers 3, 5, 8, 9, and 10) employed Quillbot to enhance the quality of students' writing by transforming simple or informal sentences into a more formal writing style, but their tool usage strategies were slightly different. Particularly, students are guided to diversify their word choices, helping them avoid repetitions and explore more ways to express their ideas (Teachers 9 and 10).

Teachers 3 and 8 both recommended using Quillbot in combination with Google Translate during the composing stage. Teacher 3 explained in the interview: "Students can use Google Translate to convert their ideas from Vietnamese to English first, and then apply Quillbot to rephrase these basic translations, turning them into more formal academic sentences." Teacher 8 also shared a similar approach: "I ask my students to first use Google Translate to express their ideas in English, and then use Quillbot to paraphrase those sentences, exploring different ways to phrase their points more naturally and formally." This integrated method was designed to help students expand their range of expressions and refine their English writing beyond direct translations.

Notably, Teacher 5 noted "In groups of four, students then use Quillbot to modify and customise their sentences, focusing on the two body paragraphs of their essays", indicating that the use of Quillbot in her lesson serves as a learning resource to help students "discover more expression" through collaborative work rather than merely acting as a shortcut for transforming text. Teacher 5 further explained this in the interview: "I want students to see paraphrasing as part of a learning process. By working together and trying out different sentence versions in Quillbot, they not only improve their writing but also learn from each other's choices. It's important to let them discuss and reflect with each other about which version sounds more formal or appropriate, and why, that's how they can learn."

In comparison, only a small group of students used Quillbot in the composing stage, as recommended by 50% of teachers, but they cannot work collaboratively due to the nature of the task. According to the screenshots, five students (IDs 5, 7, 9, 10, 13)

used Quillbot to find synonyms and paraphrase sentences, as illustrated in Figures 55, 56, 57, 58 and 59

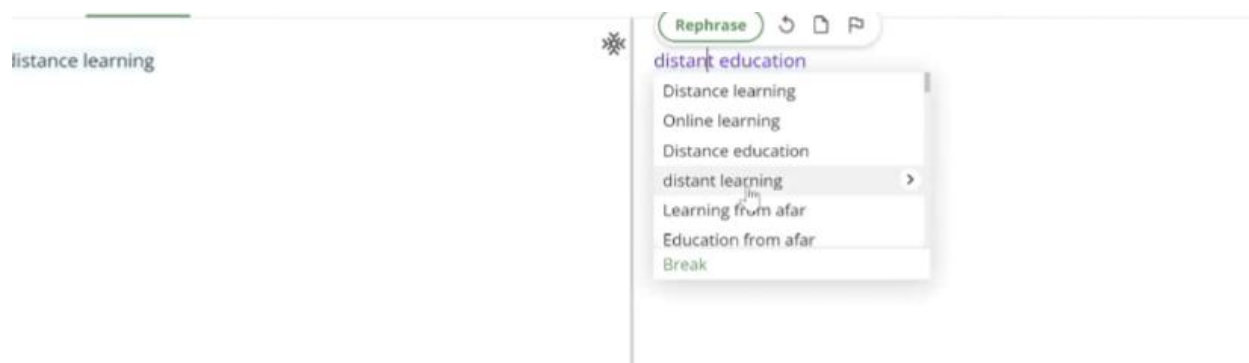


Figure 55. Student 5 used Quillbot to find the synonyms of “distance learning” and chose the word “distance education” and “learning from afar” in his essay.

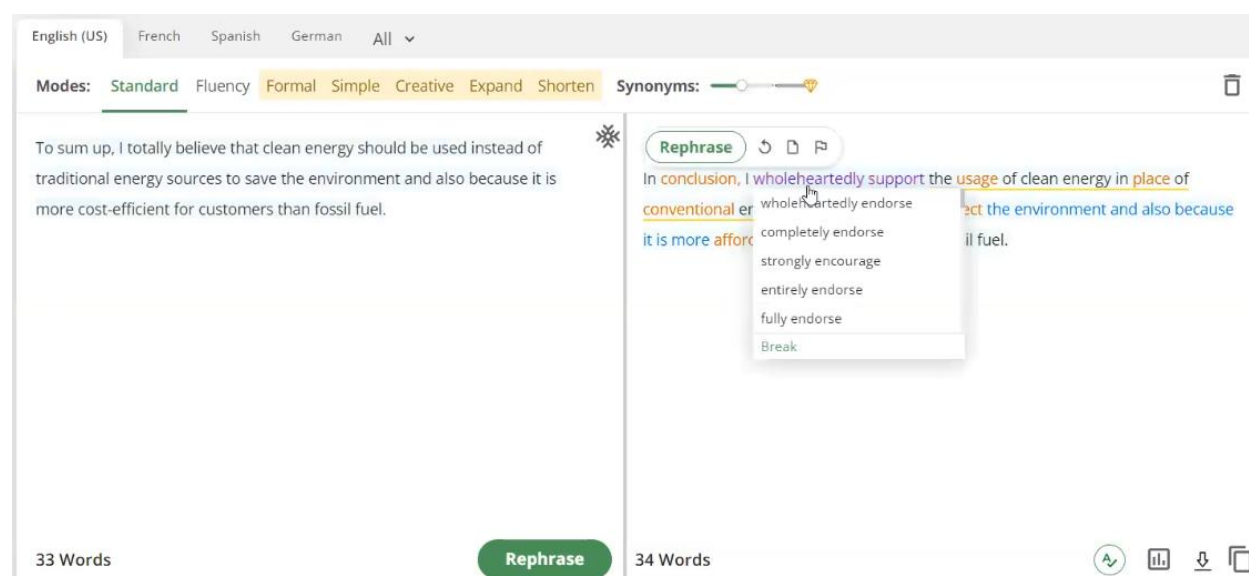


Figure 56. Student 7 paraphrased a sentence with Quillbot and selected the phrase “wholeheartedly endorse” from its output.

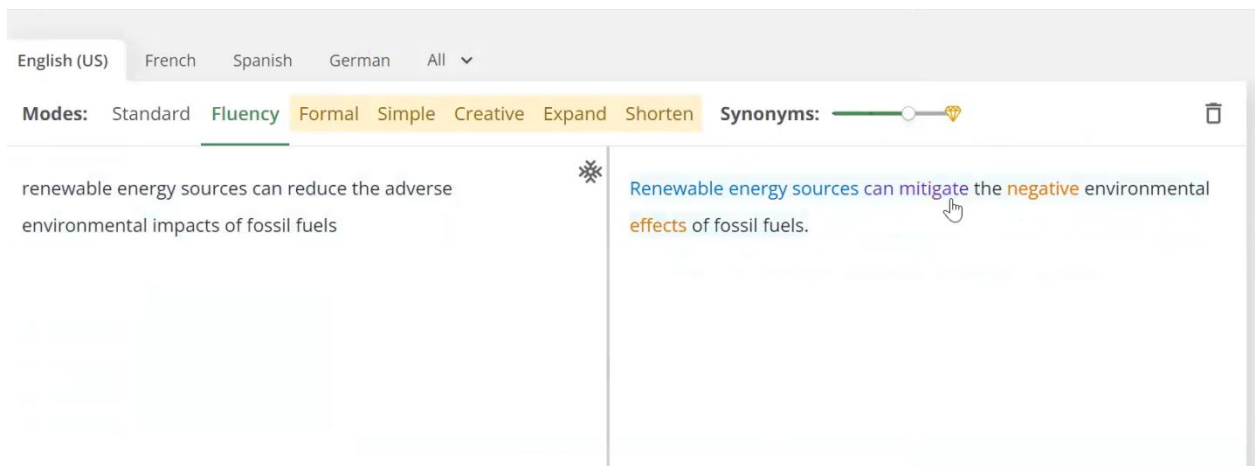


Figure 57. Student 9 used Quillbot to paraphrase the sentence that she wrote, paying attention to the word “mitigate”.

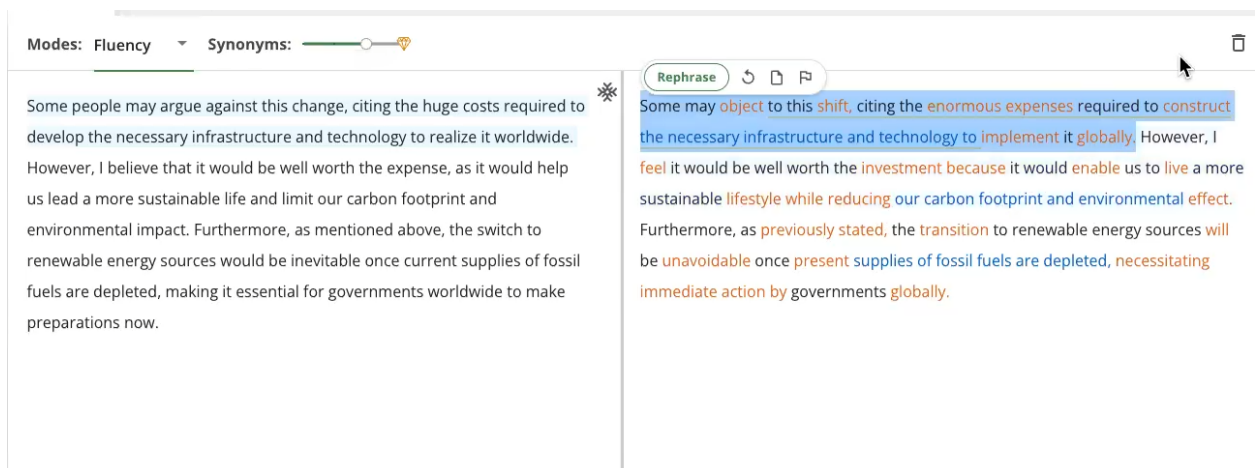


Figure 58. Student 10 used Quillbot to paraphrase the two sentences. He only chose the first sentence from the output.

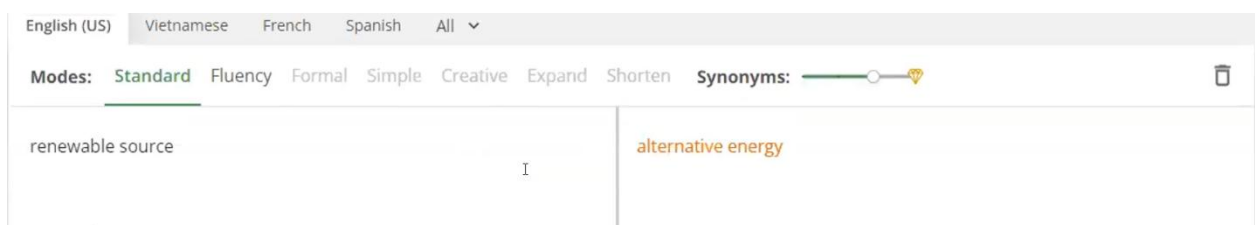


Figure 59. Student 13 used Quillbot to paraphrase the phrase “renewable source” and she selected “alternative energy”

During the interview, in response to the interview question "In the composing step, did you find the tools useful? Which tool helped you the most? How did it help you?", these five students mentioned the support of Quillbot in the composing step as follows:

[105] Yes, in the composing step, I found the tools useful. I used them to paraphrase words such as "distance learning" to "online learning" or "distance education." Additionally, I used SketchEngine's Word Sketch to find more verbs related to "attention," but I did not find a relevant word. (Student 5)

[106] In the composing step, I used Google Translate to find the English equivalent of the Vietnamese word "ủng hộ", and then I got the word "support". After that, I used Quillbot to search for the synonym and got the word "endorse.". (Student 7)

[107] I wrote: "Renewable energy sources can reduce the adverse environmental impact of fossil fuels." After Quillbot, I decided to change to: "Renewable energy sources can mitigate the negative environmental effects of fossil fuels.". I did not know the word "mitigate" so I had to use Sketch Engine to confirm later. (Student 9)

[108] Yes, I found the tools useful in the composing step. Quillbot was particularly helpful. [...]. I used Quillbot to paraphrase specific sections of my essay. For instance, Quillbot suggested changing "led to debates" to "sparked discussion," which I found to be a good suggestion. I also adopted Quillbot's version of the first sentence: "Some may object to this shift, citing the enormous expenses required to develop the necessary infrastructure and technology to implement it globally." I retained the following sentence as it was. Additionally, I changed "unavoidable" to "inevitable" and "making it essential" to "necessitating." However, instead of using Quillbot's suggestion of "immediate action," I kept my own phrase, "prompt action.". (Student 10)

[109] In the composing step, I found Quillbot useful for paraphrasing the "renewable sources" to "alternative energy" and I chose its suggestions. (Student 13)

These interview extracts show that the five students (IDs 5, 7, 9, 10, 13) used Quillbot

in ways that sometimes followed and sometimes diverged from their teachers' instructions. The lower-scoring student, ID 13, relied on Quillbot mainly for surface-level substitutions—accepting changes such as “renewable sources” to “alternative energy” without further assessment. This partly reflects the usage purpose suggested by Teachers 9 and 10 to avoid repetitions. However, Student 13 was not able to critically assess the output of this tool. By contrast, Student 10, demonstrated the reflective approach in tool usage suggested by Teacher 5 by accepting only suitable suggestions such as “sparked discussion,” “inevitable,” and “necessitating” but rejected “immediate action” in favor of “prompt action” to preserve intended nuance. Additionally, Student 7, a high-scoring participant, followed the Google Translate–Quillbot sequence recommended by Teachers 3 and 8, first translating the Vietnamese verb ủng hộ to “support” and then refining it to the more precise “endorse”. Students 9 and 5 shared the same usage of combining Quillbot and Sketch Engine. Student 9 retained Quillbot’s revision “mitigate the negative environmental effects” only after verifying the verb mitigate in Sketch Engine. Similarly, Student 5 combined Quillbot (e.g., “distance learning” to “online learning” or “distance education”) with Sketch Engine’s Word Sketch to search for verbs related to “attention”. These tool usage strategies were not mentioned in any teacher’s lesson plan, but still clearly demonstrates their cognitive engagement with the tool’s output, as these two high-scoring students did not blindly accept these outputs from Quillbot but deliberately verified them with Sketch Engine.

5.4.3 Revising

Five teachers (IDs 1, 2, 4, 6, and 7) wrote in their lesson plan that they would incorporate Quillbot in the revising stage:

[110] This is a session for individual revision. The teacher suggests that students use Quillbot to get suggestions for their writing. Students will then selectively choose the sentences they want to paraphrase. This will help improve the quality and variety of their writing. (Lesson plan, Teacher 1)

[111] Students will use Grammarly to check and note down their errors. For peer feedback, they will exchange their writing with a classmate to receive constructive comments. They can also use Quillbot to paraphrase sentences or

phrases that are repeated or too simple. The teacher will ask one or two students to volunteer and share the feedback they received with the class. Finally, the teacher will provide delayed feedback based on the overall performance of the students. (Lesson plan, Teacher 2)

[112] Use Quillbot to paraphrase, use Sketch Engine to check collocations and then peer-review. (Lesson plan, Teacher 4)

[113] Ask students to input their writing into Grammarly to check for grammar mistakes and make notes of these errors so the whole class can become aware of common mistakes. Then, have them apply the topic question and their topic sentences from their body paragraphs into Quillbot to see how they are phrased and evaluate which sentences are better, theirs or Quillbot's. Teachers should also ask students to paraphrase the topic sentence of the introduction, body paragraph 1, and body paragraph 2. (Lesson plan, Teacher 6)

[114] The teacher will pair up students to comment on two different essays. They should use Grammarly for assistance, preferably with Google Docs. If a word is repetitive, they can use Quillbot to suggest paraphrasing, allowing students to decide on appropriate synonyms. After receiving feedback, students will edit their essays based on the comments and then submit them to the teacher. (Lesson plan, Teacher 7)

Lesson plans and interviews show that five teachers (1, 2, 4, 6, and 7) used Quillbot not simply to replace students' own sentences but to prompt critical comparison between original and paraphrased versions. Teacher 1 asked students to select sentences to paraphrase with Quillbot to "improve the quality and variety of their writing." Teacher 2 recommended using Quillbot for repeated or overly simple sentences as part of a wider revision process with Grammarly and peer feedback. Teacher 4 combined Quillbot for paraphrasing with Sketch Engine for collocations and peer review. Teacher 6 explicitly asked students to compare Quillbot's output with their own topic sentences and decide which was better. Teacher 7 embedded Quillbot in peer feedback, using it to suggest alternatives for repetitive words. Collectively, these teachers emphasised using Quillbot at the revision stage to compare versions and critically evaluate automated paraphrases for language improvement.

Analysis of screenshots from eight high-scoring students (IDs 6, 7, 9, 11, 15, 18, 19 and 20) and one low-scoring student (Student 17) indicated that they used Quillbot to revise their work (Figures 60, 61, 62, 63, 64, 65, 66, 67 and 68).

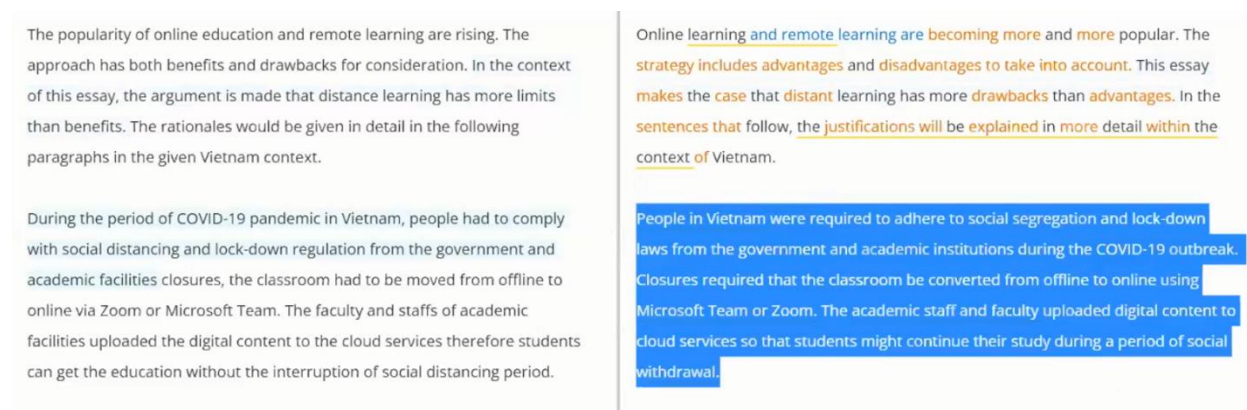


Figure 60. Student 6 paraphrased the first and second paragraphs, but he only chose the second paragraph output.



Figure 61. Student 7 used Quillbot to paraphrase a sentence and she chose the output.

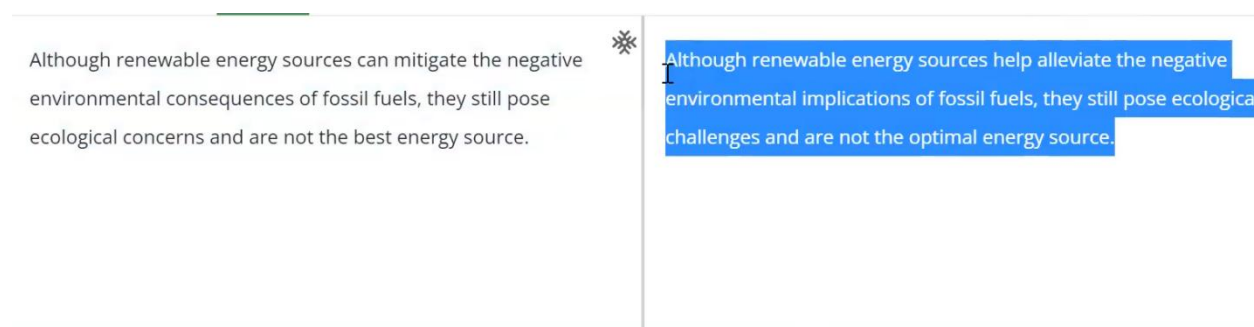


Figure 62. Student 9 used Quillbot to paraphrase her introduction sentence and used it as the conclusion.

Although fossil fuels have played an important role in people's lives since the dawn of industrial civilization, many people are of the opinion that they should be substituted by renewable energies. Personally, I quite agree with the given opinion because renewable energies are more environmentally friendly and these are endless natural sources that people can take advantage of in the future when fossil fuels are running out of.

Although fossil fuels have been a significant part of human life since the beginning of modern civilisation, many individuals believe that renewable energies should take their place. I personally agree with the viewpoint since renewable energies are more environmentally friendly and because they may be used indefinitely by humans in the future when fossil fuels run out.

Figure 63. Student 11 used Quillbot to paraphrase the introduction and chose the output.

First, the use of fossil fuels is closely associated with the rate of climate change, an alarming issue that humanity is facing. For that reason, if we keep using fossil fuels to fuel our world, there would be dire consequences. Moreover, as green energy like wind and solar come from nature, we could never use up this energy as compared to fossil fuel ones. This not only could help people to protect the environment by using environmentally friendly energy but it is also more sustainable for us long-term due to its never-run-out-of-energy characteristic.

Rephrase ↺ ↻ ↗

First, the usage of fossil fuels is intimately related to the rate of climate change, which is a grave continuing to use fossil fuels to implications. Furthermore, because wind and solar come from nature as we do with fossil fuels. This not only directly helps people conserve the environment by using ecologically friendly energy, but it is also more sustainable for us in the long run because it never runs out of energy.

deeply result,
closely
personally
close
completely
directly

Figure 64. Student 15 used Quillbot to paraphrase the first body paragraph and made changes.

People on earth develop once they are healthy.

On earth, once a person is healthy, they develop. I

Figure 65. Student 17 used Quillbot to paraphrase a sentence

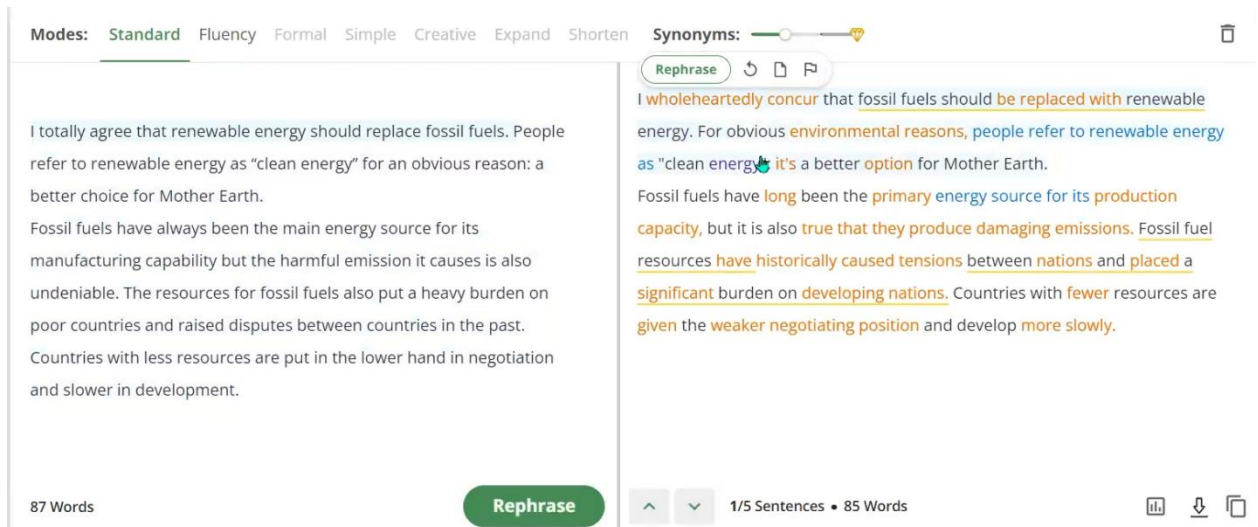


Figure 66. Student 18 used Quillbot to paraphrase two paragraphs. She selectively modified the output and chose only two sentences

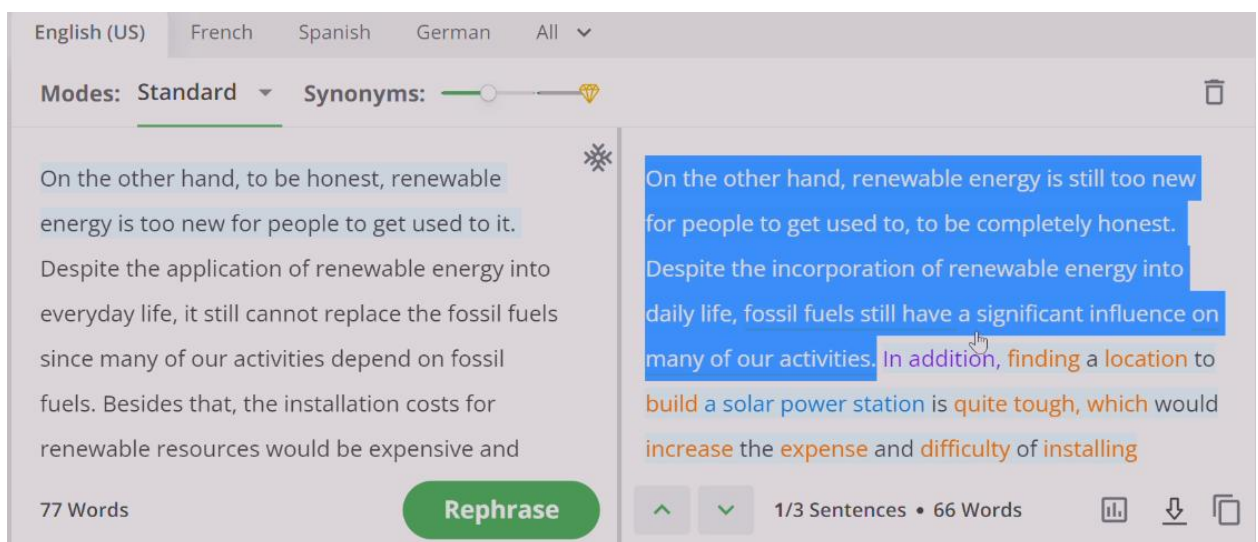


Figure 67. Student 19 used Quillbot to paraphrase two sentences and used the output.

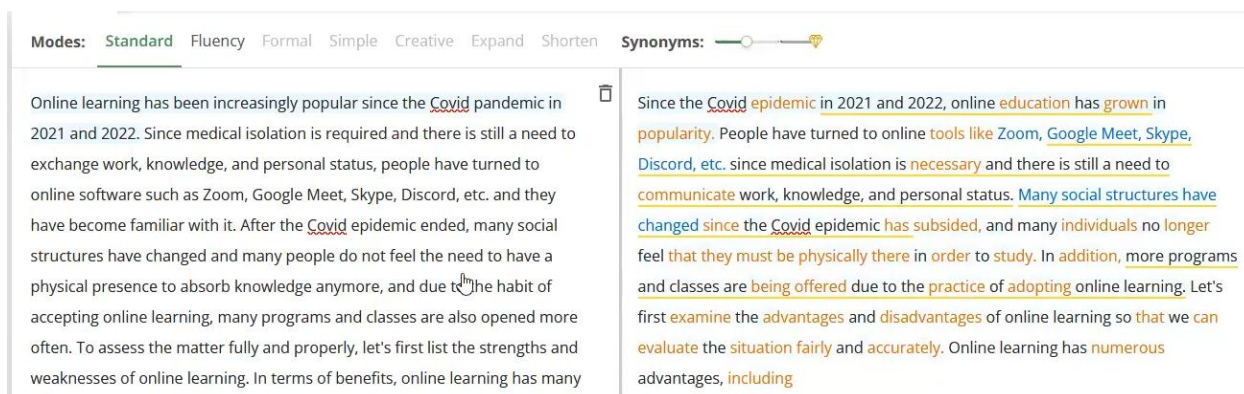


Figure 68. Student 20 used Quillbot to paraphrase the first paragraph, which was the output of Google Translate from his Vietnamese text. He chose the output but did not use the last sentence.

According to interview transcripts of Question 5: “In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?”, these nine students (IDs 6, 7, 9, 11, 15, 17, 18, 19, 20) mentioned the usage of Quillbot during revision as follows:

[115] Quillbot, I used it to paraphrase all of my paragraphs. I think it did a good job in rewriting. I replaced my original sentence “In the context of this essay, the argument is made that distance learning has more limites than benefit.” with “This essay makes the case that distance learning has more drawbacks than advantages.” because it sounds more concise, as it helped me learn that one way to rewrite is: “This essay makes the case that...”. For other parts, I did not change anything as I thought my original writing was better. (Student 6)

[116] I used Quillbot to paraphrase my original writing. I changed this sentence: “Another reason I prefer renewable energy is that it is significantly less expensive than fossil fuels.”. Quillbot also helped me paraphrased lots of my sentences, for example, from “I totally agree with this standpoint for the following reasons.” to “ I completely concur with this viewpoint for the reasons listed below.”, and I learned that “concur” is a synonym of “agree”. Quillbot also helped me rewrite the order and structure of this sentence: “First, because it does not emit as much hazardous exhaust as fossil fuels do, recyclable energy can help to prevent the environment from becoming increasingly polluted.”. I also learn more words such as: “entirely”, “conventional energy sources”.

(Student 7)

[117] *In the revising and editing step, Quillbot was the most helpful tool for checking my work. I was taught that the conclusion sentence of the essay should be a paraphrased version of the introduction. I usually have a difficult time rewriting my own sentences, but Quillbot outputs helped me to do it faster and explore more synonyms. After a few trials, I found that some phrases are even better than my original words, such as the phrases “alleviate the negative environmental implications” and “pose ecological challenges”, which seem to be more formal than my original phrase “reduce the adverse environmental impacts” and “have potential ecological risks”. (Student 9)*

[118] *During the revising and editing stage, I used Quillbot and Google Translate. I used Quillbot to paraphrase the introduction, as it provided useful synonyms to enhance my writing. For example, it suggested changing “people’s lives” to “human’s life”, “many people” to “many individuals”, and “people can take advantage of it” to “it may be used indefinitely by humans”, and I agree with these suggestions, I think they are better than my own writing, they sounds more formal and suit with the context of the environment (Student 11)*

[119] *I also used Quillbot to improve my body paragraphs after completing my essay. It allowed me to modify words I thought were not suitable and see more options. For example, I learned that “successfully implement the usage of these energies into real life” can be paraphrased as “properly execute the use of these energies in real life” and I learned the new phrase “properly execute” I also discovered that “alter” is a synonym for “change” and that “much cheaper” can be rewritten as “significantly cheaper”. (Student 15)*

[120] *I tried using Quillbot one time. It helped me paraphrase the sentence from “People on earth develop once they are healthy” to “On earth, once a person is healthy, they develop”, I think it sounded better than my previous writing, but I didn’t learn any synonyms. (Student 17)*

[121] *I used Quillbot to paraphrase the last paragraph but only made changes to the following sentences: “On the other hand, renewable energy is still too new for people to get used to, to be completely honest. Despite the*

incorporation of renewable energy into daily life, fossil fuels still have a significant influence on many of our activities.”, and I preferred this version because it offered better word choices and phrasing than my original sentences. In particular, “the incorporation of renewable energy into daily life” sounds more natural than my initial phrase, “the application of renewable energy into everyday life.” Additionally, the structure “fossil fuels still have a significant influence on many of our activities” is an improvement over “many of our activities depend on fossil fuels.”. (Student 18)

[122] For Quillbot, I used it to paraphrase the last paragraph, but I only changed these sentences: “On the other hand, renewable energy is still too new for people to get used to, to be completely honest. Despite the incorporation of renewable energy into daily life, fossil fuels still have a significant influence on many of our activities.” because I think this output contains better word choice and phrases than my original sentences, particularly the phrase “the incorporation of renewable energy into daily life” is better than my previous sentence: “the application of renewable energy into everyday life”, and this also contains a better structure: “fossil fuels still have a significant influence on many of our activities” compared to “many of our activities depend on fossil fuels”. (Student 19)

[123] Grammarly was the most helpful tool in the revising and editing step, especially for checking punctuation and articles. It was very helpful for self-learning. I think the texts generated by Quillbot sound more formal to me in general, compared to those of Google Translate. [...] I think the texts generated by Quillbot sound more formal to me in general, compared to those of Google Translate. The last sentence I think is a bit redundant, so I did not include it. (Student 20)

Among the nine students who used Quillbot at the revising stage, eight high-scoring students (6, 7, 9, 11, 15, 18, 19, 20) showed clear cognitive engagement by accepting some suggestions and rejecting others. Student 6, for example, replaced “distance learning has more limites than benefit” with “has more drawbacks than advantages” for conciseness but kept sentences she felt were better in her own version. Students 7, 9, and 11 reported learning and choosing more formal, precise items such as

“concur,” “alleviate the negative environmental implications,” “pose ecological challenges,” “many individuals,” and “human’s life” when they fit the context. Students 15, 18, and 19 highlighted improvements like “properly execute the use of these energies in real life” and “the incorporation of renewable energy into daily life,” explicitly noting gains in clarity, structure, and new phrases. Student 20 also contrasted Quillbot’s more formal tone with Google Translate and deliberately removed a redundant sentence.

Overall, these eight high-scoring students used Quillbot to refine meaning, tone, and formality rather than to make automatic changes, closely aligning with teachers’ recommended revision strategies (Teachers 1, 2, 4, 6, 7). In contrast, the low-scoring Student 17 used Quillbot only once, offered the vague comment “I think it sounded better,” and admitted, “I didn’t learn any synonyms,” indicating minimal evaluation or language learning and a clear divergence from the teacher-endorsed approach.

In summary, teacher and student use of Quillbot diverged in planning and revision, but overlapped somewhat in composing. In planning, only one of 10 teachers (Teacher 8) used Quillbot to support idea generation and vocabulary, suggesting this function was rarely highlighted in class and may explain why no students used it at this stage. During composing, five teachers (3, 5, 8, 9, 10) recommended Quillbot to upgrade simple or informal sentences into a more formal style, often in combination with tools such as Google Translate. Teacher 5 also framed Quillbot as a collaborative learning resource rather than a shortcut. Of the five students who used Quillbot while composing, four high-scoring students (5, 7, 9, 10) followed broadly similar strategies: Student 7 combined Google Translate and Quillbot like Teachers 3 and 8, while Students 9 and 10 checked Quillbot outputs with Sketch Engine. Notably, Students 9 and 5—both high scorers—went beyond teacher guidance by systematically pairing Quillbot with Sketch Engine’s Word Sketch function. Their score gains and cognitive engagement suggest this student-initiated combination is pedagogically promising and worth integrating into future instruction to support language development during composing and revision.

By contrast, the low-scoring student (ID 13) used Quillbot merely as a quick fix for word or sentence repetition. At the revision stage, only high-scoring students (IDs 6, 7, 9, 11, 15, 18, 19, and 20) evaluated Quillbot’s outputs critically, which closely mirrors

the teaching approach endorsed by five teachers (IDs 1, 2, 4, 6, and 7). In contrast, the low-scoring student (ID 17) showed minimal engagement, applying Quillbot superficially and failing to explain or learn from the revisions. Taken together, these findings reveal that students who employed this paraphrasing tool as teachers advised, especially during the composing and revising stages, scored higher on their Essay 2.

5.5. Comparative analysis of Grammarly

5.5.1 Planning

In the planning stage, Grammarly is incorporated by Teacher 9 and Teacher 10 in their lesson plan as follows:

[124] Students begin by brainstorming ideas in Vietnamese and then use Google Translate to answer the topic question. The teacher provides hints to guide their thinking, such as asking, "In what situation is writing more powerful than speaking?" (e.g., formal emails in business, formal announcements) and "In what situation is speaking more powerful than writing?" (e.g., speeches by a leader). After translating their ideas into English, students use Grammarly to check for errors in the translated texts. (Lesson plan, Teacher 9)

[125] Students read the topic and choose a side between spoken communication and written communication. Then the teacher will ask students to brainstorm in Vietnamese and then use Google Translate to translate their ideas into short sentences. For example, their sentence may look like this: "Writing can be more powerful than speaking when you need to explain complex ideas clearly. For example, writing lets you choose your words carefully and make changes, which is helpful for legal papers and contracts. Students then use Grammarly to check for spelling and grammar errors in the draft. [...] If students are unsure which output to choose from Quillbot, the teacher will guide them in checking Quillbot's suggested words using Sketch Engine and selecting words based on the frequency of collocations. Finally, the teacher will ask students to use Grammarly for a final check on spelling and grammar. Once Grammarly has reduced all spelling and grammar errors, the teacher will mark the essays, focusing on errors in Task Response (TR) and

Coherence and Cohesion (CC). This might take extra time to correct. (Lesson plan, Teacher 10)

According to examples [124] and [125], Teachers 9 and 10 use Grammarly at all stages. This indicates that these teachers consider Grammarly is useful at all stages, yet not the main one in the revising stage, because they know that its error correction function is limited and can only correct surface-level errors. Following initial brainstorming activities, both Teacher 9 and Teacher 10 incorporated Grammarly into the planning stage of writing to help students refine English sentences translated from Vietnamese. Teacher 9 further explained this in the interview as follows:

[126] I always begin with brainstorming in the students' first language because I believe well-written ideas need to be finished first before composing. Once they have their points in Vietnamese, I guide them to translate those into English, usually through Google Translate, and then immediately use Grammarly to check sentence structure and grammar. For instance, if we're discussing written versus spoken communication, I might ask, 'In what situation is writing more powerful than speaking?' This helps them think critically, and Grammarly ensures their ideas are expressed correctly from the beginning. (Teacher 9)

With a similar purpose, Teacher 10 noted that her students are instructed to generate content in Vietnamese before translating and refining it in English:

[127] I would integrate Grammarly during brainstorming. My lesson plan includes a step where students brainstorm in Vietnamese and translate their position into English before they start composing. So if a student writes, 'Writing can be more powerful than speaking when you need to explain complex ideas clearly,' I tell them to run that through Grammarly right away. It helps correct issues that could cloud meaning, especially when they're translating by themselves. I then encourage them to think of Grammarly not just as a finishing tool, but as part of their drafting toolkit, which is something they use right from the planning stage to refine both clarity and accuracy. (Teacher 10)

Teachers 9 and 10 adopted a similar pedagogical approach by incorporating Grammarly into the planning stage, immediately after students translated their

brainstormed ideas from Vietnamese into English using Google Translate. Rather than treating Grammarly as a stand-alone tool, both teachers positioned it as part of the larger instructional task for the essay. This early integration reflects their broader view of Grammarly as a formative support tool, yet only for surface-level corrections, such as spellings or basic grammar, which aims to help students articulate complex ideas more clearly in English. Nonetheless, none of the student participants reported using Grammarly at this stage, indicating a mismatch between teacher intentions and students' perceptions of the tool's usefulness during planning.

5.5.2 Composing

During the composing stage, Grammarly is incorporated by one teacher, as presented in Teacher 8's lesson plan:

[128] In the composing process, students can use Google Translate to help express their opinions in separate paragraphs. To enhance the quality of their writing, students can employ Quillbot to paraphrase their paragraphs. This is to improve writing style and clarity. After that, students should use Grammarly to review and correct any grammatical errors. (Lesson Plan, Teacher 8)

According to the lesson plan, the teacher outlined a sequence of multiple tools, in which students first use Google Translate to express their ideas in English, then apply Quillbot to paraphrase and enhance clarity, and finally use Grammarly to correct grammatical errors. Her interview response to Interview Question 14, "If you must use Grammarly to teach an IELTS Task 2 essay, what step of instruction or teaching methodology would be the most suitable for this tool?" was as follows:

[129] I would use Grammarly for brainstorming and composing. I think this tool would help with drafting sentences. However, to check the overall quality, I think my lesson should involve peer corrections and teachers' corrections. I hope that after Grammarly, they can notice their errors and remember the corrections after composing. (Teacher 8)

Accordingly, the response indicates that Teacher 8 viewed Grammarly as most effective during the early stages of writing, specifically brainstorming and composing, and the revision step should not include this tool but to use peer feedback and teacher

feedback. In other words, Teacher 8 positioned Grammarly's output as a resource for students to compare and "notice their errors and remember the corrections after composing", not as a shortcut to refine texts. By contrast, no student in this study used Grammarly during the composing stage, suggesting that they did not view it as a resource for improving language during composition.

5.5.3 Revising

Across the lesson plans and interview responses of nine teachers (Teachers 1, 2, 3, 4, 5, 6, 7, 9, and 10), Grammarly was consistently integrated into the revision stage as a tool to correct students' grammatical accuracy. This finding implies that there is a strong consensus among teachers regarding Grammarly's role as a revision tool. However, differences regarding their usage purpose emerged in their interview responses. In particular, Teachers 1, 3, and 4 shared a common teaching approach that positioned Grammarly not merely as a tool for quick correction, but as a learning aid intended to help students recognise their errors, internalise grammatical rules and improve their writing language over time. Teacher 1 explained in the interview:

[130] After students finished composing their essay, I would tell my students, 'Open Grammarly and run a final check before you hit submit.' I want that moment to feel like a self-audit, not a teacher-imposed correction. I've noticed that when students do this final pass themselves, they come to class already aware of recurring mistakes and are prepared to ask deeper questions about how to avoid them next time. (Teacher 1).

This indicates that Teacher 1 uses Grammarly as a post-composing, self-editing tool aimed at promoting independent learning and reflective revision. Teacher 3 shared this similar purpose clearly in her interview:

[131] I highly recommend the function of highlighting each mistake and suggestions in Grammarly. This feature encourages students not just to correct errors quickly, but to actually learn from them. I always remind my students to read their work carefully, double-check the mistakes with Grammarly suggestions, instead of accepting it blindly, because understanding why an error occurred is what helps them improve in the long term. (Teacher 3).

This reflects her intention to guide students toward active engagement with the tool's feedback by encouraging them to read, reflect on, and understand the rationale behind each correction.

Similarly, Teacher 4 incorporated Grammarly into the revision process with a focus on error detection and grammar awareness. In her lesson plan, she instructed students to use Grammarly during revision to identify common grammatical issues, particularly those related to "singular, plural nouns; S–V agreement" (Lesson Plan, Teacher 4). In the interview, she reinforced this as follows:

[132] During the revision stage, I always tell my students to use Grammarly, not just to fix the mistakes, but to really look at what kinds of errors they're making. I've noticed a lot of them have trouble with things like subject–verb agreement and singular versus plural nouns, so I specifically ask them to pay attention to those. I usually tell my students in class: "Don't just click accept and try to understand why these sentences or words are wrong". I also encourage them to keep a list of common errors they make, so over time, they can become more aware and hopefully avoid repeating them.

Together, the data reveal that these three teachers (Teachers 1, 3, and 4) used Grammarly exclusively during the revision stage not as a quick fix, but as a learning aid to support error noticing, promote reflective thinking, and foster language development.

Aside from using Grammarly for individual corrections, other teachers explicitly focused on using Grammarly to complement peer feedback (Teachers 2, 6, and 7) and teacher feedback (Teachers 5, 7 and 9). Teacher 2, for instance, instructed students to "begin by using Grammarly to check their work and take note of any errors", followed by "peer review and class-wide feedback sharing" (Lesson Plan, Teacher 2). Similarly, Teacher 6 asked students to "input their writing into Grammarly to check for grammar mistakes and make notes of these errors so the whole class can become aware of common mistakes" (Lesson Plan, Teacher 6). Teacher 7 integrated Grammarly into collaborative editing sessions using Google Docs: "They should use Grammarly for assistance, preferably with Google Docs. After receiving feedback, students will edit their essays based on the comments" (Lesson Plan, Teacher 7). When answering

Interview Question 14: “If you must use Grammarly to teach an IELTS Task 2 essay, what step of instructions or teaching methodology would be the most suitable for this tool?”, Teacher 7 stated:

[133] Grammarly is really helpful for catching surface-level issues like spelling, punctuation, and basic grammar. I’ve found that when students use it before submitting their work, it saves me a lot of time on minor corrections. But it doesn’t always handle more complex aspects like cohesion or word choice very well. That’s why I always tell my students to go over those areas with a classmate during peer review and to pay special attention to my feedback afterwards. Grammarly can assist, but it’s not a substitute for real human input when it comes to improving flow and clarity. (Teacher 7)

Accordingly, Teacher 7 clearly emphasised how Grammarly has limitations in handling cohesion and word choice, and thus these issues should be carefully reviewed with classmates and through teacher feedback

Teachers 5 and 9 considered Grammarly as preparation for, not a substitute for, teacher feedback. Teacher 5 stressed that Grammarly should *complement* teacher comments: in her lesson plan, “The teacher then gives comments on some essays to highlight common mistakes,” and in interview she noted it helps her “correct my students’ essays faster” and “suggest other structures or expressions.” Teacher 9 similarly used Grammarly as a pre-check for surface errors: “After spelling errors and grammar errors are removed by Grammarly, the teacher will mark the essays, looking for errors in TR and CC.” She added that Grammarly helps students “learn from their mistakes by reviewing the suggestions at each stage,” while recognising its limits for coherence and linking phrases.

Notably, Teacher 10 used Grammarly both in the planning stage and revising stage, and adopted a more distinctive approach:

[134] If students are unsure which output to choose from Quillbot, the teacher will guide them in checking Quillbot’s suggested words using Sketch Engine and selecting words based on the frequency of collocations. Finally, the teacher will ask students to use Grammarly for a final check on spelling and grammar. Once Grammarly has reduced all spelling and grammar errors, the teacher will

mark the essays, focusing on errors in Task Response (TR) and Coherence and Cohesion (CC). This might take extra time to correct. (Lesson plan, Teacher 10)

In the interview, the teacher added:

[135] At the revising stage, I want them to confirm the suggestion from Grammarly, such as students should write “focus on” and not “focus to” or “focus at”, and check if the suggestion of Grammarly aligns with the most frequent collocations found by Sketch Engine. (Interview, Teacher 10).

According to her lesson plan and interview responses, Teacher 10 explicitly instructed students to verify Grammarly’s suggestions using Sketch Engine, demonstrating a multi-tool approach and a focus on cognitive engagement with the suggestions of Grammarly during the revision stage.

Screenshots and interview responses of students indicate that 19 out of 20 used Grammarly for revision. Student 9 was the only participant who did not use Grammarly and also did not mention its role in her interview response. The students’ interview data reveal similar patterns in how Grammarly was used during the revising and editing stages: considering Grammar as a shortcut for revision. For example, Student 1 found Grammarly “particularly useful for checking overall spelling and grammar” and noted corrections such as changing “with affordable prices” to “at affordable prices.” Student 2 reported that Grammarly helped with article usage, plural forms, and verb agreement, such as correcting “applications that are available” and “setted up” to “set up.” Similarly, Student 4 observed that Grammarly corrected errors like “the mesmerizing alternatives of other fuels” to “mesmerizing alternative to other fuels.” Other students, including Students 7, 13, 14, and 15, shared similar uses of Grammarly for identifying issues with article usage, verb tense, noun plurality, and prepositional accuracy. These practices from students largely mirrored the intentions outlined in the three teachers’ lesson plans (Teachers 1, 3, and 4), where Grammarly was positioned as a final-stage revising tool aimed at ensuring grammatical accuracy before submission.

However, while these 19 students generally followed the teachers’ intended use of Grammarly, a key difference emerged in how they engaged with its feedback and

demonstrated language improvements. The high-scoring students—specifically Students 1, 3, 5, and 10—demonstrated more cognitive engagement by explaining why certain changes were helpful or explicitly preferring Grammarly’s suggestions over their original phrasing. For instance, Student 1 mentioned that Grammarly “helped correct article and preposition usage”, noting that “with affordable prices” should be “at affordable prices” and “choose the suitable time” was revised to “choose a suitable time.” Student 3 acknowledged Grammarly’s role in identifying missing articles, such as adding “a” before “electric car” and “petrol-powered car”. Student 5 highlighted a specific grammar refinement, stating that Grammarly changed “the thirst of knowledge” to “the thirst for knowledge,” which they accepted as more accurate. Notably, Student 10 demonstrated a more advanced level of cognitive engagement by selectively applying Grammarly’s suggestions rather than accepting them uncritically. For example, when Grammarly flagged the sentence “In conclusion, I completely agree with the suggestion that fossil fuels be replaced by their renewable counterparts” as incorrect, Student 10 chose to keep the original, stating that Grammarly “did not offer any better suggestion.” This rejection shows the student’s confidence in their language judgment and an ability to critically assess the tool’s feedback, aligning with the teachers’ goal of promoting metalinguistic awareness. However, no student followed the same teaching approach of Teacher 10 in cross-verifying Grammarly’s output with Sketch Engine to confirm collocational accuracy—for example, evaluating whether “focus on” is more appropriate than “focus to”. This highlights a gap between the learning-oriented tool usage purpose promoted by teachers and the primarily corrective, single-tool reliance observed in student practices. Even for the four high-scoring students (IDs 1, 3, 5 and 10), while being more reflective in their acceptance or rejection of Grammarly’s suggestions, they did not verify Grammarly’s feedback through corpus consultation. This gap suggests that these students may not yet possess the digital literacy or metalinguistic awareness needed to fully exploit the complementary functions of a corpus tool, or perhaps the limited time in this context was a factor as well.

By contrast, the low-scoring students (IDs 4, 11, 12, 13, 16 and 17) relied on Grammarly primarily for surface-level corrections and consistently accepted its suggestions with minimal critical reflection. For instance, Student 11 dismissed Grammarly’s usefulness, saying, “Grammarly also required me to buy a subscription

and it did not give me any useful corrections,” yet offered no evaluation of its free features or whether any suggestions were accepted or rejected. Student 12 also described a passive experience, noting, “I only see some highlighted phrases,” and admitted to not making changes beyond the visual cues. These responses suggest that low-scoring students tended to treat Grammarly as an instant, automatic tool to use after composing rather than a tool for learning or language development. Unlike higher-scoring students who reflected on why changes were needed, these low-scoring learners lacked cognitive engagement and made little effort to assess the relevance or quality of the tool’s feedback. Student 16 stated that: “With Google Translate helping me with my whole essay, I don’t even have to think in English,” and used Grammarly only for a final grammar check, without commenting on specific changes or their appropriateness.

Notably, these three low-scoring students (Students 4, 13 and 17) showed some level of engagement with Grammarly. Student 4 stated, “During the revising and editing step, I used Grammarly to check errors. Grammarly was helpful, for example, it changed ‘the mesmerizing alternatives of other fuels’ to ‘mesmerizing alternative to other fuels’ and added ‘the’ before ‘equipment’.” However, he did not question or critically assess these suggestions, indicating a reliance on Grammarly’s outputs without further reflection. Student 17 stated that “Grammarly were the most helpful tools during the revising and editing step. Grammarly helped with corrections like “The Earth” and pointing out errors that I need to write in plural form, such as from “human” to “humans”, correct article usage from “the lack of water” to “a lack of water”, and I learned that I made a spelling mistake with “breathe”.” Student 13 stated in her interview:

[136] *In the revising and editing step, Grammarly was the most helpful tool for checking grammar and spelling. It helped me notice mistakes, especially with plural nouns. For example, I wrote “I partly agree with this ideas” but Grammarly corrected it to “I partly agree with this idea” It also fixed “too much traditional sources” to “too many traditional sources” when talking about fossil fuels. I also made some mistakes with verbs. For example, I wrote “it completely disappear” but Grammarly changed it to “it completely disappears” Another mistake was with prepositions. I wrote “put a financial burden for some nations” but it*

corrected it to "put a financial burden on some nations" These corrections helped me improve my writing a lot. (Student 13)

However, compared high-scoring students' usage, Students 4, 13 and 17 did not question or reject any outputs from Grammarly, or use another tool such as Sketch Engine to verify it, as recommended by Teacher 10.

Other differences were also observed between students' actual use of Grammarly and the ways teachers intended it to be integrated into writing instruction. Specifically, no students reported using Grammarly as a complement to peer feedback (as mentioned by Teachers 2, 6, and 7) or teacher feedback (as mentioned by Teachers 7, 5 and 9). However, these differences are likely due to the design constraints of the current study. Because the research setting did not allow students to participate in peer review or receive teacher feedback during the writing and revision stages, they were unable to engage with Grammarly in the reflective approach recommended by these teachers.

In summary, the data reveal both similarities and mismatches between teachers' pedagogical intentions and students' actual practices with Grammarly. While this automated writing evaluation tool was widely adopted by teachers, its intended function varied across the different stages of writing—planning, composing, and revising. Teachers 9 and 10 integrated this tool during the planning stage to help students refine ideas translated from Vietnamese and ensure grammatical accuracy at the planning stage, highlighting that Grammarly should only be used for surface-level corrections. One teacher employed Grammarly during the composing stage with a similar purpose, encouraging students to use it as a tool for identifying and internalising their grammar mistakes while drafting. In the revision stage, Grammarly's use showed the strongest consensus, with nine teachers incorporating it not just for surface-level error correction but also to promote grammar awareness, metalinguistic reflection, and long-term language development. Collectively, these findings suggest that despite a minority of teachers used Grammarly for planning and composing for quick surface-level corrections, the strongest consensus among all 10 teachers was its role to promote students' grammar awareness, encourage reflections on their gaps, and facilitate long-term language learning.

In contrast to teachers' usage of Grammarly across all three writing stages, students' usage patterns showed a narrower view of Grammarly. Although 19 out of 20 students reported using the tool, they did so only in the final stages of writing to address surface-level issues such as grammar, punctuation, and spelling. This difference was most pronounced in the planning and composing stages, as no students reported using Grammarly as an early-stage writing aid, as intended by Teachers 9, 10, and 8. During the revision phase, teacher and student usage show the most similarities. However, there were differences in the level of engagement among students. A minority of high-scoring students (Students 1, 3, 5, and 10) demonstrated critical reflection by selectively applying Grammarly's suggestions and showing awareness of the grammar rules behind the corrections. In contrast, low-scoring students (IDs 4, 11, 12, 13, 16 and 17) used Grammarly more passively, accepting suggestions automatically without showing evidence of understanding or evaluating the feedback. However, none of the 20 student participants reported engaging in the use of combining tools to evaluate Grammarly's output and improve revision skills, as proposed by Teacher 10, such as verifying Grammarly's suggestions with Sketch Engine. This absence of usage reveals a gap in students' general understanding of how to use the corpus tool's potential for supporting revision. Overall, these findings suggest that students who used Grammarly in ways similar to those recommended by teachers during the revising stage—specifically, by critically engaging with the tool's feedback rather than accepting it passively—achieved higher Essay 2 scores compared to those who were not able to assess, question or reject the tool's feedback.

5.6 Summary of findings in tool usage between teachers and students

Regarding RQ2a—"What specific strategies do teachers recommend for using each of the four writing tools in IELTS essay preparation?"—the analysis reveals that teachers generally shared a reflective approach to tool usage, integrating multiple tools and consistently emphasising the cognitive engagement in each tool's output across all three writing stages. During the planning stage, 70% of teachers recommended using Sketch Engine to identify collocations and generate topic-specific vocabulary, and 90% of teachers suggested employing Google Translate to translate Vietnamese ideas into English to support vocabulary development. Quillbot was rarely suggested for planning, with only one teacher (10%) recommending its use to enrich

ideas and vocabulary. Grammarly was proposed by 20% of teachers at this stage as a tool to refine initial translated drafts. In the composing stage, Quillbot featured more prominently, with 50% of teachers recommending using this tool to formalise the writing style and to help students expand vocabulary, often in combination with Google Translate to ensure greater accuracy. In contrast, the other three tools were less commonly recommended during this stage. Only one teacher (10%) suggested using Google Translate as a starting point to convert Vietnamese phrases into English drafts, while Grammarly was noted by just 10% of teachers as a tool to promote awareness of grammatical accuracy during drafting. Sketch Engine was not recommended by any teachers for composing activities, which may indicate that they perceived its role as more relevant to planning and revision rather than to real-time drafting. In the revising stage, Grammarly was the most widely recommended tool, with 90% of teachers advising its usage not just for error correction but also to promote error awareness and self-reflection on language improvement among students. Quillbot was recommended by 50% of teachers as a means to support students' language improvement, particularly by encouraging students to evaluate the tool's suggestions and compare them to students' original writing. Sketch Engine was suggested by 20% for checking collocations to ensure word choice accuracy, and Google Translate was not used by any teacher during revision. Overall, although the ten teacher participants integrated the four tools in different ways across the three writing stages, they consistently emphasised the need for students to have behavioural and cognitive engagement with the tools. In particular, students need to evaluate all tool outputs critically, comparing with their existed language knowledge, cross-checking a tool's outputs by combining other tools (such as using Sketch Engine or Grammarly to refine outputs of Google Translate, or use Sketch Engine to verify the suggestions of Quillbot), and using these four tools as resources for language learning rather than shortcuts for text translations or paraphrasing.

Regarding RQ2b—"What strategies do students employ when using the four writing tools?"— data analysis reveals that students employed a range of strategies when using the four writing tools during the three stages. In the planning stage, only a small number of students (15%) used Sketch Engine's Concordance function to explore collocations and develop ideas, but no students used Google Translate, Quillbot, or Grammarly. During the composing stage, Google Translate was the most frequently

used tool by 70% of students, primarily serving as a vocabulary aid or to translate Vietnamese phrases into English, and in two cases (Students 16 and 20), to produce fully translated drafts of entire essays. Quillbot was employed by 25% of students to rephrase sentences and improve language variety and style, with a few students integrating it with Google Translate for more refined output. Sketch Engine was used by 15% of students to enhance word choice by checking collocations and context-specific usage. In the revising stage, Grammarly was widely adopted by nearly all students (95%) to correct spelling, grammar, and punctuation errors. Quillbot was also used during revision by 45% to refine sentences and improve academic writing style, while Sketch Engine was employed by a small percentage of students (20%) to check phrase usage and ensure naturalness in writing. Notably, a small number of students (10%) used Google Translate for backward translation (English to Vietnamese) during revision, even though this was not a recommended practice by teachers. Overall, these findings illustrate varied patterns of tool usage among students, with Google Translate and Grammarly being the most widely used for revision, and Quillbot and Sketch Engine serving as complementary resources for enhancing vocabulary and refining writing quality.

Regarding RQ2c—“Did students whose tool usage shared similarities with teachers’ strategies achieve higher Essay 2 scores compared to those whose tool usage differed?”—the analysis indicates that students who used Sketch Engine, Quillbot, and Grammarly similarly with teachers’ recommendation mostly belonged to the high-scoring group (6.5 or higher). In the planning stage, alignment meant using Sketch Engine to build topic-specific vocabulary and check collocations (recommended by 70% of teachers) and, where needed, using Google Translate to convert Vietnamese ideas into English to support lexical planning (90%); Quillbot (10%) and Grammarly (20%) played only minor roles at this point. In composing, closer alignment involved drawing on Quillbot (50%) to formalise academic style and expand vocabulary—often alongside Google Translate for accuracy checks (10%)—while not relying on Sketch Engine for real-time drafting, and using Grammarly for grammar error detection (10%). In revision, alignment was strongest when students prioritised Grammarly (90%) for error detection and reflection, complemented by Quillbot (50%) for evaluating alternative phrasings and by targeted checks in Sketch Engine (20%) to verify collocations; notably, teachers did not endorse Google Translate for revision (0%).

Across stages, teachers consistently urged students to scrutinise all outputs, cross-check suggestions across tools (e.g., use Sketch Engine or Grammarly to refine Google Translate outputs, or use Sketch Engine to verify Quillbot's paraphrases), and treat these tools as supports for learning rather than shortcuts for generating texts.

Aligning with these abovementioned tool usage strategies from teachers, high-scoring students employed Sketch Engine during planning and engaged critically with Quillbot and Grammarly in the composing and revising stages, thereby reflecting a multi-tool, evaluative approach. Some students further extended these strategies: for example, Student 7 refined Google Translate outputs through Quillbot, Student 5 combined Quillbot with Sketch Engine to improve collocation accuracy, and Students 18 and 19 used Google Translate for back-translation when revising their self-written English texts. The fact that several high-scoring students extended teacher-recommended strategies suggests that effective tool use in L2 writing is not only a matter of following instructions but also of developing cognitive engagement. These learners demonstrated the ability to evaluate tool outputs, combine multiple tools for complementary purposes, and repurpose functions (e.g., back-translation) as diagnostic techniques. These findings indicate that higher writing performance is associated with students' ability to use tools creatively and strategically, rather than in a mechanical, single-function manner, underscoring the need for teachers to design lessons that cultivate such flexible tool usage, especially with Quillbot, Sketch Engine and Google Translate.

By contrast, low-scoring students (IDs 12, 13, 16, 17) tended to use the tools superficially for instant text generation or surface-level fixes, often accepting Sketch Engine or Quillbot suggestions without sufficient evaluation—at odds with teachers' emphasis on verification and reflection. Divergence was most pronounced for Google Translate: although teachers positioned it primarily as a planning scaffold, students typically used it as a dictionary or for wholesale translation without rigorous revision, and teachers did not endorse its use in revision. Taken together, these patterns suggest that evaluative, stage-appropriate, and cross-checked tool use—rather than mere adoption, especially of Google Translate—characterises higher performers, while persistent misalignment complicates simple causal claims about tool use and writing scores. However, this relationship between teacher-aligned use and higher

Essay 2 scores was not consistently observed in the use of Google Translate. While teachers predominantly used Google Translate in the planning stage to support brainstorming and idea translation, the 20 student participants did not adopt this practice. Moreover, during composing stage, few students combined Google Translate with Quillbot as recommended by teachers; most used it merely as a dictionary. Some students translated entire essays directly from Vietnamese, a method teachers did not endorse. Additionally, during revision, certain students used Google Translate for back-translation, an approach neither recommended by teachers nor acknowledged in student reflections (Students 18 and 19). In the revision stage, some students used Google Translate for backward-translation, an approach neither recommended by teachers (0%) nor described in students' reflections. Nevertheless, for Students 18 and 19, backward-translation appears to have been beneficial, as they successfully revised their self-written English and showed gains in Essay 2 scores

To conclude, regarding the main research question RQ2—"What similarities and differences exist between teachers' recommended strategies for using the four writing tools in IELTS essay preparation and students' actual usage patterns, and how are these similarities and differences reflected in students' essay writing performance?", the data reveal both similarities and differences between teachers' recommended strategies and students' actual usage patterns of the four writing tools—Sketch Engine, Google Translate, Quillbot, and Grammarly—across planning, composing, and revising stages, which has a connection to students' Essay 2 writing scores. Teachers promoted a process-oriented approach in tool usage that emphasised multiple tool integration and cognitive engagement. In particular, they recommend Sketch Engine for vocabulary-building in planning, Quillbot and Google Translate for composing support, and a combined use of Quillbot, Sketch Engine, and Grammarly for revision to foster cognitive engagement and language development. High-scoring students tended to mirror this tool usage in practice, using Sketch Engine and Quillbot critically during planning and revision and combining tools strategically to refine writing style. They also used Grammarly not merely for error correction but for self-reflection on language accuracy, aligning with teachers' emphasis on metacognitive awareness. Combining with the findings in Chapter 4, high-scoring students also benefited more from these four tools in terms of vocabulary acquisition and grammar improvement. In contrast, low-scoring students often bypassed planning altogether, used Google

Translate as a quick translation tool rather than a means of idea generation, relied on Grammarly and Quillbot passively, and treated them as final-stage fixes without engaging in deeper revision. Notably, no student fully adopted teachers' recommended integrated use of Sketch Engine with Grammarly, highlighting a missed opportunity for cognitive engagement and language development. These patterns indicate that alignment with teachers' reflective, multi-tool strategies was observed among students with cognitive engagement and higher writing scores, while divergence or superficial tool use was observed among students with lower writing scores. Therefore, the data suggest that a reflective, process-oriented approach to tool usage—emphasising the integration of multiple tools and cognitive engagement of these four writing tools, as recommended by teachers—was a factor in students' writing performance.

5.7 Findings in teachers' and students' evaluations of the four tools

Research questions

RQ3: How do teachers and students perceive the effectiveness and limitations of the four writing tools

RQ3a: What are the strengths and limitations of each tool, as evaluated by teachers and students?

RQ3b: How do students' evaluations relate to their writing performance?

RQ3c: What additional features do teachers and students expect beyond the tools' current capabilities?

Section 5.7 aims to answer RQ3: "How do teachers and students perceive the effectiveness and limitations of the four writing tools, and how do students' evaluations relate to their writing performance?" and its sub-questions RQ3a: "What are the strengths and limitations of each tool, as evaluated by teachers and students?", RQ3b: "How do students' evaluations relate to their writing performance?", and RQ3c: "What additional features do teachers and students expect beyond the tools' current

capabilities?” To answer these questions, the researcher first identified recurring themes for each tool in interview responses, then categorised teachers’ and students’ evaluations into positive and negative feedback (Section 5.7.1). These evaluations were then cross-referenced with students’ writing score changes and Essay 2 results (as presented in Table 16, Chapter 4, Section 4.2) to examine potential relationships. Importantly, these evaluations were analysed not only for their pedagogical implications but also as indicators of emotional engagement, which the literature has highlighted as a critical factor alongside behavioural and cognitive engagement in tool-mediated writing (see Chapter 2, Section 2.2).

Next, in Section 5.7.2, the researcher analysed recommendations from both groups, focusing on similar tools and functions they suggested. These tool recommendations were also compared with students’ writing scores to explore whether specific preferences have a relationship with writing performance. Finally, Section 5.7.3 presents a summary and discussion of the findings.

5.7.1 Comparative analysis of teachers' and students' tool evaluations

To address RQ3a, “What are the strengths and limitations of each tool, as evaluated by teachers and students?”, the researcher applied a thematic coding approach to the interview data. Feedback from ten teacher participants was first categorised into two overarching themes: “Positive feedback”, representing perceptions of the tools’ usefulness, and “Negative feedback”, capturing perceived limitations, errors, or student dissatisfaction. Within each category, recurring sub-themes were identified to illustrate shared evaluative patterns.

For example, under the “Positive feedback” theme, teachers often highlighted Sketch Engine’s value for brainstorming by providing frequent collocations that could guide idea generation and paragraph development:

Sketch Engine is helpful for brainstorming because it shows the most frequent collocations of a word, which gives students natural language patterns to work with. For example, if they’re writing about ‘education,’ they can quickly see high-frequency collocations like ‘higher education,’ ‘education system,’ or ‘access to education.’ This helps spark ideas and gives them direction for developing their arguments or structuring their paragraphs. (Teacher 3)

In contrast, the “Negative feedback” theme captured concerns about Sketch Engine’s limitations for planning without teacher scaffolding:

There is not enough data to brainstorm arguments for the given topic, so prior knowledge or experience is still needed, and teachers must prepare a word cloud beforehand. (Teacher 7)

Using the same coding methodology, the researcher then analysed responses from 20 student participants. Their evaluations were also sorted into Positive and Negative Feedback categories, with common themes extracted across the dataset.

For instance, under “Positive feedback”, students noted that Sketch Engine could expand their idea pool through its Concordance function:

For brainstorming, I found Sketch Engine useful. I used its Concordance function to find more ideas for ‘distance learning’. I found some terms that appeared in the concordance examples, such as ‘home study’ and ‘open learning courses,’ which helped me with more ideas about its advantages. (Student 8)

Conversely, under “Negative feedback”, students expressed dissatisfaction with the tool’s reliance on frequency rather than contextual appropriateness:

Sketch Engine was not very helpful with word collocations because the output for each word was based solely on frequency. I did not find a satisfying word that goes well with ‘attention’. (Student 5)

Following this thematic analysis of Positive and Negative Feedback, Table 40 below summarises the comparative evaluations of Sketch Engine by teachers and students, including participant IDs and percentages to highlight cross-group trends.

Table 40. Comparative findings from teachers' and students' evaluations of Sketch Engine

Feedback	Feedback Detail	Teachers (n=10)	Students (n=20)
Positive	Useful for planning ideas	4 (40%)	2 (10%)
Positive	Useful for finding collocations and concordances to revise writing	7 (70%)	6 (30%)
Negative	Complex interface and overwhelming data	9 (90%)	7 (35%)
Negative	Limited data for brainstorming ideas	3 (30%)	12 (60%)
Negative	Limited data for collocation	1 (10%)	3 (15%)
Negative	Limited data for academic concordances	1 (10%)*	1 (5%)
Negative	Lack of vocabulary level assessment for collocations	3 (30%)	1 (5%)

The comparison analysis between teacher and student evaluations of Sketch Engine reveals both similarities and differences. A higher proportion of teachers (40%) found the tool useful for idea planning, whereas only 10% of students (Students 6 and 8) reported the same view. Similarly, 70% of teachers valued Sketch Engine for its ability to identify collocations and provide concordance lines to support language refinement and revision. In contrast, only 30% of students (Students 3, 6, 8, 9, 12,18) evaluated this function as beneficial. These differences in findings indicate that teachers hold more positive perceptions of Sketch Engine's potential for idea planning and language development than students do. In practice, however, student usage was very limited. As discussed in Chapter 5, Section 2, only two high-scoring students (IDs 6 and 8) employed Sketch Engine during planning with the similar purpose of extracting phrases from concordance lines to support idea development. The majority of students either bypassed the planning stage or relied on other tools, revealing a clear gap between teachers' pedagogical intentions for Sketch Engine and the ways students engaged with—or failed to engage with—its functions.

In terms of negative feedback, 90% of teachers expressed concern about the tool's complexity, citing its interface and dense data as overwhelming, especially for lower-proficiency learners. However, only 35% of students (Students 1, 2, 12, 14, 16, 18, 19) shared this concern, likely because many did not engage with the tool as extensively

or struggled to navigate it altogether. More notably, while only 30% of teachers felt the tool was unhelpful for brainstorming, a significantly larger proportion of students (60%) commented on the same limitation. This mismatch points to different expectations: teachers may view the tool as a language resource that the students need to critically engage with and compare with their self-written texts. Meanwhile, students expect more direct, instant support, especially for ideas. A small proportion of participants—10% of teachers (Teacher 10) and 15% of students (Students 5, 9, and 13)—reported that the tool provided limited or unsatisfactory collocational data, suggesting that frequency-based outputs did not always yield meaningful or context-appropriate word combinations. Additionally, 20% of teachers (Teachers 3 and 9) and 5% of students (Student 6) found the concordance results insufficiently academic, indicating that the examples drawn from corpora were not always suitable for formal writing contexts. More prominent negative feedback among teachers was the absence of vocabulary level information, with 30% (Teachers 3, 4, and 6) pointing out that Sketch Engine's outputs of collocations does not refer to the Common European Framework of Reference for Languages (CEFR) levels of each collocation, which the teachers believe to make this tool beneficial for language learning. In contrast, only one student (5%) mentioned this issue, which may suggest that most students were either less aware of CEFR levels or did not expect the tool to provide that kind of support.

When comparing students' evaluations of Sketch Engine with their writing scores, Students 6 and 8, who were the 10% that provided positive feedback on Sketch Engine's usefulness for planning, were high-scoring students and showed score improvement in their Essay 2. This suggests that the usage of Sketch Engine to search for a keyword or key phrases during the planning stage positively affected these students' perceptions and writing scores. However, both high-scoring and low-scoring students gave negative feedback for this corpus tool, thus data are insufficient to support any generalisable or causal link between this corpus tool's evaluation and writing scores. The follow up discussion about the disconnect between the positive academic literature on corpus tools and their actual practice will be presented in Chapter 6, Section 6.2.

Following the analysis of Sketch Engine, the researcher turns to Google Translate, comparing evaluations from both teachers and students. Table 41 below presents these comparative findings.

Table 41. Comparative findings from teachers' and students' evaluations of Google Translate.

Category	Feedback Detail	Teachers (n=10)	Students (n=20)
Positive	Easy to use; helpful in finding the meaning of words or sentences	6 (60%)	18 (90%)
Negative	Can generate contextually inaccurate translations	5 (50%)	4 (20%)
Negative	Overreliance may lead to lazy thinking or over-dependence on technology	2 (20%)	0 (0%)

According to Table 41, a notable similarity is the overall positive feedback, particularly regarding the tool's usefulness for vocabulary support and sentence-level translation. Specifically, 60% of teachers found Google Translate helpful in assisting learners with finding word meanings and basic sentence construction, while an even larger proportion of students (90%, so 18 out of 20), reported similar benefits. This strong student endorsement suggests that the tool is widely accessible and easy to use, meeting their immediate needs while composing. However, the lower percentage for teachers might show they're more cautious, thinking beyond short-term use to how the tool affects students' language improvement over time.

When it comes to negative feedback, teachers expressed more concerns than students, particularly about translation accuracy and reliance on this tool. Half of the teachers (50%) observed that Google Translate often produces contextually inaccurate translations, cautioning that such outputs could lead to language misuse and hinder the development of precise academic writing skills. In contrast, only 20% of students (Students 1, 5, 7, and 15) raised similar concerns. Furthermore, while 20% of teachers expressed concerns about students becoming overly reliant on Google Translate, they specifically warned that such overreliance might discourage independent thinking and hinder the development of productive language skills. This

reflects a pedagogical awareness of the potential long-term effects of passive tool use. However, the absence of such concerns among students suggests they may be unaware that habitual, uncritical use of translation tools could impede their language learning.

Among the 90% of students who gave positive feedback for Google Translate, both high- and low-scoring learners were represented—specifically, seven low-scoring (Students 4, 11, 12, 13, 16, 17, 19) and nine high-scoring students (Students 1, 2, 3, 6, 7, 8, 9, 10, 20)—indicating that favorable perceptions were shared across different proficiency levels. Students 5 (high-scoring, with score improvement) and 12 (low-scoring, with a score decrease) did not report any positive feedback, likely due to no tool use during their composition of Essay 2 (see Chapter 4, Section 3). Moreover, 20% of students (Students 1, 5, 7, and 15) who provided negative feedback about Google Translate, and who specifically noted that it can produce incorrect or contextually inaccurate translations, were high-scoring students. This supports the findings presented in Chapter 4, Section 3, which highlighted higher levels of critical assessment of the tool’s output among high-scoring students. Nevertheless, one of them—Student 1—experienced a slight decrease in Essay 2 score, while the other three showed score increases. Collectively, these findings are insufficient to support any generalisable or causal link between students’ evaluations of Google Translate and their writing scores.

The next analysis turns to Quillbot. Table 42 displays comparative findings from teachers' and students' evaluations of Quillbot:

Table 42. Comparative findings from teachers' and students' evaluations of Quillbot

Category	Feedback Detail	Teachers (n=10)	Students (n=20)
Positive	Helps students learn synonyms and paraphrase to different grammar structures	7 (70%)	12 (60%)
Negative	Outputs can sometimes be irrelevant and unusable	4 (40%)	6 (30%)
Negative	Did not work or loaded slowly	0 (0%)	2 (10%)

The evaluations of Quillbot in Table 37 reveal generally positive feedback from both teachers and students, with some differences in emphasis and concerns. For instance, 70% of teachers (7 out of 10) viewed the tool positively, particularly for its ability to help students learn synonyms and paraphrase using different grammatical structures. This aligns closely with 60% of students (12 out of 20) who also found Quillbot beneficial for rephrasing and improving sentence variety. Both groups also gave similar negative feedback, as 40% of teachers and 30% of students reported that the tool occasionally generated irrelevant or unusable output, raising concerns about reliability and the need for critical assessment of this tool. Additionally, 10% of students (Students 1 and 7) reported technical issues such as slow loading or functionality problems, though no teachers mentioned such issues.

Among the 60% of students who gave positive feedback on Quillbot (12 out of 20), the majority—10 students—were high-scoring, while only two were low-scoring. This suggests that high-performing students were more likely to critically recognise and articulate the tool's merits, particularly in relation to lexical improvement. Additionally, six out of the seven students (IDs 1, 3, 7, 10, 11, 18 and 19) who gave negative feedback about the irrelevant output of Quillbot were high-scoring (IDs 1, 3, 7, 10, 11, and 18). This highlights the pattern observed in Chapter 4, Section 3: that high-performing students in Essay 2 were more likely to critically assess tool-generated output, recognising its limitations and using it selectively. However, changes in students' writing scores did not consistently align with their evaluations of the tool. For instance, three students who gave negative feedback (Students 1, 3, and 12) experienced a decline in their Essay 2 scores, whereas the remaining five students, who also expressed negative views (Students 7, 10, 11, 18, and 19), demonstrated score improvements. These mixed results suggest that there is no clear link between students' evaluations of Quillbot and their writing scores.

Regarding Grammarly, Table 43 displays comparative findings from teachers' and students' evaluations of this tool.

Table 43. Comparative findings from teachers' and students' evaluations of Grammarly

Feedback	Feedback Detail	Teachers (n=10)	Students (n=20)
Positive	Provides fast grammar corrections	10 (100%)	14 (70%)
Positive	Reduces marking time for teachers	2 (20%)	0 (0%)
Negative	Suggestions are robotic or focus too much on sentence reduction (unhelpful for IELTS)	3 (30%)	3 (15%)
Negative	Limited correction ability regarding more complex errors	10 (100%)	9 (45%)

The evaluation of Grammarly revealed similar positive evaluations from both teachers and students. All 10 teachers (100%) rated Grammarly positively for providing fast grammar corrections, while 70% of students (14 out of 20) also viewed the tool positively, particularly for its assistance in revising grammar and sentence structure. Furthermore, while a small number of teachers (20%) valued Grammarly for reducing their marking workload, no students mentioned this advantage, suggesting that teachers see the tool as helpful but not as a replacement for human feedback. In contrast, students tend to focus only on Grammarly's instant grammar correction functions and may lack awareness of the broader pedagogical functions of such a tool.

However, a closer examination of the negative feedback reveals several important concerns. Three teachers (30%) and three students (15%) criticised Grammarly's suggestions as "Overly robotic and overly focused on sentence reduction", which may negatively affect writing scores. Below are excerpts of their negative feedback on this theme:

[137] There are unnecessary corrections. Grammarly tends to break down small sentences and create a robotic style by removing relative clauses and changing word order. (Teacher 2)

[138] It can sometimes suggest removing the inclusion of a relative clause. I think this is not beneficial for language learning since it reduces the grammar variety. The tool also offers limited support for improving cohesion. It rarely

provides useful feedback on linking phrases or how students can better connect ideas between sentences and paragraphs. (Teacher 6)

[139] Some of Grammarly's corrections are a bit "dry" for the writing style that is expected in the IELTS test. For instance, it often flags long sentences with numerous noun phrases as issues. This could negatively impact the Grammatical Range and Accuracy criterion of the IELTS exam. Moreover, some suggestions of it are too "fixed" and might not be accurate all the time. For example, "go shopping" and "do shopping" are both correct in terms of grammar and semantics, but Grammarly would always identify the phrase "do shopping" as an error and change it to "go". (Teacher 10)

[140] I would like tools that can suggest ideas. Grammarly sometimes provides informal suggestions for IELTS writing, so it would be helpful if it could choose a more formal writing style. (Student 10)

[141] Grammarly usually identifies the use of passive voice as incorrect, yet I was taught that the passive voice is important for an IELTS essay. (Student 12)

[142] Not a challenge, but I found that Grammarly sometimes suggests removing extra words, which can be confusing [...] For example, changing "to" to "in order to" was a suggestion from Grammarly. (Student 15)

Accordingly, a key criticism among these responses is that the tool tends to overly simplify texts and produce robotic suggestions by breaking down complex sentences, removing relative clauses, and removing certain prepositions. Teacher 2, for instance, noted that such changes diminish the natural flow and syntactic variety of students' writing. Similarly, Teacher 6 emphasised that Grammarly's suggestions often reduce grammatical richness and provide little support for cohesion, such as linking phrases or transitions between ideas. These structural simplifications may negatively affect students' scores, especially in the Grammatical Range (GA) and Accuracy and Coherence and Cohesion (CC) criteria. Teacher 10 reinforced this concern by explaining that Grammarly flags complex constructions, such as noun phrases, as errors, and consistently enforces rigid lexical choices—such as correcting "do shopping" to "go shopping"—despite both being grammatically acceptable. From the students' perspective, the issues echoed similar themes. Student 10 mentioned

Grammarly's tendency to suggest informal or inappropriate phrasings for academic writing. Student 12 noted that the tool often discourages the use of passive voice, despite its relevance in formal writing tasks like the IELTS. Student 15 commented that Grammarly's suggestions, such as replacing "to" with "in order to," can be confusing or unnecessary, and that the tool offers little assistance in generating content or developing ideas. Collectively, these responses suggest that Grammarly is limited to surface-level corrections and often misaligns with the specific and grammatical requirements in IELTS essay writing. Furthermore, given the small proportion of responses from teachers (30%) and students (15%), this finding suggests that only a minority of participants in this study claimed that Grammarly should not be used in isolation for revision—particularly when preparing for the IELTS essay format. Instead, a more widely reported concern among all ten teachers (100%) was Grammarly's limited ability to correct more complex grammar errors, such as relative clauses, verb tense consistency, and parallel structures. However, this negative feedback was expressed by only 45% of students (IDs 1, 2, 4, 7, 10, 12, 13, 17, and 18). This discrepancy suggests that more than half of student participants may lack the metalinguistic awareness required to critically assess the accuracy of automated feedback.

When comparing students' evaluations of Grammarly with their writing performance, the 70% of students who gave positive feedback included both high-scoring students (Students 1, 2, 5, 7, 8, 14, 15, 18, and 20) and low-scoring students (Students 4, 13, 17, and 19). Among these 13 students, only one (Student 1) experienced a decline in their Essay 2 score, while the remaining 12 showed score improvements, suggesting that the tool's perceived usefulness was shared across proficiency levels and may have provided support during the writing process. By contrast, students who gave negative feedback also represented a mix of proficiency levels, including high-scoring students (Students 1, 2, 7, 10, 18) and low-scoring students (Students 4, 12, 13, 17). Their writing score changes were also varied: three students (Students 1, 12, 17) experienced a decline in performance, while the other six showed improvements (Students 2, 4, 7, 10, 13, 18). This mixed pattern suggests that recognising Grammarly's limitations does not consistently correspond to writing score changes, and that evaluations of the tools' strengths and weaknesses are not exclusive to any specific proficiency level.

In summary, the findings for RQ3a: “What are the strengths and limitations of each tool, as evaluated by teachers and students?” suggest that while evaluations of the four digital writing tools were broadly positive, teachers and students differed in important ways, even as some areas of alignment emerged. Sketch Engine was viewed more positively by teachers, who highlighted its value for idea planning and collocation searches, though most students found it too complex and limited for brainstorming; only two high-scoring students used it effectively in planning, showing slight score gains. Google Translate received widespread praise from both groups for accessibility and vocabulary support, though teachers were more cautious, noting risks of inaccuracy and overreliance, while students largely overlooked these concerns. Quillbot was perceived by teachers and high-scoring students as helpful for paraphrasing and lexical variety, yet both groups noted occasional irrelevant outputs and technical issues, with mixed links to writing performance. Grammarly was the most positively evaluated, unanimously endorsed by teachers and widely valued by students for fast grammar corrections, though teachers and some students highlighted its inability to address complex errors and its tendency toward robotic, oversimplified suggestions. Across all tools, teachers tended to adopt a more pedagogical and critical stance, while students emphasised simple interface and immediate benefits, with high-scoring learners showing greater critical awareness than low-scoring peers.

Regarding RQ3b: “How do students’ evaluations relate to their writing performance?”, the findings reveal that students’ perceptions of the four digital writing tools—whether favourable or critical—did not consistently align with measurable changes in their writing scores. High-scoring and low-scoring students alike expressed both positive and negative views, suggesting that evaluations alone are not reliable indicators of performance outcomes. This highlights that effective tool usage was not determined by students’ emotional engagement with the tools. Rather, improvements in writing were more closely associated with behavioural and cognitive engagement—that is, how strategically and critically the tools were employed, as revealed in Chapter 4.

5.7.2 Additional tool recommendations from teachers and students

Section 5.7.2 addresses RQ3c: “What additional features do teachers and students expect beyond the tools’ current capabilities?” To answer this question, the researcher analysed teachers’ responses to Interview Question 18: “Do you recommend any other

tools for teaching the IELTS Task 2 essay?” and Interview Question 19: “Do you recommend any other tools for students to practice the IELTS Task 2 essay at home?”, as well as students’ responses to Interview Question 8: “Do you recommend or would use any other tools for this IELTS essay writing practice? Why?” Teachers’ and students’ answers were categorised into themes based on the names of the recommended tools and intended usage purposes. Table 44 below presents the tools recommended by teachers beyond the four examined in this study, along with their suggested usage, the number of teachers who made each recommendation, corresponding percentages, and teacher IDs.

Table 44. A summary of teachers’ recommendations of other tools and their usage purposes

Tool’s Name	Purpose	Teachers (n=10)
ChatGPT	Brainstorming ideas	4 (40%)
ChatGPT and Poe	Advanced writing corrections (lexical choice, coherence and cohesion)	2 (20%)
MindMeister	Brainstorming ideas	1 (10%)
Flax, COCA, Ozdic, OneLook Thesaurus	Finding collocations	2 (20%)
Cambridge or Oxford dictionary	Checking for word meanings	1 (10%)

Data from Table 44 indicates that 40% of teachers (Teachers 1, 2, 7 and 8) recommended Generative AI tools like ChatGPT or Poe to assist with brainstorming ideas, while 20% (Teachers 9 and 10) expressed a need for a more advanced writing correction function that is embedded in Generative AI. Below are their interview responses:

[143] Yes, I would recommend using ChatGPT as a support tool for teaching the IELTS Task 2 essay, especially during the brainstorming phase. Many students struggle with generating ideas quickly under time pressure, and this tool can really help them overcome that initial block. I also encourage students to use it to explore how arguments can be developed. For instance, they can ask, ‘What are some reasons people support investment in public transport?’ or ‘Can you give examples to support the view that private car use should be

reduced?’ ChatGPT can offer a range of ideas that students may not think of themselves, and they can then select, adapt, and personalise those points in their own writing. Of course, I always emphasise the importance of critical thinking. They shouldn’t copy full paragraphs or sentences, but instead treat the tool as a brainstorming partner—to help generate ideas, outline arguments, and even practice paraphrasing. It’s all about training them to engage with the tool actively and thoughtfully, not passively. (Teacher 1)

[144] Students can focus on the skills that tools cannot perform, such as planning ideas. Students can use Google search for planning, and collocation dictionaries like Flax, COCA, and Ozdic can use corpora from other countries. Students can also use ChatGPT to explore more ideas, but I would not encourage them to just ask the AI to “write the essay” and copy the answer. (Teacher 2)

[145] I would suggest using ChatGPT to support students with brainstorming ideas for IELTS Task 2 essays. Many students find it difficult to generate relevant arguments or examples, especially when they are unfamiliar with the topic. I often show them how to use ChatGPT to explore different perspectives. For example, if the question is about whether university education should be free, I guide students to ask the tool for reasons in support and against the idea. This helps them understand both sides of the argument and decide which one to develop in their essay. (Teacher 7)

[146] I think ChatGPT can be a valuable tool to help students come up with more arguments on a wider range of IELTS Task 2 topics. Some students tend to rely on memorising vocabulary, and when they face unfamiliar questions. So I would guide them to use ChatGPT to explore the topic more deeply. For example, if the essay asks about the benefits of investing in space research, students can ask ChatGPT for reasons why space programs might be important for society, and it could suggest ideas related to scientific innovation, environmental monitoring, or international cooperation. I would instruct students to ask the right kinds of prompts, like ‘What are the advantages and disadvantages of [topic]?’ or ‘What are three reasons to support this opinion?’ This helps students think more critically about how to structure their argument.

Students should summarise those points in their own words and using the tool only as a starting point. (Teacher 8)

[147] I would suggest using ChatGPT to support students with more advanced corrections, particularly in areas like vocabulary enhancement and the use of linking words. Many students tend to repeat basic vocabulary or rely on simple connectors like 'and,' 'but,' or 'because' in their essays. I show them how to use ChatGPT to suggest more precise or academic alternatives. For instance, if a student writes 'people like watching movies because they are fun,' I guide them to ask ChatGPT how to improve the sentence. It might suggest: 'people enjoy watching films due to their entertainment value and emotional appeal.' This helps expand their lexical range and increase formality. (Teacher 9)

[148] I recommend that my students use ChatGPT to generate more ideas and arguments. It's an all-in-one AI tool that can help save time. The key factor will be crafting the right prompts, step-by-step, to compose the essay more effectively. (Teacher 10)

Accordingly, these six teachers highlighted how Generative AI can support students to explore multiple perspectives and develop structured arguments through well-crafted prompts. Teachers 1, 7 and 8, for instance, noted that by asking specific questions—such as the advantages and disadvantages of a topic—students are encouraged to think more critically and structure their responses more logically. Furthermore, there was a similar caution among teachers against passive or uncritical use of ChatGPT. Teachers 1, 2, and 8 explicitly warned students not to copy entire AI-generated responses but to instead adapt and personalise the content while maintaining academic integrity. This reinforces the idea that teachers see ChatGPT as a tool to scaffold student thinking, not to replace it. Beyond idea generation, Teacher 9 extended the role of ChatGPT to lexical enhancement and discourse-level cohesion, describing how Generative AI could help students replace overused vocabulary and basic connectors with more formal and varied expressions, directly contributing to improvements in Lexical Resource and Coherence and Cohesion—two key IELTS scoring criteria. Collectively, these insights suggest that, even at the time of data collection in 2023, when Generative AI tools such as ChatGPT were still relatively new and very different from how they are now, teachers were already

beginning to accept that AI could support not only surface-level grammar correction but also more cognitively demanding aspects of writing, such as idea development and coherence and cohesion, but they consistently saw it as a tool to scaffold student thinking, not to replace it

5.7.2.1 Teachers' suggested alternatives to AI

By contrast, a similar percentage of teachers (40% in total) recommended more traditional, non-AI tools for language resources, including 10% who opted for a non-AI drawing tool called Mindmeister to assist students with brainstorming, 20% referred to other corpus tools (Flax, COCA, Ozdic, OneLook) over Sketch Engine for their simpler interface, and 10% favoured reputable online dictionaries such as Cambridge or Oxford for checking word meanings over Google Translate. Below are examples of their responses:

[149] Students can focus on the skills that tools cannot perform, such as planning ideas. Students can use Google search for planning, and collocation dictionaries like Flax, COCA, and Ozdic can use corpora from other countries. (Teacher 2)

[150] I would recommend OneLook Thesaurus. It can find all parts of speech of a word, its synonyms, words that go with it, as well as slang or idioms used with this word. (Teacher 3)

[151] I would also recommend students use a tool named Mindmeister for presenting their outline and improving critical thinking. (Teacher 5)

[152] I would stay close to the traditional methods: a dictionary (Cambridge dictionary or Oxford dictionary) because they can give me the level of each word. And I would use the Google search engine to look for concordances. (Teacher 6)

These interview responses from the four teachers reveal a shared belief that writing development should go beyond simply relying on automated tools and instead integrate non-AI language resources that encourage active cognitive engagement. Teacher 2 emphasises the importance of students developing skills that cannot be

fully replicated by mere data exploration, such as idea planning, and recommends complementary tools—like Google Search for content research and collocation dictionaries such as FLAX, COCA, and Ozdic—to support authentic language use drawn from diverse corpora. Similarly, Teacher 3’s recommendation of OneLook Thesaurus reflects an intention to help students explore vocabulary more flexibly and contextually. Teacher 5 recommended a non-AI tool called MindMeister for brainstorming. This suggestion indicates an awareness that effective writing begins with well-structured ideas, and that digital mind-mapping tools can help students organise their thoughts systematically while engaging in higher-order thinking. Teacher 6’s preference for traditional dictionaries and Google Search for concordance searches suggests a pedagogical approach grounded in more authoritative resources. However, none of the four teachers explicitly raised concerns about the use of Generative AI or about the risk of students becoming overly dependent on such tools. This suggests that issues commonly associated with Generative AI, such as reduced opportunities for learners to practise independent language production, the negative effects of hallucinated outputs, and difficulties in maintaining academic integrity (Deng et al., 2025; Dwivedi et al., 2023), had not yet become central to teachers’ pedagogical thinking at the time these interviews were conducted in 2023, when tools like ChatGPT were still relatively new and less advanced than they are today.

In summary, these insights indicate a growing acceptance among teachers that Generative AI tools like ChatGPT can play a valuable role in language learning, not only by handling surface-level grammar corrections but also by supporting deeper cognitive tasks such as idea development, coherence, and cohesion, with 60% of teachers viewed such tools as useful for helping students generate arguments, explore multiple perspectives, and improve their writing. The other 40% of teachers preferred traditional non-AI tools, such as mind-mapping software, other corpus-based search tools, and reputable online dictionaries. These choices reflect a desire to use tools that promote critical thinking, autonomy, and a deeper understanding of language. Collectively, these insights reveal a shared perspective among teachers—that the goal of using either Generative AI or traditional tools is not merely to produce a polished final product, but to foster critical thinking and language learning. In other words, the emphasis is on how students engage with tools to develop ideas, recognise their gaps and learn, rather than simply generating high-quality essay products. Notably, none of

the ten teacher participants raised concerns about potential plagiarism arising from students' use of Generative AI outputs, suggesting that ethical implications were either under-recognised in the year 2023. One possible explanation is that teachers perceived classroom use as relatively controlled and monitored, particularly in an IELTS writing context that does not require citation of sources. However, this lack of concern becomes problematic if students later transfer these AI-supported practices to higher education writing tasks that demand explicit referencing and adherence to academic integrity norms (see also Chapter 6, Section 6.3 for further implications).

5.7.2.2 Students' recommendations of other tools

While teachers' evaluations highlight a balance between embracing Generative AI and relying on traditional tools, with limited attention to issues of academic integrity, it is equally important to consider how students themselves approach tool use. The following section, therefore, turns to students' recommendations of additional tools, outlining their purposes and rationales. Table 45 below presents students' answers to other tools, along with their usage purpose from Interview Question 8: "Do you recommend or would use any other tools for this IELTS essay writing practice? Why?".

Table 45. Summary of other tool recommendations from students and their purposes

Tool's Name	Purpose	Students (n=20)
ChatGPT / Google Bard / Bing AI	Brainstorming ideas and giving feedback	11 (55%)
Google Search	Brainstorming ideas	3 (15%)
Cambridge Dictionary	Checking word meanings	2 (10%)
Ozdic	Finding collocations	1 (5%)
Thesaurus Dictionary	Finding synonyms	4 (20%)

Data from Table 45 reveal that the most frequently mentioned tool type was Generative AI, including ChatGPT, Google Bard and Bing AI, recommended by 55% of students (11 out of 20), with the purpose of assisting in brainstorming ideas and giving feedback. This is a similar proportion of students as teachers' (40%), highlighting the growing demand from both teachers and students to make use of the capabilities of Generative AI. Among these 10 students, two were low-scoring (Students 4 and 14),

whereas the remaining eight were high-scoring (Students 3, 5, 6, 7, 8, 9, 10, 15, and 20). Below are 11 students' responses who mentioned the use of ChatGPT:

[153] I recommend brainstorming ideas with ChatGPT. Additionally, tools that can organise the arguments in a paragraph with cohesion and coherence would be beneficial. (Student 3)

[154] I think I would use ChatGPT for brainstorming ideas and check for errors, I think it can correct my essay better than Grammarly. (Student 4)

[155] Yes, I recommend using a Thesaurus dictionary to enhance vocabulary variety and refine word choice, particularly when revising topic sentences or paraphrasing the question. I also find ChatGPT helpful for generating ideas and developing arguments, but I use it carefully by drafting specific prompts. For example, instead of asking a general question like "Write an essay for this topic" I might ask, "What are three reasons why governments should invest in renewable energy?" or "Can you provide examples to support the idea that public transport reduces environmental pollution?" These prompts help me gather relevant points without copying full responses. I then select useful ideas and rephrase them in my own words to fit the IELTS essay structure. (Student 5)

[156] I recommend using ChatGPT for generating ideas and getting more structures and ideas for the essay. I would also use it to assess the band score of my essay, though I trust its evaluation about 50%. (Student 6)

[157] I recommend using ChatGPT to ask for arguments and extract ideas. Additionally, any tools that can evaluate the text would be helpful. It is important to check evaluations from different sources. (Student 7)

[158] I want to try Bing AI to help correct my writing, especially for coherence and word choice. However, I would still review its suggestions carefully, since not all AI-generated corrections are contextually accurate. (Student 8)

[159] I recommend using ChatGPT to ask for arguments, not for language. It helps with ideas to express in English and makes me confident that I will

improve. (Student 9)

[160] I recommend using ChatGPT to suggest ideas and provide more wording options based on experience. Additionally, Google Bard can be used to generate more ideas. (Student 10)

[161] I would opt for ChatGPT to explore more writing ideas. (Student 14)

[162] I recommend using ChatGPT because it is an all-in-one tool. I paste my own ideas into ChatGPT and ask it to "make this idea more developed". The more specific the input, the better the output. I do not use it to evaluate the whole essay and do not trust it completely for that purpose. Grammarly is good for checking grammar, while ChatGPT paraphrases better than Quillbot and is better with contexts. (Student 15)

[163] I recommend using ChatGPT to generate ideas and then selecting relevant arguments to put into Quillbot for rewriting. However, one should not rely on tools for long-term practice. (Student 20)

These interview responses show that almost all 11 students viewed idea generation as the main benefit of Generative AI at the planning stage of IELTS Task 2. In particular, nine high-performing students (Students 3, 5, 6, 7, 8, 9, 10, 15, 20) not only recommended using ChatGPT to brainstorm and develop ideas but also described increasingly purposeful and strategic ways of doing so. Student 3, for example, used ChatGPT for “brainstorming ideas” and emphasised the need for support with organising arguments “with cohesion and coherence.” Student 5 illustrated more cognitive engagement by using specific prompts (e.g. asking for three reasons or concrete examples) rather than requesting a full essay, and by selecting and rephrasing useful ideas to fit the IELTS essay structure.

Other students adopted similarly selective approaches. Student 6 used ChatGPT for ideas and approximate band-score estimates but reported trusting its evaluation only “about 50%.” Student 7 used it “to ask for arguments and extract ideas,” explicitly rejecting the option of having the whole essay generated. Student 8 stressed the need to “review its suggestions carefully,” recognising that not all AI-generated corrections are contextually accurate. Student 9 drew a clear boundary between content and

language, recommending ChatGPT “to ask for arguments, not for language,” while still acknowledging its role in helping express ideas and build confidence. Student 15 described pasting in her own ideas and asking the tool to “make this idea more developed in language,” underlining the importance of prompt quality and critical judgement.

Student 10 similarly reported using ChatGPT for both ideas and wording “based on experience,” and also mentioned Google Bard, suggesting a comparative, multi-source approach. Student 20 outlined a sequential, multi-tool strategy: “generate ideas with ChatGPT, then select relevant arguments to put into Quillbot for rewriting,” assigning distinct functions to each tool. Collectively, excerpts [153–163] show that these students were actively and explicitly planning how to integrate Generative AI into their essay-writing practices in strategic and critical ways. However, it remains uncertain how much genuine language learning occurs in these processes, given that the tools still assume substantial responsibility for both idea development and linguistic expression (see Chapter 6, Section 6.2 for further discussion). By contrast, the responses of Students 4 and 14—both low-scoring—were brief and lacked indications of cognitive engagement. Student 4 stated, “I think I would use ChatGPT for brainstorming ideas and check for errors. I think it can correct my essay better than Grammarly,” while Student 14 simply noted, “ChatGPT to explore more writing ideas,” reinforcing the tool’s role in the ideation phase. Although both students identified similar purposes for using ChatGPT, their responses did not demonstrate how they would evaluate, refine, or combine its output with other tools. This stands in contrast to high-scoring students, who articulated more deliberate and reflective strategies for integrating Generative AI into their writing process.

Overall, students’ recommendations for using Generative AI tools, particularly among high-scoring students, broadly aligned with teacher perspectives, with 40% of teachers also endorsing ChatGPT as a brainstorming assistant to address planning difficulties. However, when these stated intentions are compared with students’ actual tool use across writing stages (Table 38), there is little evidence that such practices were fully realised in Essay 2. Only two students (IDs 6 and 8) used Sketch Engine for planning, the rest of 18 students did not spend time planning or use any of the tools for this purpose, suggesting that tool-mediated planning was largely absent in practice. This

highlights a gap between students' perceived and actual use of Generative AI, at its early-stage in 2023, and raises the concern that students who already avoided investing time in planning and exploring ideas with tools may be even more inclined to skip the planning stage when working with Generative AI. Without investing such time for planning, students are more likely to adopt AI-generated content uncritically, which may ultimately hinder their ability to generate original ideas, organise arguments, and evaluate tool outputs.

In contrast to the 50% of students who recommended ChatGPT for idea generation, traditional tools were mentioned by fewer student participants—Google Search (15%), Thesaurus (20%), Cambridge and Oxford dictionaries (10%), and Ozdic (5%). Below are these seven students' responses:

[164] Yes, I recommend using a Thesaurus dictionary to enhance vocabulary variety and refine word choice, particularly when revising topic sentences or paraphrasing the question. [...] (Student 5)

[165] I recommend using the Cambridge dictionary and Thesaurus because they are more trusted. I use the closest suggestions in bold but mostly depend on my knowledge. (Student 11)

[166] I recommend using a collocation dictionary (OZdict), Thesaurus, and Google Search to search for sample readings to generate ideas. (Student 12)

[167] I recommend using the Oxford Dictionary (Eng-Eng) for checking word meanings and ChatGPT for exploring and generating more ideas. ChatGPT can be compared with Google Translate for similar quality suggestions. (Student 13)

[168] If I had more time, I would use a dictionary such as Thesaurus for better word choices. (Student 16)

[169] I recommend using Google Search for more ideas and searching for more sample essays to learn from them. (Student 17)

[170] I recommend using Google Search for articles and sample essays to gather more ideas and vocabulary on the topic of renewable energy. (Student 19)

According to these responses, only Student 11 answered that “I recommend using the Cambridge dictionary and Thesaurus because they are more trusted”. In contrast, the remaining six students who recommended traditional tools—such as Google Search, Ozdic, Oxford Dictionary, and thesauruses—did not elaborate on why they chose these resources. Their responses focused only on tool functions (e.g., finding synonyms or accessing sample essays) but lacked specific justification related to trustworthiness, effectiveness, or ease of use. This suggests that for most students who recommended these traditional tools may have adopted them out of habit, accessibility, or general familiarity, rather than a critically assessed option. Furthermore, among the seven students who recommended traditional tools, five were low-scoring students (Students 12, 13, 16, 17, and 19), and two students were high-scoring (Student 11 and Student 5). Students 11, 13, 16 and 19 had score improvements. By contrast, Students 12 and 17, who valued these tools, struggled with their Essay 2’s writing topic and had a decrease in their Essay 2 scores, and also showed a more superficial engagement compared to high-scoring students (See findings in Chapter 4, Sections 4.3 and 4.4). In contrast, Student 11, a high-scoring student, demonstrated a more strategic and autonomous approach—using the Cambridge Dictionary and Thesaurus selectively while primarily relying on personal knowledge, suggesting that traditional tools, when used critically, might also link with higher writing scores.

Collectively, these results point to differing patterns of tool preference between the two student groups. A larger number of low-scoring students reported using traditional, non-AI tools such as online dictionaries or corpus resources, yet they tended to offer only brief and somewhat superficial accounts of how they employed these tools. In contrast, fewer high-scoring students mentioned relying on these non-AI resources and instead expressed greater interest in experimenting with emerging Generative AI tools for purposes such as idea generation, planning, and refining their writing. Another notable finding is that none of the students expressed concerns about plagiarism when using content generated by Generative AI, even though many described relying on

these tools for idea generation and writing support. For example, Student 3 shared, “I recommend brainstorming ideas with ChatGPT. Additionally, tools that can organise the arguments in a paragraph with cohesion and coherence would be beneficial,” without mentioning how to ensure originality in the final text. Similarly, Student 4 stated, “I think I would use ChatGPT for brainstorming ideas and check for errors,” and Student 6 noted, “I recommend using ChatGPT for generating ideas and getting more structures and ideas for the essay,” yet neither referred to concerns about copying or proper attribution. Although Student 5 described drafting specific prompts and rephrasing ideas—“I might ask, ‘What are three reasons why governments should invest in renewable energy?’ ... I then select useful ideas and rephrase them in my own words to fit the IELTS essay structure”—this was framed as a strategy for improving writing rather than explicitly avoiding plagiarism. Likewise, other students, such as Student 9, who stated, “I recommend using ChatGPT to ask for arguments, not for language,” and Student 15, who explained, “I paste my own ideas into ChatGPT and ask it to ‘make this idea more developed,’” did not raise any concerns about plagiarism. Their comments point to two distinct uses—argument generation and idea elaboration—yet in practice these functions often come hand-in-hand, making it difficult to separate them. This blurring of roles highlights a key challenge of GenAI: its capacity to “do everything at once”, unlike the more function-specific four tools. These findings, however, must be interpreted with caution given the rapid development of GenAI capabilities. In 2023, tools such as ChatGPT were still relatively new and often viewed as only partially accurate, which had encouraged the more selective, idea-focused, and critical use reported by Students 3, 4, 5, 6 and 15. As GenAI tools continue to advance and are perceived as more reliable, students may increasingly favour AI-generated text over their own ideas, which can lead to weakening cognitive engagement, discouraging independent idea generation, and heightening risks of over-reliance (see Chapter 6, Section 6.3 for further discussion of these risks).

To sum up, in the context of 2023, these responses indicate that although students were already integrating GenAI into their writing process, they did not yet fully appreciate the risk of unintended plagiarism when incorporating AI-generated content into their work. This pattern also reflects teachers’ attitudes at that time, suggesting that both teachers and students tended to overlook the possibility that AI-generated text could lead to unintended plagiarism, particularly when students reproduced

content without adequate paraphrasing or critical evaluation.

5.7.3 Summary of findings in tools' evaluations and recommendations between teachers and students

To answer the first part of the main research question RQ3: “How do teachers and students perceive the effectiveness and limitations of the four writing tools?”, the comparative analysis of teacher and student evaluations of the four writing tools—Sketch Engine, Google Translate, Quillbot, and Grammarly—revealed both similarities and differences.

Regarding Sketch Engine, findings revealed that a high proportion of teachers positively evaluated this corpus tool for planning ideas (40%), searching for collocation or concordance (70%), while only 10% and 30% of students, respectively, shared those views. Negative feedback on the tool's complex interface was also more common among teachers (90%) than students (35%). More students found Sketch Engine unhelpful for brainstorming than teachers (60% compared to 30%). Relatively similar percentages from both groups also reported limitations in the tool's collocation outputs, with 10% of teachers and 15% of students expressing such concerns, as well as limitations in its academic concordance outputs, noted by 10% of teachers and 5% of students. Another piece of negative feedback is the lack of CEFR-level assessment for each collocation output, with a higher percentage of teachers (30%) than students (5%) mentioning this.

Regarding Google Translate, both teachers and students gave positive feedback for its instant vocabulary and sentence-level translation, with a notably higher proportion of students (90%) evaluating the tool positively compared to teachers (60%). Negative feedback was more common among teachers, with 50% highlighting translation inaccuracies and 20% warning against overreliance on this machine translation tool. In comparison, only 20% of students—all high-scoring—mentioned the contextually inaccurate translations of this tool, and none expressed concerns about overreliance, suggesting that high-scoring students demonstrated greater critical awareness of Google Translate's limitations compared to low-scoring students.

Evaluations of Quillbot were generally positive among both teachers and students, with 70% of teachers and 60% of students praising its usefulness for paraphrasing and

improving sentence variety. In terms of negative feedback, similar proportions of teachers (40%) and students (30%) noted that the tool occasionally generated irrelevant or unusable output. Notably, 10% of students experienced technical issues with the tool, though this was not reported by teachers.

Evaluations of Grammarly revealed generally positive feedback from both teachers and students, with 100% of teachers and 70% of students positively evaluating its fast grammar correction function. Furthermore, while 20% of teachers valued Grammarly for easing their marking workload, indicating that teachers view it as a minor step rather than a substitute for human feedback. In contrast, no students noted this benefit, suggesting that they might overlook Grammarly's broader pedagogical role. Negative feedback was also raised by both groups—30% of teachers and 15% of students—regarding Grammarly's overly robotic suggestions and tendency toward sentence simplification, which may limit students' ability to demonstrate grammatical range (GA), a key criterion for achieving higher IELTS band scores. Moreover, while all teachers acknowledged Grammarly's limited ability to correct complex grammar, only 45% of students recognised this shortcoming, indicating a lack of critical awareness of the tool's correction function among some students across all levels of proficiency.

Regarding tool recommendations, both teachers and students identified Generative AI chatbots such as ChatGPT, Google Bard, and Bing AI as useful resources for supporting various aspects of essay writing. From the teachers' perspective, 40% recommended using ChatGPT for idea generation and language revision. These teachers also emphasised the need for cognitive engagement rather than passive copying the output. From the students' perspective, 55% (11 out of 20) recommended Generative AI for brainstorming and feedback purposes. Among these, the nine high-scoring students demonstrated cognitive engagement by crafting specific prompts, comparing multiple outputs, and rewriting the information in their own words. In contrast, the two low-scoring students in this group provided brief responses and showed little evidence of cognitive engagement, as they did not explain how they would evaluate or refine the AI-generated output.

Regarding tools outside the technology category of Generative AI, 40% of teachers recommended non-AI resources such as mind-mapping software and online dictionaries to promote learner autonomy and deeper language understanding. A

smaller group of students also reported using similar non-AI tools, including Google Search (15%), a thesaurus (20%), Cambridge and Oxford dictionaries (10%), and Ozdic (5%). Notably, among the seven students who favoured these traditional resources, six of them mentioned tool usage purposes, such as finding synonyms (Thesaurus) or accessing sample essays (Google Search). However, their responses showed limited explanations related to their trust in the tools or perceptions of their effectiveness. This pattern suggests that these students' usage of traditional tools may stem more from a preference for familiar tools than from cognitive engagement and critical assessment.

To answer the second part of the main research question (RQ3: "How do students' evaluations relate to their writing performance?"), data from students' evaluations and writing scores in Chapter 4 suggest no clear or consistent relationship between students' evaluations of each of the four tools and their writing outcomes. Regarding Sketch Engine, although two high-scoring students appreciated its potential for idea generation and vocabulary development, a majority of students from both high- and low-scoring groups also described its outputs as complicated, irrelevant, or unhelpful. In the case of Google Translate, students' evaluations were largely positive across proficiency levels, with both high- and low-scoring students expressing favourable views. However, negative feedback on translation inaccuracies was only mentioned by high-scoring students, reflecting a more critical assessment of this group. Similarly, among students who gave positive feedback for Quillbot, the majority were high-scoring, and most of those who gave critical feedback about poor output were also high-scoring, suggesting that high-scoring students tended to assess the output of this paraphrasing tool more critically. In the case of Grammarly, students' evaluations and writing performance showed mixed patterns. Both high- and low-scoring students offered positive feedback, with most experiencing improved scores, suggesting that the tool's perceived usefulness extends across proficiency levels. Conversely, negative feedback also came from students of varying proficiency, with some seeing declines in performance while others improved.

Nevertheless, some patterns in interview responses appear to connect students' tool recommendations with their writing outcomes. Among the 11 students who recommended using Generative AI for brainstorming and feedback, nine high-scoring

students demonstrated cognitive engagement by crafting specific prompts, exploring multiple AI-generated outputs, and rewriting the information in their own words, whereas the two low-scoring students in this group provided brief responses and showed little evidence of such engagement, offering no explanation of how they would evaluate or refine the AI-generated content. These findings suggest a connection between students' writing performance and the depth of their interaction with Generative AI. High-scoring students appear not only to value these tools but to use them strategically, applying critical thinking to adapt and integrate AI outputs into their writing. In contrast, low-scoring students seem less able or less inclined to engage with the tools critically, indicating that merely having access to AI support is not enough to improve writing outcomes. Moreover, differences in students' recommendations of traditional writing tools seem connected to their writing performance. A greater number of low-scoring students (5) than high-scoring students (2) reported using non-AI tools, such as online dictionaries or corpus resources; however, their descriptions of how they used these tools were often brief and superficial. In contrast, only two high-scoring students recommended using other traditional tools, and one of them also expressed interest in exploring emerging Generative AI tools.

Overall, findings in Chapter 5, Section 7 indicate that teachers generally viewed the four writing tools more positively and critically than students, especially valuing Sketch Engine for idea planning and language refinement, while students found its interface challenging and less helpful for brainstorming. Google Translate was praised by both groups for quick translations, but teachers were more cautious about inaccuracies and overreliance, concerns noted mainly by high-scoring students. Both Quillbot and Grammarly received positive feedback for language support, though teachers were more aware of their limitations in handling complex language and potential impacts on writing quality. Finally, while both teachers and students recommended Generative AI tools for writing, high-scoring students reported engaging more critically and strategically, whereas lower-scoring students and users of traditional tools showed less critical awareness or explanation of their tool use.

Although students' emotional engagement with the four writing tools is not sufficient to establish a direct link with measurable writing outcomes, there are some indications of a connection between students' writing performance and both their assessment of

the limitations of the four tools as well as their suggestions for engaging with Generative AI. High-scoring students not only recommended this new type of tool but also would use it strategically, demonstrating critical thinking as they adapt and integrate AI-generated content into their writing. In contrast, lower-scoring students appeared to be less equipped or less inclined to engage critically with these tools. Collectively, these findings highlight the more pronounced impact of other factors discussed in Chapter 4, including language proficiency, as well as students' cognitive engagement—actively exploring and using tools—and critical assessment—the ability to evaluate and verify the accuracy of tool outputs—in their writing processes. Additionally, data comparisons show that many students, regardless of proficiency, did not consistently use the four tools during the planning stage, indicating a gap between their intended and actual practices. This gap may reflect the cognitive demands of using these tools and raises concerns that students who already avoid planning could skip this step entirely with Generative AI, risking uncritical adoption of AI content and undermining essential writing skills like idea development, argument structuring, and language revision. Another concern is the limited awareness of plagiarism risks associated with Generative AI, suggesting that, in the context of 2023, both teachers and students may have underestimated how readily AI-generated text could result in unintended plagiarism.

CHAPTER 6: DISCUSSION

Introduction

Chapter 6 brings together the main findings from Chapters 4 and 5 and interprets them in relation to the theoretical framework and prior research on corpus tools, machine translation, paraphrasing tools, automated writing evaluation, and Generative AI. Section 6.1 addresses the three research questions by synthesising quantitative and qualitative results. It shows how the tools influenced scores on the four IELTS writing criteria, how tool use differed between higher- and lower-scoring students, and how internal and external factors shaped both gains and declines, relating these patterns to the Output Hypothesis, the Noticing Hypothesis, and cognitive engagement. Section 6.2 focuses on the comparison of teachers' recommended strategies with students' actual tool use across three stages: planning, composing, revising, and discusses convergences, mismatches, and growing interest in Generative AI. Finally, Section 6.3 presents implications and future directions.

6.1. Discussion of key findings

6.1.1 Findings on students' writing scores and tool usage

To answer RQ1: "What is the impact of the four writing tools on students' IELTS essay writing scores?" and RQ1a: "Does the use of the writing tools result in an increase in the overall writing scores of students? If so, what aspects of IELTS writing assessment (i.e., Task Achievement, Coherence and Cohesion, Lexical Resource, Grammatical Range and Accuracy) are most affected by students' use of the tools?", the quantitative findings (Table 19, Section 4.1.2) at the group level indicate that the intervention involving four writing tools in Essay 2 generally had a positive impact on students' overall writing scores, with paired t-tests revealing a statistically significant improvement ($p = .0026$). Regarding the four IELTS writing criteria, this study found significant improvements in LR and GA, but not in TR and CC. Despite the group-level score improvements, no participant achieved the highest Band 9 score in any criterion. This suggests that the four tools under study still fall short in supporting the sophisticated language use required for achieving native-like writing proficiency, and could not support more complex criteria like TR and CC. Although prior studies have yet to provide similar findings in this level of IELTS as they used other methods for

writing assessment, this similar limitation in achieving the maximum “holistic writing scores” among student participants was noted by Huang and Wilson (2021) with the usage of an AWET called Pigai.

This positive impact at a group level on holistic writing scores aligns with prior studies that have revealed the generally positive impact on English writing outcomes, particularly in vocabulary and grammar using each of the four types of writing tools, including corpus tools (CTs) (Chan & Liou, 2005; Gilmore, 2009; Hu & Deng, 2023; Li, 2017; Wu, 2016; Yoon & Jo, 2014), machine translation tools (MTTs) like Google Translate (Cancino & Panes, 2021; Kol et al., 2018; Lee, 2020; Lee & Briggs, 2021), automated paraphrasing tools (APTs) such as Quillbot (Amyatun & Kholis, 2023; Truong, 2023; Zhao, 2024), and different types of automated writing evaluation tools (AWETs) including Grammarly (Barrot, 2021; Dikli & Bleyle, 2014; Ghuftron & Rosyida, 2018; Grimes & Warschauer, 2010; Huang & Wilson, 2021; Kellogg et al., 2010; Liao, 2016; Link et al., 2014; Palermo & Thomson, 2018; Palermo & Wilson, 2020; Rababah & Talafha, 2024; Sanosi, 2022; Wang et al., 2012; Ware, 2014; Wei et al., 2023). Nevertheless, unlike earlier research reporting consistent writing score increase among all participants after using CTs (Chan & Liou, 2005; Gilmore, 2009; Hu & Deng, 2023; Li, 2017; Sun & Wang, 2003; Wu, 2016), APTs (Truong, 2023), MTTs (Lee, 2020) or AWETs (Dikli & Bleyle, 2014; Ghuftron & Rosyida, 2018; Rababah & Talafha, 2024; Sanosi, 2022), this study revealed variability in post-intervention outcomes at the individual level: with a minority (4 out of 20 students) experiencing score decreases.

Furthermore, Pearson correlation analysis (Table 23) revealed that the usage of Sketch Engine was moderately and positively associated with Essay 2 scores ($r = .529$, $p = .016$), while Google Translate, Quillbot and Grammarly demonstrated minimal correlations with Essay 2 scores. The finding points to Sketch Engine as the tool most consistently associated with better performance, aligning with its role as a corpus-based tool that fosters more authentic language use. By contrast, the other three tools may depend more on how they are used rather than on tool accessibility, which could explain the lack of clear correlations. These findings complement earlier research on the positive impact of corpus tools, with improvements in collocational knowledge, grammar, and lexical accuracy (Boulton, 2009; Chan & Liou, 2005;

Gilmore, 2009; Hu & Deng, 2023; Li, 2017; Wu, 2016; Yoon & Jo, 2014). The positive correlation between Sketch Engine and writing scores supports studies that emphasise the pedagogical value of concordance consultation for developing authentic language use (Crosthwaite et al., 2019; Fang et al., 2021). The mismatch between pedagogic value of corpus tools and their actual use among learners is discussed in Section 6.2.2.

6.1.2 Findings on students' tool engagement, writing score outcomes and language learning

6.1.2.1 Tool engagement and writing score outcomes

To answer RQ1b: “How do high-scoring and low-scoring students differ in their ability to engage with the tools and use them to support their language development?”, using the IELTS Band 6.5 as a benchmark, the researcher classified 20 students into low-scoring (Essay 2 score below 6.5, $n = 7$) and high-scoring (Essay 2 score of 6.5 and above, $n = 13$) groups. This classification was intended to facilitate a comparison of how students utilised the four tools in Essay 2, enabling a deeper understanding of whether specific tool usage patterns were associated with higher or lower writing outcomes. The Pearson correlation analysis (Chapter 4, Section 4.1.3, Table 20) revealed a moderately strong and statistically significant positive relationship between scores on Essay 1 and Essay 2 ($r = .759$, $p < .001$). This result indicates that while the tools contributed to overall score improvements, they did not substantially alter students' relative performance rankings. In other words, students who initially performed at lower levels in Essay 1 largely remained in the low-scoring group, while higher-scoring students maintained their relative standing in Essay 2. Collectively, quantitative findings at the group-level analysis indicate that having access to these four tools does not guarantee improved writing scores. Instead, the effectiveness of such tools depends on how critically and strategically learners engage with them, as well as on learners' existing language proficiency and writing skills. These results resonate with prior empirical studies reviewed in Chapter 2. For example, corpus-based research (e.g., Boulton, 2009; Chan & Liou, 2005; Gilmore, 2009) similarly highlighted gains in lexical and grammatical accuracy through corpus consultation. The moderate correlation between Sketch Engine usage and Essay 2 performance aligns with findings that corpus tools are especially beneficial for high-scoring learners

who can strategically interpret concordance lines (Crosthwaite et al., 2019; Lin, 2021). By contrast, the weak correlations for Google Translate, Quillbot, and Grammarly suggest that while these tools supported some aspects of writing, their effects were less pronounced without cognitive engagement. This nuance is reflected in earlier studies reviewed in Chapter 2, which show that MTTs and APTs often help with fluency and vocabulary breadth (Kol et al., 2018; Tsai, 2019), but may not directly improve accuracy or coherence unless paired with active evaluation and revision. Furthermore, the finding that low-performing students in Essay 1 generally remained in the lower-scoring group, while higher-performing students maintained their relative advantage in Essay 2, diverges from previous studies suggesting that MTTs could help narrow the performance gap between proficiency levels (Chon et al., 2021; Lee, 2021).

Qualitative data from the group-level analysis (Chapter 4, Section 4.3) reveal clear differences between high- and low-scoring students in Essay 2, particularly in the depth of their cognitive engagement with the four tools and the extent of their reflections on language learning, especially vocabulary development and grammatical improvement. The behavioural and cognitive engagement, as coded in this study as Cognitive Engagement (CE) theme, denotes students' active, strategic, and reflective use of the four tools, encompassing behavioural actions (e.g., frequency and purpose of use, cross-checking, revision), as mentioned in Fredricks et al. (2004), Hiver et al. (2021a) and Svalberg (2009). Observations from screen captures during the composing process of Essay 2 and thematic analysis from students' interview responses show that high-scoring students in Essay 2 demonstrated behavioural engagement by exploring multiple tools, experimenting with different features, and, more importantly, using one tool to cross-check or refine outputs of another. For instance, some high-scoring students used Google Translate to check whether Quillbot's paraphrased texts convey their intended meaning. High-scoring students also used Sketch Engine to actively search for collocations or concordances of outputs from Quillbot. By contrast, low-scoring students demonstrated a more superficial degree of behavioural engagement, as they tended to use each tool solely for its main function, such as checking whether a word appeared in a corpus with Sketch Engine, composing by translating entire texts from Vietnamese to English using Google Translate, or modifying texts with Quillbot. Low-scoring students did not actively employ one tool to verify, compare, or refine the outputs generated by another.

The Cognitive Engagement (CE) theme in the context of this study also refers to students' cognitive engagement, or their critical assessment (e.g., noticing gaps, evaluating alternatives, reasoning about form–meaning fit) of the tools' outputs. This theme draws on the Output and Noticing hypotheses, which explain how active production and attention to linguistic form support learning (Schmidt, 1990, 2001; Swain, 1995). High-scoring students, who achieved 6.5 and above, demonstrated more cognitive engagement while using tools. In particular, these students verified tool outputs by cross-referencing multiple platforms, rejected inaccurate suggestions, and showed awareness of tool limitations—behaviours strongly associated with effective language learning strategies in prior studies (Zhao, 2023; Hyland & Hyland, 2019). By contrast, low-scoring students did not consider the contextual appropriateness of the output in their responses and, most notably, did not reject any suggestions from any tools, indicating a lack of cognitive engagement from low-scoring students.

Collectively, these findings on behavioural and cognitive engagement corroborate earlier research on corpus tools, which shows that higher-proficiency learners possess greater corpus literacy and vocabulary knowledge, enabling them to manage the cognitive demands of large datasets, while lower-level learners often experience cognitive overload that limits productive use (Crosthwaite & Steeples, 2022; Daskalovska, 2015; Franken, 2012; Saeedakhtar, 2020). A similar pattern emerges in studies of MTTs, where advanced learners tend to input full sentences and critically evaluate translations, whereas less proficient peers frequently accept outputs (Ahn & Chung, 2020; Cancino & Panes, 2021; Chung & Ahn, 2021; Xu, 2022). Likewise, findings in Chapter 4 align with research on AWETs indicating that while low-proficiency learners benefit from AWE feedback at a surface level but often struggle to interpret its outputs (Chou et al., 2016; Gao, 2021; Saricaoglu & Bilki, 2021; Shang, 2019; Shen et al., 2023; Wei et al., 2023; Xu & Zhang, 2022; Yousofi, 2022), whereas high-proficiency learners were able to evaluate, refine, and strategically apply AWE feedback (Cavaleri & Dianati, 2016; Koltovskaia, 2020; Liao, 2016; Zhang & Hyland, 2018). However, these findings contrast with the conclusions of Chon et al. (2021), who argued that tools like Google Translate can help reduce the performance gaps, helping less experienced writers achieve proficiency levels comparable to those of skilled writers through L1 translation. The results also contrast with Chan and Liou's

(2005) study on corpus tools, which reported that low-proficiency learners achieved significantly greater gains than advanced students.

6.1.2.2 Tool engagement and language learning

Findings from Chapter 4, Section 4.3 reveal that high- and low-scoring students (in relation to Essay 2) differed in how they reflected on their language learning experiences after using the tools. High-scoring students, who demonstrated more instances of cognitive engagement, specifically reported learning new synonyms and collocations through Sketch Engine, Quillbot, and Google Translate. They also became more aware of their grammar and spelling mistakes thanks to Grammarly and used its corrections as learning opportunities. Notably, high-scoring students also demonstrated how they identified a new grammatical structure from Quillbot to help them paraphrase their sentences. In contrast, low-scoring students offered less detailed reflections in their interview responses, with few references to how they learn new vocabulary or improve their grammar after using the four tools. Taken together, these group-level qualitative findings complement the quantitative results, showing that while the four writing tools can enhance specific aspects of writing performance—particularly vocabulary and grammar—access to the tools alone is insufficient, and not all students experienced meaningful learning progress; rather, the tools’ effectiveness on language learning was mediated by students’ behavioural and cognitive engagement. These findings align with research on corpus tools, which showed that higher-proficiency learners display greater corpus literacy and vocabulary knowledge, enabling them to manage the cognitive demands of concordance analysis, while lower-level learners often struggle with surface-level searches and experience cognitive overload (Crosthwaite et al., 2019; Crosthwaite & Steeples, 2022; Daskalovska, 2015; Fang et al., 2021; Franken, 2012; Lin, 2021; Saeedakhtar, 2020). Similarly, findings align with previous empirical research on MTTs, that higher-proficiency students are more likely to input full sentences and critically evaluate outputs, while lower-proficiency learners tend to accept translations uncritically (Ahn & Chung, 2020; Chung & Ahn, 2021; Lee, 2020; Xu, 2022). However, they diverge from studies on AWETs that emphasise the difficulties faced by low-proficiency learners, who often struggle to interpret AWE feedback explanations and therefore fail to fully benefit from the corrections provided (Gao, 2021; Shang, 2019; Yousofi, 2022),

because none of the participants in this study struggle to interpret Grammarly's suggestions.

6.1.2.3 Factors contributing to score decreases

In addressing RQ1c: “What factors contribute to decreases in students’ writing scores when using these tools?”, alongside group-level analysis in writing scores, the individual-level analysis (Table 21) shows that after the tool intervention, 15 of 20 students improved their Essay 2 scores, one showed no change, and four declined. This subset included both high-scoring ($n = 2$) and low-scoring ($n = 2$) students. Findings from four case studies (Chapter 4, Sections 4.4.1 to 4.4.5) highlight behavioural and cognitive engagement as key internal factors shaping outcomes. The two low-scoring students (12 and 17) accepted tool suggestions uncritically, leading to misused vocabulary and phrases—especially from Sketch Engine, Google Translate, and Grammarly—which produced awkward expressions, grammatical errors, and lower scores. The two high-scoring students (1 and 3) engaged more critically, but their essays still contained subtle lexical and coherence issues that Grammarly failed to detect, resulting in slight score decreases. These cases suggest that even proficient learners can be disadvantaged when relying on tool feedback without sufficient cognitive engagement. In addition to these internal factors, one student was affected by a technical issue with QuillBot, and three (3, 12, 17) reported topic unfamiliarity and time pressure. However, these external challenges were less decisive, since other students improved despite similar conditions. In summary, findings from case study analyses reinforce the group-level findings: the effectiveness of writing tools hinges on behavioural and cognitive engagement, and insufficient engagement can lower scores—even for high-scoring learners. These results are consistent with broader research showing that it is learners’ engagement and proficiency, rather than tool availability alone, that determine writing outcomes (Leńko-Szymańska, 2014; Qiu, 2024; Yoon, 2016).

6.2 Comparison of teachers’ and students’ tool usage, evaluations and recommendations

6.2.1 Comparison of teachers’ and students’ tool usage

Building on findings from Chapter 4 that tool availability alone was insufficient for score gains and language learning, the study analysed teachers' lesson plans and interviews to compare their recommended strategies with students' actual tool use (Chapter 5). This comparison aimed to explore how teachers would integrate these four tools to better support writing and language learning, and the extent to which these strategies were reflected among students, to address RQ2: "What similarities and differences exist between teachers' recommended strategies for using the four writing tools in IELTS essay preparation and students' actual usage patterns, and how are these similarities and differences reflected in students' essay writing performance?".

Regarding teachers' usage of the four tools (in addressing RQ2a), the analysis of teachers' lesson plans and interviews from 10 teacher participants revealed a common theme: teachers encouraged cognitive engagement, particularly and critical evaluation of the tools' output, viewing the four tools as language resources to compare with students' writing rather than merely as instruments for instant text generation, paraphrasing, or correction, and rather than focusing solely on improving writing scores. The stage-specific use of the four tools by teachers in this study both reflects and extends insights from prior research. In the planning stage, the strong emphasis on Google Translate for translating Vietnamese ideas into English (90%) resonates with studies highlighting its widespread adoption as a vocabulary scaffold and idea generator in L2 writing (Chandra & Yuyun, 2018; Alhaisoni & Alhaysony, 2017). Similarly, teachers' recommendation of Sketch Engine (70%) for collocational exploration aligns with earlier findings that corpus tools can enhance lexical awareness and support idea generation when introduced at the pre-writing stage (Bunting, 2013; Seidlhofer, 2000; Lin, 2019; Ma et al., 2022a; Naismith, 2016). The limited uptake of Quillbot and Grammarly by a minority of teachers during planning adds more data to the growing body of literature, positioning paraphrasing tools and automated writing evaluation systems as potentially valuable in supporting brainstorming and early drafting.

In the composing stage, teachers' relatively high usage of Quillbot (50%) to formalise style and expand vocabulary mirrors findings that automated paraphrasing can assist in developing sentence variety and academic tone, which mirrors the approach of teachers in Zhao's study (2022). Yet, their minimal recommendations for Google

Translate, Grammarly, and Sketch Engine at this stage diverge from research showing that many learners actively employ these tools for sentence construction, error-checking, and lexical refinement while drafting (Yoon, 2016; Crosthwaite et al., 2019). This discrepancy suggests that while teachers viewed composing as a stage requiring stylistic sophistication, students may rely more heavily on translation and correction tools for scaffolding basic sentence formation.

In the revising stage, teachers' overwhelming preference for Grammarly (90%) reflects prior studies identifying automated writing evaluation tools as effective for raising error awareness and promoting learner autonomy in revision (Li, Link, & Hegelheimer, 2015; Ranalli, 2018). The moderate recommendation of Quillbot (50%) further supports findings that paraphrasing tools can encourage students to critically evaluate alternatives and refine their writing (Dinh & Stracke, 2021). By contrast, the limited reliance on Sketch Engine (20%) and the absence of Google Translate diverge from findings that learners themselves often use corpus tools for checking accuracy or revising drafts (Alm & Watanabe, 2021; Bridle, 2019; Chambers & O'Sullivan, 2004; Gaskell & Cobb, 2004; Mueller & Jacobsen, 2015; O'Sullivan & Chambers, 2006; Quinn, 2015; Todd, 2001; Yoon, 2016) and use MTTs as a source of errors during the revising stage (Anderson, 1995; French, 1991; Kliffer, 2005; Niño, 2004, 2008; Richmond, 1994; Shei, 2002; Somers, 2003). Taken together, these findings confirm earlier research showing that teachers frame corpus tools primarily as language resources for learning rather than shortcuts, emphasising the need for behavioural and cognitive engagement (Bunting, 2013; Burton, 2003; Lin, 2019; Ma et al., 2022a; Naismith, 2016; Seidlhofer, 2001; Somers et al., 2006). At the same time, the results extend the literature by highlighting how teachers allocate distinct pedagogical roles to each tool across different writing stages, a stage-based perspective that is less explicitly developed in previous studies.

Findings for RQ2b show that students used the four tools unevenly across writing stages, in ways that both align with and diverge from prior research. In the planning stage, only a few engaged with Sketch Engine. This reflects earlier research showing a clear divide between teachers and students: teachers often instruct students to use corpus tools during planning and revision (Charles, 2018; Lin, 2019; Ma et al., 2022a; Römer, 2011; Todd, 2001), whereas students mainly employ them for grammar and

vocabulary revision (Bridle, 2019; Yoon, 2008, 2016). Sketch Engine's relatively low use during revision contrasts with prior studies showing that ESL learners frequently consult corpora to resolve lexical and grammatical problems (Bridle, 2019; Lai & Chen, 2015; Yoon, 2008, 2016). For instance, Lai and Chen (2015) found that, after training, 14 students routinely used online corpora and dictionaries in timed writing to check vocabulary usage, collocations, grammar patterns, sentence structure, and spelling when dictionaries were insufficient. Yoon's (2008, 2016) high-proficiency participants similarly integrated corpus tools such as COCA and JustTheWord into drafting and revision to refine lexical choices, prepositions, and collocations, which increased their confidence. Conroy (2010) reported that about 40% of trained students used concordancers and Google while writing to address lexico-grammatical issues and build vocabulary, and Bridle (2019) showed that Chinese learners relied on a corpus to choose appropriate synonyms, with most corpus-based revisions proving successful.

During composing, this study found Google Translate to be dominant, especially for vocabulary and sentence translation, which aligns with prior research identifying MTTs as students' primary aids for lexical decision-making and sentence construction (Chandra & Yuyun, 2018; Farzi, 2016; Tsai, 2019). The moderate use of Quillbot for rephrasing and style refinement also resonates with previous studies showing that automated paraphrasing tools can support lexical variety and academic tone (Truong, 2023; Chen et al., 2013). In revising, Grammarly's widely adoption in this study mirrors earlier findings that AWETs are widely used for error correction, grammar checking, and surface-level editing (Barrot, 2021; Chapelle et al., 2015; Lavolette et al., 2015). The additional use of Quillbot for revision by students aligns with research showing its potential to improve sentence fluency and reduce repetition (Alzubi et al., 2024; Marzuki et al., 2023).

Regarding findings for RQ2c, the similar tool usage approach from teachers was reflected mostly in the group of high-scoring students. In particular, high-scoring students were actively combining tools, such as using Sketch Engine for idea development or vocabulary exploration, using Quillbot to generate synonyms or phrases with their original writing, and then using Google Translate to confirm the meaning of these outputs. Findings in Chapter 5 (Sections 5.1 to 5.5) further reveal

that high-scoring students followed teacher-recommended strategies with Sketch Engine, Google Translate, and Quillbot. They also extended these strategies by refining Google Translate outputs, combining Quillbot and Sketch Engine to improve collocation accuracy, and using back-translation with Google Translate. These results suggest that higher writing performance is linked to flexible and strategic tool use rather than mechanical reliance, which has important implications for lesson design. Low-scoring students, by contrast, did not follow this approach because they used the tools merely as a shortcut for text generation, translation, paraphrasing, and correction, without the ability to critically assess their contextual appropriateness or recognise their own errors in language. Moreover, low-scoring students relied heavily on Google Translate, with some translating entire sentences or essays from Vietnamese to English without behavioural engagement, a practice that teachers explicitly discouraged, and used Quillbot primarily for quick paraphrasing rather than for enhancing academic tone or precision as intended. Rather than revising tool outputs, low-scoring students tended to treat these outputs as final products, which limited their linguistic development and resulted in more frequent errors in grammar, coherence, and context. More importantly, insights from low-scoring students revealed that their approaches to using the tools, which differed from those recommended by teachers, did not contribute to their language learning progress. For instance, findings reveal that one low-scoring student translated an entire essay from Vietnamese into English, a tool usage approach that was not recommended by any teacher, but still achieved higher Essay 2 scores. However, this student showed no evidence of cognitive engagement or assessment and demonstrated no evidence of vocabulary or grammatical improvement in their responses. This finding underscores a subtle but important issue: although some low-scoring students in this study achieved higher writing scores, they still exhibited limited progress in actual language learning due to their lack of behavioural engagement with the four tools and cognitive engagement with the outputs. While this specific gap is new, prior studies on machine translation and paraphrasing tools have raised related concerns, cautioning that uncritical reliance can limit learning benefits and lead to issues such as patchwriting or inappropriate language transfer (Cancino & Panes, 2021; Groves & Mundt, 2015; Rogerson & McCarthy, 2017; Shi, 2004).

Additionally, the comparative analysis of teachers' and students' tool usage also revealed a gap between how teachers advised students to use writing tools and how students put them into practice. While teachers encouraged careful planning and the use of resources like the corpus tool to support idea development and language choices, many students skipped the planning stage altogether. Instead, only two students made use of the four tools for planning, while most of them proceeded directly to writing their essays, potentially missing opportunities to enhance their writing quality and language learning. The mismatches between teachers' and students' tool usage reflect earlier research showing distinct practices: teachers often employ corpus tools during planning and revision to support instruction (Ma et al., 2022a; Todd, 2001), while students typically rely on them during composing and editing to address grammar and vocabulary choices (Bridle, 2019; Yoon, 2016a). These findings align with studies showing that students used MTTs mainly for composing (Chandra & Yuyun, 2018; Dahmash, 2020; Musk, 2022) and revising (Chandra & Yuyun, 2018). However, they contrast with Clifford et al. (2013), who reported MTT use across all stages of writing—pre-writing, composing, and revising. Moreover, this mismatch parallels earlier research on AWETs, where teachers have adopted them primarily for structured, benchmark-oriented tasks, while students engage with them more variably, often at the surface level during composing and revising (Chodorow et al., 2010; Chapelle et al., 2015; Roscoe et al., 2017).

6.2.2 Comparison of teachers' and students' tool evaluations

In relation to RQ3: "How do teachers and students perceive the effectiveness and limitations of the four writing tools" and RQ3a: "What are the strengths and limitations of each tool, as evaluated by teachers and students?", findings indicated that teachers generally expressed more positive yet critical views than students, particularly valuing Sketch Engine for idea planning and lexical development. This aligns with prior research showing that teachers frequently integrate corpus tools into pre-writing stages to enhance collocational awareness and grammatical accuracy (Bunting, 2013; Seidlhofer, 2001; Ma et al., 2022a). However, students in this study were less enthusiastic, citing interface complexity and limited brainstorming value, echoing earlier findings that learners—especially those with lower proficiency—often struggle with interpreting concordance lines and can feel overwhelmed by corpus data

(Crosthwaite et al., 2019; Leńko-Szymańska, 2014; Yoon, 2016a, 2016b). This mismatch reveals a possible disconnect between theory and practice when it comes to language teaching using corpora and corpus tools.

Both teachers and students positively evaluated Google Translate for quick vocabulary and sentence-level translations, but teachers were more cautious, raising concerns about translation inaccuracies and the risk of overreliance, while students largely praised its convenience without recognising such limitations. This difference in perceptions partly reflects earlier studies showing that teachers and educators often express criticisms related to academic integrity and linguistic quality (Burton, 2003; Somers et al., 2006), while also aligning with studies that highlighted more favourable student perceptions of MTTs, particularly regarding vocabulary support and writing efficiency (Alhaisoni & Alhaysony, 2017; Alm & Watanabe, 2021; Chung & Ahn, 2022; Garcia & Pena, 2011; Nguyen, 2022).

Quillbot was positively evaluated by both groups for its usefulness in paraphrasing and sentence variation. This finding resonates with experimental studies that demonstrated Quillbot's effectiveness in improving coherence, grammar, and cohesion in student essays (Amyatun & Kholis, 2023; Truong, 2023). Nevertheless, both teachers and students acknowledged that outputs could be irrelevant or awkward, echoing Rogerson and McCarthy's (2017) observation that paraphrasing tools sometimes generate "word salad" or fail to capture meaning without user intervention.

Grammarly was generally praised by teachers for offering fast corrections, yet they consistently noted its inability to handle complex grammar and its tendency to oversimplify student writing. These criticisms parallel previous studies showing that AWETs can only support surface-level grammar correction and often fail to capture more complicated aspects of writing (McGee, 2006; Kukich, 2000; Li & Casanave, 2012; Prentice & Kinden, 2018). Students, however, expressed fewer concerns than teachers, reflecting earlier reports that learners tend to appreciate AWE tools for their immediacy and clarity of feedback, even if pedagogical depth is lacking (Barrot, 2023; Chen et al., 2013).

6.2.3 Comparison of students' evaluations and writing score outcomes

To answer RQ3b: “How do students' evaluations relate to their writing performance?”, the study examined students' evaluations (emotional engagement) with the four tools to explore possible links with writing outcomes. Results showed no clear connection between students' positive or negative evaluations and their subsequent score changes, as both high- and low-scoring students expressed broadly similar feedback (Section 5.6). Notably, this relationship between tool perceptions and score outcomes has received little to no attention in previous research, as Chapter 2 (Section 2.7) highlighted, with only one recent study by Qiu (2024) reported a lack of connection between students' attitudes toward corpus tools and their usage patterns or writing results, whereas other studies have yet to investigate this relationship. This underscores the originality of the present finding, pointing to a critical gap in the literature: whether and how affective evaluations of digital writing tools translate into measurable writing improvements.

Additionally, some patterns in students' evaluations reaffirm the findings in Chapter 4 (Section 4.1), highlighting that high-scoring students evaluated the tools more critically compared to low-scoring students. For instance, some high-scoring students praised Sketch Engine for planning and idea development, and their evaluations were connected to their improved writing scores, whereas lower-scoring students with decreased scores often found it overwhelming or unhelpful. Additionally, the quantitative results in Table 20 (Section 4.1.3) indicate that Sketch Engine use was associated with improved writing scores, as demonstrated by the Pearson correlation analysis results ($r = .529$, $p = .016$). This aligns with prior corpus-based research showing that concordance consultation can enhance lexical and grammatical accuracy in L2 writing (Boulton, 2009a; Chan & Liou, 2005; Gilmore, 2009). Moreover, several studies emphasise that corpus tools tend to be most beneficial for higher-proficiency learners who can engage critically with large datasets and interpret concordance lines strategically, whereas lower-proficiency learners often struggle with surface-level searches and misinterpretations (Crosthwaite et al., 2019; Fang et al., 2021; Lin, 2021; Pérez-Paredes et al., 2012). Taken together, these findings suggest that while only a few students in this study made extensive use of Sketch Engine, the tool can nevertheless be effective in enhancing writing performance—particularly for high-

proficiency learners with the corpus literacy and cognitive engagement skills needed to exploit its potential. Regarding Google Translate, only high-scoring students in this study expressed negative feedback with Google Translate, particularly noting its contextual inaccuracies and risks of overreliance. Similarly, awareness of limitations in Quillbot and Grammarly was found only among high-scoring students, whereas lower-scoring peers did not report such concerns. This finding aligns with earlier studies suggesting that lower-proficiency learners cannot recognise and critique the shortcomings of MTTs and AWE tools when errors disrupt meaning (Chon et al., 2021; Xu, 2022), but it aligns with findings that higher-proficiency learners are generally more capable of evaluating tool outputs critically and identifying inaccuracies (Ahn & Chung, 2020; Zhang & Hyland, 2018).

6.2.4 Comparison of teachers' and students' tool recommendations

In terms of recommendations for additional tools, both teachers and students in this study reported growing interest in GenAI tools (e.g., ChatGPT, Bing AI). Given that data collection took place in 2023, when classroom uptake of GenAI was still new, teachers emphasised the need for cognitive engagement and cautioned against passive dependence. This position aligns with emerging research that recognises GenAI's usefulness for planning and modelling tasks while also foregrounding risks related to accuracy, misuse, plagiarism, and potential impacts on assessment validity and critical thinking (Cotton et al., 2023; Deng et al., 2025; ElSayary, 2023; Jeon & Lee, 2021; Kim & McGill, 2024; Perkins, 2023; Rukiati et al., 2023; Yang & Chen, 2023). Notably, however, whereas prior studies consistently report strong teacher concerns about plagiarism and academic integrity (Cotton et al., 2023; Derakhshan & Ghiasvand, 2024; Dimitrov, 2023; Mabuan, 2024; Rahma & Fithriani, 2024; Yang & Chen, 2023), neither teachers nor students in this study raised plagiarism as an issue, pointing to a consequential gap in ethical awareness.

Although both high- and low-scoring students reported using ChatGPT, only the high-scoring group demonstrated thoughtful and strategic engagement, while low-scoring students provided limited reflections and offered less detail about how they used the tool. This finding aligns with prior evidence that students tend to view Generative AI more positively than teachers, appreciating its role in idea generation, reducing routine tasks, and correcting language (Adams et al., 2023; Barrett & Pack, 2023; Dai et al.,

2023; Duah & McGivern, 2024; Masoudi, 2024; Yang & Chen, 2023), but it also echoes recent observations that proficiency shapes engagement with ChatGPT—higher-proficiency students distribute their use more strategically across writing stages, while lower-proficiency students engage less critically (Duong & Chen, 2025).

Finally, while teachers and students also mentioned traditional non-AI tools such as mind-mapping software, dictionaries, and corpus tools, students rarely provided detailed justifications for these recommendations. This contrasts with prior research highlighting the value of such tools for developing learner autonomy and deeper language awareness (Boulton, 2009a, 2009b; Chan & Liou, 2005; Gilmore, 2009; Li, 2017; Wu, 2016; Yoon & Jo, 2014).

6.3 Implications

Findings from this study provide implications for ESL/EFL students, teachers, and tool developers interested in integrating the four types of tools examined—corpus tools (Sketch Engine), machine translation tools (Google Translate), automated paraphrasing tools (Quillbot), and automated writing evaluation tools (Grammarly)—as well as emerging Generative AI technologies into writing instruction. For ESL/EFL students, while these four tools can certainly help with exploring the English language, translating, building vocabulary, correcting spelling and grammar errors, they do not automatically lead to higher writing scores or language learning progress (i.e. new vocabulary acquisition and grammar improvement) if used uncritically. The students who achieved the highest scores in this study were those who actively engaged with the tools: they experimented with different features, compared outputs from multiple tools, checked whether the tools' suggestions fit the essay topic and context, and made decisions about what to keep or change in their writing. In contrast, students who simply copied translations, accepted paraphrased sentences and Grammarly's corrections without considering meaning and context, tended to achieve lower scores, demonstrated a lower degree of cognitive engagement, and articulated fewer reflections on their language learning progress. Therefore, students should view these tools not as shortcuts for producing instant essays, but as resources to support their thinking and writing processes, and it is important for students to critically question tool outputs, cross-check information, and reflect on how these tools can help them learn new words or recognise their grammar errors. Moreover, while Generative AI tools can

quickly produce entire essays, relying too heavily on them can prevent students from developing important skills like brainstorming ideas, organising arguments, and learning how to express themselves clearly and accurately in English. Students should instead use such tools to support specific language tasks like finding new vocabulary, checking grammar, or exploring different ways to phrase ideas. By becoming more critical, reflective, and strategic users of technology, students can improve their writing scores and make real progress in their language learning. However, avoiding overreliance on Generative AI tools is challenging for students, as these tools offer multiple functions—such as generating, paraphrasing, and editing—all within a single platform. In many cases, students cannot disable the generative features, making it difficult to use the tools selectively. This shift has also affected previously simpler tools, like Grammarly and Quillbot, which has included integrated AI capabilities that blur the line between support and full content generation since March 2023 (Grammarly, 2023; Routh, 2024).

The gaps between teachers' recommended tool usage strategies and students' actual practices highlight the need for more explicit and targeted instruction on how to integrate digital tools across all writing stages, particularly the planning phase. Additionally, teachers should prioritise explicit instruction and modelling of cognitive engagement strategies for effectively combining these four tools. They should also design practical classroom activities that train students in behavioural and cognitive engagement, including verifying tool outputs, comparing revisions, and articulating the reasons behind their choices. For example, teachers could demonstrate how to use Sketch Engine to explore collocations and develop topic-specific vocabulary, following the approach of high-scoring students, who used concordance lines to generate precise phrases for essay planning. Teachers might also illustrate how to integrate multiple tools strategically, such as employing Quillbot to paraphrase sentences and then checking the contextual appropriateness of new vocabulary in Sketch Engine, a process that high-scoring students used not only to diversify their vocabulary and grammar but also to ensure contextual appropriateness. Findings from Chapter 5 also highlight an additional strategy used by high-scoring students, which was absent from the lesson plans of the ten teacher participants but could be valuable for instruction: employing Google Translate for backward translation—from English into Vietnamese—as a revision step to check coherence and meaning accuracy. To

integrate this practice, teachers could incorporate exercises where students translate their English drafts back into Vietnamese to verify whether their intended meaning is accurately conveyed, mirroring the techniques used by high scorers during revision. Findings also indicate that teachers widely appreciated Grammarly for fast grammar corrections but also acknowledged its shortcomings with complex grammar and its tendency to oversimplify writing—a limitation students noticed less frequently. This underscores the need for teachers to help students use Grammarly critically and to supplement it with instruction on advanced grammar and varied writing styles, especially to meet the specific requirements of IELTS writing. By embedding such examples into teaching, teachers can help students move beyond passive tool use and cultivate the critical thinking skills essential for effective writing and meaningful language learning.

Educators and developers must recognise that technology should support, not replace, the learning process. Consequently, tools should help students stay cognitively engaged, develop metacognitive awareness, and facilitate genuine language learning. However, it should be noted that the four tools examined in this study were initially function-specific, but the integration of Generative AI has increasingly blurred these distinctions. For instance, tools like Quillbot and Grammarly, originally designed for paraphrasing and grammar correction, has also included content generation features since December 2024 (Routh, 2024). To maintain their educational value, developers should remove content-generated feature, and incorporate features that promote reflection and critical thinking such as simplified corpus search options, side-by-side comparisons of original and AI-translated or paraphrased texts, and prompts that encourage learners to critically assess and verify suggestions. These enhancements can help ensure that AI tools remain effective supports for language learning rather than shortcuts for producing texts.

Moreover, to address the gap identified in this study, where students struggled with idea development despite access to large English corpora, teachers and developers should explore ways for AI tools to better support idea generation and planning without overwhelming students or overshadowing their own ideas. A practical approach is to use step-by-step critical-thinking prompts that elicit brief student inputs. Alternatively, developers should help teachers monitor students' writing behaviours, enabling them

to provide more effective feedback on students' writing progress. Together, these insights call for a collaborative effort between pedagogical practice and tool design to create educational environments where writing tools are leveraged to foster both writing scores and deeper language learning.

Beyond practical and cognitive considerations, the study also uncovered a notable ethical gap. Despite clear evidence that some students used Google Translate to translate entire essays (Chapter 4, Section 4.3) and described how they would use Generative AI tools (Chapter 5, Section 5.7.2.1), neither teachers nor students expressed concerns regarding plagiarism or overdependence on Generative AI. Part of this lack of concern may reflect the nature of the English proficiency test-preparation context, where tasks do not require source use or referencing and are completed under timed, controlled conditions, so plagiarism issues are less strictly foregrounded. However, if the ethical implications of Generative AI-assisted writing remain unclear or underemphasised, students might mistakenly believe that such practices are generally permissible. This highlights an urgent need for more explicit ethics education in language teaching and test-preparation contexts, including the IELTS writing test instruction, as the test does not require the use of citations, and has drawn criticisms for not aligning with the academic writing standards expected at the university level (Green, 2007; Moore & Morton, 2005; Read & Hayes, 2003).

Teachers should integrate discussions about academic integrity, authorship, and responsible technology usage into writing instruction, helping students understand that while tools can support learning, blindly submitting tool-generated texts may cross ethical boundaries and, in other contexts, constitute plagiarism. Institutions should consider creating clear guidelines and policies that define acceptable versus unacceptable uses of translation tools and AI writing assistants, tailored specifically to language learning environments where students may assume that using such tools is harmless. Finally, tool developers also have a role to play in promoting ethical use. However, as these four tools increasingly integrate advanced Generative AI capabilities, the boundaries between drafting, paraphrasing, and correcting have become increasingly blurred. This convergence of functions makes it more difficult to isolate specific tool uses and provide ethical guidance. Such complexity underscores the need for collaboration between developers and educators in designing user

experiences that foster transparency, accountability, and informed decision-making. One practical approach could involve integrating built-in prompts, disclaimers, or guidance to remind users to double-check the credibility of information sources after each AI-generated or AI-translated content.

CHAPTER 7: CONCLUSION

7.1. Purpose and research questions

This thesis examined the role of four digital writing tools—Sketch Engine (corpus tool), Google Translate (machine translation tool), Quillbot (automated paraphrasing tool), and Grammarly (automated writing evaluation tool)—in supporting Vietnamese university students' IELTS Writing Task 2 essay writing. Drawing on the Output Hypothesis, the Noticing Hypothesis, and multidimensional concepts of engagement, the study investigated not only the impact of these tools on learners' writing scores and language learning but also how teachers and students integrated them into writing processes, engaged with them, and perceived their value. The study was guided by three main research questions and their sub-questions:

RQ1: What is the impact of the four writing tools on students' IELTS essay writing scores?

RQ1a: Does the use of the writing tools result in an increase in the overall writing scores of students? If so, what aspects of IELTS writing assessment (i.e., Task Achievement, Coherence and Cohesion, Lexical Resource, Grammatical Range and Accuracy) are most affected by students' use of the tools?

RQ1b: How do high-scoring and low-scoring students differ in their ability to engage with the tools and use them to support their language development?

RQ1c: What factors contribute to any observed decrease in students' writing scores when using these tools?

RQ2: What similarities and differences exist between teachers' recommended strategies for using the four writing tools in IELTS essay preparation and students' actual usage patterns, and how are these similarities and differences reflected in students' essay writing performance?

RQ2a: What specific strategies do teachers recommend for using each of the four writing tools in IELTS essay preparation?

RQ2b: What strategies do students employ when using the four writing tools?

RQ2c: Did students whose tool usage shared similarities with teachers' strategies achieve higher Essay 2 scores compared to those whose tool usage differed?

RQ3: How do teachers and students perceive the effectiveness and limitations of the four writing tools

RQ3a: What are the strengths and limitations of each tool, as evaluated by teachers and students?

RQ3b: How do students' evaluations relate to their writing performance?

RQ3c: What additional features do teachers and students expect beyond the tools' current capabilities?

7.2 Summary of key findings

Findings for RQ1 and RQ1a show that the use of the four digital writing tools—Sketch Engine, Google Translate, Quillbot, and Grammarly—generally improved students' IELTS Essay 2 scores, with paired t-tests confirming a significant overall gain ($p = .0026$). Significant improvements were observed in GA, LR, but not in TR and CC (Table 19). However, no student reached the Band 9 level, indicating that the tools alone cannot support native-like proficiency. While group-level results align with prior research highlighting the benefits of corpus, translation, paraphrasing, and evaluation tools, this study revealed greater variability at the individual level, as four out of twenty students recorded score decreases. Correlation analysis further indicated that only Sketch Engine use was moderately and positively associated with higher scores ($r = .529$, $p = .016$), reinforcing its pedagogical value in promoting authentic language use. In contrast, Google Translate, Quillbot, and Grammarly showed minimal correlations with score gains, suggesting that their effectiveness depends more on students' critical and strategic use rather than tool availability.

Findings for RQ1b (Chapter 4) show clear differences between high- and low-scoring students in their cognitive engagement with the four writing tools and the resulting

impact on their IELTS Essay 2 performance. Using Band 6.5 as a benchmark, students were divided into low- (n = 7) and high-scoring groups (n = 13). While overall scores improved, correlation analysis revealed that students largely maintained their relative performance levels from Essay 1 to Essay 2, suggesting that tool access alone did not alter proficiency-based differences. High-scoring students engaged more deeply, both behaviourally and cognitively: they experimented with multiple tools, cross-checked outputs (e.g., verifying Quillbot paraphrases with Google Translate or Sketch Engine), rejected inaccurate suggestions, and showed awareness of tool limitations. In contrast, low-scoring students tended to rely on single tools for basic functions, often accepting outputs uncritically without considering contextual appropriateness. Moreover, high- and low-scoring students differed in how they reflected on language learning after using the four tools. High scorers demonstrated stronger cognitive engagement, reporting gains in vocabulary and collocations through Sketch Engine, Quillbot, and Google Translate, as well as greater awareness of grammar and spelling errors via Grammarly, which they treated as learning opportunities. Some also identified new grammatical structures from Quillbot to improve paraphrasing. In contrast, low scorers gave less detailed reflections and made fewer links between tool use and vocabulary or grammar development.

Findings for RQ1c show that while most ESL students improved their Essay 2 scores after using the tools, four students experienced decreases. Case study analysis (Chapter 4, Sections 4.4.1 to 4.4.4) revealed that these four students accepted tool suggestions uncritically, misused vocabulary or grammar from Sketch Engine, Google Translate, and Grammarly, leading to awkward phrasing and errors. The two high-scoring students (Students 1 and 3), though more critical in their engagement, also saw slight declines when tool feedback failed to detect subtle lexical or coherence issues. External factors such as technical problems, topic unfamiliarity, and time pressure contributed in some cases but were less decisive, as other students improved under similar conditions. Overall, the results highlight that both proficiency and the depth of behavioural and cognitive engagement determine tool effectiveness, and that insufficient cognitive engagement can reduce scores even for high-proficient ESL writers.

Building on findings that tool availability alone does not guarantee writing improvement, the study compared teachers' recommended strategies with students' actual tool use. In relation to findings in RQ2a, teachers generally encouraged critical evaluation of tool outputs, framing them as resources to support idea development, vocabulary, and revision rather than shortcuts for generating text. They also assigned distinct roles to different tools at each stage of writing: Google Translate and Sketch Engine during planning, Quillbot during composing, and Grammarly during revision.

Findings for RQ2b suggested that ESL students used the four tools differently from teachers across three writing stages. In planning, very few engaged with Sketch Engine and none with the other tools. During composing, most relied on Google Translate for vocabulary or full-text translation, with smaller numbers using Quillbot for rephrasing and Sketch Engine for collocations. In revising, Grammarly was used by nearly all students to correct errors, with Quillbot supporting style refinement and Sketch Engine helping check phrase usage. A minority of high-scoring students also used Google Translate for back-translation, despite this not being recommended. Overall, Google Translate and Grammarly dominated usage, while Quillbot and Sketch Engine played more complementary roles in vocabulary enhancement and writing refinement. These findings also reveal a gap between teachers' recommendations and students' practices, as many students, both high-scoring and low-scoring, bypassed the planning stage and underused tools like Sketch Engine for idea development.

Finding for RQ2c suggests that high-scoring students largely mirrored and extended these strategies by combining tools—using Sketch Engine to check Quillbot outputs, refining Google Translate translations, or employing back-translation for coherence—showing that flexible and cognitive engagement supported stronger outcomes. By contrast, low-scoring students tended to use the tools superficially, often relying on Google Translate for whole-sentence or essay translation or using Quillbot for quick paraphrasing, while accepting outputs uncritically and overlooking context or errors. Although some low scorers achieved score gains, their lack of behavioural and cognitive engagement indicates limited language learning in their interview responses.

Findings for RQ3 and RQ3a show both overlap and divergence in how teachers and students evaluated the four writing tools. Teachers held more positive yet critical views, particularly valuing Sketch Engine for lexical development and planning, while

students found it less useful due to interface complexity and difficulty interpreting concordance lines. Both groups perceived Google Translate as beneficial for quick translations, but teachers highlighted risks of inaccuracy and overreliance, whereas students focused mainly on its convenience. Quillbot was well regarded for both paraphrasing and sentence variety, though both acknowledged that outputs could sometimes be awkward or inaccurate. Grammarly received strong praise from teachers for providing fast corrections, but was also criticised for oversimplifying and missing complex grammar, concerns students raised far less often.

Findings for RQ3b show no clear link between students' positive or negative evaluations of the four tools and their writing score changes, as both high- and low-scoring students gave similar feedback. This gap in the literature highlights the originality of the study, since little prior research has examined how emotional engagement with tools translates into measurable performance. However, some patterns emerged: high-scoring students tended to evaluate tools more critically, praising Sketch Engine for planning and idea development, while low-scoring students often found it overwhelming. Quantitative results confirmed that Sketch Engine was significantly associated with score gains, supporting prior research on its value for enhancing lexical and grammatical accuracy, especially among higher-proficiency learners with corpus literacy skills. By contrast, only high scorers noted the limitations of Google Translate, Quillbot, and Grammarly, while low scorers did not recognise such issues. This suggests that awareness of tool limitations is more strongly associated with higher Essay 2 outcomes, who also showed deeper engagement, whereas surface-level reliance among lower scorers limited their ability to identify inaccuracies or make effective use of tool feedback.

Finally, findings for RQ3c show that both teachers and students recommended expanding tool use to include Generative AI chatbots like ChatGPT and BingAI. Teachers acknowledged their potential for planning and modelling but stressed the need for cognitive engagement, warning against overreliance—concerns consistent with earlier research. Teachers and students also suggested traditional tools such as dictionaries, corpus tools, and mind-mapping software, though students gave little justification for these. However, unlike most prior studies, neither teachers nor students raised plagiarism as an issue, highlighting a gap in ethical awareness. Both

high- and low-scoring students reported using ChatGPT, but only high scorers demonstrated critical and strategic engagement, while low scorers used it less critically and provided limited reflections. This pattern suggests that proficiency influences how effectively learners can exploit Generative AI, echoing the findings observed with the four writing tools.

7.3 Contributions

7.3.1 Theoretical contribution

By integrating the Output Hypothesis, the Noticing Hypothesis, and multidimensional engagement frameworks, this study advances theoretical understandings of how digital tools mediate L2 writing. It demonstrates that productive language learning occurs not through tool use alone but through cognitive engagement with tool outputs—particularly when learners reject, refine, and reason about alternatives. This aligns with Swain's (1985, 1995) claim that language learning is promoted through output modification, as students in this study benefited most when they actively evaluated and adjusted tool-generated suggestions rather than accepting them passively. Similarly, the findings reinforce Schmidt's (1990, 2001) Noticing Hypothesis by showing that learners who noticed gaps between their own language and tool outputs were more likely to improve lexical and grammatical accuracy.

The study also contributes to engagement theory by empirically illustrating how behavioural, cognitive, and emotional dimensions integrated in students' digital tool use. High-scoring students displayed behavioural engagement by experimenting with multiple tools, cognitive engagement by evaluating and cross-checking outputs, and emotional engagement through reflective comments about their progress, reflecting Fredricks et al.'s (2004) and Hiver et al.'s (2021a) multidimensional models of engagement. Low-scoring students, in contrast, often lacked these deeper forms of engagement, relying on single tools uncritically and offering little evidence of noticing or output modification. These findings reaffirm that engagement is not merely an outcome of tool use but a mediating factor that shapes whether and how digital tools contribute to language learning.

7.3.2 Empirical contribution

This study makes several empirical contributions to research on digital writing tools and L2 writing. First, it demonstrates that access to these four tools alone does not ensure improved writing outcomes; rather, gains were observed when students engaged behaviourally and cognitively—experimenting with features, comparing outputs, checking context, and reflecting on revisions. This finding provides empirical evidence that a student’s cognitive engagement with tools, rather than passive reliance, is a key determinant of tool effectiveness. Second, the study documents how the integration of Generative AI has blurred distinctions between corpus tools, paraphrasers, grammar checkers, and evaluative systems, offering new data that challenge traditional classifications of digital tools and point to their convergence as multifunctional platforms. Third, results show that tool use can promote not only accuracy and fluency but also metacognitive development, as learners demonstrated decision-making, evaluation, and reflection—empirical patterns that connect technology use with deeper cognitive processing. Finally, the finding that neither students nor teachers expressed concern about plagiarism or overreliance on Generative AI provides novel evidence of an ethical gap in test-preparation contexts, where academic integrity issues may be overlooked. Together, these contributions add new empirical insights into how tool engagement, proficiency, and context shape writing performance, metacognition, and ethical awareness in L2 learning.

7.3.3 Pedagogical contribution

This study shows that digital writing tools support learning only when students engage with them critically and strategically. High scorers explored multiple tools, cross-checked outputs, and reflected on their use, while low scorers relied on direct translations, paraphrasing, or corrections without evaluating context. This highlights the need for teachers to provide explicit instruction on how to question tool outputs, refine suggestions, and combine resources effectively during planning, composing, and revising.

Findings also suggest stage-specific pedagogical uses: Google Translate for idea scaffolding, Sketch Engine for collocations, Quillbot for sentence variety, and Grammarly for revision. High scorers applied these roles flexibly, while low scorers

treated outputs as final products. Teachers should therefore emphasise tool limitations and show learners how to supplement automated feedback with their own judgement. Instruction should also be differentiated by proficiency, with lower-level learners supported through guided searches and structured exercises, and higher-level learners encouraged to develop autonomy and stylistic precision.

Notably, Generative AI tools such as ChatGPT can add further challenges in language teaching. While students welcomed their convenience, only high scorers used them critically, and neither teachers nor students expressed concern about plagiarism. This calls for ethics education in writing instruction, clarifying appropriate and responsible use. As AI functions increasingly merge into existing tools, teachers must also train students to navigate multifunctional platforms thoughtfully, for example, through demonstrations, collaborative output comparisons, and reflective tasks.

Overall, the study contributes pedagogically by emphasising that digital tools should be framed not as shortcuts but as resources for reflection, cognitive engagement, and deeper language learning. Explicit strategy instruction, proficiency-sensitive support, and ethics education are essential to ensure that tools enhance both writing performance and long-term language development.

7.3.4 Methodological contribution

By triangulating test scores, screen recordings, interviews, and lesson plans, the study demonstrates a multi-perspective approach to capturing engagement. This methodological contribution illustrates how behavioural, cognitive, and emotional dimensions can be analysed holistically in tool-mediated writing contexts.

7.4 Limitations of the study

7.4.1 Context limitations

This study was conducted within the specific context of IELTS writing preparation, which imposes unique demands, including strict time constraints, particular task types, and high-stakes assessment pressures. As a result, the findings may not fully generalise to other English writing tasks. Moreover, the research took place in a particular educational and cultural environment, and the insights gained may reflect

local teaching practices, attitudes toward technology, and examination-focused learning strategies that may not be directly applicable in other regions or educational systems. Additionally, the study examined tool usage exclusively within timed writing conditions, which may have limited participants' opportunity to explore certain tools more deeply, particularly those requiring more extensive interaction, such as corpus searches in Sketch Engine. Therefore, while the study provides valuable insights into tool use in an IELTS preparation context, caution should be exercised in extending these conclusions to broader language learning or writing instruction settings.

The study encountered practical challenges related to abrupt tool access issues experienced by individual participants during the writing task. For instance, Student 1 reported that Quillbot became temporarily inaccessible due to server errors, preventing them from paraphrasing certain sentences as planned. Similarly, Student 12 faced difficulties accessing Grammarly, with the tool failing to load properly during their writing session. These unexpected technological disruptions may have influenced the students' writing performance, introducing additional stress, breaking their writing flow, and limiting their ability to compose or refine their essays using the given tools. Consequently, while these disruptions affected only a small number of participants (2 out of 20), they introduced an element of variability into the results and suggest that future studies and teaching practices should account for potential tool instability, possibly by preparing contingency plans or alternative resources to mitigate the impact of sudden technological failures.

7.4.2 Participant limitations

This study has several limitations that should be considered when interpreting the findings. The relatively small sample size of 20 students and 10 teachers limits the generalisability of the results to broader populations or different educational contexts. The participant group was relatively homogeneous in terms of educational context and cultural background, which may limit the applicability of the results to more diverse populations. The research was also conducted specifically within the scope of the IELTS test-preparation context, which may not fully reflect tool usage patterns in other types of writing tasks.

7.4.3 Methodological limitations

While the study combined quantitative data with qualitative insights from interviews and screen captures, some findings rely on self-reported data, which may be affected by memory bias. Moreover, although this study involved comparing students' performance on two essays—one written without tools and another with access to four writing tools—it does not follow a traditional control vs. experimental group design. In a typical control-experimental setup, two separate groups are used: one group (the control) completes the task without intervention. In contrast, a different group (the experimental) completes it with the intervention. In contrast, this study adopts a within-subjects design, where the same participants experience both conditions. This design enables a direct comparison of each student's performance with and without tool support, controlling for individual differences in writing ability. However, it also introduces potential limitations, such as practice effects or differences in topic familiarity. For instance, in Chapter 4, three students (Students 3, 12, and 17) reported being less familiar with the “renewable energy” topic in Essay 2 than with the “distance learning” topic in Essay 1. This suggests that, for these cases, writing score changes may partly reflect varying levels of topic familiarity rather than the impact of tool usage. However, a further limitation relates to the IELTS test itself, which assumes that all candidates can respond effectively to any given topic, regardless of their prior knowledge or familiarity. In this study, some students reported limited familiarity with the “renewable energy” topic in Essay 2, yet still achieved strong performance by leveraging the available tools. This suggests that while topic unfamiliarity can disadvantage some candidates, tool support may mitigate this effect for others, raising questions about the fairness of the test's assumption of equal accessibility across topics.

Another limitation is the lack of investigation into long-term language learning outcomes, as the study primarily measured immediate writing performance and reflections on learning, rather than sustained skill development. Additionally, while the study identified an ethical gap in students' attitudes toward the use of translation tools and Generative AI, it did not explore the underlying causes of this gap. This is largely because the study focused on usage patterns and writing performance within the constraints of the IELTS test format, which does not involve source use or citation.

Finally, participants in this study had access only to the free version of the tools, which typically impose feature restrictions and usage interruptions. Such constraints likely affected how students engaged with the tools and may have attenuated measurable benefits. Therefore, these findings are unlikely to generalise fully to contexts with tool subscription access.

7.5 Future research directions

Building on the limitations identified in this study and recognising the unique constraints of the IELTS preparation context, future research should broaden its scope to investigate writing tool use across more essay topics or writing genres such as academic, professional, and creative writing, thereby allowing insights into how tool use varies across different genres, audiences, and purposes. Further studies should also recruit larger and more diverse participant groups, incorporating learners of different proficiency levels, age ranges, cultural backgrounds, and language experiences to enhance generalisability. Experimental designs featuring control groups are also recommended to isolate the specific impact of tool use on writing outcomes, while longitudinal research could examine the long-term effects of tools on language development and potential dependence. Incorporating neurocognitive methods such as brain imaging or physiological measures like eye-tracking could provide deeper insights into the cognitive processes and attentional patterns involved when learners interact with writing tools, helping to clarify how these tools affect thinking, planning, and decision-making during writing. Notably, this study was conducted in 2023, just before the widespread adoption of more advanced Generative AI tools, offering a valuable opportunity to examine individual writing tools in relative isolation. As these tools continue to evolve, increasingly integrating functions such as translation, paraphrasing, editing, and content generation into single platforms, it becomes more challenging to disentangle the specific effects of each function. Therefore, a key strength of this study lies in its ability to isolate and examine the impact of distinct tool types, as demonstrated by the findings on statistically significant effects of individual tools on students' writing score improvements presented in Chapter 4. The study design also revealed that while the tools had a measurable impact on performance, the changes were not overly drastic; a clear and significant correlation was observed between Essay 1 and Essay 2 written by the same student

($r = .759$, $p < .001$). This suggests that using the four tools separately influenced writing outcomes without completely masking individual ability, an important consideration for teachers seeking to assess genuine student performance. By contrast, if students were using fully integrated Generative AI tools, the gap between essays written with and without such tools would likely be much larger, potentially masking differences between low- and high-proficiency learners and making it more difficult to track individual writing development over time. When building on this study's approach of disentangling the effects of tool functions, future research could adopt longitudinal designs to examine how sustained use of specific, functionally distinct tools (with no content-generated function) over time influences students' writing development, and compare outcomes with those resulting from the use of all-in-one, content-generated AI tools.

7.6 Concluding statement

This thesis demonstrates that the four categories of digital writing tools (corpus tools, machine translation tools, automated paraphrasing tools, and automated writing evaluation tools) can enhance ESL writing performance, but only when learners engage with them critically, strategically, and reflectively. Simply providing access to these tools is insufficient; meaningful improvement depends on the quality of engagement, which is shaped by teacher guidance and learner agency. By positioning the findings within broader debates in applied linguistics and digital education, the study underscores the continuing importance of human cognitive engagement in a context increasingly influenced by Artificial Intelligence. As AI-based tools continue to advance, this research highlights the importance of positioning technology as a support, rather than a substitute for the learner's active role in language development.

REFERENCES

- Abasi, A.R., Akbari, N., & Graves, B. (2006). Discourse appropriation, construction of identities, and the complex issue of plagiarism: ESL students writing in graduate school. *Journal of Second Language Writing*, 15, 102–117.
- Abdel Latif, M. M. M. (2021). Corpus literacy instruction in language teacher education: Investigating Arab EFL student teachers' immediate beliefs and long-term practices. *ReCALL*, 33(1), 34–48. doi:10.1017/S0958344020000129
- Adams, D., Chuah, K.-M., Devadason, E., & Azzis, M. S. A. (2023). From novice to navigator: Students' academic help-seeking behaviour, readiness, and perceived usefulness of ChatGPT in learning. *Education and Information Technologies*, 29, 13617–13634. <https://doi.org/10.1007/s10639-023-12427-8>
- Ahn, S., & Chung, E.S. (2020). Students' perceptions of the use of online machine translation in L2 writing. *Multimedia-Assisted Language Learning*, 23(2), 10–35
- Aiken, M., & Balan, S. (2011). An analysis of Google Translate accuracy. *Translation Journal*, 16(2). Retrieved from <http://www.bokorlang.com/journal/56google.htm>
- Aiken, M., & Wong, Z. (2006). Spanish-to-English translation using the Web. Southwest Decision Sciences Institute. Available from <http://www.swdsi.org/swdsi06/Proceedings06/Papers/IBT04.pdf>
- Alammar, A., & Amin, E.A. (2023). EFL Students' Perception of Using AI Paraphrasing Tools in English Language Research Projects. *Arab World English Journal*, 14(3) 166-181. DOI: <https://dx.doi.org/10.24093/awej/vol14no3.11>
- Alhaisoni, E., & Alhaysony, M. (2017). An Investigation of Saudi EFL University Students' Attitudes towards the Use of Google Translate. *International Journal of English Language Education*, 5(1), 72. <https://doi.org/10.5296/ijele.v5i1.10696>
- Ali, K., & Alireza, F. (2014). The effect of computer-assisted translation on L2 learners' mastery of writing. *International Journal of Research Studies in Language Learning*, 3, 29–44.
- Allen, L. K., Jacovina, M. E., & McNamara, D. S. (2016). Computer-based writing

instruction (Chapter 21). In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (2nd ed., pp. 316-330). Routledge. <https://files.eric.ed.gov/fulltext/ED586512.pdf>

Alm, A., & Watanabe, Y. (2021). Online machine translation for L2 writing across languages and proficiency levels. *Australian Journal of Applied Linguistics*, 5(3), 135–157. <https://doi.org/10.29140/ajal.v5n3.53si3>

Alotabi, A.N., & Alzu'bi, M. (2022). Computer-assisted language learning (CALL): Using paragraph punch software in developing EFL paragraph writing skills. *Journal of Language and Linguistic Studies*, 18(1), 323-332. Doi: 10.52462/jlls.184

Al-Othman, A. A. M. (2024). EFL students develop cognitive and metacognitive self-regulated writing strategies using automated feedback: A case study. *Theory and Practice in Language Studies*, 14(5), 1525–1536. <https://doi.org/10.17507/tpsls.1405.26>

Alzubi, A. F., Mohammad, T., & Khan, S. I. (2024). Evaluating the effectiveness of QuillBot in improving students' paraphrasing skills: Teachers' voices. *Journal of Theoretical and Applied Information Technology*, 102(6), 2556-2567.

Amyatun, R. L., & Kholis, A. (2023). Can Artificial Intelligence (AI) like QuillBot AI Assist Students' Writing Skills? Assisting Learning to Write Texts using AI. *ELE Reviews: English Language Education Reviews*, 3(2), 135–154. <https://doi.org/10.22515/elereviews.v3i2.7533>

Anderson, D. D. (1995). Machine translation as a tool in second language learning. *CALICO Journal*, 13(1), 68–97

Ansorge, L., Ansorgeová, K., & Sixsmith, M. (2021). Plagiarism through Paraphrasing Tools—The Story of One Plagiarized Text. *Publications*, 9(4), 48. <https://doi.org/10.3390/publications9040048>

Bacha, N. (2001). Writing evaluation: what can analytic versus holistic essay scoring tell? *System*, 29(3), 371–383.

Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing*

language assessments and justifying their use in the real world. Oxford University Press.

Bagheri, M. S., & Riasati, M. J. (2016). EFL graduate students' IELTS writing problems and students' and teachers' beliefs and suggestions regarding writing skill improvement. *Journal of Language Teaching and Research*, 7(1), 198-209. <https://doi.org/10.17507/jltr.0701.23>

Bagheri, M. S., & Zare, M. (2009). The role of using IELTS model essays in improving learners' writing and their awareness of writing features. *Journal of English Language Studies*, 1(1), 115–130.

Bahri, H., & Mahadi, T. (2016). Google translate as a supplementary tool for learning Malay: A case study at Universiti Sains Malaysia. *Advances in Language and Literary Studies*, 7(3), 161–167. <http://dx.doi.org/10.7575/aiac.all.v.7n.3p.161>

Bai, L., & Hu, G. (2016). In the face of fallible AWE feedback: how do students respond? *Educational Psychology*, 37(1), 67–81. <https://doi.org/10.1080/01443410.2016.1223275>

Bailey, C., & Withers, J. (2018). What Can Screen Capture Reveal About Students' Use of Software Tools When Undertaking a Paraphrasing Task?. *Journal of Academic Writing*, 8(2), 176–190. <https://doi.org/10.18552/joaw.v8i2.456>

Bailey, D., & Lee, A. R. (2020). An exploratory study of Grammarly in the language learning context: An analysis of test-based, textbook-based, and Facebook corpora. In *Proceedings of EUROCALL 2020 Conference* (pp. 25–31). Research-publishing.net. <https://doi.org/10.14705/rpnet.2020.48.1151>

Bailey, J. (2018, March 8). *A brief history of article spinning*. *Plagiarism Today*. <https://www.plagiarismtoday.com/2018/03/08/a-brief-history-of-article-spinning>

Baisa, V., & Suchomel, V. (2014). SkELL - web interface for English language learning. In *Eighth Workshop on Recent Advances in Slavonic Natural Language Processing* (pp. 63-70). Retrieved from <https://nlp.fi.muni.cz/raslan/2014/12.pdf>

Ball, R. V. (1989). Computer-assisted translation and the modern languages

curriculum. *TheCTISS File*, 8, 52–5

Baralt, M., Gurzynski-Weiss, L., & Kim, Y. (2016). Engagement with language: How examining learners' affective and social engagement explains successful learner-generated attention to form. In Sato, M., & S. Ballinger (Eds.), *Peer interaction and second language learning. Pedagogical potential and research agenda* (pp. 209–240). Amsterdam: John Benjamins.

Barkaoui, K. (2010). Variability in ESL Essay Rating Processes: The Role of the Rating Scale and Rater Experience. *Language Assessment Quarterly*, 7(1), 54–74. <https://doi.org/10.1080/15434300903464418>

Barrett, A., & Pack, A. (2023). Not quite eye to A.I.: Student and teacher perspectives on the use of generative artificial intelligence in the writing process. *International Journal of Educational Technology in Higher Education*, 20(1), 59. <https://doi.org/10.1186/s41239-023-00427-0>

Barrot, J. S. (2021). Using automated written corrective feedback in the writing classrooms: effects on L2 writing accuracy. *Computer Assisted Language Learning*, 36(4), 584–607. <https://doi.org/10.1080/09588221.2021.1936071>

Barrot, J. S. (2022). Integrating Technology into ESL/EFL Writing through Grammarly. *RELC Journal*, 53(3), 764–768. <https://doi.org/10.1177/0033688220966632>

Barrot, J. S. (2023a). Using automated written corrective feedback in the writing classrooms: Effects on L2 writing accuracy. *Computer Assisted Language Learning*, 36(4), 584–607. <https://doi.org/10.1080/09588221.2021.1936071>

Barrot, J. S. (2023b). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, Article 100745. <https://doi.org/10.1016/j.asw.2023.100745>

Barrot, J. S., & 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>

Bauer, B. A. (1981). *A study of the reliabilities and the cost-efficiencies of three methods of assessment for writing ability*. Champaign: University of Illinois.

Becker, A. (2011). Examining rubrics used to measure writing performance in U.S. intensive English programs. *The CATESOL Journal*, 22(1), 113–117.

Benson, P. (2011). *Teaching and researching autonomy in language learning*. Pearson.

Bhagat, R., & Hovy, E. (2013). What is a paraphrase? *Computational Linguistics*, 39(3), 463–472. https://doi.org/10.1162/COLI_a_00166

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Birch, S., & Ladd, G. (1997). The teacher–child relationship and children’s early school adjustment. *Journal of School Psychology*, 35, 61–79.

Bloch, J. (2009). The design of an online concordancing program for teaching about reporting verbs. *Language Learning & Technology*, 13(1), 59–78.

Bostič Bishop, N. (2011). Model essay as a feedback tool in Task 2 of the IELTS writing exam instruction for Slovene students. *ELOPE: English Language Overseas Perspectives and Enquiries*, 8(1), 91–105. <https://doi.org/10.4312/elope.8.1.91-105>

Boulton, A. (2009a). Testing the limits of data-driven learning: Language proficiency and training. *ReCALL*, 21(1), 37–54. <https://doi.org/10.1017/S0958344009000068>

Boulton, A. (2009b). Corpora for all? Learning styles and data-driven learning. In M. Mahlberg, V. González-Díaz, & C. Smith (Eds.), *Proceedings of the 5th Corpus Linguistics Conference*. Lancaster University. <http://ucrel.lancs.ac.uk/publications/cl2009>

Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning*, 60(3), 534–572. <https://doi.org/10.1111/j.1467-9922.2010.00566.x>

Boulton, A. (2012). What data for data-driven learning? *The EUROCALL Review*:

Proceedings of the EUROCALL 2011 Conference, 20, 23–27. Retrieved from <https://files.eric.ed.gov/fulltext/ED544438.pdf>

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(23), 48–393. <https://doi.org/10.1111/lang.12224>

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology*, 25(3), 66–89. <http://hdl.handle.net/10125/73450>

Bowen, N. E. J. A., & Nanni, A. (2021). Piracy, playing the system, or poor policies? Perspectives on plagiarism in Thailand. *Journal of English for Academic Purposes*, 51, 100992. <https://doi.org/10.1016/j.jeap.2021.100992>

Braun, S. (2006) “ELISA: a pedagogically enriched corpus for language learning purposes,” in S. Braun, K. Kohn and J. Mukherjee (eds) *Corpus Technology and Language Pedagogy*, Frankfurt am Main: Peter Lang, pp. 25-47.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp063oa>

Breyer, Y. (2009). Learning and teaching with corpora: Reflections by student teachers. *Computer Assisted Language Learning*, 22(2), 153–172. <https://doi.org/10.1080/09588220902778328>

Bridle, M. A. (2019). Learner use of a corpus as a reference tool in error correction: Factors influencing consultation and success. *Journal of English for Academic Purposes*, 37, 52-69. <https://doi.org/10.1016/j.jeap.2018.11.003>

Briggs, N. (2018). Neural machine translation tools in the language learning classroom: Students’ use, perceptions, and analyses. *The JALT CALL Journal*, 14(1), 3–24. <https://doi.org/10.29140/jaltcall.v14n1.j221>

British Academic Written English Corpus (BAWE) (No date) <https://the.sketchengine.co.uk/open/>

British Council. (2017, September 8). *IELTS numbers rise to three million a year*. British Council. <https://www.britishcouncil.org/contact/press/ielts-numbers-rise-three->

million-year

British Council. (2019). *IELTS grows to 3.5 million a year*. British Council Vietnam. <https://www.britishcouncil.vn/en/about/press/ielts-grows-35-million-year>

British Council. (2023). *IELTS writing band descriptors*. https://takeielts.britishcouncil.org/sites/default/files/ielts_writing_band_descriptors.pdf

British Council. (n.d.). *IELTS scores explained*. Take IELTS. <https://takeielts.britishcouncil.org/teach-ielts/test-information/ielts-scores-explained>

British National Corpus (BNC) (n.d.) <https://www.sketchengine.co.uk/bnc>

Brossell, G. (1986). Current research and unanswered questions in writing assessment. In K. Greenberg, H. Wiener, & R. Donovan (Eds.), *Writing assessment: Issues and strategies* (pp. 168–182). New York, NY: Longman.

Brown, J. D., & Hudson, T. (2002). *Criterion-referenced language testing*. Cambridge applied linguistics series. Cambridge: Cambridge University Press.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). *Language models are few-shot learners*. *arXiv preprint arXiv:2005.14165*. <https://arxiv.org/abs/2005.14165>

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review*, 31(1), 21–32.

Bruner, J. S. (1966). *Toward a Theory of Instruction*. Cambridge, MA: Harvard University Press.

Bunting, J. D. (2013). An investigation of language teachers' explorations of the use of corpus tools in the English for Academic Purposes class. PhD Dissertation. Georgia State University. Retrieved on 5 March 2017 from http://scholarworks.gsu.edu/alesl_diss/26.

Burton, C. (2003). Idea: The Online Translator: Implementing National Standard 4.1. *Hispania*, 86(2), 320–321. <https://doi.org/10.2307/20062865>

Buschek, D., Zürn, M., & Eiband, M. (2021). The impact of multiple parallel phrase suggestions on email input and composition behaviour of native and non-native

English writers. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (CHI '21, Article 732, pp. 1–13). Association for Computing Machinery. <https://doi.org/10.1145/3411764.3445372>

Bygate, M., & Samuda, V. (2009). Creating pressure in task pedagogy: The joint roles of field, purpose, and engagement within the interaction approaches. In Mackey, A., & C. Polio (Eds.), *Multiple perspectives on interaction: Second language research in honor of Susan M. Gass* (pp. 90–116). Abingdon: Routledge.

Callison-Burch, C. & Fournoy., R.S. (2001). A program for automatically selecting the best output from multiple machine translation engines. *Proceedings of MT Summit VIII*, pages 63-66, Retrieved from <https://aclanthology.org/2001.mtsummit-papers.12.pdf>

Cambridge English. (2020). Teacher Guide for Writing C1 Advanced. Retrieved from https://assets.cambridgeenglish.org/schools/CER_6647_V1d_JUL20_Teacher-Guide-for-Writing-C1_Advanced.pdf

Cambridge English. (2023). C2 Proficiency Teacher's Handbook. Retrieved from <https://www.cambridgeenglish.org/Images/168194-c2-proficiency-teachers-handbook.pdf>

Cancino, M., & Panes, J. (2021). The impact of Google Translate on L2 writing quality measures: Evidence from Chilean EFL high school learners. *System*, 98, 102464.

Carr, N. (2008). Is Google Making Us Stupid? *Teachers College Record: The Voice of Scholarship in Education*, 110(14), 89-94. <https://doi.org/10.1177/016146810811001427>

Carter, D., & Inkpen, D. (2012). Searching for poor quality machine translated text: Learning the difference between human writing and machine translations. In L. Kosseim & D. Inkpen (Eds.), *Advances in Artificial Intelligence. Canadian AI 2012. Lecture Notes in Computer Science* (Vol. 7310, pp. 46–57). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-30353-1_5

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), 223–236. Retrieved from <http://www.journal.aall.org.au/index.php/jall/article/view/393/246>

Celce-Murcia, M., & Larsen-Freeman, D. (1983). *The Grammar Book*. Newbury House.

Chambers, A. (2005). Integrating corpus consultation in language studies. *Language Learning & Technology*, 9(2), 111–125. <https://doi.org/10.64152/10125/44022>

Chambers, A. (2007). Popularising corpus consultation by language learners and teachers. In E. Hidalgo, L. Quereda, & J. Santana (Eds.), *Corpora in the foreign language classroom* (pp. 3–16). Rodopi. <https://doi.org/10.1163/9789401203906>

Chambers, A. (2019). Towards the corpus revolution? Bridging the research-practice gap. *Language Teaching*, 52(4), 460–475. <https://doi.org/10.1017/S0261444819000089>

Chambers, A., & Bax, S. (2006). Making CALL work: Towards normalisation. *System*, 34(4), 465–479. <https://doi.org/10.1016/j.system.2006.08.001>

Chambers, A., & O'Sullivan, I. (2004). Corpus consultation and advanced learners' writing skills in French. *ReCALL*, 16(1), 158–172. <https://doi.org/10.1017/S0958344004001211>

Chan, S. T. S., Lo, N. P. K., & Wong, A. M. H. (2024). Enhancing university level English proficiency with generative AI: Empirical insights into automated feedback and learning outcomes. *Contemporary Educational Technology*, 16(4), ep541. <https://doi.org/10.30935/cedtech/15607>

Chan, T.-P., & Liou, H.-C. (2005). Effects of web-based concordancing instruction on EFL students' learning of verb-noun collocations. *Computer Assisted Language Learning*, 18(3), 231–250. <https://doi.org/10.1080/09588220500185769>

Chandra, S. O., & Yuyun, I. (2018). The use of Google Translate in EFL essay writing. *LLT Journal: A Journal on Language and Language Teaching*, 21(2), 228–238

Chang, J.-Y. (2014). The use of general and specialized corpora as reference sources for academic english writing: A case study. *ReCALL*, 26, 243-259. <http://>

[dx.doi.org/10.1017/S0958344014000056](https://doi.org/10.1017/S0958344014000056)

Chang, P., Chen, P.-J., & Lai, L.-L. (2022). Recursive editing with Google Translate: The impact on writing and error correction. *Computer Assisted Language Learning*, 1–26. <https://doi.org/10.1080/09588221.2022.2147192>

Chang, W.-L., & Sun, Y.-C. (2009). Scaffolding and web concordancers as support for language learning. *Computer Assisted Language Learning*, 22, 283-302. <http://dx.doi.org/10.1080/09588220903184518>.

Chapelle, C. A., & Sauro, S. (2019). *The handbook of technology and second language teaching and learning*. Hoboken, NJ: John Wiley & Sons.

Chapelle, C. A., Cotos, E., & Lee, J. (2015). Validity arguments for diagnostic assessment using automated writing evaluation. *Language Testing*, 32(3), 385-405. <https://doi.org/10.1177/0265532214565386>

Charge, N. & Taylor, L.B. (1997). Recent Developments in IELTS. *ELT Journal*, 51(4), 374. Retrieved October 30, 2025 from <https://www.learntechlib.org/p/84323/>.

Charles, M. (2012). 'Proper vocabulary and juicy collocations': EAP students evaluate do-it-yourself corpus-building. *English for Specific Purposes*, 31(2), 93–102. <https://doi.org/10.1016/j.esp.2011.12.003>

Charles, M. (2014). Getting the corpus habit: EAP students' long-term use of personal corpora. *English for Specific Purposes*, 35, 30–40. <https://doi.org/10.1016/j.esp.2013.11.004>

Charles, M. (2018). Corpus Tools for Writing Students. In *The TESOL Encyclopedia of English Language Teaching* (pp. 1–7). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118784235.eelt0554>

Charles, M., & Hadley, G. (2022). Autonomous corpus use by graduate students: A long-term trend study (2009–2017). *Journal of English for Academic Purposes*, 56, 101095. <https://doi.org/10.1016/j.jeap.2022.101095>

Chastain, K. (1990). Characteristics of graded and ungraded compositions. *Modern Language Journal*, 11, 367-383

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 & Technology*, 12(2), 94–112.

Chen, H. J. H. (2011). Developing and evaluating a web-based collocation retrieval tool for EFL students and teachers. *Computer Assisted Language Learning*, 24(1), 59–76. <https://doi.org/10.1080/09588221.2010.526945>

Chen, M. H., Huang, S. T., Chang, J. S., & Liou, H. C. (2013). Developing a corpus-based paraphrase tool to improve EFL learners' writing skills. *Computer Assisted Language Learning*, 28(1), 22–40. <https://doi.org/10.1080/09588221.2013.783873>

Chen, M., & Flowerdew, J. (2018). A critical review of research and practice in data-driven learning (DDL) in the academic writing classroom. *International Journal of Corpus Linguistics*, 23(3), 335–369. <https://doi.org/10.1075/ijcl.16130.che>

Chen, M., Tang, Q., Wiseman, S., & Gimpel, K. (2019). Controllable paraphrase generation with a syntactic exemplar. In A. Korhonen, D. Traum, & L. Màrquez (Eds.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 5972–5984). Association for Computational Linguistics. <https://doi.org/10.18653/v1/P19-1599>

Cheng, W., Warren, M., & Xu, X.-F. (2003). The language learner as language researcher: Putting corpus linguistics on the timetable. *System*, 31(2), 173–186

Chenoweth, N. A., & Hayes, J. R. (2003). The inner voice in writing. *Written Communication*, 20(1), 99–118. <https://doi.org/10.1177/0741088303253572>

Chodorow, M., Gamon, M., & Tetreault, J. (2010). The utility of article and preposition error correction systems for English language learners: Feedback and assessment. *Language Testing*, 27(3), 419–436. <https://doi.org/10.1177/0265532210364391>

Chon, Y. V., Shin, D., & Kim, G. E. (2021). Comparing L2 learners' writing against parallel machine-translated texts: Raters' assessment, linguistic complexity and errors. *System*, 96, 102408. <https://doi.org/10.1016/j.system.2020.102408>

Chou, H.-C. C., Moslehpour, M., & Yang, C.-Y. (2016). My Access and writing error

corrections of EFL college pre- intermediate students. *International Journal of Education*, 8(1), 144–161. <https://doi.org/10.5296/ije.v8i1.9209>

Christenson, S.L., Reschly, A.L., & Wylie, C. (Eds.) (2012). *Handbook of research on student engagement*. New York: Springer

Christoph A. Hafner, Wing Yee Jenifer Ho, (2020). Assessing digital multimodal composing in second language writing: Towards a process-based model, *Journal of Second Language Writing*, Volume 47, 2020,100710, ISSN 1060-3743, <https://doi.org/10.1016/j.jslw.2020.100710>.

Chung, E. S. (2020). The effect of L2 proficiency on post-editing machine translated texts. *Journal of Asia TEFL*, 17(1), 182-193. Retrieved from <https://www.proquest.com/scholarly-journals/effect-l2-proficiency-on-post-editing-machine/docview/2394755031/se-2>

Chung, E. S., & Ahn, S. (2022). The effect of using machine translation on linguistic features in L2 writing across proficiency levels and text genres. *Computer Assisted Language Learning*, 35(9), 2239–2264. <https://doi.org/10.1080/09588221.2020.1871029>

Chung, G. K. W. K., & O'Neil, H. F., Jr. (1997, December). *Methodological approaches to online scoring of essays* (CSE Technical Report No. 461). Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, University of California, Los Angeles. Retrieved from <https://files.eric.ed.gov/fulltext/ED418101.pdf>

Clark, T., East, M., Elder, C., & Shaw, S. (2023). *IELTS writing scales: Review and update—Summary overview*. British Council. Retrieved from <https://s3.eu-west-2.amazonaws.com/ielts-web-static/production/Research/ielts-writing-scales-review-and-update-summary-overview-clark-et-al-2023.pdf>

Clifford, J., Merschel, L., & Munné, J. (2013). Surveying the landscape: What is the role of machine translation in language learning? *The Acquisition of Second Languages and Innovative Pedagogies*, 10, 108–121. doi: 10.7203

Cobb, T. (1997). Is there any measurable learning from hands-on concordancing?

System, 25(3), 301–315. [https://doi.org/10.1016/S0346-251X\(97\)00024-9](https://doi.org/10.1016/S0346-251X(97)00024-9)

Cobb, T. (1999). Breadth and depth of lexical acquisition with hands-on concordancing. *Computer Assisted Language Learning*, 12(4), 345–360. <https://doi.org/10.1076/call.12.4.345.5699>

Cobb, T. (2007) Computing the vocabulary demands of L2 reading. *Language Learning & Technology*, 11(3): 38–63.

Cohen, A. D. (1986). Mentalistic measures in reading strategy research: Some recent findings. *English for Specific Purposes*, 5, 131–145.

Cohen, A. D. (1994). *Assessing language ability in the classroom*. Boston: Heinle & Heinle.

Cohen, A. D., & Cavalcanti, M. C. (1987). Viewing feedback on compositions from the teacher's and the student's perspective. *ESpecialist*, 16, 13–28.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2017a). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Routledge.

Cohen, L., Manion, L., & Morrison, K. (2017b). Validity and reliability. In *Research methods in education* (8th ed., pp. [chapter 14]). Routledge. <https://doi.org/10.4324/9781315456539-14>

Colby, R. (2014). Writing and assessing procedural rhetoric in student-produced video games. *Computers and Composition*, 31, 43–52. <https://doi.org/10.1016/j.compcom.2013.12.003>

Coleman, D., & Strafield, S., & Hagan, A. (2003). The attitudes of IELTS stakeholders: Student and staff apperception of IELTS in Australian, UK, and Chinese tertiary institutions. *IELTS Australia Research*, 5, 20-34

Collins, A. M., & Loftus, A. E. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82, 407–428.

Connell, J. P., & Wellborn, J. G. (1991). Competence, autonomy, and relatedness: A motivational analysis of self-system processes. In M. Gunnar & L. A. Sroufe (Eds.),

Minnesota Symposium on Child Psychology (Vol. 23). Chicago: University of Chicago Press.

Conroy, M. A. (2010). Internet tools for language learning: University students taking control of their writing. *Australasian Journal of Educational Technology*, 26(6). <https://doi.org/10.14742/ajet.1047>

Coombe, C., Davidson, P., O'Sullivan, B., & Stoyhoff, S. (2012). *The Cambridge guide to second language assessment*. New York: Cambridge University Press.

Cooper, T. (2014). Can IELTS writing scores predict university performance? Comparing the use of lexical bundles in IELTS writing tests and first-year academic writing. *Stellenbosch Papers in Linguistics Plus*, 42(0), 63. <https://doi.org/10.5842/42-0-155>

Correa, M. (2011). Academic dishonesty in the second language classroom: Instructors' perspectives. *Modern Journal of Language Teaching Methods*, 1, 65–79.

Correa, M. (2014). Leaving the “peer” out of peer-editing: Online translators as a pedagogical tool in the Spanish as a second language classroom. *Latin American Journal of Content and Language Integrated Learning*, 7(1), 1–20. <https://doi.org/10.5294/laclil.2014.7.1.1>

Cotos, E., Link, S., & Huffman, S. (2017). Effects of DDL technology on genre learning. *Language Learning and Technology*, 21(3), 104–130. <http://llt.msu.edu/issues/october2017/cotoslinkhuffman.pdf>.

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 1–13.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press. http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf.

Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*. Cambridge

University Press.

Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*. Cambridge University Press.

Crossley, S. A., Roscoe, R., & McNamara, D. S. (2014). What is successful writing? An investigation into the multiple ways writers can write successful essays. *Written Communication*, 31(2), 184–214

Crosthwaite, P., & Steeples, B. (2022). Data-driven learning with younger learners: Exploring corpus-assisted development of the passive voice for science writing with female secondary school students. *Computer Assisted Language Learning*, 37(7–8), 1663–1687. <https://doi.org/10.1080/09588221.2022.2068615>

Crosthwaite, P., Wong, L. L. C., & Cheung, J. (2019). Characterising postgraduate students' corpus query and usage patterns for disciplinary data-driven learning. *ReCALL*, 31(3), 255–275. doi:10.1017/S0958344019000077

Cumming, A. (1990). Metalinguistic and ideational thinking in second language composing. *Written Communication*, 7(4), 482–511. <https://doi.org/10.1177/07410883900007004003>

Cumming, A., Kantor, R., & Powers, D. E. (2001). Scoring TOEFL essays and TOEFL 2000 prototype writing tasks: An investigation into raters' decision making and development of a preliminary analytic framework. Princeton, NJ. Retrieved from <https://www.ets.org/Media/Research/pdf/RM-01-04.pdf>.

Cumming, A., Kantor, R., & Powers, D. E. (2002). Decision making while rating ESL/EFL writing tasks: A descriptive framework. *The Modern Language Journal*, 86(1), 67–96. <https://doi.org/10.1111/1540-4781.00137>.

Cummings, A. (1998). Theoretical perspectives on writing. *Annual Review of Applied Linguistics*, 18, 61–78.

Dahmash, N. (2020). 'I Can't Live Without Google Translate': A Close Look at the Use

of Google Translate App by Second Language Learners in Saudi Arabia. *Arab World English Journal*, 11 (3) 226 -240.

Dai, Y., Lai, S., Lim, C. P., & Liu, A. (2023). ChatGPT and its impact on research supervision: Insights from Australian postgraduate research students. *Australasian Journal of Educational Technology*, 39(4), 74–88. <https://doi.org/10.14742/ajet.8843>

Dang, C. N., & Dang, T. N. Y. (2021). The predictive validity of the IELTS test and contribution of IELTS preparation courses to international students' subsequent academic study: Insights from Vietnamese international students in the UK. *RELC Journal*, 54(1), 84–98. <https://doi.org/10.1177/0033688220985533>

Dao, P., Nguyen, M. X. N. C., & Iwashita, N. (2019). Teachers' perceptions of learner engagement in L2 classroom task-based interaction. *The Language Learning Journal*, 49(6), 711–724. <https://doi.org/10.1080/09571736.2019.1666908>

Darwin, Rusdin, D., Mukminatien, N., Suryati, N., Laksmi, E. D., & Marzuki. (2024). Critical thinking in the AI era: An exploration of EFL students' perceptions, benefits, and limitations. *Cogent Education*, 11(1). <https://doi.org/10.1080/2331186X.2023.2290342>

Daskalovska, N. (2013). Corpus-based versus traditional learning of collocations. *Computer Assisted Language Learning*, 28(2), 130–144. <https://doi.org/10.1080/09588221.2013.803982>

Davies, M. (2008). Corpus of Contemporary American English. Retrieved from <http://corpus.byu.edu/coca>

Dawson, P. (2017). Assessment rubrics: Towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education*, 42(3), 347–360. <https://doi.org/10.1080/02602938.2015.1111294>

Dembsey, J. M. (2017). Closing the Grammarly gaps: A study of claims and feedback from an online grammar program. *The Writing Center Journal*, 36(1), 63–100. <https://www.onlinewritingcenters.org/wp-content/uploads/2019/01/2017-Dembsey-WCJ-36.1.pdf>

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers and Education*, 227. <https://doi.org/10.1016/j.compedu.2024.105224>

Derakhshan, A., & Ghiasvand, F. (2024). *Is ChatGPT an evil or an angel for second language education and research? A phenomenographic study of research-active EFL teachers' perceptions*. *International Journal of Applied Linguistics*, 34, 1246–1264. <https://doi.org/10.1111/ijal.12561>

Diab, R., & Balaa, L. (2011). Developing detailed rubrics for assessing critique writing: impact on EFL university students' performance and attitudes. *TESOL Journal*, 2(1), 52–72.

Diederich, P. B. (1974). *Measuring growth in English*. Urbana, IL: National Council of Teachers of English.

Diederich, P. B., French, J. W., & Carlton, S. T. (1961). Factors in judgments of writing ability (Research Bulletin No. RB-61-15). *Princeton: Educational Testing Service* <https://doi.org/10.1002/j.2333-8504.1961.tb00286.x>.

Dikli, S., & Bleyle, S. J. (2014). Automated writing evaluation: Writing into the future. *Assessing Writing*, 21, 18-37. <https://doi.org/10.1016/j.asw.2014.03.004>

Dinneen, C. (2021). Students' use of digital translation and paraphrasing tools in written assignments on Direct Entry English Programs. *English Australia Journal*, 37(1), 40–51. Retrieved from <https://files.eric.ed.gov/fulltext/EJ1341751.pdf>

Doherty, S., & Kenny, D. (2014). The design and evaluation of a Statistical Machine Translation syllabus for translation students. *The Interpreter and Translator Trainer*, 8(2), 295–315. doi:10.1080/1750399X.2014.937571

DOI: <https://dx.doi.org/10.24093/awej/vol11no3.14>

Dörnyei, Z., & Kormos, J. (2000). The role of individual and social variables in oral task performance. *Language Teaching Research*, 4(3), 275-300. <https://doi.org/10.1177/136216880000400305>

Dörnyei, Z., & Kormos, J. (2000). The role of individual and social variables in oral task

performance. *Language Teaching Research*, 4, 275–300.

Doyle-Jones, C. (2019). “Technology Brings Learning to Life”: Planning Writing Opportunities through Multimodal and Digital Resources. *Language and Literacy*, 21(2), 52–68. <https://doi.org/10.20360/langandlit29468>

Drechsel, J. (1999). Writing into silence: Losing voice with writing assessment technology. *Teaching English in the Two-Year College*, 26, 380–387

Duah, J. E., & McGivern, P. (2024). How generative artificial intelligence has blurred notions of authorial identity and academic norms in higher education, necessitating clear university usage policies. *International Journal of Information and Learning Technology*, 41(2), 180–193. <https://doi.org/10.1108/IJILT-11-2023-0213>.

Ducar, C., & Schocket, D. H. (2018). Machine translation and the L2 classroom: Pedagogical solutions for making peace with Google Translate. *Foreign Language Annals*, 51(4), 779–795. <https://doi.org/10.1111/flan.12366>

Duong, T. N. A., & Chen, H. L. (2025). An AI Chatbot for EFL Writing: Students’ Usage Tendencies, Writing Performance, and Perceptions. *Journal of Educational Computing Research*, 63(2), 406–430. <https://doi.org/10.1177/07356331241312363>

Durrant, P. (2014). Corpus frequency and second language learners' knowledge of collocations. *International Journal of Corpus Linguistics*, 19(4), 443-477.

Dwivedi, Y., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., ... Wright, R. (2023). “So what if ChatGPT wrote it?”: Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Technology and Management*, 71(102642), Article 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>

Ebadi, S., Gholami, M., & Vakili, S. (2022). Investigating the Effects of Using Grammarly in EFL Writing: The Case of Articles. *Computers in the Schools*, 40(1), 85–105. <https://doi.org/10.1080/07380569.2022.2150067>

Ebrahimi, A., & Faghih, E. (2017). Integrating corpus linguistics into online language teacher education programs. *ReCALL*, 29(1), 120–135. doi:10.1017/S0958344016000070

Eckes, T. (2008). Rater types in writing performance assessments: A classification approach to rater variability. *Language Testing*, 25(2), 155-185. <https://doi.org/10.1177/0265532207086780>

Educational Testing Service. (n.d.). GRE General Test: Prepare for the Test (Analytical Writing) Scoring. Retrieved from <https://www.ets.org/gre/test-takers/general-test/prepare/content/analytical-writing/scoring.html#accordion-06cf390d3c-item-0e9cac930c>

Educational Testing Service. (n.d.). TOEFL iBT Writing Rubrics. Retrieved from <https://www.ets.org/pdfs/toefl/toefl-ibt-writing-rubrics.pdf>

Elola, Idoia & Oskoz, Ana. (2017). Writing with 21st century social tools in the L2 classroom: New literacies, genres, and writing practices. *Journal of Second Language Writing*. 36. 10.1016/j.jslw.2017.04.002.

ElSayary, A. (2023). An investigation of teachers' perceptions of using ChatGPT as a supporting tool for teaching and learning in the digital era. *Journal of Computer Assisted Learning*, 40(3), 931–945. <https://doi.org/10.1111/jcal.12926>

Erdosy, M. U. (2004). *Exploring variability in judging writing ability in second language: A study of four experienced raters of ESL compositions* (TOEFL Research Report RR 03-17). Princeton: Educational Testing Service. Retrieved from <https://onlinelibrary.wiley.com/doi/pdfdirect/10.1002/j.2333-8504.2003.tb01909.x>

Erstad, O. (2011). Citizens navigating in literate worlds: The case of digital literacy. In Michael Thomas (Ed.), *Deconstructing digital natives. Young people, technology and exchange*. *Language Awareness* 20(4). 343–357.

Fahim, M., & Bijani, H. (2011). The effect of rater training on raters' severity and bias in second language writing assessment. *Iranian Journal of Language Testing*, 1(1), 1-16.

Fan, N. (2023). Exploring the Effects of Automated Written Corrective Feedback on EFL Students' Writing Quality: A Mixed-Methods Study. *SAGE Open*, 13(2). <https://doi.org/10.1177/21582440231181296>

Fang, L., Ma, Q. & Yan, J. (2021). The effectiveness of corpus-based training on collocation use in L2 writing for Chinese senior secondary school students. *Journal of China Computer-Assisted Language Learning*, 1(1), 80-109. <https://doi.org/10.1515/jccall-2021-2004>

Farr, F. (2008). Evaluating the use of corpus-based instruction in a language teacher education context: Perspectives from the users. *Language Awareness*, 17(1), 25-43.

Farzi, R. (2016). *Taming translation technology for L2 writing: Documenting the use of free online translation tools by ESL students in a writing course* (Doctoral dissertation, University of Ottawa). University of Ottawa. <https://doi.org/10.20381/ruor-5181>

Fauziah, H., Basthomi, Y., & Yannuar, N. (2024). Concordancers vs. other tools: Comparing their roles in students' English language retention. *TESL-EJ*, 29(1), Article EJ113A2. <https://tesl-ej.org/wordpress/issues/volume29/ej113/ej113a2/>

Ferris, D.R., & Hedgcock, J.S. (2005). *Teaching ESL composition: Purpose, process, and practice* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Sage Publications.

Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research*, 59, 117–142.

Finn, J. D., Pannozzo, G. M., & Voelkl, K. E. (1995). Disruptive and inattentive/withdrawn behavior and achievement among fourth graders. *Elementary School Journal*, 95, 421–454.

Firat, M. (2023). What ChatGPT means for universities: Perceptions of scholars and students. *Journal of Applied Learning and Teaching*, 6(1). <https://doi.org/10.37074/jalt.2023.6.1.22>

Fitria, T. N. (2021). QuillBot as an online tool: Students' alternative in paraphrasing and rewriting of English writing. *Englisia Journal of Language Education and*

Humanities, 9(1), 183-196. <https://doi.org/10.22373/ej.v9i1.10233>

Floridi, L., Chiriatti, M. (2020). Its Nature, Scope, Limits, and Consequences. *Minds & Machines*, 30, 681–694. <https://doi.org/10.1007/s11023-020-09548-1>

Flower, L. (1989). Cognition, context and theory building. *College Composition and Communication*, 40, 282–311.

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, 32(4), 365–387. <https://doi.org/10.2307/356600>

Flowerdew, J. (2014). Corpus-based approaches to language description for specialized academic writing. *Language Teaching*, 50(1), 90–106. <https://doi.org/10.1017/S0261444814000378>

Flowerdew, J., & Miller, L. (1995). On the notion of culture in L2 lectures. *TESOL quarterly*, 29(2), 345-373.

Flowerdew, L. (2015a). Data-driven learning and language learning theories: Whither the twain shall meet. In: A. Lenko-Szymanska & A. Boulton (Eds.), pp. 15-36.

Flowerdew, L. (2015b). Using corpus-based research and online academic corpora to inform writing of the discussion section of a thesis. *Journal of English for Academic Purposes*, 20, 58–68. <https://doi.org/10.1016/j.jeap.2015.06.001>

Fountain, A., & Fountain, C. (2009). A new look at translation: Teaching tools for language and literature. In *Empowerment through collaboration: Dimension* (pp. 1–15). Atlanta, GA: Southern Conference on Language Teaching

Franken, M. (2012). The nature and scope of student search strategies in using a web derived corpus for writing. *The Language Learning Journal*, 42(1), 85–102. <https://doi.org/10.1080/09571736.2012.678013>

Frankenberg-Garcia, A. (2005). Pedagogical uses of monolingual and parallel concordances. *ELT Journal*, 59(3), 189-198.

Frankenberg-Garcia, A. (2012). Raising teachers' awareness of corpora. *Language Teaching*, 45, 475-489.

Frankenberg-Garcia, A., Lew, R., Roberts, J., Rees, G., & Sharma, N. (2019). Developing a writing assistant to help EAP writers with collocations in real time. *ReCALL*, 31(1), 23-39. doi:10.1017/S0958344018000150

Fredholm, K. (2014). Effects of online translation on morphosyntactic and lexicalpragmatic accuracy in essay writing in Spanish as a foreign language. *CALL Design: Principles and Practice - Proceedings of the 2014 EUROCALL Conference, Groningen, The Netherlands*, 96–101. <http://doi.org/10.14705/rpnet.2014.000201>

Fredholm, K. (2015). "Online Translation Use in Spanish as a Foreign Language Essay Writing: Effects on Fluency, Complexity and Accuracy." *Revista Nebrija de Lingüística Aplicada* 18.

Fredholm, K. (2019). "Effects of Google Translate on Lexical Diversity: Vocabulary Development among Learners of Spanish as a Foreign Language." *Revista Nebrija de Lingüística Aplicada a La Enseñanza de Lenguas* 13 (26): 98–117

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>

Fredricks, J.A., McColskey, W. (2012). The Measurement of Student Engagement: A Comparative Analysis of Various Methods and Student Self-report Instruments. In: Christenson, S., Reschly, A., Wylie, C. (eds) *Handbook of Research on Student Engagement*. Springer, Boston, MA. https://doi.org/10.1007/978-1-4614-2018-7_37

French, R. J. (1991). Machine translation. In Brierley, W. & Kemble, I. R. (Eds.), *Computers as a tool in language learning* (pp. 55–70). Ellis Horwood Limited

Frodesen, J. (2007). Linguistic challenges of summary and paraphrase. Paper presented at the 2007 Symposium on Academic Language Development. San Francisco, CA.

Gain, A., Rao, M., & Bhat, K. S. (2019). Usage of Grammarly – Online grammar and spelling checker tool at the Health Sciences Library, Manipal Academy of Higher Education, Manipal: A study. *Library Philosophy and Practice*, Article 2611. <https://digitalcommons.unl.edu/libphilprac/2611/>

Gao, J. (2021). Exploring the feedback quality of an automated writing evaluation system Pigai. *International Journal of Emerging Technologies in Learning (IJET)*, 16(11), 322. <https://doi.org/10.3991/ijet.v16i11.19657>

Garcia, I., & Pena, M. I. (2011). Machine translation-assisted language learning: writing for beginners. *Computer Assisted Language Learning*, 24(5), 471–487. <https://doi.org/10.1080/09588221.2011.582687>

Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System*, 32(3), 301–319. <https://doi.org/10.1016/j.system.2004.04.001>

Ghalib, T. K., & Hattami, A. A. (2015). Holistic versus analytic evaluation of EFL writing: a case study. *English Language Teaching*, 8(7), 225–236.

Ghasemi, H., & Hashemian, M. (2016). A comparative study of Google Translate translations: An error analysis of English- to- Persian and Persian- to- English translations. *English Language Teaching*, 9(3), 13-17. <https://doi.org/10.5539/elt.v9n3p13>

Ghufron, M. A., & Rosyida, F. (2018). The role of grammarly in assessing English as a Foreign Language (EFL) writing. *Lingua Cultura*, 12(4), 395-403. <https://doi.org/10.21512/lc.v12i4.4582>

Giessler, R. (2024). EFL writers' cognitive engagement with AWE feedback. *Language Awareness*, 33(2), 428–445. <https://doi.org/10.1080/09658416.2023.2269088>

Gilmore, A. (2009). Using online corpora to develop students' writing skills. *ELT Journal*, 63(4), 363–372. <https://doi.org/10.1093/elt/ccp010>

Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 359–370). Routledge. <https://doi.org/10.4324/9780203856949.ch26>

Gilquin, G., & Laporte, S. (2021). The use of online writing tools by learners of English: Evidence from a process corpus. *International Journal of Lexicography*, 34(4), 472-492. <https://doi.org/10.1093/ijl/ecab012>

Gilquin, Gaëtanelle & Granger, Sylviane. (2010). How can data-driven learning be

used in language teaching?

Godshalk, F. I., Swineford, F., & Coffman, W. E. (1966). *The measurement of writing ability*. New York: College Entrance Examination Board.

Godwin-Jones, R. (2015). Contributing, creating, curating: Digital literacies for language learners. *Language Learning & Technology*, 19(3), 8–20.

Godwin-Jones, R. (2017). Data-informed language learning. *Language Learning & Technology*, 21(3), 927. <https://doi.org/10.125/44629>

Godwin-Jones, R. (2022). Partnering with AI: Intelligent writing assistance and instructed language learning. *Language Learning & Technology*, 26(2), 5–24. <http://doi.org/10.125/73474>

Grammarly. (n.d.). *Grammarly Writing Assistant*. Retrieved October 2, 2024, from <https://www.grammarly.com/>

Grammarly. (2023, March 8). *Ushering in a new era of communication assistance with generative AI*. Grammarly. Retrieved October 2, 2025, from <https://www.grammarly.com/blog/product/grammarlygo-augmented-intelligence/>

Green, A. (2007). *IELTS washback in context: Preparation for academic writing in higher education* (Vol. 25). Cambridge University Press.

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), 4–44

Groves, M., & Mundt, K. (2015). Friend or foe? *English for Specific Purposes*, 37(0), 112-121. <https://doi.org/10.1016/j.esp.2014.06.002>

Groves, M., & Mundt, K. (2021). A ghostwriter in the machine? Attitudes of academic staff towards machine translation use in internationalised Higher Education. *Journal of English for Academic Purposes*, 50, 100957. <https://doi.org/10.1016/j.jeap.2021.100957>

Guichon, N., & Cohen, C. (2016). Multimodality and CALL. In F. Farr, & L. Murray

(Eds.), *The Routledge handbook of language learning and technology* (pp. 509–521).

Guo, Q., Feng, R., & Hua, Y. (2021). How effectively can EFL students use automated written corrective feedback (AWCF) in research writing? *Computer Assisted Language Learning*, 35(9), 2312–2331. <https://doi.org/10.1080/09588221.2021.1879161>

Gutiérrez., G.A. & Gilmore.A. (2021). A mixed methods case study on the use and impact of web-based lexicographic tools on L2 writing, *Computer Assisted Language Learning*, DOI: 10.1080/09588221.2021.1987273

Habeeba, L. S., & Muhammedb, M. R. (2020). Investigating the Effectiveness of Google Translate Among Iraqi Students. *International Journal of English Linguistics*, 12(12), 1-13.

Hafner, C. A., & Miller, L. (2011). Fostering learner autonomy in English for science: A collaborative digital video project in a technological learning environment.

Hafner, C. A., Chik, A., & Jones, R. H. (2013). Engaging with digital literacies in TESOL. *TESOL Quarterly*, 47, 812–815. <http://dx.doi.org/10.1002/tesq.136>.

Hakiki, G. N. R. (2021). EFL students' perception of the use of Grammarly application in writing class. *Eduvelop: Journal of English Education and Development*, 4(2), 99–106. <https://doi.org/10.31605/eduvelop.v4i2.891>

Hamp-Lyons, L. (1990). 'Second language writing assessment issues' in B. Kroll (ed.). *Second Language Writing: Research Insights for the Classroom*. New York: Cambridge University Press.

Hamp-Lyons, L. (2007). Editorial: worrying about rating. *Assessing Writing*, 12,1–9.

Hayes, J. R., & Berninger, V. W. (2014). Cognitive processes in writing: A framework. In B. Arfé, J. Dockrell, & V. Berninger (Eds.), *Writing development in children with hearing loss, dyslexia, or oral language problems: Implications for assessment and instruction* (pp. 3–15). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199827282.003.0001>

Hayes, John. 1996. A new framework for understanding cognition and affect in writing. In Michael

Hayes, John. 2012a. Modelling and remodeling writing. *Written Communication* 29. 369–388.

Hayes, John. 2012b. Evidence from language bursts, revision, and transcription for translation and its relation to other writing processes. In Michel Fayol, Denise Alamargot and Virginia

Heift, T. (2010). Prompting in CALL: A longitudinal study of learner uptake. *Modern Language Journal*, 94(2), 198–216. <https://doi.org/10.xxxx/yyyy>

Henry, A., & Thorsen, C. (2020). Disaffection and agentic engagement: ‘Redesigning’ activities to enable authentic self-expression. *Language Teaching Research*, 24, 456–475

Herring, S. C., Scheidt, L. A., Wright, E., & Bonus, S. (2005). Weblogs as a bridging genre. *Information Technology and People*, 18(2), 142–171. <http://dx.doi.org/10.1108/09593840510601513>

Herrington, A., & Moran, C. (2001). What happens when machines read our students’ writing? *College English*, 63, 480–499 <https://doi.org/10.2307/378891>

Hirvela, A., & Du, Q. (2013). “Why am I paraphrasing?”: Undergraduate ESL writers’ engagement with source-based academic writing and reading. *Journal of English for Academic Purposes*, 12(2), 87-98. <https://doi.org/10.1016/j.jeap.2012.11.005>

Hirvela, A., Hyland, K. & Manchón, R. (2016). 2. Dimensions in L2 writing theory and research: Learning to write and writing to learn. In R. Manchón & P. Matsuda (Ed.), *Handbook of Second and Foreign Language Writing* (pp. 45-64). Berlin, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9781614511335-005>

Hiver, P., Al-Hoorie, A. H., Vitta, J. P., & Wu, J. (2021a). Engagement in language learning: A systematic review of 20 years of research methods and definitions. *Language Teaching Research*, 28(1), 201-230. <https://doi.org/10.1177/13621688211001289> (Original work published 2024)

Hiver, P., Al-Hoorie, A.H., & Mercer, S. (Eds.) (2021b). *Student engagement in the language classroom*. Clevedon: Multilingual Matters

Hiver, P. & Dao, P. (2025). From task motivation to L2 learning: Understanding links through learners' task engagement. *TESOL Quarterly*.
<https://doi.org/10.1002/tesq.3410>

Hong, S. C. (2010). EFL learners' consciousness-raising through a corpus-based approach. *English Teaching*, 65(1), 57–86.
<https://doi.org/10.15858/engtea.65.1.201003.57>

Horst, M., Cobb, T., & Nicolae, I. (2005). Expanding academic vocabulary with an interactive on-line database. *Language Learning & Technology*, 9(2), 90–110.
<https://doi.org/10.64152/10125/44021>

Hosenfeld, C. (1977). A preliminary investigation of the reading strategies of successful and unsuccessful second language learners. *System*, 5, 110–123.

Howard, R. M. (1995). Plagiarisms, authorships, and the academic death penalty. *College English*, 57(7), 788–806.

Hu, K., & Deng, Y. (2023). The effect of corpus-based writing practices on EFL learners' lexical diversity and lexical sophistication. In C. Li, S. K. S. Cheung, F. L. Wang, A. Lu, & L. F. Kwok (Eds.), *Blended learning: Lessons learned and ways forward* (Lecture Notes in Computer Science, Vol. 13978, pp. 115–127). Springer.
https://doi.org/10.1007/978-3-031-35731-2_9

Huang, Ping-Yu. & Tsao, Nai-Lung. (2021). Using collocation clusters to detect and correct English L2 learners' collocation errors, *Computer Assisted Language Learning*, 34:3, 270-296, DOI: 10.1080/09588221.2019.1607880

Huang, Y., & Wilson, J. (2021). Using automated feedback to develop writing proficiency. *Computers and Composition*, 62, 102675.
<https://doi.org/10.1016/j.compcom.2021.102675>

Huang, Z. (2014). The effects of paper-based DDL on the acquisition of lexicogrammatical patterns in L2 writing. *ReCALL*, 26(2), 163–183.
<https://doi.org/10.1017/S0958344014000020>

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.

Huot, B. (1996). Toward a New Theory of Writing Assessment. *College Composition and Communication*. <https://doi.org/10.2307/358601>

Hutchins, J. (2006). *Machine translation: History*. In K. Brown (Ed.), *Encyclopedia of language & linguistics* (2nd ed., pp. 375–389). Elsevier. <https://doi.org/10.1016/B0-08-044854-2/00937-8>

Hutchins, W. J. (2000). *The first decades of machine translation: Overview, chronology, sources*. In W. J. Hutchins (Ed.), *Early years in machine translation* (pp. 1–16). Amsterdam: John Benjamins.

Hyland, F. (1998). The impact of teacher written feedback on individual writers. *Journal of Second Language Writing*, 7(3), 255- 286.

Hyland, K. (2003). *Second Language Writing* (Cambridge Language Education). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511667251

Hyland, K. (2003). *Second language writing*. Cambridge University Press.

Hyland, K. (2004). *Genre and second language writing*. Anna Arbor, MI: University of Michigan press

Hyland, K., & Hyland, F. (Eds.). (2019). *Feedback in Second Language Writing: Contexts and Issues* (2nd ed.). Cambridge: Cambridge University Press.

IDP IELTS. (2024). *How to understand question types in Task 2*. Retrieved from <https://ielts.idp.com>

IDP IELTS. (2024). *IELTS Writing Task 1 Question Types: A Comprehensive Guide*. Retrieved from <https://ielts.idp.com>

IELTS (2024). *Academic test: Sample test questions*. IELTS. Retrieved October 24, 2024, from <https://ielts.org/take-a-test/preparation-resources/sample-test-questions/academic-test>

IELTS Advantage. (2024). *IELTS Writing Task 2: Everything you need to know*. Retrieved from <https://www.ieltsadvantage.com>

IELTS Asia. (2019). *Required IELTS scores for studying in the U.S*. Retrieved from

<https://www.ieltsasia.org/hk/en/study-in-us/required-score>

IELTS Charlie. (2024). *IELTS Writing Task 2: Question types*. Retrieved from <https://ieltscharlie.com>

IELTS Liz. (2024). *IELTS Writing Task 2 Essay Topics*. Retrieved from <https://ieltsliz.com>

IELTS Material. (2024). *IELTS Writing Exam – Format, Question Types & Tips*. Retrieved from <https://ieltsmaterial.com>

IELTS-Writing.info. (n.d.). *Writing samples — Task 1 and Task 2 | IELTS test*. Retrieved November 6, 2025, from <https://www.ielts-writing.info/EXAM/ielts-writing/>

Imran, M., & Almusharraf, N. (2023). Teaching innovation in university education: Case studies and main practices: edited by Jose Ramon Saura, Hershey, PA, IGI Global, 2022, 330 pp., \$165 (paperback), ISBN 9781668444412. *The Social Science Journal*, 62(1), 236–238. <https://doi.org/10.1080/03623319.2023.2201973>

International English Language Testing System (IELTS). (n.d.). *IELTS scoring in detail*. IELTS. Retrieved from <https://ielts.org/take-a-test/your-results/ielts-scoring-in-detail>

Jabak, O. (2019). Assessment of Arabic–English translation produced by Google Translate. *International Journal of Linguistics, Literature and Translation (IJLLT)*, 2(4), 248-257. <https://ssrn.com/abstract=3441139>

Jacobs, H. L., Zingraf, S. A., Wormuth, D. R., Hartfiel, V. F., & Hughey, J. B. (1981). *Testing ESL composition: A practical approach*. Rowley, MA: Newbury House.

Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies*, 28(12), 15873–15892. <https://doi.org/10.1007/s10639-023-11834-1>

Johns, T. (1991) From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In: Johns, T. and King, P. (eds.), *Classroom Concordancing*. *English Language Research Journal*, 4:27–45.

Johns, T. F. (1994). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 293–313). Cambridge University Press.

Jolley, J. R., & Maimone, L. (2015). Free online machine translation: Use and perceptions by Spanish students and instructors. In A. J. Moeller (Ed.), *Learn languages, explore cultures, transform lives: Selected papers from the 2015 Central States Conference on the Teaching of Foreign Languages* (pp. 181–200). Central States Conference on the Teaching of Foreign Languages.

Jones, M. (2009). Back-translation: The latest form of plagiarism. In *Proceedings of the 4th Asia Pacific Conference on Educational Integrity* (pp. 1–7). Wollongong, Australia.

Kasneci, E., Sessler, K., Kühl, N. et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>

Kaur, J., & Hegelheimer, V. (2005). ESL students' use of concordance in the transfer of academic word knowledge: An exploratory study. *Computer Assisted Language Learning*, 18, 287–310.

Kaye, P. (2009). Translation activities in the language classroom. *Teaching English*. Retrieved May 10, 2019, from <https://www.teachingenglish.org.uk/article/translationactivities-language-classroom>.

Kazemzadeh, A., & Kashani, A. (2014). The effect of computer-assisted translation on L2 learners' mastery of writing. *International Journal of Research Studies in Language Learning*, 3, 29–44.

Keck, C. (2006). The use of paraphrase in summary writing: A comparison of L1 and L2 writers. *Journal of Second Language Writing*, 15(4), 261–278. <https://doi.org/10.1016/j.jslw.2006.09.006>

Keck, C. (2010). How do university students attempt to avoid plagiarism? A grammatical analysis of undergraduate paraphrasing strategies. *Writing & Pedagogy*,

Keck, C. (2014). Copying, paraphrasing, and academic writing development: A re-examination of L1 and L2 summarization practices. *Journal of Second Language Writing*, 25, 4–22. <https://doi.org/10.1016/j.jslw.2014.05.005>

Kellogg, R., Whiteford, A., & Quinlan, T. (2010). Does automated feedback help students learn to write? *Journal of Educational Computing Research*, 42, 173-96.

Kennedy, C., & Miceli, T. (2010). Corpus-assisted creative writing: introducing intermediate Italian learners to a corpus as a reference resource. *Language Learning & Technology*, 14, 28–44.

Kern, R. G. (2000). Literacy and language teaching. Oxford, UK: Oxford University Press Retrieved from <http://trove.nla.gov.au/version/40355894>.

Kilgarrriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten Years On. *Lexicography*, 1(1), 7–36. <https://doi.org/10.1007/s40607-014-0009-9>

Kim, H., & McGill, A. L. (2024). AI - induced dehumanization. *Journal of Consumer Psychology*, jcpy.1441. <https://doi.org/10.1002/jcpy.1441>

Kim, S. (2019). Playing with machine translation in language classroom: Affordances and constraints. *Multimedia-Assisted Language Learning*, 22(2), 9–28. doi:1015702/mall.2019.22.2.

Kim, Y., Le, T.-L. v., Kim, D., Lee, M., & Lee, S.-J. (2024). *Design Opportunities for Explainable AI Paraphrasing Tools: A User Study with Non-native English Speakers*. 1061–1083. <https://doi.org/10.1145/3715336.3735740>

Kirchhoff, K., Turner, A. M., Axelrod, A., & Saavedra, F. (2011). Application of statistical machine translation to public health information: A feasibility study. *Journal of the American Medical Informatics Association*, 18(4), 473–478. <https://doi.org/10.1136/amiajnl-2011-000176>

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery,

problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Kliffer, M. (2005). An experiment in MT post-editing by a class of intermediate/advanced French majors. In *Proceedings EAMT 10th Annual Conference* (pp. 160–165).

Kliffer, M. (2008). Post-editing machine translation as an FSL exercise. *Porta Linguarum*, 9, 53–67.

Knobel, M., & Lankshear, C. (Eds.), (2007). *A new literacies sampler* New York, NY: Peter Lang Retrieved from http://everydayliteracies.net/files/NewLiteraciesSampler_2007.pdf.

Knoch, U. (2009). The assessment of academic style in EAP writing: The case of the rating scale. *Melbourne Papers in Language Testing*, 13(1), 35–62. Retrieved from https://arts.unimelb.edu.au/_data/assets/pdf_file/0011/3518687/13_1_2_Knoch.pdf

Knoch, U. (2011). Rating scales for diagnostic assessment of writing: What should they look like and where should the criteria come from? *Assessing Writing*, 16(2), 81–96. <https://doi.org/10.1016/j.asw.2011.02.003>

Kobayashi, H., & Rinnert, C. (1992). Effects of the first language on second language writing: Translation versus direct comparison. *Language Learning*, 42(2), 183–215

Koehn, P. (2010). *Statistical machine translation*. Cambridge: Cambridge University Press.

Koh, W.-Y. (2017). Effective applications of automated writing feedback in process-based writing instruction. *English Teaching*, 72(3), 91–118. <https://doi.org/10.15858/ENGTEA.72.3.201709.91>

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537–550. <https://doi.org/10.1177/00336882231162868>

Kol, S., Schcolnik, M., & Spector-Cohen, E. (2018). Google Translate in Academic Writing Courses? *The EuroCALL Review*, 26(2), Article 2. <https://doi.org/10.4995/eurocall.2018.10140>

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>

Kondo-Brown, K. (2002). A FACETS analysis of rater bias in measuring Japanese second-language writing performance. *Language Testing*, 19(1), 3-31.

Koo, K. (2006). Effects of using corpora and online reference tools on foreign language writing: A study of Korean learners of English as a second language. (Unpublished doctoral dissertation). University of Iowa, Iowa City, IA.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163

Kroll, B. (1998). Assessing writing abilities. *Annual Review of Applied Linguistics*, 18, 219–240.

Kukich, K. (2000). Beyond Automated Essay Scoring. *IEEE Intelligent Systems*, 15(5), 22–27

Kuo, S. (2007). Which rubric is more suitable for NSS liberal studies? Analytic or holistic? *Educational Research Journal*, 22(2), 179–199.

Kurniati, E. Y., & Fithriani, R. (2022). Post-graduate students' perceptions of QuillBot utilization in English academic writing class. *Journal of English Language Teaching and Linguistics*, 7(3), 437-451. <https://doi.org/10.21462/jeltl.v7i3.852>

Lai, S. L., & Chen, H. J. (2015). Dictionaries vs concordancers: actual practice of the two different tools in EFL writing. *Computer Assisted Language Learning*, 28(4), 341-363. doi:10.1080/09588221.2013.839567

Lambert, C., & Zhang, G. (2019). Engagement in the use of English and Chinese as foreign languages: The role of learner-generated content in instructional task design. *The Modern Language Journal*, 103, 391–411.

Lambert, C., Philp, J., & Nakamura, S. (2017). Learner-generated content and engagement in second language task performance. *Language Teaching Research*,

21(6), 665–680.

Landers, R. (2015). Computing intraclass correlations (ICC) as estimates of interrater reliability in SPSS. *The Winnower*, 2, e143518.81744. <https://doi.org/10.15200/winn.143518.81744>

Larson-Guenette, J. (2013). It's just reflex now: German language learners' use of online resources. *Die Unterrichtspraxis/Teaching German*, 46(1), 62–74. doi:10.1111/tger. 10129

Laufer, B. (2011) The contribution of dictionary use to the production and retention of collocations in a second language. *International Journal of Lexicography*, 24(1): 29–49. <https://doi.org/10.1093/ijl/ecq039>

Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics*, 29(4), 694e716.

Lavolette, E., Polio, C. & Kahng, J. (2015). The accuracy of computer-assisted feedback and students' responses to it. *Language Learning & Technology*, 19(2), 50–68. <https://doi.org/10.64152/10125/44417>

Lee, C., Wong, K. C. K., Cheung, W. K., & Lee, F. S. L. (2009). Web-based essay critiquing system and EFL students' writing: A quantitative and qualitative investigation. *Computer Assisted Language Learning*, 22(1), 57-72.

Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes*, 25(1), 56–75. <https://doi.org/10.1016/j.esp.2005.02.010>

Lee, S. M. (2019). The impact of using machine translation on EFL students' writing. *Computer Assisted Language Learning*, 33(3), 157–175. <https://doi.org/10.1080/09588221.2018.1553186>

Lee, S. M. (2021). The effectiveness of machine translation in foreign language education: a systematic review and meta-analysis. *Computer Assisted Language Learning*, 36(1–2), 103–125. <https://doi.org/10.1080/09588221.2021.1901745>

Lee, S.-M., & Briggs, N. (2021). Effects of using machine translation to mediate the revision process of Korean university students' academic writing. *ReCALL*, 33(1), 18–33. <https://doi.org/10.1017/S0958344020000191>

Lei, H., Cui, Y., & Zhou, W. (2018). Relationships between student engagement and academic achievement: A meta-analysis. *Social Behavior and Personality: An International Journal*, 46(3), 517–528. <https://doi.org/10.2224/sbp.7054>

Leńko-Szymańska, A. (2014). Is this enough? A qualitative evaluation of the effectiveness of a teacher-training course on the use of corpora in language education. *ReCALL*, 26(2), 260–278. doi:10.1017/S095834401400010X

Levy and Sarah Ransdell (eds.), *The science of writing*, 1–27. Mahwah, NJ: Lawrence Erlbaum

Li, J., & Zhang, X. (2020). The effects of Pigai on Chinese EFL learners' writing performance and self-efficacy. *Frontiers in Psychology*, 11, 752793. <https://doi.org/10.3389/fpsyg.2021.752793>

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, S. (2017). Using corpora to develop learners' collocational competence. *Language Learning & Technology*, 21(3), 153–171. Retrieved on 24 February 2022 from <http://www.lltjournal.org/item/3018>

Li, Y., & Flowerdew, J. (2007). Shaping Chinese novice scientists' manuscripts for publication. *Journal of Second Language Writing*, 16(2), 100–117. doi:10.1016/j.jslw.2007.05.001

Li, Y., & Casanave, C. P. (2012). Two first-year students' strategies for writing from sources: Patchwriting or plagiarism? *Journal of Second Language Writing*, 21(2), 165–180. <https://doi.org/10.1016/j.jslw.2012.03.002>

Li, Z., Feng, H.-H., & Saricaoglu, A. (2017). The Short-Term and Long-Term Effects of AWE Feedback on ESL Students' Development of Grammatical Accuracy. *CALICO*

Journal, 34(3), 355–375. <https://www.jstor.org/stable/90014700>

Liang, L. (2025). The impact of using DeepL Translator on Chinese EFL students' story writing. *Journal of China Computer-Assisted Language Learning*, 5(1), 56-93. <https://doi.org/10.1515/jccall-2024-0009>

Liao, H.-C. (2016a). Enhancing the grammatical accuracy of EFL writing by using an AWE-assisted process approach. *System*, 62, 77–92. <https://doi.org/10.1016/j.system.2016.02.007>

Liao, H.-C. (2016b). Using automated writing evaluation to reduce grammar errors in writing. *ELT Journal*, 70(3), 308–319. <https://doi.org/10.1093/elt/ccv058>

Liao, M., & Tseng, C. (2010). Students' Behaviors and Views of Paraphrasing and Inappropriate Textual Borrowing in an EFL Academic Setting. *Journal of Pan-Pacific Association of Applied Linguistics*, 14, 187-211.

Lin, M. (2021). Effects of data-driven learning on college students of different grammar proficiencies: A preliminary empirical assessment in EFL classes. *SAGE Open*, 11(3), 1–12. <https://doi.org/10.1177/21582440211029936>

Lin, M. H. (2019). Becoming a DDL teacher in English grammar classes: A pilot study. *The Journal of Language Learning and teaching*, 9(1), 70–82

Lin, M. H., & Lee, J.-Y. (2015). Data-driven learning: Changing the teaching of grammar in EFL classes. *ELT Journal*. 69(3), 264-274. <http://dx.doi.org/10.1093/elt/ccv010>.

Lin, M. H., & Lee, J.-Y. (2019). Pedagogical suitability of data-driven learning in EFL grammar classes: A case study of Taiwanese students. *Language Teaching Research*, 23(5), 541–561

Lin, M.C., & Liou, H.C. (2009). Expansion of EFL academic vocabulary for writing via web-enhanced lexical instruction. *English Teaching & Learning*, 33(2), 95–146.

Lin, M.H. (2021). Effects of data-driven learning on college students of different grammar proficiencies: A preliminary empirical assessment in EFL classes. *SAGE Open*, 11(3), 1-15. <https://doi.org/10.1177/21582440211029936>

Lin, M.H., & Lee, J. (2019). Pedagogical suitability of data-driven learning in EFL grammar classes: A case study of Taiwanese students. *Language Teaching Research*, 23, 541 - 561.

Link, S., Dursun, A., Karakaya, K., & Hegelheimer, V. (2014). Towards best ESL practices for implementing automated writing evaluation. *CALICO Journal*, 31(3), 323–344. <https://doi.org/10.11139/cj.31.3.323-344>

Link, S., Mehrzad, M., & Rahimi, M. (2020). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*, 35(4), 605–634. <https://doi.org/10.1080/09588221.2020.1743323>

Liou, H.-C., & Liu, S.-Y. (2021). Exploring the relationships between English writing motivation and uptake of corpus-aided corrective feedback: A longitudinal study. In M. Charles, & A. Frankenberg-Garcia (Eds.), *Corpora in ESP/EAP writing instruction: Preparation, exploitation, analysis* (pp. 57–77). Routledge.

literacy: Tools to support learning across disciplines. In *Adolescent literacy in the era of*

Little, D. (1991). *Learner autonomy 1: Definitions, issues, and problems*. Authentik.

Liu, D. and P. Jiang. (2009). 'Using a Corpus-Based Lexicogrammatical Approach to Grammar Instruction in EFL and ESL Contexts.' *The Modern Language Journal* 93.1: 61–78.

Liu, D., & Han, L. (2015). Using corpora to teach grammar to future English teachers: A case study. *System*, 52, 77-89.

Liu, K., Kwok, H. L., Liu, J., & Cheung, A. K. F. (2022). Sustainability and Influence of Machine Translation: Perceptions and Attitudes of Translation Instructors and Learners in Hong Kong. *Sustainability*, 14(11), 6399. <https://doi.org/10.3390/su14116399>

Lloyd-Jones, R. (1977). Primary trait scoring. In C. R. Cooper & L. Odell (Eds.), *Evaluating writing: Describing, measuring, judging* (pp. 3–32). Urbana, IL: National Council of Teachers of English.

- Lo, C. K., Hew, K. F., & Jong, M. S. Yung. (2024). The influence of ChatGPT on student engagement: A systematic review and future research agenda. *Computers and Education*, 219. <https://doi.org/10.1016/j.compedu.2024.105100>
- Lo, S. (2023). Neural machine translation in EFL classrooms: learners' vocabulary improvement, immediate vocabulary retention and delayed vocabulary retention. *Computer Assisted Language Learning*, 38(3), 592–611. <https://doi.org/10.1080/09588221.2023.2207603>
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press.
- Lu, J., & Zhang, Z. (2013). Assessing and supporting argumentation with online rubrics. *International Education Studies*, 6(7), 66–77.
- Lumley, T. (2002). Assessment criteria in a large-scale writing test: What do they really mean to the raters? *Language Testing*, 19(3), 246–276.
- Luo, Q., & Zhou, J. (2017). Data-driven learning in second language writing class: A survey of empirical studies. *International Journal of Emerging Technologies in Learning (iJET)*, 12(3), 182–196. <https://doi.org/10.1093/applin/amy012>
- Luo, Q., Chen, L., Yu, D., & Zhang, K. (2023). The Mediating Role of Learning Engagement Between Self-Efficacy and Academic Achievement Among Chinese College Students. *Psychology Research and Behavior Management*, 16, 1533–1543. <https://doi.org/10.2147/PRBM.S401145>
- Luo, Y. and 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. doi: 10.4236/ojml.2017.74015
- Luton, L. (2003). If the computer did my homework, How come I didn't get an "A"? *French Review*, 76(4), 766-770. <https://doi.org/10.2307/3133085>
- Lytvyn, M. (2022). *A history of innovation at Grammarly. Grammarly Blog.*
- Ma, Q., Tang, J., & Lin, S. (2022b). The development of corpus-based language

pedagogy for TESOL teachers: A two-step training approach facilitated by online collaboration. *Computer Assisted Language Learning*, 35(9), 2731–2760. <https://doi.org/10.1080/09588221.2021.1895225>

Ma, R. (E.), Cheung, L. M. E., & Yang, J. (2022a). Teacher paths for developing corpus-based language pedagogy: A case study. *Computer Assisted Language Learning*. Advance online publication. <https://doi.org/10.1080/09588221.2022.2040537>

Mabuan, R. A. (2024). *ChatGPT and ELT: Exploring teachers' voices*. *International Journal of Technology in Education*, 7(1), 128–153. <https://orcid.org/0000-0001-6871-2432>

Maimone, L., & Jolley, J. Free Online Machine Translation: Use and Perceptions by Spanish Students and Instructors .

Manchon, R. M. (2016). Language and L2 writing: Learning to write and writing to learn in academic contexts. In K. Hyland, & P. Shaw (Eds.), *The Routledge handbook of English for academic purposes* (pp. 139e151). Routledge.

Marzuki, , Widiati, U., Rusdin, D., Darwin, , & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education*, 10(2). <https://doi.org/10.1080/2331186X.2023.2236469>

Masoudi, H. A. (2024). *Effectiveness of ChatGPT in improving English writing proficiency among non-native English speakers*. *International Journal of Educational Sciences and Arts*, 3(4), 62–84. <https://doi.org/10.59992/IJESA.2023.v3n4p2>

McCarthy, B. (2004). Does online machine translation spell the end of take-home translation assignments? *CALL-EJ online*, 6.1. Retrieved from <http://www.clec.ritsumei.ac.jp/english/callejonline/9-1/mccarthy.html>

McCarthy, K. S., Roscoe, R. D., Allen, L. K., Likens, A. D., & McNamara, D. S. (2022). Automated writing evaluation: Does spelling and grammar feedback support high-quality writing and revision? *Assessing Writing*, 52, 100608. <https://doi.org/10.1016/j.asw.2022.100608>

McDowell, L. (2020). Error analysis: A methodological exploration and application. In P. Clements, A. Krause, & R. Gentry (Eds.), *Teacher efficacy, learner agency*. Tokyo: JALT. <https://doi.org/10.37546/JALTPCP2019-53>

McGee, T. (2006). Taking a spin on the Intelligent Essay Assessor. In P. F. Ericsson, & R. H. Haswell (Eds.), *Machine scoring of student essays: Truth and consequences* (pp. 79–92). Logan, UT: Utah State University Press.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30-46. <https://doi.org/10.1037/1082-989X.1.1.30>

_Mckay, Sandra Lee. (2006). *Researching Second Language Classrooms*. Abingdon, Oxon: Routledge.

McNamara, D. S., Crossley, S. A., Roscoe, R. D., Allen, L. K., & Dai, J. (2015). A hierarchical classification approach to automated essay scoring. *Assessing Writing*, 23, 35–59. <https://doi.org/10.1016/j.asw.2014.09.002>

Medvedev, G. (2016). Google Translate in teaching English. *Journal of Teaching English for Specific and Academic Purposes*, 4(1), 181-193. <http://espeap.junis.ni.ac.rs/index.php/espeap/article/view/318>

Mercer, S. (2019). Language Learner Engagement: Setting the Scene. In: Gao, X. (eds) *Second Handbook of English Language Teaching*. Springer International Handbooks of Education. Springer, Cham. https://doi.org/10.1007/978-3-030-02899-2_40

Mizumoto, A., & Chujo, K. (2016). Who is data-driven learning for? Challenging the monolithic view of its relationship with learning styles. *System*, 61, 55e64.

Mohammad, T., Alzubi, A. A. F., Nazim, M., & Khan, S. I. (2023). EFL paraphrasing skills with QuillBot: Unveiling students' enthusiasm and insights. *Journal of Pedagogical Research*, 7(5), 359-373. <https://doi.org/10.33902/JPR.202324645>

Moon, S., & Oh, S. Y. (2018). Unlearning overgenerated be through data-driven learning in the secondary EFL classroom. *ReCALL*, 30(1), 48–67.

<https://doi.org/10.1017/S0958344017000246>

Moore, N. S., & MacArthur, C. A. (2016). Student use of automated essay evaluation technology during revision. *Journal of Writing Research*, 8(1), 149–175. <https://doi.org/10.17239/jowr-2016.08.01.05>

Moore, T., & Morton, J. (2005). Dimensions of difference: A comparison of university writing and IELTS writing. *Journal of English for Academic Purposes*, 4(1), 43–66. <https://doi.org/10.1016/j.jeap.2004.02.001>

Morphy, P., & Graham, S. (2012). Word processing programs and weaker writers/readers: A meta-analysis. *Reading and Writing*, 25(3), 641–678. <https://doi.org/10.1007/s11145-010-9292-5>

Mueller, C. M., & Jacobsen, N. D. (2016). A comparison of the effectiveness of EFL students' use of dictionaries and an online corpus for the enhancement of revision skills. *ReCALL*, 28(1), 3–21. doi:10.1017/S0958344015000142

Müller, A. (2015). The differences in error rate and type between IELTS writing bands and their impact on academic workload. *Higher Education Research & Development*, 34(6), 1207–1219. <https://doi.org/10.1080/07294360.2015.1024627>

Mundt, K., M. Groves, and A. Csernus. (2015). "Exploring Opportunities: The Impact of Google Translate on EAP Teaching." Submitted for publication.

Musk, N. (2022). Using online translation tools in computer-assisted collaborative EFL writing. *Classroom Discourse*, 13(2), 119–144. <https://doi.org/10.1080/19463014.2021.2025119>

Myers, M. J. (2000). Voice recognition software and a hand-held translation machine for second-language learning. *Computer Assisted Language Learning*, 13(1), 29–41. doi: 10.1076/0958-8221(200002)13:1;1-k;ft029

Na, C. D., & Nhat., Chi., M, N. X. (2017). Paraphrasing in academic: A case study of Vietnamese learners of English. *Language Education in Asia*, 8(1), 9-24. https://doi.org/10.5746/LEiA/17/V8/I1/A02/Na_Mai

Naismith, Ben. (2016). Integrating corpus tools on intensive CELTA courses. *ELT*

Journal, Volume 71, Issue 3, p. 273–283, <https://doi.org/10.1093/elt/ccw076>

New London Group (1996). *A pedagogy of multiliteracies: Designing social futures*. Harvard Educational Review, 66(1), 60–92. New York, NY: Routledge.

Newmann, F., Wehlage, G. G., & Lamborn, S. D. (1992). The significance and sources of student engagement. In F. Newmann (Ed.), *Student engagement and achievement in American secondary schools* (pp. 11–39). New York: Teachers College Press.

Nguyen Hoai (2024, April 20). *Nearly 50 universities announced admission based on IELTS certificate*. Retrieved from <https://www.vietnam.vn/en/gan-50-dai-hoc-cong-bo-xet-tuyen-bang-chung-chi-ielts>

Nguyen, B.V., & Ngo, T.C.T. (2021). Using the internet for self-study to improve translation for English-majored seniors at Van Lang University. *International Journal of TESOL & Education*, 1(2), <http://eoi.citefactor.org/10.11250/ijte.01.02.007>

Nguyen, H., T. (2022) Vietnamese EFL Learners' Perceptions of the Use of Google Translate in Learning English, *International Journal of Education, Learning and Development*, Vol. 10, No.6, pp.41-53

Nguyen, L. Q., & Le, H. V. (2022). *Improving L2 learners' IELTS Task 2 writing: The role of model essays and noticing hypothesis*. Language Testing in Asia. <https://doi.org/10.1186/s40468-022-00206-0>.

Nguyen, Trang & Khau, Anh. (2021). Investigating Teachers' and Graduates' Remarks on the Translation of English Modals into Vietnamese Provided by Google Translate. 21. 172-185. 10.17576/gema-2021-2104-09.

Niño, A. (2004). Recycling MT: A Course on Foreign Language Writing via MT post-editing.. Paper presented at CLUK (Computational Linguistics United Kingdom 7th Annual Research Colloquium), 6th and 7th January 2004 in the University of Birmingham, United Kingdom. <http://www.mt-archive.info/CLUK-2004-Nino.pdf>

Niño, A. (2008). Evaluating the use of machine translation post-editing in the foreign language class. *Computer Assisted Language Learning*, 21(1), 29–49. <https://doi.org/10.1080/09588220701865482>

Niño, A. (2009). Machine translation in foreign language learning: Language learners' and tutors' perceptions of its advantages and disadvantages. *ReCALL*, 21(2), 241–258. doi:10.1017/S0958344009000172

Nunan, D. (1992). *Research methods in language learning*. Cambridge, UK: Cambridge University Press. (Chapter 7)

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge, UK: Cambridge University Press.

O'Neill, R., & Russell, A. M. T. (2019). Stop! Grammar time: University students' perceptions of the automated feedback program Grammarly. *Australasian Journal of Educational Technology*, 35(1), 42-56. <https://doi.org/10.14742/ajet.3795>

O'Sullivan, I., & Chambers, A. (2006). Learners' writing skills in French: Corpus consultation and learner evaluation. *Journal of Second Language Writing*, 15(1), 49–68. <https://doi.org/10.1016/j.jslw.2006.01.002>

Oghigian, K. and Chujo, K. (2010) An effective way to use corpus exercises to learn grammar basics in English. *Language Education in Asia*, 1: 200–214.

O'Keeffe, A. (2021). Data-driven learning – a call for a broader research gaze. *Language Teaching*, 54(2), 259-272. doi:10.1017/S0261444820000245

O'Neill, E. (2016). *Measuring the impact of online translation on FL writing scores*. *IALLT Journal of Language Learning Technologies*, 46(2), 1–39. <https://doi.org/10.17161/iallt.v46i2.8560>

O'Sullivan, I. (2007). Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy. *ReCALL*, 19, 269. <http://dx.doi.org/10.1017/S095834400700033X>.

Owen, C. (1996) Do concordances require to be consulted? *ELT Journal*, 50(3): 219–224.

Palermo, C., & Thomson, M. M. (2018). Teacher implementation of Self-Regulated Strategy Development with an automated writing evaluation system: Effects on the argumentative writing performance of middle school students. *Contemporary*

Palermo, C., & Wilson, J. (2020). Implementing automated writing evaluation in different instructional contexts: A mixed-methods study. *Journal of Writing Research*, 12(1), 63-108. <https://doi.org/10.17239/jowr-2020.12.01.04>

Park, J. (2019). Limitations of Grammar Correction Tools: A Case Study. *Journal of Second Language Writing*, 38, 115-121.

Park, K. (2012) Learner-corpus interaction: A locus of microgenesis in corpus-assisted L2 writing. *Applied Linguistics*, 33(4): 361–385.

Payne, J. S., & Ross, B. M. (2005). Synchronous CMC, working memory, and L2 oral proficiency development. *Language Learning & Technology*, 9(3), 35-54.

Pearson, W. S. (2024). Affective, behavioural, and cognitive engagement with written feedback on second language writing: a systematic methodological review. In *Frontiers in Education* (Vol. 9). Frontiers Media SA. <https://doi.org/10.3389/feduc.2024.1285954>.

Pecorari, D. (2003). Good and original: Plagiarism and patchwriting in academic second-language writing. *Journal of Second Language Writing*, 12(4), 317–345. [https://doi.org/10.1016/S1060-3743\(03\)00052-5](https://doi.org/10.1016/S1060-3743(03)00052-5)

Pennington, M.C. (1993). A critical examination of word processing effects in relation to L2 writers. *Journal of Second Language Writing*, 2, 227-255.

Pérez-Paredes, P., Sánchez-Tornel, M., & Calero, J. M. A. (2012). Learners' search patterns during corpus-based focus-on-form activities: A study on hands-on concordancing. *International Journal of Corpus Linguistics*, 17(4), 482-515.

Perkins, M. (2023). Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice*, 20(2). <https://doi.org/10.53761/1.20.02.07>

Pham, A. T., Nguyen, Y. N. N., Tran, L. T., Huynh, K. D., Le, N. T. K., & Huynh, P. T. (2022). University Students' Perceptions on the Use of Google Translate: Problems

and Solutions. *International Journal of Emerging Technologies in Learning (iJET)*, 17(04), pp. 79–94. <https://doi.org/10.3991/ijet.v17i04.28179>

Philp, J., & Duchesne, S. (2016). Exploring engagement in tasks in the language classroom. *Annual Review of Applied Linguistics*, 36, 50–72.

Piaget, J. (1952). *The Origins of Intelligence in Children*. New York: International Universities Press.

Piaget, J. (1970). *Science of Education and the Psychology of the Child*. New York: Viking.

Pitukwong, K., & Saraiwang, S. (2024). Exploring the effectiveness of digital writing tools on Thai EFL students' writing. *Contemporary Educational Technology*, 16(3), ep519. <https://doi.org/10.30935/cedtech/14808>

Platt, E., & Brooks, F. (2002). Task engagement: A turning point in foreign language development. *Language Learning*, 52, 365–400.

Prentice, F. M., & Kinden, C. E. (2018). Paraphrasing tools, language translation tools and plagiarism: An exploratory study. *International Journal for Educational Integrity*, 14(1), 11. <https://doi.org/10.1007/s40979-018-0036-7>

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Erlbaum.

Qi, D., & Lapkin, S. (2001). Exploring the role of noticing in a three-stage second language writing task. *Journal of Second Language Writing*, 10, 277–303.

Qiu, X. (2024). Exploring the effect of corpus-based writing instruction on learner-corpus interaction in L2 revision: A study of Chinese EFL disciplinary writers. *TESOL Journal*, 58, 1108–1137. <https://doi.org/10.1002/tesq.3308>

Quah, C. K. (2006). *Translation and technology*. Basingstoke: Palgrave Macmillan.

QuillBot. (2024). *About QuillBot - Write without limits*. QuillBot. Retrieved October 24, 2024, from <https://www.quillbot.com>

QuillBot. (n.d.). *QuillBot Paraphrasing Tool*. Retrieved October 2, 2024, from

<https://quillbot.com/>

Quinn, C. (2015). Training L2 writers to reference corpora as a self-correction tool. *ELT Journal*, 69(2), 165-177. <https://doi.org/10.1093/elt/ccu062>

Rababah, L. M., & Talafha, D. K. (2024). Unlocking writing potential: Assessing the impact of Grammarly on Jordanian EFL students' writing proficiency. *Theory and Practice in Language Studies*, 14(5), 1486-1492. <https://doi.org/10.17507/tpls.1405.21>

Rahma, A., & Fithriani, R. (2024). *The potential impact of using ChatGPT on EFL students' writing: EFL teachers' perspective*. Indonesian EFL Journal / Indonesian EFL Learning Innovation (various listings show same article), 10(1), 11–20 <https://doi.org/10.25134/ieflj.v10i1.9222>

Rakhmanina, L., & Serasi, L. (2022). Utilizing Quillbot paraphraser to minimize plagiarism in students' scientific writing. *Novateur Publications*, 26-33. <http://novateurpublication.org/index.php/np/article/view/5>

Ranalli, J., Link, S., & Chukharev-Hudilainen, E. (2017). Automated writing evaluation for formative assessment of second language writing: Investigating the accuracy and usefulness of feedback as part of argument-based validation. *Educational Psychology*, 37(1), 8–25. <https://doi.org/10.1080/01443410.2015.1136407>

Ray, P. P. (2023). *ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope*. *Internet of Things and Cyber-Physical Systems*, 3, 121–154. <https://doi.org/10.1016/j.iotcps.2023.04.003>

Read, J., & Hayes, B. (2003). The impact of IELTS on preparation for academic study in New Zealand. *IELTS Research Reports*, 4, 153–206.

Reeve, J. (2012). A self-determination theory perspective on student engagement. In Christenson, S.L., Reschly, A.L., & C. Wylie (Eds.), *Handbook of research on student engagement (pp.149–172)*. New York: Springer

Reid, J.M. (1994). Responding to ESL students' texts: The myths of appropriation. *TESOL Quarterly*, 28, 273–292.

Reinhardt, J., Warner, C., & Lange, K. (2014). Digital games as practices and texts: New literacies and genres in an L2 German classroom. In J. P. Guikema, & L.

Reinhardt, J., Warner, C., & Lange, K. (2014). Digital games as practices and texts: New literacies and genres in an L2 German classroom. In J. P. Guikema, & L. Williams (Eds.), *Digital literacies in foreign and second language education* (pp. 159–178). San Marcos, TX: CALICO.

Riazi, M., Shi, L., & Haggerty, J. (2018). Analysis of the empirical research in the journal of second language writing at its 25th year (1992–2016). *Journal of Second Language Writing*, 41, 41-54. <https://doi.org/10.1016/j.jslw.2018.07.002>

Riazi, M., Shi, L., & Haggerty, J. (2018). Analysis of the empirical research in the journal of second language writing at its 25th year (1992–2016). *Journal of Second Language Writing*, 41, 41-54. <https://doi.org/10.1016/j.jslw.2018.07.002>

Richards, J. C., & Schmidt, R. (2002). *Longman Dictionary of Language Teaching and Applied Linguistics* (3rd ed.). United Kingdom: Pearson Education Limited. Retrieved from http://www.saint-david.net/uploads/1/0/4/3/10434103/linguistic_term_dictionary.pdf

Richards, K. (2003). Qualitative inquiry in TESOL. New York: Palgrave MacMillan. (Chapter 2)

Richmond, I. (1994). Doing it backwards: Using translation software to teach target language grammaticality. *Computer Assisted Language Learning*, 7(1), 65-78

Roca de Larios, J., Manchón, R. M., Murphy, L., & Marín, J. (2008). *The foreign language writer's strategic behaviour in the allocation of time to writing processes*. *Journal of Second Language Writing*, 17(1), 30–47. <https://doi.org/10.1016/j.jslw.2007.08.005>

Roe, J., & Perkins, M. (2022). What are automated paraphrasing tools and how do we address them? A review of a growing threat to academic integrity. *International Journal for Educational Integrity*, 18, Article 15. <https://doi.org/10.1007/s40979-022-00109-w>

Roe, J., & Perkins, M. (2022). What are automated paraphrasing tools and how do we

address them? A review of a growing threat to academic integrity. *International Journal for Educational Integrity*, 18, Article 15. <https://doi.org/10.1007/s40979-022-00109-w>

Roehr-Brackin, K. (2018). *Metalinguistic awareness and second language acquisition*. New York: Routledge.

Rogerson, A. M., & McCarthy, G. (2017). *Using internet-based paraphrasing tools: Original work, patchwriting or facilitated plagiarism?* *International Journal for Educational Integrity*, 13, Article 2. <https://doi.org/10.1007/s40979-016-0013-y>

Rogerson, A. M., & McCarthy, G. (2017). Using internet-based paraphrasing tools: Original work, patchwriting or facilitated plagiarism? *International Journal for Educational Integrity*, 13(2). <https://doi.org/10.1007/s40979-016-0013-y>

Römer, U. (2006). *Discourse in the professions: Perspectives from corpus linguistics*. *Studies in Second Language Acquisition*, 28(4), 646–648. <https://doi.org/10.1017/S0272263106300290>

Römer, U. (2006). Pedagogical applications of corpora: some reflections on the current scope and a wish list for future developments. *Zeitschrift für Anglistik und Amerikanistik* 54/2: 121–34.

Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics*, 31, 205–225. <https://doi.org/10.1017/S0267190511000055>

Rosário P, Högemann J, Núñez JC, Vallejo G, Cunha J, et al. (2019) The impact of three types of writing intervention on students' writing quality. *PLOS ONE* 14(7): e0218099. <https://doi.org/10.1371/journal.pone.0218099>

Roscoe, R. D., Wilson, J., Johnson, A. C., & Mayra, C. R. (2017). Presentation, expectations, and experience: Sources of student perceptions of automated writing evaluation. *Computers in Human Behavior*, 70, 207–221. <https://doi.org/10.1016/j.chb.2016.12.076>

Routh, N. (2024, December 9). *What is Generative AI? | Meaning & Examples*.

QuillBot. <https://quillbot.com/blog/ai-writing-tools/what-is-generative-ai/>

Rukiati, E., Wicaksono, J. A., Taufan, G. T., & Suharsono, D. D. (2023). AI on learning english: Application, benefit, and threat. *Journal of Language, Communication, and Tourism*, 1(2), 32–40. <https://doi.org/10.25047/jlct.v1i2.3967>

Russell, S., & Norvig, P. (2010). *Artificial intelligence: A modern approach* (3rd ed.). Prentice Hall.

Sachs, R., & Polio, C. (2007). Learners' uses of two types of written feedback on an L2 writing revision task. *Studies in Second Language Acquisition*, 29, 67–100

Saeedakhtar, A. (2020). The primary and secondary effects of data-driven learning on high- and low-intermediate learners' Knowledge of collocations. *Issues in Language Teaching*, 9(2), 311-346. doi: 10.22054/ilt.2021.52953.505

Saleh Mahdy Mohammed, O. S. M. M., Samad, S. S., & Mahdi, H. S. (2020). The attitudes of professional translators and translation students towards computer-assisted translation tools in Yemen. *Dil ve Dilbilimi Çalışmaları Dergisi*, 16(2), 1084–1095. <https://doi.org/10.17263/jlls.759371>

Sang Y., Hiver P. (2021). Engagement and companion constructs in language learning: Conceptualizing learners' involvement in the L2 classroom. In Hiver P., Al-Hoorie A.H., S. Mercer (Eds.), *Student engagement in the language classroom* (pp. 17–37). Clevedon: Multilingual Matters.

Sanosi, A. B. (2022). The impact of automated written corrective feedback on EFL learners' academic writing accuracy. *Journal of Teaching English for Specific and Academic Purposes*, 301–317. <https://doi.org/10.22190/JTESAP2202301S>

Santos, D. & A. Frankenberg-Garcia (2007). The corpus, its users and their needs: A user-oriented evaluation of COMPARA. *International Journal of Corpus Linguistics* 12.3, 335–374.

Saricaoglu, A., & Bilki, Z. (2021). Voluntary use of automated writing evaluation by content course students. *ReCALL*, 33(3), 265–277. <https://doi.org/10.1017/S0958344021000021>

Schaefer, E. (2008). Rater bias patterns in an EFL writing assessment. *Language Testing*, 25(4), 465-493. <https://doi.org/10.1177/0265532208094273>

Scherer, D. L. (1985). *Measuring the measurements: A study of evaluation of writing. An annotated bibliography* (ERIC Document Reproduction Service No. ED260015)

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge University Press.

Schmidt, R.W. (1990) The role of consciousness in second language learning. *Applied Linguistics* 11, 129–58.

Schrader, C., & Kalyuga, S. (2020). Linking students' emotions to engagement and writing performance when learning Japanese letters with a pen-based tablet: An investigation based on individual pen pressure parameters. *International Journal of Human-Computer Studies*, 135, 102374. <https://doi.org/10.1016/j.ijhcs.2019.102374>

Scott, M., & Tribble, C. (2006). Textual patterns: Keywords and corpus analysis in language education. Amsterdam, The Netherlands: John Benjamins.

Seidlhofer, B. (2000) 'Operationalizing intertextuality: using learner corpora for learning' in L. Burnard and A. McEnery (eds.) *Rethinking Language Pedagogy from a Corpus Perspective*, pp. 207–24. New York: Peter Lang.

Seo, J. Y. (2024). Exploring the Educational Potential of ChatGPT: AI-Assisted Narrative Writing for EFL College Students. *Language Teaching Research Quarterly*, 43. <https://doi.org/10.32038/ltrq.2024.43.01>

Seow, A. (2002). The writing process and process writing. In J. C. Richards & W. A. Renandya (Eds.), *Methodology in language teaching* (pp. 315–320). Cambridge University Press.

Shabani, E. A., & Panahi, J. (2020). *Examining consistency among different rubrics for assessing writing*. *Language Testing in Asia*, 10, Article 12. <https://doi.org/10.1186/s40468-020-00111-4>.

Shabani, E. A., & Panahi, J. (2020). Examining consistency among different rubrics

for assessing writing. *Language Testing in Asia*, 10(12).
<https://doi.org/10.1186/s40468-020-00111-4>

Shaffer, C. (1989). A comparison of inductive and deductive approaches to teaching foreign languages. *The Modern Language Journal*, 73(4), 395–403.
<https://doi.org/10.1111/j.1540-4781.1989.tb05319.x>

Shang, H. F. (2019). Exploring online peer feedback and automated corrective feedback on EFL writing performance. *Interactive Learning Environments*, 30(1), 4–16. <https://doi.org/10.1080/10494820.2019.1629601>

Shaw, S., & Weir, C. J. (2007). *Examining writing: Research and practice in assessing second language writing* (Studies in Language Testing 26). UCLES/Cambridge University Press.

Shei, C.-C. (2002). Teaching MT through pre-editing: Three case studies. In *6th EAMT Workshop Teaching Machine Translation* (pp. 89–98), Manchester.

Shen, C., Shi, P., Guo, J., Xu, S., & Tian, J. (2023). From process to product: writing engagement and performance of EFL learners under computer-generated feedback instruction. *Frontiers in Psychology*, 14. <https://doi.org/10.3389/fpsyg.2023.1258286>

Sheppard, F. (2011). Medical writing in English: The problem with Google Translate. *Presse Médicale*, 40(6), 565–566. <https://doi.org/10.1016/j.lpm.2011.02.024>

Shermis, M. D., Burstein, J. C., & Bliss, L. (2004). The impact of automated essay scoring on high stakes writing assessments. *Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA*

Shi, L. (2004). Textual borrowing in second-language writing. *Written Communication*, 21, 171–200

Shin, D. S., Cimasko, T., & Yi, Y. (2020). Development of metalanguage for multimodal composing: A case study of an L2 writer's design of multimedia texts. *Journal of Second Language Writing*, 47, 100714. <https://doi.org/10.1016/j.jslw.2020.100714>

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420-428. <https://doi.org/10.1037/0033->

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

Sinclair, J. (2005). Corpus and text: Basic principles. In M. Wynne (Ed.), *Developing linguistic corpora: A guide to good practice* (pp. 1–16). Oxford: Oxbow Books.
<https://doi.org/20.500.14106/dlc>

Singh, V., Singh, A. K., & Sonkar, S. K. (2023). Awareness and use of paraphrasing tools by the research scholars of BBAU and CUSB: A survey. *Alexandria: The Journal of National and International Library and Information Issues*, 32(2–3), 77–89.
<https://doi.org/10.1177/09557490231217716>

Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effect of teacher behavior and student engagement across the school year. *Journal of Educational Psychology*, 85, 571–581

Skinner, E.A., Kindermann, T.A., & Furrer, C. (2009). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement*, 69, 493–525.

Socher, R., Huang, E. H., Pennin, J., Manning, C. D., & Ng, A. Y. (2011). Dynamic pooling and unfolding recursive autoencoders for paraphrase detection. In *Advances in Neural Information Processing Systems (NIPS)*. Granada, Spain.

Somers, H. (2011). Machine translation: History, development, and limitations. In K. Malmkjær, & K. Windle (Eds.), *The Oxford handbook of translation studies* (pp. 427e440). Oxford University Press.

Somers, H. (2012). Computer-Assisted Language Learning and Machine Translation. In *The Encyclopedia of Applied Linguistics*, C.A. Chapelle (Ed.).
<https://doi.org/10.1002/9781405198431.wbeal0175>

Somers, H. (Ed.). (2003). *Computers and translation: A translator's guide*. John Benjamins Publishing Company

- Somers, H., Gaspari, F., & Niño, A. (2006). Detecting inappropriate use of free online machine translation by language students: A special case of plagiarism detection. In V. Hansen & B. Maegaard (Eds.), *Proceedings of the 11th Annual Conference of the European Association for Machine Translation*. European Association for Machine Translation. <https://aclanthology.org/2006.eamt-1.6/>
- Stapleton, P. (2005). Using the Web as a Research Source: Implications for L2 Academic Writing. *The Modern Language Journal*, 89, 177–189.
- Stapleton, P., & Kin, B. L. K. (2019). Assessing the accuracy and teachers' impressions of Google translate: A study of primary L2 writers in Hong Kong. *English for Specific Purposes*, 56, 18-34. <https://doi.org/10.1016/j.esp.2019.07.001>
- Stasimioti, M., & Sosoni, V. (n.d.). *Undergraduate Translation Students' Performance and Attitude vis-à-vis Machine Translation and Post-editing: Does Training Play a Role?* 12.
- Steding, S. (2009). Machine translation in the German classroom: Detection, reaction, prevention. *Die Unterrichtspraxis*, 42(2), 178-189.
- Stevens, V. (1991) Concordance-based vocabulary exercises: A viable alternative to gap-fillers. In: Johns, T. and King, P. (eds.), *Classroom Concordancing*. *English Language Research Journal*, 4:47–61
- Suci, I. P. P., & Syafryadin. (2024). Students' perception on the use of Grammarly in writing English journals. *Journal of Applied Linguistics and Literacy*, 8(2), 198–211. <https://doi.org/10.29140/jaltcall.v17n2.336>
- Suhono, S., Zuniati, M., Pratiwi, W., & Hasyim, U. A. A. (2020). Clarifying Google Translate Problems of Indonesia-English Translation of Abstract Scientific Writing. *EAI* (24-25), 113.
- Sulistyaningrum, S. D. (2021). Utilizing online paraphrasing tools to overcome students' paraphrasing difficulties in literature reviews. *Journal of English Language Studies*, 6(2), 229. <https://doi.org/10.30870/jels.v6i2.11582>
- Sun, D. Y. (2017). Application of post-editing in foreign language teaching: Problems

and challenges. *Canadian Social Science*, 13(7), 1–5. <https://doi.org/10.3968/9698>

Sun, Y., & Wang, L. (2003). Concordancers in the EFL Classroom: Cognitive Approaches and Collocation Difficulty. *Computer Assisted Language Learning*, 16, 83–94.

Sun, Y.C. (2007). Learner perceptions of a concordancing tool for academic writing. *Computer Assisted Language Learning*, 20(4), 323–343.

Svalberg, A. M.-L. (2009). Engagement with language: Interrogating a construct. *Language Awareness*, 18(3–4), 242–258. <https://doi.org/10.1080/09658410903197264>

Svalberg, A.M.L. (2017). Researching language engagement: Current trends and future directions. *Language Awareness*, 27, 21–39.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Rowley, MA: Newbury House.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 125–144). Oxford University Press.

Swales, J. (2002). “Integrated and fragmented worlds: EAP materials and corpus linguistics”. In J. Flowerdew (ed). *Academic Discourse*. Harlow: Longman, p.150-164

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Talai, T. and Fotovatnia, Z. (2012) Data-driven learning: A student-centered technique for language learning. *Theory and Practice in Language Studies*, 2(7): 1526–1531.

Tan, E., & Carnegie, S. (2022). “It’s not plagiarism, it’s a bad use of power phrasing”: Assessment of home and international student (mis)understandings of citation practice. *Innovations in Education and Teaching International*, 59(3), 285–295. <https://doi.org/10.1080/14703297.2020.1844779>

Thompson, P. A. (2006). Assessing the contribution of corpora to EAP practice. In Z. Kantaridou, I. Papadopoulou, & I. Mahili (Eds.), *Motivation in learning language for specific and academic purposes*. Thessaloniki, Greece: University of Macedonia.

Thurstun, J. and Candlin, C. (1998) Concordancing and the teaching of the vocabulary of academic English. *English for Specific Purposes*. 17(3): 267–280.

Tieu, L. T. C., & Baker, J. R. (2022). Using model essays in conjunction with noticing as a feedback instrument in IELTS writing preparation. *Innovation in Language Learning and Teaching*, 17(2), 380–392. <https://doi.org/10.1080/17501229.2022.2049799>

Todd, R. W. (2001). Induction from self-selected concordances and self-correction. *System*, 29(1), 91-102

Tran–Thanh, V. (2025). Recent trends in English Language education in Vietnam: Navigating Policy, Practice, and Societal Shifts. *English Today*, 1–6. doi:10.1017/S0266078424000506

Tribble, C. (2015). Teaching and language corpora: Using Sketch Engine with students. *International Journal of Corpus Linguistics*, 20(3), 375-404.

Truong, H. H. (2023). Using QuillBot for enhancing EFL learners' essay writing skills. *HNUE Journal of Science*, 68(3), 41–50. Retrieved from <https://vjol.info.vn/index.php/DHSP-GD/article/download/84079/72866/>

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

Tsai, S.-C. (2019). Using Google Translate in EFL drafts: A preliminary investigation. *Computer Assisted Language Learning*, 32(5–6), 510–526. <https://doi.org/10.1080/09588221.2018.1527361>

Tsai, S.-C. (2022). Chinese students' perceptions of using Google Translate as a translingual CALL tool in EFL writing. *Computer Assisted Language Learning*, 35(5–6), 1250–1272. <https://doi.org/10.1080/09588221.2020.1799412>

Turner, K. H., & Katic, E. K. (2009). The Influence of Technological Literacy on

Students' Writing. *Journal of Educational Computing Research*, 41(3), 253-270.
<https://doi.org/10.2190/EC.41.3.a>

UK Home Office. (2024). *Knowledge of English for a student visa*. Gov.uk. Retrieved from <https://www.gov.uk/student-visa/knowledge-of-english>

UCLES. (2006, January 26). *UCLES Group annual report 2004–2005*. University of Cambridge Local Examinations Syndicate. Retrieved from <https://www.cambridgeassessment.org.uk/Images/140026-annual-report-04-05.pdf>

Urdan, T. C. (2022). *Statistics in plain English* (5th ed.). Routledge.
<https://doi.org/10.4324/9781003006459>

Uysal, H. H. (2009). A critical review of the IELTS writing test. *ELT Journal*, 64(3), 314–320. <https://doi.org/10.1093/elt/ccp026>

van Rensburg, A., Snyman, C., & Lotz, S. (2012). Applying Google Translate in a higher education environment: Translation products assessed. *Southern African Linguistics and Applied Language Studies*, 30(4), 511–524.
<https://doi.org/10.2989/16073614.2012.750824>

Veal, L. R., & Hudson, S. A. (1983). Direct and indirect measures for large-scale evaluation of writing. *Research in the Teaching of English*, 17, 285

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wahle, J., Ruas, T., Foltynnek, T., Meuschke, N., & Gipp, B. (2021). Identifying machine based plagiarism. *arXiv preprint arXiv:2103.11909*. <https://arxiv.org/abs/2103.11909>

Wang, H., Wu, H., He, Z., Huang, L., & Church, K. (2021). Progress in machine translation. *Engineering*, 7(6), 718–729. <https://doi.org/10.1016/j.eng.2021.03.023>

Wang, W., & Wen, Q. (2002). L1 use in the L2 composing process: An exploratory study of 16 Chinese EFL writers. *Journal of Second Language Writing*, 11(3), 225-246.

Wang, Y. J., Shang, H. F., & Briody, P. (2012). Exploring the impact of using

automated writing evaluation in English as a foreign language university students' writing. *Computer Assisted Language Learning*, 26(3), 234–257. <https://doi.org/10.1080/09588221.2012.655300>

Ware, P. (2014). Feedback for Adolescent Writers in the English Classroom *Exploring Pen-and-Paper, Electronic, and Automated Options. Writing & Pedagogy*, 6(2), 223-249. <https://doi.org/10.1558/wap.v6i2.223>

Ware, P., Kern, R. & Warschauer, M. (2016). 14. The development of digital literacies. In R. Manchón & P. Matsuda (Ed.), *Handbook of Second and Foreign Language Writing* (pp. 307-328). Berlin, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9781614511335-017>

Warschauer, M., & Grimes, D. (2008). Automated Writing Assessment in the Classroom. *Pedagogies: An International Journal*, 3(1), 22–36. <https://doi.org/10.1080/15544800701771580>

Warschauer, M., & Grimes, D. (2008). Automated Writing Assessment in the Classroom. *Pedagogies: An International Journal*, 3(1), 22–36. <https://doi.org/10.1080/15544800701771580>

Warschauer, M., & Ware, P. (2006). Automated writing evaluation: defining the classroom research agenda. *Language Teaching Research*, 10(2), 157-180. <https://doi.org/10.1191/1362168806lr190oa>

Wei, P., Wang, X., & Dong, H. (2023). The impact of automated writing evaluation on second language writing skills of Chinese EFL learners: a randomized controlled trial. *Frontiers in Psychology*, 14, 1249991. <https://doi.org/10.3389/fpsyg.2023.1249991>

Wette, R. (2010). Evaluating Student Learning in a University-Level EAP Unit on Writing Using Sources. *Journal of Second Language Writing*, 19, 158-177. <https://doi.org/10.1016/j.jslw.2010.06.002>

White, E. M. (1985). *Teaching and assessing writing*. San Francisco, CA: Jossey-Bass.

Widdowson, H.G. (2000). On the limitations of linguistics applied. *Applied Linguistics*

21(1), 3-25.

Williams, L. (2006). Web-based machine translation: A tool for promoting electronic literacy and language awareness. *Foreign Language Annals*, 39, 565-578. <https://doi.org/10.1111/j.1944-9720.2006.tb02276.x>

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>

Wirantaka, A., & Fijanah, M. (2021). Effective use of Google Translate in writing. In *Proceedings of the 10th Annual International Conference on Language and Literature Studies* (ASSEHR, Vol. 10). Atlantis Press. <https://doi.org/10.2991/assehr.k.211227.003>

Wong, T., & Lee, J. (2016). Corpus-based learning of Cantonese for Mandarin speakers. *ReCALL*, 28(2), 187–206. <https://doi.org/10.1017/S0958344015000257>

Woo, D. J., Susanto, H., Yeung, C. H., Guo, K., & Fung, A. K. Y. (2024). Exploring AI-Generated text in student writing: How does AI help? *Language Learning & Technology*, 28(2), 183–209. <https://doi.org/10.64152/10125/73577>

Wu, H., Li, S., Zheng, J., & Guo, J. (2020). Medical students' motivation and academic performance: the mediating roles of self-efficacy and learning engagement. *Medical Education Online*, 25(1). <https://doi.org/10.1080/10872981.2020.1742964>

Wu, Hsiao-ping & Garza, Esther. (2014). Types and Attributes of English Writing Errors in the EFL Context—A Study of Error Analysis. *Journal of Language Teaching and Research*. 5. 10.4304/jltr.5.6.1256-1262.

Wu, Y. J. (2016). The effects of utilizing corpus resources to correct collocation errors in L2 writing: Students' performance, corpus use, and perceptions. In S. Papadima-Sophocleous, L. Bradley, & S. Thouèsny (Eds.), *CALL communities and culture – short papers from EUROCALL 2016* (pp. 479–484). Research-publishing.net. <https://doi.org/10.14705/rpnet.2016.eurocall2016.610>

- Xu, J. (2022). Proficiency and the use of machine translation: A case study of four Japanese learners. *L2 Journal*, 14(1), 77-104. <https://doi.org/10.5070/L214151328>
- Xu, J., & Zhang, S. (2022). Understanding AWE Feedback and English Writing of Learners with Different Proficiency Levels in an EFL Classroom: A Sociocultural Perspective. *Asia-Pacific Education Researcher*, 31(4), 357–367. <https://doi.org/10.1007/s40299-021-00577-7>
- Yang, T. C., & Chen, J. H. (2023). Pre-service teachers' perceptions and intentions regarding the use of chatbots through statistical and lag sequential analysis. *Computers and Education: Artificial Intelligence*, 4. <https://doi.org/10.1016/j.caeai.2022.100119>
- Yoon, C. (2016a). Concordancers and dictionaries as problem-solving tools for ESL academic writing. *Language Learning & Technology*, 20(1), 209–229. <https://doi.org/10.64152/10125/44453>
- Yoon, C. (2016b). Individual differences in online reference resource consultation: Case studies of Korean ESL graduate writers, *Journal of Second Language Writing*, Volume 32, 67-80. <https://doi.org/10.1016/j.jslw.2016.04.002>.
- Yoon, H. (2008). More than a linguistic reference: The influence of corpus technology on L2 academic writing. *Language Learning and Technology*, 12(2), 31–48. <http://hdl.handle.net/10125/44142>.
- Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, 13(3), 257–283. <https://doi.org/10.1016/j.jslw.2004.06.002>
- Yoon, H., & Jo, J. W. (2014). Direct and indirect access to corpora: An exploratory case study comparing students' error correction and learning strategy use in L2 writing. *Language Learning & Technology*, 18(1), 96–117. <http://ilt.msu.edu/issues/february2014/yoontojo.pdf>
- Yoon, H.-J. (2021). Interactions in EFL argumentative writing: Effects of topic, L1 background, and L2 proficiency on interactional metadiscourse. *Reading and Writing*, 34(3), 705–725. <https://doi.org/10.1007/s11145-020-10085-7>

- Yousofi, R. (2022). *Grammarly deployment (in)efficacy within EFL academic writing classrooms: An attitudinal report from Afghanistan*. *Cogent Education*, 9(1), Article 2142446. <https://doi.org/10.1080/2331186X.2022.2142446>
- Zamel, V. (1982). Writing: The Process of Discovering Meaning. *TESOL Quarterly*, 16(2), 195–209. <https://doi.org/10.2307/3586792>
- Zamel, V. (1983). The Composing Processes of Advanced ESL Students: Six Case Studies. *TESOL Quarterly*, 17(2), 165–187. <https://doi.org/10.2307/3586647>
- Zaretskaya, A., Pastor, G. C., & Seghiri, M. (n.d.). *Integration of Machine Translation in CAT Tools: State of the Art, Evaluation and User Attitudes*. 14.
- Zhang Q, Wang DY, Voelker GM. (2014). DSpin: detecting automatically spun content on the web. *NDSS* <https://doi.org/10.14722/ndss.2014.23004>
- Zhang, X. (2021). Digital writing tools and L2 writing: An evaluation of their effectiveness. *Journal of Second Language Writing*, 52, 100816. <https://doi.org/10.xxxx/yyyy>
- Zhang, Z. (2017). Student engagement with computer-generated feedback: A case study. *ELT Journal*, 71(3), 317–328. <https://doi.org/10.1093/elt/ccw089>
- 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>
- Zhao, C. (2025). Effects of Interactive Feedback and Learning Proficiency on Chinese EFL Learners' Affective Engagement. *Journal of Linguistics and Communication Studies*, 4(2), 25–39. <https://doi.org/10.56397/jlcs.2025.04.03>
- Zhao, X. (2022). Leveraging Artificial Intelligence (AI) Technology for English Writing: Introducing Wordtune as a Digital Writing Assistant for EFL Writers. *RELC Journal*, 54(3), 890-894. <https://doi.org/10.1177/00336882221094089> (Original work published 2023)

APPENDICES

Appendix A: Online questionnaire for student participant recruitment

USING AI TOOLS FOR IELTS ESSAY WRITING PRACTICE

This research is being conducted by Nguyen Hoai Ly Nga—a PhD student at the University of Birmingham and a lecturer at the University of Economics, Hochiminh City (UEH). The supervisor of this project is Dr. Amanda Patten [REDACTED].

This short survey is designed to find out about Vietnamese students' usage of technology and tools that support the learning of written English. It will take approximately 5 minutes to complete the survey. Another aim of the survey is to identify suitable participants for a larger project on the usage and usefulness of technology and Artificial Intelligence (AI) tools in teaching argumentative writing.

You will be able to register your interest in hearing more about this project as part of the survey.

Students participating in the larger project will be compensated.

All data gathered from this survey will be secured and remain confidential. You may withdraw your data from this survey within 30 days from completion by emailing the researcher.

If you have any questions or concerns, please contact the researcher via this email:

[REDACTED]

Your time in completing this short survey is greatly appreciated.

Participant consent

Please read the following statements:

If you agree to them please check the box and press the NEXT button to register your consent and for the survey to begin.

☐ I agree

Are you above 18 years old, are currently an (undergraduate or postgraduate) student and have already taken the official IELTS test?

- Yes

- No

Please provide your most recent IELTS Writing Score: _____

AI writing tools are technology software or applications that can search for language resources, translate a sentence, paraphrase a sentence, identify spelling errors, correct grammatical errors and evaluate written texts. Examples of AI language tools are: Google Translate, SketchEngine, Grammarly, Quillbot, ProWritingAid, etc.

In general, what is your opinion about using AI tools for practicing English writing skill?

- Extremely negative

- Somewhat negative

- Neutral

- Somewhat positive

- Extremely positive

What are the AI writing tools that you have used or heard about?

Have you ever used these tools to practice writing IELTS Task 2 individually (without the help of a teacher)?

Online translation tools

- Yes

- No

Grammar-checker tools

- Yes
- No

Online dictionary

- Yes
- No

Paraphrasing tools

- Yes
- No

Corpus tools

- Yes
- No

From your experience, which tool can paraphrase a sentence? (If your answer is not in the list, please type the name of the tool in the box marked “Other”)

- Thesaurus
- Reverso
- QuillBot
- I don't know
- Other: _____

From your experience, which tool can suggest grammar corrections? (If your answer is not in the list, please type the name of the tool in the box marked “Other”)

- Reverso
- Thesaurus
- Wordtune
- Ginger
- I don't know

- Other: _____

From your experience, which tool can check the collocations of a word? (If your answer is not in the list, please type the name of the tool in the box marked "Other")

- Thesaurus
- Sketchengine
- Grammarly
- QuillBot
- I don't know

- Other: _____

IELTS TASK 2

Now please share with me your IELTS Task 2 learning experience.

What steps in writing IELTS Task 2 do you find easy?

- Brainstorming ideas
- Composing
- Paraphrasing sentences
- Revising / Editing

What steps in writing IELTS Task 2 do you find difficult?

- Brainstorming ideas
- Composing
- Paraphrasing sentences
- Revising / Editing

What topics of IELTS Task 2 essay do you have the most confidence with?

- Education
- Technology
- Language & culture
- Health
- Music & Arts
- Environment
- Sports
- Jobs and workplace
- Crime & law
- Transportation
- Tourism

What types of IELTS Task 2 questions do you have the most confidence with?

- To what extent do you agree or disagree
- Discuss both sides and give your opinion

- What are the advantages and disadvantages

- What are the causes and problems

- What are the causes and solutions

- What are the problems and solutions

Our project

Getting involved in our larger project

We are looking for students who are interested in writing in English with technology tools, to participate in a larger project. The primary aim of the study is to investigate the usage and usefulness of technology and Artificial Intelligence (AI) tools for individual writing practices of argumentative writing (IELTS Task 2 essay) among ESL students.

Students participating in this study will receive a compensation of 300,000 VND for a 2-hour session.

If you are interested in being contacted for the study, please give us your contact information.

Please be noted that only participants who satisfy specific requirements after survey completion will be contacted.

Would you be happy to hear more about this project?

- Yes

- No

Thank you for participating in the survey.

Please provide your name:

Please provide your email address so that I can contact you later:

Appendix B: Online questionnaire for teacher participant recruitment

This research is being conducted by Nguyen Hoai Ly Nga—a PhD student at the University of Birmingham and a lecturer at the University of Economics, Hochiminh City (UEH). The supervisor of this project is Amanda Patten [REDACTED].

This short survey is designed to find out about teachers' use and awareness of technology and tools that support the learning of written English. It will take approximately 5 minutes to complete the survey. Another aim of the survey is to identify suitable participants for a larger project on the usage and usefulness of technology and Artificial Intelligence (AI) tools in teaching argumentative writing. You will be able to register your interest in hearing more about this project as part of the survey. Teachers participating in the larger project will be compensated. All data gathered from this survey will be secured and remain confidential. You may withdraw your data from this survey within 30 days from completion by emailing the researcher.

If you have any questions or concerns, please contact the researcher via this email:

[REDACTED]

Your time in completing this short survey is greatly appreciated.

Participant consent

Please read the following statements:

If you agree to them please check the box and press the NEXT button to register your consent and for the survey to begin.

☐ I agree

Using AI to teach writing

AI writing tools are technology software or applications that can search for language resources, translate a sentence, paraphrase a sentence, identify spelling errors, correct grammatical errors and evaluate written texts. Examples of AI language tools are: Google Translate, SketchEngine, Grammarly, Quillbot, Criterion, Write&Improve, etc.

In the following questions, please share with me your awareness or usage of AI tools.

Using AI to teach writing

Have you ever used AI tools in your writing lessons/curriculum?

- Yes

- No

Please list some AI tools for improving English writing that you have used or heard about:

In general, what is your attitude towards using technology and AI tools in teaching writing?

- Strongly positive

- Positive

- Neutral / No opinion

- Negative

- Strongly negative

From your experiences, which of these tools can help you check word collocations—combinations of words that are closely affiliated with each other?

For example, word collocations of "take" can be: "take part", "take a chance", "take time", etc.

If it is not in this list, please type the name of the tool that you use.

- Grammarly

- Corpus tools (SketchEngine, English corpora, JTW, etc.)

- QuillBot

- Google Translate

- I don't know

- Other: _____

From your experiences, which of these tools can be used to automatically check the grammar of a written paragraph? (If it is not in this list, please type the name of the tool that you use.)

- Thesaurus

- Corpus tools (SketchEngine, English corpora, JTW, etc.)

- QuillBot

- Grammarly

- I don't know

- Other: _____

From your experiences, which of these tools can be used to paraphrase a sentence? (If it is not in this list, please type the name of the tool that you use.)

- QuillBot

- Write&Improve

- Thesaurus

- Corpus tools (SketchEngine, English corpora, JTW, etc.)

- I don't know

- Other: _____

From your experiences, which of these tools can be used to evaluate a piece of written

paragraph in terms of grammar and word usage? (Choose all that are correct.)

- Thesaurus
- Criterion
- Grammarly
- SketchEngine
- Write&Improve
- Turnitin

What do you think are the usages of a corpus tool? (Choose all that are correct.)

- To search for ideas and meanings of a sentence
- To check the frequency of a word from a collection of texts
- To check the meaning of a word
- To check collocations / concordances of a word
- To check prepositions that usually go with a word

How do you feel about using Google Translate or similar machine translation tools as a part of your IELTS Task 2 writing lesson?

- Very positive
- Maybe
- No, I would not use it

How do you feel about using Automated Writing Evaluation (AWE) tools that can evaluate and give feedback to a written paragraph as a part of your IELTS Task 2 writing curriculum?

- Very positive

- Maybe

- No, I would not use it

How do you feel about using AI automatic paraphrase tools as a part of your IELTS Task 2 writing lesson?

- Very positive

- Maybe

- No, I would not use it

Getting involved in our larger project

We are looking for teachers who are interested in teaching English as a Second Language (ESL) with technology tools, to participate in a larger project. The primary aim of the study is to investigate the usage and usefulness of technology and Artificial Intelligence (AI) tools in teaching argumentative writing (IELTS Task 2 essay) among ESL teachers.

Teachers participating in this study will receive a compensation of 800,000 VND for a 90-minute session, including lesson planning, demo teaching, and a follow-up interview. If you are interested in being contacted for the study, please give us your contact information and tell us more about your teaching experience in the next section.

Would you be happy to hear more about this project?

- Yes

- No

Contact details and teaching experience

Thank you for your interest in participating. Please provide your contact information so I can contact you later.

Please provide your name:

Please provide your email address:

How long have you been teaching the IELTS Task 2 Writing skill?

- Less than 6 months
- 6 months–less than 1 year
- 1–2 years
- 3–4 years
- More than 4 years

What are the outcome levels of IELTS writing classes that you have taught? (Check all that apply.)

- Basic (below 5.0)
- Standard (from 5.0 to 6.0)
- Intermediate (from 6.0 to 6.5)
- Advanced (above 6.5)

Appendix C. The 20 essays composed without tools (Essay 1)

STUDENT 1

The question of whether fossil fuels should be replaced by green alternative energy sources or not has caused heated debate. Despite certain initial financial drawbacks, such an approach can ensure the sustainable development of humanity and, thus, must be promoted.

Firstly, completely replacing fossil fuels with renewable energy may result in great expenses. Since fossil fuels are still dominant energy sources at the moment and the

majority of equipment still utilize this form of energy, switching to greener alternatives may require tremendous changes to such equipment, resulting in extra costs. Enterprises and households, therefore, may be reluctant to such changes. In addition, renewable energy sources are still developing and their manufacturing costs are still high. Consequently, their exorbitant prices may become unaffordable to deprived regions where fossil fuels may be more economical.

Despite the aforementioned economical downsides, switching from non-renewable energy to greener ones is crucial for environmental protection and natural resources preservation. First, when fossil fuels are used, they create toxic gases that pollute the air and are detrimental to the existence of countless creatures. Therefore, should renewable energy be prioritized, it would curb rising emission and safeguard air quality. Moreover, fossil fuels are finite and they are being depleted at an unprecedented rate. If the usage of those energy sources continues to rise, future generation might suffer from severe energy shortages. Renewable form of energy, however, can be harnessed from nature and can be much more sustainable that accommodates human's demand for energy while not exacerbating environmental degradation.

In conclusion, although fossil fuels still play an essential role in modern society and replacing them may incur certain financial losses, renewable energy is crucial for the sustainability of humanity in the long term.

STUDENT 2

With the industrialization in global scale, energy has been played a essential role in development, which is most made up by fossil fuels. However, fossil fuels are said to be blamed for causing climate change, creating more devastating disasters. Therefore, many advocates suggest that we should alter how we make energy by switching to renewable energy. I totally agree with that.

To begin with, fossil fuels destroy is said to be destroying the environment. Fuels such as gas, petroleum or coals require to be combusted in order to harvest its energy. The end product is not only Carbon Dioxide, a greenhouse gas, but also many other gases which cause acid rains. By avoiding burning fuels, green energy can reduce the carbon footprint and keep the environment clean. Moreover, if we use solar or wind energy, we can save more land by using the existed one comparing to mining a wide area for coals like we have been doing. Many of solar plants nowadays are intergrated with farmland, in which crops such as tomato is planted while the solar panels are installed on top.

Additionally, renewable energy have an infinite potential comparing to fossil fuels which is limited. Which the development of technology, green energy have been more effecient and harvested an significant amount in the past decades. By the year 2021, renewable energy has accounted for 40% in Germany total generated electricity. Also, other neutral carbon fuels might be the alternatives for fossil fuels. Ethanol or E-gas derives from corn or biogas from algea are being researched.

In conclusion, in my opinion, renewable not only helps us fight against climate change but also have an infinite potential for usage and we should totally switch to a greener path.

STUDENT 3

With the advancement of cutting-edge technology, the increasing popularity of distance learning has been witnessed in different educational environments. This trend brings about both merits and demerits to students and teachers in which, I believe, the former far outweighs the latter.

Admittedly, there are inevitable limitations of distance learning, which mostly takes place in the form of an online classroom. First, face-to-face interaction, which is the key to effective communication, cannot be adopted in this form of learning. Although students can be asked to turn on their camera and actively contribute to the lesson, many of them refused to do so for they feel a sense of embarrassment and unwillingness from such a compelling force. As a result, a dynamic atmosphere or sufficient acquisition of knowledge is less likely to happen in a distance learning class. Second, when students are isolated at home learning on their own through a screen, there is hardly any source of extrinsic and intrinsic motivation for them to exert themselves. Little commitment and attachment to the their teacher and friends also means little attempt to be competitive or diligent to perform at their best in class. Learning without such motivation is likely to result in unfavorable academic results and most importantly, unenjoyable learning process.

Having said that, the aforementioned disadvantages can be dealt with if students have proper awareness of independent learning, which can fully be assisted through the benefits of distance learning. The most significant advantage of learning from home is its being time-and cost-saving. This also means students can set aside more resources to invest in their study. For example, instead of spending time going to school, they use it in a more meaningful way for lesson revision or preparation; instead of paying high tuition fee for a traditional class, they use it more wisely to buy materials or assisting tools. Another positive feature of this tech-driven learning environment is that it enables unimaginable possibilities of various teaching and learning styles with the use of advanced digital platforms. As long as teachers and students can take full advantage of these functions, any problem, regardless of less interaction or motivation, can be tackled radically since the huge information and interesting learning activities they gain are already the biggest advantage over any other drawbacks.

In conclusion, the benefits that distance learning generates are much more remarkable than its disadvantages. This all depends on the capabilities of teachers and students to make the best out of technology to carry out this modern form of classroom.

STUDENT 4

It is no exaggeration that there has remained a topic of a debate as to whether it is better to apply télé-education. While some proponents are convinced that this trend can offer tremendous benefits for learners, its drawbacks cannot be ignored. In the following paragraph, bothe pros and cons of this situation will be mentioned and discuss before I draw the positive points are more enormous.

On the one hand, there are numerous advantages in attending virtual classroom. Firstly, learning via Internet can be time-consuming for both schools and students. For instance, students can learn anything and anywhere provided that they have a computer connecting to the Internet. In terms of teachers in school, they can deliver a lesson to all classes by video instead of teaching the same things again and again. Secondly, online education saves costs for the students and schools. For example, establishments can be reduced heat costs or decline buget for investing in facilities since expanded campus is not necessary. Additionally, e-learning provides opportunities for people who have lived far from the college, the university such as the rural area. The fact shows that many inhabitants in remote rural areas still have chances to enhance their degree thanks to this advanced technology.

On the other hand, online classes offer more emphasis on the analytic aspect of learning and there is less instructor and student contact. It is undeniable that it ignores realistic learning elements and encorages passive learning. For exmample, learner at school how to make new friends, be polite, deal with failure, and most importantly, compete but distance learning is not. It is clear that the students percieve it to be less successful than classroom instruction. For instance, according to a new survey via Internet, a larger percentage of students considered classroom instructions to be much more successful than digital education. Another drawback is that onluene classes cannot handle the thousands of students who attempt to participate in debates. Furthermore, if digital education is intended for disciplines that need preparation, it can be challenging.

In conclusion, the ever-increasing availibility of digital education has gained traction among young adults for many tears and is viewed as helpful by many and less sucessful by others. The focus should be placed on its efficacy in order to identity it as the most cost-effective learning method other than classroom instruction.

STUDENT 5

In recent years, the idea of replacing the old, more damaging fossil fuels with newer, cleaner sources of energy has been a topic of interests to many people, even leaders of countries around the world. However, the author disagrees with the suggestion for realistically, at the time being, even though in more developed countries it is easier to implement renewable energy into their daily life, in the rest of the less advanced countries, renewable energy can still be luxury and is rather hard to get access to.

It is easy to see why people are getting more and more excited about making clean energy the main energy source in everyday life. The benefits of using solar, tidal, wind energy have been extensively studied and are now essentially public knowledge. As the name suggests, the energy sources in question are not only clean, but also basically infinite, making them much more preferable choices compared to fossil fuels. In addition to that, it is now easier than ever to bring renewable energy into your house. A study conducted by the University of Texas shows that the price of a solar panel has reduced by 95% since 2002. And even if it is expensive to install solar panels in the first place, it is also widely known that the economic of this choice would work out in the long run,

too.

Having said that, clean energy is only cheap and easily accessible in more developed countries in the world like the US, the UK, or western countries. Basically only in countries that have a strong economy or have had great initiatives regarding the use of renewable energy. But once we look at countries that are not as developed both financially and technologically, the dream of completely replacing fossil fuels with clean energy sources seems like a distant dream. A rather painstakingly clear example would be the use of electric cars versus normal cars. In the US, where companies like Tesla is strong and has established an extensive network of charging stations across the country, it is easier and more reasonable to purchase an electric car, despite its expensive cost. However, in other countries, namely Vietnam, the price of an electric car is not only high but there still exists a lack of charging stations, or at the very least, not enough so that the owners of the new electric cars can feel safe and comfortable using one in their everyday life.

In conclusion, even though the idea of replacing fossil fuels with renewable energy is definitely an honourable one, with the current situations, it is undeniable that we are still quite far from making that dream a reality. This shows that more progress and work still needed to be made so that one day everyone can utilize clean energy sources in their day-to-day life.

STUDENT 6

The topic of whether renewable energy should replace fossil fuels is still debatable among academic community and people in the globe. In personal opinion related to this topic, fossil fuels are likely to be replaced by renewable energy in long future for sustainability and it would take a journey to transform the habit of using fossil fuels with renewable energy in daily life.

In modern society, common vehicles and machinery such as motorbikes, cars, power backup plants, etc... are using fossil fuels as main source of energy to operate. Notwithstanding the efficiency and effectiveness, fossil fuels are named one of reason for high carbon foot print on planet. Moreover, fossil fuels has limit resource, hence, it is inevitable that one day it will reach the end point of harvest.

As a result and consideration of sustainability in environment and economic factors, the renewable energy appears as alternative solution for traditional fossil fuels. For example, since 2022, people has been more aware of environment issue and willing to move from gasoline-powered engine cars to electric cars like Tesla series. The reason is more than just environmental concerns, it is also economic as the gasoline prices increase due to the scarcity which make the electricity bill for charging a Tesla seems lower than the gasoline bill.

In conclusion, with the high awareness of people about environment and the inevitability of the end of fossil fuels resource in near future, the renewable energy will become the suitable alternative source of energy replacing fossil fuels. The challenge to science and economic community is to lower the cost of production of renewable energy and apply the renewable energy in daily life operation similar with the time moving from steam engine to gasoline engine.

STUDENT 7

Distance learning has become more and more common worldwide, especially since the Covid-19 pandemic. This essay will discuss the advantages and disadvantages of this new trend in global education.

When it comes to the benefits of online learning, one must admit that low cost is its prime strong point. Because distance education does not require physical facilities such as lecture halls, libraries and campus, the tuition will decrease considerably. The second merit of online schooling is its ability to reach a wider range of learners, in comparison with in-person education. Distance learners will merely need a reliable internet connection to join the class. Therefore, an education program will probably be attended by millions of learners worldwide.

However, distance learning do have its drawbacks. Educators are concerned about the poorer quality of this kind of learning, because of several reasons. One reason is the looser control of instructors over their learners. Online education organizers expect their learners to be self-disciplined, which is not always the case. Another negative point of distance learning is the lack of contact between teachers and learners, and among learners themselves. As social interactions are important parts of education programs, the limited socializing of online learning may prevent students from attending such courses.

To conclude, I believe that the positives of distance learning outweigh its negatives. It is very advantageous that online learning is available to far more learners than physical education. Knowledge is significant to both individuals' and countries' development. And in cases of required limited social contact such as the Covid-19 pandemic, online learning is the only solution for education to continue. As a result, governments should invest more on establishing or consolidating their current distance educational systems to meet the requirements of the future.

STUDENT 8

Some people believe that fossil fuels should be replaced by alternative energy. Personally, I partly agree with this statement.

On the one hand, there are several reasons why people should use renewable energy instead of fossil fuels. First, it reduces polluted environment. The amount of carbon emission will not be released when using alternative energy so it reduces air pollution. In addition, polluted water might be lessened because there are not any oil spill cases. Second, renewable energy is endless so it can be used for long-term. For example, wind energy can provide electric power for a whole city without fear of power cutting.

However, I still believe that alternative energy cannot replace fossil fuels due to these reasons. Firstly, installation fee of renewable energy is quite expensive which can create a heavy financial burden on the government and companies. For example, the

average cost of installing a wind turbine might be up to 4 million dollars and require a wind farm with hundreds of turbines to supply power. Secondly, some certain machines still need fossil fuels to run such as vehicles. Therefore, it is hard to get rid of fossil fuels.

In brief, although many people think that renewable energy should replace fossil fuels because of its advantages, I believe that it is difficult for countries to replace fossil fuels.

STUDENT 9

Education has recently witnessed the uprising trend of distance learning as a result of development in technology. Besides many advantages brought by this new type of learning, there are also some concerns about its accessibility and quality in comparison with formal education. Although regarding some of the disadvantages, I strongly believe that the advantages of distance learning outweigh its drawbacks.

One of the most significant advantages of distance learning is that it enhances the accessibility to education. To some people who are suffering from disability or mental disorders, they might not have the school environment that adequately meet their condition's demands. With distance learning, they can adjust the study load according to their needs and focus on learning rather than dealing with other problems such as discrimination arising at school. Moreover, distance learning also bridges the gap between school and students under some circumstances that include social distancing. During the COVID-19 pandemic, many schools turned to e-learning to keep up the learning process while still maintain safety for both teachers and students. Another advantage is that distance learning provides a wider range of choices for people when learning. With e-learning platforms like Coursera or edX, people can now explore different horizons of knowledge according to their interests from their home without having to spend a long time in pursuing a formal degree.

Some people may argue that the requirement of Internet for distance learning is an obstacle for people having financial problems or living in poor conditions. However, there have been many charity organizations devoting to provide free Internet for people in harsh conditions, not to mention some community centres with computers that could be used for distance learning. The quality of distance learning's programs are ensured as they are provided by qualified institutions.

In conclusion, distance learning has bridged the gap of education and students under many circumstances. Despite some drawbacks, I believe that the advantages of distance learning are undeniable.

STUDENT 10

Online learning is gaining in popularity, especially during the COVID-19 pandemic. In my opinion, this trend has more benefits than drawbacks.

Admittedly, there are several disadvantages to distance learning. Firstly, it can be a less personal experience than learning in a traditional classroom. Without face-to-face interactions, for example, relationships between teachers and students rarely amount

to anything more than a pay-to-use service. Another common problem with online learning is that internet connection is not always reliable, making the teaching and learning quality less than desirable.

Nevertheless, the advantages offered by this form of learning are far more significant. The convenience it brings allows people to learn from anywhere and eliminates the need to travel back and forth to attend their classes, which can be especially helpful for those residing in remote, rural areas. With online learning, overseas courses have also become more accessible, allowing people to pursue their education without the undue financial pressure of extra travel, food, and accommodation expenses. Also thanks to distance learning, even seemingly unprecedented disruptors such as the COVID pandemic could not stop education from taking place, allowing teachers to keep their jobs and students to safely learn from the comfort of their own home.

In conclusion, despite its limitations, I believe that distance learning brings significant benefits to individuals and society as a whole and therefore should be further developed.

STUDENT 11

There is a new way of learning named distance learning, in which teachers and students participate in virtual classes by using some technological devices such as computers and headsets rather than physically attend to onsite classes. While there are some drawbacks, I still believe that this way of learning can bring various benefits.

There are some disadvantages for learners and educational institutions when it comes to distance learning. Firstly, students who have problems in concentration may find it difficult to effectively study online lesson as they can be distracted from their surrounding environment such as noise, other entertaining programs on their own computer, etc. It is obviously that students can focus on their lesson better in their traditional class under teacher's direct instruction and observation. Secondly, educational institutions utilizing distance learning in their teaching program may need to put a great effort in assessment. According to a recent survey in testing, students tend to cheat by opening another tab when doing online tests. If the institutions want to mitigate this problem, they have to invest on some applications to control students' screens while they are taking online test, which might be at a high cost.

Although there are some negative sides as mentions, I still believe that distance learning has been a huge development in the history of education as it has offered a range of benefits. It can not be denied that distance learning has significantly reduced costs in education. With distance learning, there is no need for students to travel a long distance to hire accommodation to study the courses that they desire. Moreover, the fees of online courses are normally more reasonable than those of the traditional classes because the costs for hiring venues or facilities are cut off .This helps enable students to afford them even when they don't have enough financial support from their families. This means a lot when distance learning has created more learning opportunities for people, especially those who cannot afford expensive courses and conditions for completing them. Another benefit that distance learning can offer us is the flexibility. With it, students become more active in choosing studying time. They can also review their lessons anytime by watching the recordings of the class. This is very

ideal for students who have busy schedules or working at the same time with studying.

In conclusion, despite of some issues related to testing assurance and students' concentration, distance learning is a promising way of learning because it reduces costs and creates more studying opportunities for everyone.

STUDENT 12

Distance learning have been popular for over two decades. Particularly since the outbreak of Covid-19 pandemic, distance learning have in fact become a remedy for educational institutions worldwide to continue their teaching and learning practice. Learners all over the world take this mode of learning with different attitude due to the fact that beside benefits of distance learning, it still has certain drawbacks. This paper will prove that the positive sides of distance learning far outweigh its negative sides.

Distance learning is a trendy learning mode for several reasons. First, it can bring the course content beyond the classroom wall. Instead of travelling to school, teachers and learners of distance learning courses can sit comfortably in front of the computer and conduct the class sessions. That has not only reduced the time for lecturers and learners to show up physically in classes but also make courses taken charged by a lecturer of a US university accessible to learners all over the world. Students in Vietnam or South Africa can attend a course on Data Science delivered by a lecturer in New York University. Second, the tempo-and-spacio flexibility of distance learning has made education possible regardless of the condition. Take for example, during Covid-19 pandemic, distance learning became a remedy for schools and universities worldwide to continue their teaching and learning activities. It can be said that distance learning seems the most flexible mode of learning thanks to the mentioned advantages.

Distance learning, however, still has certain disadvantages. One obvious minus point of this learning mode is its total dependance on the quality of equipment and Internet access. If for example, the electricity at the lecturer's house is cut off at the middle of the lesson, the session has to stop. For synchronously interactive courses, when the Internet access of some students is not stable, the online discussion with the lecturer and his classmates can become an annoying experience. Second, distance learning is sometimes criticized for not being able to provides the class with small-group interaction opportunities. For instance, when we study distance learning course, it is not as easy to do pairwork or group work as in the offline classes.

In conclusion, distance learning has altered the mode of teaching and learning in many positive ways such as providing unlimited and equal learning opportunities for learners of all age groups in all parts of the world, reducing the traveling time to get to classes, and making teaching and learning happen in almost severe situations such as pandemics. However, distance learning has its own limitation in depending on the quality of techonolical facilities and Internet access and less interation opportunities between learners. However, it can be seen that the benefits of distance learning far outweigh its deficiencies.

STUDENT 13

In recent years, online learning has been common because of its benefits. From my

perspective, this has both pros and cons, and the disadvantages outweigh the advantages.

It is undeniable that learning from home helps learners save time and money effectively. Firstly, because people can learn everywhere, they do not spend time on going to schools or universities. Therefore, it can save their time, and they will have more time to do other tasks. In addition, they do not have to hire a accommodation to live if they learn a school which is far from their school. Thus, they can save money. For example, some young people choose to attend distance programmes of universities instead of studying abroad. This helps them save a lot of money and they still have time for working.

Although distance learning has benefits for learners, it also has some drawbacks. When learners learn from video lessons, it lacks interactions between students and teachers. As a result, some people feel bored and have no motivation to maintain their learning. In addition, if they do not understand lessons or have problems during their learning, they cannot ask teachers immediately and they have to wait for a long time. In some situations, as learners sit in front of screens for a long period to study, it can cause some diseases related to their eyes and back. For instance, after attending some online courses a few weeks, some people give up their learning since they realise that it is not effective that they expect.

In conclusion, distance learning brings some benefits for learners but it still has drawbacks. Therefore, people should consider carefully before learning from home.

STUDENT 14

There are some concerns about which kind of energy will be the future of human being. There might be traditional source like fossil fuel which we extract from earth or alternative one like wind, solar, hydro source of energy. From my point of view, there is no doubt that there will have a big change in how people use energy.

First of all, burning source of energy brings us heat to generate power for running different types of machine, but what is burned will be abandoned into ash and smoke. This brings to a conclusion that energy from fuel is limited and it depends on amount of that energy in our Earth.

Secondly, there is a rocket trend in population that we all know our Earth excess 8 bil soon, the more human beings the more energy consuming we need. As a result, even we know that human exist 10 thousand years ago, and we have half of fossil fuel left, but there is no fool that people will have 10 thousand years left. The later we change how we use energy, the sooner we get near our extinction day.

Thirdly, renewable energy consist of different kinds of energy which exist around human like solar, geothermal, wind, wave of sea and they are abundant, so there is no need in concerning about amount of these source of energy. In the other hand, there is a high cost in researching and applying those source of energy, but after that, the cost for maintaining these equipment can be less than the amount of money we need to invest in solving our environment effected by air emission of fossil fuel.

To sum up, there is no doubt that we should replace tradional kind of energy into

alternative energy as a consequent of development of human beings. There are many extinct species before us so change or end will depends on our next step from today.

STUDENT 15

With the advancement of technology, remote learning has become more and more prevalant. Although this trend has certain drawbacks, I believe that the advantages are of greater importance.

Opponents of online learning might argue there are some disadvantages regarding the lack of communication and supervision between teachers and students. Concerning the former, online learning often lack students and teachers engagement as they can only see each other through the screens. This could lead to some misunderstanding due to the fact that human often communicate through body language and tone of voice more than verbal language. As a result, when teachers give instructions, they might not aware how students react and therefore unable to measure their understanding and performance. Regarding the latter idea, online class often lack disciplined aspect compared to offline class. This is because students often are not willing to turn on their cameras during the class and often make an excuse such as their cameras are broken or their wifi connection are weak. As a result, teachers might face a challenge when they want to supervise students.

Nevertheless, I believe that the perks of distant learning is far greater for several reasons. Firstly, online learning allows people to learn at the comfort of their own home as they do not need to experience the hassle of commuting to school or waking up on time. Moreover, as the class is conducted online, people now have a chance to choose courses from a variety of school. For example, we now can learn courses from pretigious school such as Harvard or Yale within a click of our mouse. This would bring tremendous benefits for those who previous did not have an opportunity to access such precious resources. Another benefit would be the flexibility that distant courses offer. People who are adults or even our grandparents could register for a learning program as it allows them to study at anytime they want. This not only benefits the individuals but also the governments as they now have a well-educated workforce and citizens.

In conclusion, although there are some drawbacks of online learning, I still belive that they pale in comparison with the advantages for the aforementione reasons.

STUDENT 16

In my opinion, distance learning's advantages does not outweigh its disadvantages. Because in some art subjects like guitar or drawing you should learn face to face with teacher to get professional level. Not only learning face to face, you'd better learn 1 by 1 (1 mentor with 1 student).

For example, my brother comes to Saigon from France to learn Rock Guitar. I introduced him to a master of rock. My bro learns 3 days a week with the master. The master teaches him very carefully from the beginning: notes, tempo, chords, harmony

and improvising. The master even took him to the guitar shop to buy a new one and gave advices that he should not buy a guitar with whammy bar, cause it is very difficult to change the strings. The master also gave him advice to buy an amplifier with reasonable price and good quality also.

My bro is eager to learn and the master teaches him step by step, looks at his performance and reaction to adjust his mistakes, even counts tempo for him. He makes progress very quickly after 3 lessons because of learning face to face. His mom asked me if he can continue learn guitar online when he comes back to France. The master said that learning online is very inconvenient and very hard to get professional level. Because it is very hard to explain through a screen of laptop or desktop. Furthermore, the wifi often lag and when the teacher says something, it takes some minutes for the student to get the signals and in that waiting time he may ask again or another question. That will make them confuse and not match with each other. Besides, there are some techniques that you should be taught face to face, it will be much better.

STUDENT 17

Nowadays, the modern life gives us many chances to choose different learning ways as we want. Therefore, we can choose distance learning instead of learning in person. Although distance learning has many advantages, especially in the Covid pandemic, it still contains many disadvantages that prevent learners from self-development.

Firstly, distance learning reduces the interactions between human and human, which is very essential in the normal life. At usual, people interact with each other not only by talking through voice, but also physical contacts such as hugging, looking each other, body languages... Distance learning somehow makes people more passive just by looking at the screens with no direct talking and laughing with the lecturer. This boring enviroment will decrease the learners' motivation of studying.

Secondly, distance learning reduces learners' focus and attention in learning. In the traditional way, people sit in class and there will be rules such as "no phones" to keep learners from distracting. By contrast, in distance learning, people usually do multitasking like texting, playing games, chatting while studying and therefore, they cannot focus on the lecture and studying. Their learning becomes inefficient.

Thirdly, distance learning depends mostly on the Internet. Therefore, in the rural areas where the Internet is not stable, it will be big problem for the learners choosing this way of studying. The covid pandemic shows that half of the lectures given were interrupted by the wifi. Therefore, many students find it hard to understand clearly what the teachers delivered. Students also felt it harder to slow down the lecture because they couldn't raise their hands or interact with teachers directly.

In conclusion, the better option of learning is going to class and getting the knowledge directly because it is not only knowledge given but also the interactions and comprehension. Distance learning is really helpful in the pandemic time but it cannot completely replace the direct learning in which the learners can really focus on their studying and develop themselves.

STUDENT 18

Having so many advantages like saving cost, accessibility, bridge geographical distance, schedule flexibility, reduce social anxiety, etc; distance learning is getting more popular,

making education more accessible. Even so, on-site learning still have the strong benefits that distance learning can't replace.

Human interaction. Just like how a video call can't beat a warming hug, learning through a monitor screen lessen the chances for student to hone their soft skills in socializing.

While face to face teamwork helps student learn to observe & self-control to give proper reaction in seconds, distance learning delay reacting time and provide a free-to-explode environment

which at first seems to be stress-free but later revealed to be vulnerable. Like how distance learning reduce social anxiety but on-site learning help students to cope with it better & better.

Hence the failing of adaptability. Schooling is not just about getting knowledge, it's also about putting oneself out there into a small society, learning the beauty of diversity, experiencing

and contributing as part of a whole.

Last but not least: self-control - the number one factor that decide the success of distance learning. While giving students a freedom to schedule their own learning, it gives huge challenge

especially to younger students. With traditional schooling system, there is an urgency to push students into taking action. Distance learning shows them a freedom to take it easy.

Not to mention how distance learning is not suitable for experimental subjects, with just the 3 points above, distance learning is great for supporting education but can never be a replacement in my opinion.

STUDENT 19

Due to the COVID-19 pandemic outbreak, distance learning has become more and more popular since meeting in person in classrooms is impossible during social distancing. Distance learning has both advantages and disadvantages, however, in my opinion, the disadvantages clearly outweigh the advantages.

The first advantage that can be stated are: with the help of nowadays technology, learning in distance is so much easier and more convenient than before when we do

not need to travel far from our home to attend classrooms or seminars; especially during social distancing due to the pandemic, people cannot travel or even go outside. Moreover, distance learning is also less costly. To be more specific, educational institutions do not have to rent a place to teach; they do not have to pay a lot for the electricity or water management fee. Therefore, it would be much less of hard works and expenses that they need to deal with.

Otherwise, distance learning also have side effects that cannot be a substitute for direct learning at schools. The lack of interaction between students and teachers would be a problem, which can lead to the slack off of studying. It would be inevitable for students to focus on studying since teachers have very limited controls over them. Far from my experiences, I do not find online classroom is efficient, when it is very hard for me to focus on what the teachers saying. Studies have shown that direct interaction is very important for human to stay connected and build deeper relationship. In addition, distance learning may mean that sometimes we do tests online, leading to cheating, which will not affect truthfully about actual studying.

To conclude, I find that distance learning surely has a number of advantages thanks to the development of technology, however, direct learning, in my perspectives, is more beneficial and more efficient towards studying.

STUDENT 20

Fossil fuels have been the main source of energy of human civilization since the dawn of mankind. The evolve of fossil fuels developed from the very first stage of using wood, into coal and finally petroleum, gas, oil in the modern days.

At first it has been pleasant using those sources of energy, but as time goes on, especially after the WW2 with the rapid developement of modern industry and with its population fast growth the world started to have doubt and worries on fossil fuels: its quantity is dropping and it is causing environment issues. With those issues appearing, human start to research for other sources of energy and scientists introduced "renewable energy". such as solar energy, wind energy, tide energy, geothermal energy, etc and many claim that renewable energy should replace fossil fuels. In my opinion, I agree that renewable energy should replace fossil fuels, as much as possible to our capacity and knowledge.

First of all, fossil fuels cause many negative environment effect such as global warming, acid rain, air pollution and those effects are irreversible and uncontrollable. For exemple in my country nowadays due to heavy traffic which emit all kind of by-product of fossil fuels burning process, the rain already started to tastesour and irritate the eyes when caught, which never happened about 30 years ago (an early and mild stage of acid rain). Secondly, fossil fuels are non-replacable and non-recycleable, so with the current rate of usage of the whole world, they will be depleted in the near future.

As a result, starting from now human should concentrate on researching and developing technology for renewable energy as hard as possible. With the current stage of technology renewable energy such as solar or wind power still cannot replace fossil

fuels because of their high price compared to low ratio of usage. Sometimes the post-products of renewable energy are even more harmful than fossil fuels itself (solar power panels, for example.)

To conclude, I agree that fossil fuels should be replaced with renewable energy, as fast and as thorough as possible. We must concentrate on the research on renewable energy to make them cheaper, more efficient and without harmful post products.

Appendix D. The 20 essays composed with tools (Essay 2)

STUDENT 1

Remote learning has become ubiquitous in modern society. Despite drawbacks in terms of decreased real-life interactions and increased distractions, the financial and social benefits of such a mode of learning are far more significant.

Due to its inherent convenience and lack of real-life interactions, online learning may distract undisciplined students and worsen their communication skills. Remote learners may be tempted by numerous distractions on the internet that would be prohibited under conventional classroom conditions. As a consequence, students who lack self-control may squander their time and not fully focus on online lessons, resulting in declining academic results. In addition, since online learners may suffer from insufficient social interactions, their communication skills may be exacerbated. After the Covid-19 pandemic in Vietnam, for example, a number of students have reported difficulties when re-interacting with their peers in real life.

Despite the aforementioned downsides, remote learning offers much greater merits for learners worldwide. Thanks to technological advances, students from all over the world can now have access to prestigious educational institutions regardless of geographic distance. Off-campus learning has empowered them to pursue a world-class education at affordable prices, especially for underprivileged pupils. In addition, distance learning saves precious time for learners. Not having to spend time commuting, students can enjoy extended relaxation time which can enhance their mental health and academic performance. Moreover, online learning offers students with flexibility as they can choose a suitable time for their education, boosting their productivity and learning effectiveness.

In conclusion, although distance learning may be distracting and detrimental to the communication skills of some individuals, its economic and social upsides to learners are much more considerable.

STUDENT 2

With the development of technology, classrooms are now digitalized and can be attended online. However, there are benefits and drawbacks to remote studying. In my point of view, the disadvantage overshadows the other.

On the one hand, having an online education can hinder many students. One of the many problems is its difficulty in concentration. Without any discipline, students can be distracted by more enjoyable websites or applications that are available instead of taking the lesson seriously. Subsequently, this will result in lower efficiency and make studying less effective. Moreover, it can negatively affect users both physically and mentally. For instance, staring at the screen for a long amount of time can cause myopic, dry eyes or neck pain. Online study is also said to set up a physical boundary between classmates preventing them from socializing which will negatively affect students' mental in the long term.

On the other hand, online studying can have some bright sides. By studying online, less transport is required, which can benefit those who are far from school. Using the internet can also make studying more fun. With a wide range of tools and webs, these can facilitate teachers and students in innovatively approaching subjects.

In conclusion, even though online studying can facilitate many different tools and does not need to move a lot, it is not worth taking compared to the many hindrances it creates including its negative effect on mental and physical health.

STUDENT 3

The use of fossil fuels as an energy resource has created many problems for the natural environment. Therefore, there has been an increasing demand for an alternative source, namely renewable energy, to mitigate fossil fuels' dangerous outcomes. From my perspective, I totally agree with this solution because of the benefits that these clean energies bring about.

First and foremost, the amount of fossil fuels is limited, whereas the amount of renewable energy is infinite. The more we use fossil fuels, the more they become depleted, which results in their continually rising cost to prevent people from using them up. Gradually, people can no longer afford to buy fossil fuels in order to generate energy for their daily needs. By contrast, natural energy resources like solar or wind power can be harnessed from endless force. Therefore, people do not need to worry about facing the risk of their depletion. Some may argue that the cost of installation of a system to collect this type of energy is higher than the current cost of fossil fuels. However, in the long run, it will be much more cost-effective as the energy from nature is automatically converted into electricity without any further expense for by-processes.

Furthermore, it is obvious that the use of fossil fuels leaves detrimental impacts on the environment while renewable energy is totally clean and produces no harmful by-products. The burning of coal to generate energy, for instance, releases harmful emissions to the environment, which increases the carbon footprint and contributes to global warming. Another typical example of the adverse effect of fossil fuels is that the exhaust fumes from those means of transport that use petrol are among the factors leading to air pollution. Meanwhile, solar power generates energy directly from the heat and light of the sun without any by-products; and an electric car is far more environmentally friendly than a petrol-powered car for no exhaust fumes being produced.

In conclusion, due to their limitless amount and clean generating process, renewable resources deserve to be practically applied in daily life and mass production for the

sustainable development of our globe. The government may need to fund the beginning stage of installation but this can bring about positive effects in the long term.

STUDENT 4

It is no exaggeration that there has remained a topic of debate as to whether it is better to use renewable sources. While some proponents are convinced that this trend can offer tremendous benefits for people, its drawbacks cannot be ignored. In my opinion, humans should manage the need for these sources because they both have their necessities, such as renewable sources (solar, wind, air power) that are not polluting the environment.

On the one hand, the first aforementioned view, in which some folks argue that resources like solar and air power are mesmerizing alternatives to other fuels though. In addition, there are a plethora of merits of those fuels, for example, these are free of cost for everyone, easy to access and their superior benefit is, these are pollution-free. Although, with the use of these resources nations can protect their environment.

On the other hand, there is another group of people, who hold the opposite view. The community should continue the practice of fossil fuels because there are some reasons for it. One of the reasons is, the equipment that we need for renewable sources is highly expensive so everyone cannot afford them. Secondly, air and sunlight sources depend on the weather, but on the other side people can get fossil fuels in any condition and a person does not need to buy any expensive equipment for it. For instance, in Vietnam, climate conditions remain extremely challenging approximately every year, so the government would not rely on these resources.

In conclusion, in my perspective, humankind should manage the usage of all fuels because these all have significant characteristics like renewable fuels are cheap and unpolluted but not be accessible in atrocious climate conditions whenever fossil fuels are available all seasons.

In conclusion, while renewable energy is cheap and unpolluted, it cannot be accessible in atrocious climate conditions. In my perspective, the government should carefully manage the usage of all kinds of fuels because they have both upsides and downsides.

STUDENT 5

To satisfy the thirst for knowledge of studious students across the world, in 2008, MIT posted a series of videotaped lectures from some of its most famous courses onto its website. Several universities started to follow suit and then came the birth of online education platforms like FutureLearn and EdX that provide quality courses to students from anywhere in the world that have access to the Internet. And after the turning point that is COVID-19, distance learning is now as normal a term as a term can ever be, which leads to the question of whether the benefits of online learning can outweigh its shortcomings. The author believes that the pros of distance learning, its many conveniences, do outweigh its cons, the lack of interaction.

When it comes to distance education, one of the hardest obstacles to overcome is how interaction with the teachers and fellow students is non-existence. Besides the weekly whole-class discussion, where everyone can ask, answer, and comment on each other's topics, there is undoubtedly a lack of communication and interaction compared

to an in-person classroom. This would easily lead to study sessions that can be seen as boring and not interactive, thus requiring a high level of discipline from the students themselves. A report from EdX, one of the leading online education platforms, in 2018 shows that around 65% of students participating in massive open online courses (MOOCs) failed to finish the courses. Although to be fair, this is on the students rather than distance learning itself, it does show difficulties presented to online educators: to think of a way to manage and hold the attention of learners.

On the other side of the coin, it is undeniable that distance learning has opened up doors like never before to students across the globe. With its flexible timing and the fact that students can study from anywhere that the Internet can reach, courses like MOOCs have proved to be a game-changer in the education field. Furthermore, thanks to platforms like FutureLearn, EdX, and many more popping up, students can now get access to courses provided by the best universities in the world for free. Such platforms now also provide pathways, a set of courses that can help students get real certificates that will get them real jobs in the real world. Thanks to distance learning, schools, and teachers have constantly been trying to renew themselves, to challenge themselves to make studying more interesting, more interactive to reach a wider audience of students.

To sum up, besides the fact that learning from afar is not as fun and engaging as learning in a real class, the benefits and opportunities distance learning brings about without a doubt make up for its flaws. And with online learning only becoming more popular and more normal for students, the author believes things will only go up from here.

STUDENT 6

The popularity of online education and remote learning is rising. The approach has both benefits and drawbacks for consideration. This essay makes the case that distance learning has more drawbacks than advantages. The rationales would be given in detail in the following paragraphs in the given Vietnam context.

During the period of the COVID-19 pandemic in Vietnam, people were required to adhere to social segregation and lock-down regulation from the government and academic facilities closures. Hence, the classroom had to be moved from offline to online via Zoom or Microsoft Team. The faculty and staff of academic facilities uploaded the digital content to the cloud services therefore students can get an education without the interruption of the social distancing period.

In the beginning, it was very helpful that students can go to class and learn without going out of their homes. All they needed was a computer, a solid wifi connection, and a selection of appropriate software. The advantages include maintaining social distance and keeping up with test schedules while saving time and energy on the commute from home to school.

However, the disadvantages of remote learning emerged in both technology and human factors. Regarding technology, the WIFI connection in Vietnam is not stable and sufficient to handle large group conference calls which results in frequent lagging and disconnection in online classes. In terms of human factors, because of staying at home for a long period to study online, students began to show symptoms of mental illness,

dependence, and cheating during online examination time due to a lack of face-to-face supervision and administration from the faculty and staff of academic facilities.

In conclusion, the disadvantages of distance learning may outweigh the benefits if the approach was not properly used.

STUDENT 7

Environmentalists, who are concerned about global pollution, have been requesting that governments should switch to renewable energy from oil and gas. I completely concur with this viewpoint for the reasons listed below.

First, because it does not emit as much hazardous exhaust as fossil fuels do, recyclable energy can help to prevent the environment from becoming increasingly polluted. Exhaust from fuel vehicles and machines can be attributed to the heavy pollution in developed countries over the past years. As a result, nowadays, vehicle manufacturers are trying to produce hybrid or electric bikes and cars to reduce exhaust to the least possible amount.

Another reason I prefer renewable energy is that it is significantly less expensive than fossil fuels. While fossil resources are limited and lay deep under the earth's surface, which requires higher costs to exploit, clean energy can be produced with unlimited free resources, namely wind, solar light, and seawater. As investors do not have to pay much for the input, the product cost will be lower.

In conclusion, I entirely support the usage of clean energy in place of conventional energy sources in order to protect the environment and also because it is more affordable for consumers than fossil fuel.

STUDENT 8

Distance learning appears to be gaining popularity. Despite some obvious disadvantages of this trend, I believe its benefits outweigh drawbacks.

On the one hand, there are some major drawbacks when students study at a distance. Firstly, it decreases effectiveness because students easily get distracted by the surrounding environment. For example, young learners find it hard to concentrate fully on the lesson because they can stop themselves from using social media or playing games while studying at a distance. In addition, students' comprehension cannot be checked immediately by teachers. Secondly, students need to have basic elements such as computers, electronics, and the Internet to learn at a distance. This may be a financial burden for poor students.

On the other hand, I believe that the benefits are more significant than the disadvantages because of those reasons. Distance learning is very easy to access, especially for the poor. People from rural areas where schools are few can easily access education. Moreover, with only a laptop and the Internet, people can learn anywhere they want. For example, several people can go to a coffee shop to study instead of going to school which is far away from their house. Moreover, students do not need to participate in extracurricular activities so they have more time to discover the course more deeply. In addition, The tuition fee of distance learning may be much more affordable than the regular fee so people can easily access a variety of fields.

In brief, I would argue that the advantages of distance learning are more significant than the disadvantages.

STUDENT 9

The alarming conditions of natural resources have called for a change in sources of energy. As a result, many people believe that fossil fuels should be replaced by renewable energy sources. In my opinion, although renewable energy sources can reduce the adverse environmental impacts of fossil fuels, they still have potential ecological risks and are not the most optimal energy source.

It's hard to deny that renewable energy sources can mitigate the negative environmental effects of fossil fuels. Solar power or water power do not emit harmful emissions or produce toxic sewage that can pollute the surroundings. It has been reported that the air quality in the capital of the Netherlands has increased by 40 percent since the vast majority of solar panels were installed. Besides, while fossil fuels depend on limited natural resources which are gradually depleted, renewable energy as its own name states, is "renewable" and almost unlimited.

On the other hand, current renewable energy sources still pose ecological risks. First, the process of installing wind turbines or solar panels is criticized for producing more carbon footprints than the amount that they help reduce, not to mention the maintenance and replacement process. Another problem is industrial waste since most of the materials used for solar panels, turbines, or water wheels are non-degradable. Moreover, some wind farms are built in the wrong places which threatens the migratory flow of native birds as the turbines are a deadly danger to birds if hit.

In conclusion, although renewable energy sources help alleviate the negative environmental implications of fossil fuels, they still pose ecological challenges and are not the optimal energy source.

STUDENT 10

The environmental impact and finite nature of the fossil fuels we commonly use have sparked discussions over their replacement. Some believe that renewable energy would be an ideal alternative, and I completely concur with this suggestion.

The first reason for my agreement is that this switch could help protect the environment. Toxic emissions generated by the combustion of fossil fuels have long been recognized to have a negative influence on air quality, affecting the health of not only humans but also wildlife. These gasses also contribute directly to the greenhouse effect, leading to global warming, climate change, and other related environmental problems. Another important reason is that these fuels are not infinite. If people continue to consume them at their current pace, these resources will soon be exhausted, and the switch in question would inevitably ensue.

Some may object to this shift, citing the enormous expenses required to develop the necessary infrastructure and technology to implement it globally. However, I believe that it would be well worth the investment, as it would help us lead a more sustainable life by limiting our carbon footprint and environmental impact. Furthermore, as previously stated, the transition to renewable energy sources would be unavoidable once current supplies of fossil fuels are depleted, necessitating prompt action and preparation from governments worldwide.

In conclusion, I completely agree with the suggestion that fossil fuels be replaced by their renewable counterparts. If our dependence on these sources is not reduced and

no action is taken to replace them, their exhaustion will soon lead to a global energy crisis and irreversible environmental degradation.

STUDENT 11

Although fossil fuels have been a significant part of human life since the beginning of modern civilization, many individuals believe that renewable energies should take their place. I personally quite agree with the viewpoint since renewable energy is more environmentally friendly and because it may be used indefinitely by humans in the future when fossil fuels run out.

It is clear that renewable energy can bring various benefits to human beings due to its inexhaustibility and environmental friendliness. Firstly, no one can deny that many forms of sustainable power, such as solar energy and thermal energy, are endless and long-lasting. Unlike fossil fuels which are going to run out in the near future, renewable energy can provide an endless source of power. Secondly, the production and usage of this new kind of energy do not harm the environment. Taking cars running on electricity as an example, they do not create CO₂ emissions, unlike those using petrol, which enables travelers to protect the environment more effectively, than have a healthier atmosphere.

Additionally, there is an urgent need to replace fossil fuels such as gas, petrol, etc since they are going to be exhausted soon and have created many problems for the environment. According to a recent report by OPEC, the reserves of oil in UAE, which is the largest in the world, are going to be empty by 2050. This has raised a lot of concerns as energy is vital in every aspect of the world. Furthermore, the use of fossil fuels in industrial production and daily transportation has caused enormous harm to the environment due to the great deal of CO₂ emission to the air causing air pollution and water pollution. Therefore, sooner or later, this old form of power should be substituted by cleaner and safer ones.

In conclusion, with the benefits that renewable energies can provide and the need of replacing fossil fuels to protect the environment, I think that these promising kinds of energies should be deployed more not only in the production process but in people's daily lives. Although the procedure of installing renewable energy is costly and requires knowledge, it can bring huge value in the long-term prospect if it is invested effectively.

STUDENT 12

When people find different sources of alternative energies such as wind power, and thermal energy, many people think that these power sources should replace fossil fuels. This paper will prove how this proposition should be partially agreed on.

The search and use of alternative sources of energy should be promoted for several reasons. First, new sources of energy such as wind power, thermal power, or solar power are renewable energies. They can be created by unlimited sources in nature such as the wind, heat, or sunlight. That is the reason why more and more facilities are created and improved to set up the infrastructure for the widespread production, transmission, and use of these energy sources. In Vietnam, for instance, there is an increasing amount of wind turbines established along the provinces in the Central and sea areas. In addition, wind or solar power is proven to release less exhausted fumes than fossil fuels such as coal or oil and gas. That motivates not only the governments but also the producers, vendors, and users to prefer these sources of energy to petroleum or coal.

However, regardless of the economic and environmental benefits of renewable energies, they should never totally replace traditional energy sources such as coal or

petroleum. Concerning the availability of underground coal mines, and oil wells offshore, it would be a terrible waste of natural resources if they are left unused. In addition, taking all the infrastructures and labor workforce living on the execution of exploitation, production, and transmission of these energy sources into consideration, it could hardly believe that why we can stop all these activities immediately without a transition period. Finally, the provision of power from alternative sources needs time and effort; therefore, we still need to partially rely on traditional power sources for a long time.

In conclusion, even though renewable sources of energy can bring many benefits, they should not replace fossil fuels.

STUDENT 13

These days, because of environmental problems, some people suggest that humans should use renewable sources instead of fossil fuels. From my perspective, I partly agree with this idea, but people have to use fossil fuels in some situations.

It is obvious that fossil fuels such as coal and oil will be exhausted in the near future. In particular, people use too many traditional sources for machines and transport, so these sources are running out. Therefore, some scientists believe that people should create alternative energy to replace traditional sources before it completely disappears in the future. In addition, fossil fuels are one of the main causes which lead to environmental pollution. This is because vehicles release a large amount of CO₂ emission every day which causes air pollution and respiratory diseases for people. Therefore, using green energy is necessary to protect the environment.

In contrast, I think that changing renewable power such as solar or wind put a financial burden on some nations because this requires a great of money to build and develop equipment that can create energy. For example, the cost of solar panels is quite costly that some small factories cannot install them. Moreover, some machines need fossil fuel for working, and people have not created new energy yet to replace them. Therefore, people have to depend on fossil fuels until they will create alternative ones. In conclusion, although changing from fossil fuels to renewable energy has some benefits, it also causes some problems for the government. I think that people will not use any fossil fuels in the near future.

STUDENT 14

Thanks to recent innovations in how knowledge is being taught, people are having a variety of sources for learning. For example, there is an up trend in remote studying nowadays. Do the benefits of online education outweigh the drawbacks?

There is no doubt that experience is for sharing. However why we are repeating so many courses and teaching day by day as we have done since thousands of years ago if we could have a great source of informative understanding as an internet source? Everything could be stored at one time and transferred to all students numerous times with no fluctuation. Even students need to pay fees for online courses but overall, the amount of money and time they invest in distance studying is much less than in offline courses. There is no waste in a realistic world, no transportation, no waiting time.

In contrast, learning offline requires much material such as space, and books, and students should follow a fixed schedule. Perhaps, this might be a drawback on the one hand but a promotion for better studying on the other hand. Talent pays 10% in

excellence but discipline does the rest. Learning remotely without any observation from teachers could bring learners an enjoyable atmosphere but there should be control in the result.

In conclusion, virtual learning is the key point for future learning, but students must follow a rule and tight deadlines in achieving their targets. Many examinees pass entrance exams with really high marks due to their learning individually, but many other examinees fail because of their parents' negligence in monitoring and observing.

STUDENT 15

Some people argue that we should opt for green energy such as wind and hydropower instead of choosing fossil fuels. This essay entirely agrees with this idea for several reasons.

Supporters of this school of thinking may argue that there is no need to alter present sources of energy because they are significantly cheaper and require less investment and effort. This is because, in order to employ green and sustainable energy, countries must invest a significant amount of money in research and experimentation in order to properly execute the use of these energies in real life.

However, I strongly agree with the idea of replacing non-renewable energy with sustainable green ones for several reasons. First, the usage of fossil fuels is closely associated with the rate of climate change, which is a grave concern for humanity. As a result, continuing to use fossil fuels to power our world will have disastrous implications. Furthermore, because green energy sources such as wind and solar come from nature, we will never exhaust this energy as we do with fossil fuels. This not only helps people conserve the environment by using ecologically friendly energy, but it is also more sustainable for us in the long run because it never runs out of energy.

In conclusion, even though I understand why some people do not want to replace fossil fuels with other energy, I still disagree with them for the above reasons.

STUDENT 16

I think renewable energy will be the future of the world. Due to the indiscriminate use of natural resources by humans, these precious energies gradually fall into a state of depletion.

In addition, the development of industries that use a lot of natural energy also causes emissions of environmental pollution, the greenhouse effect, and climate change, melting ice, etc. These things, if continued, will destroy the world.

Renewable energy will overcome these risks and provide a better living environment for people as well as save valuable resources for future generations.

The development of green energy and renewable raw materials and fuels also contributes to raising people's consciousness in protecting the environment and the earth we live in.

Explosive consumerism (e.g., over-shopping of unnecessary things, the huge amount of waste by fast fashion brands) not only depletes resources used as raw materials for production but also shows moral degradation and lack of restraint of people.

The use of traditional energy leads to scarce energy resources; increasing prices lead to wars and competition for territory and sea to usurp oil and coal energy resources. Using green energy will reduce this situation and evoke the good moral sense of mankind.

STUDENT 17

In recent years, the quality of the earth is decreasing in an alarming way. Humans exploit fossil fuels, which is harming the original environment around us. It is time to think about using renewable energy instead of fossil fuels because of their benefits.

Firstly, renewable energy like solar power helps us to reduce coal burning. Therefore, it prevents tons of smoke into the environment. It helps to save the cleaner air for people to breathe. Thanks to it, humans are saved from many nasal diseases. On earth, once a person is healthy, they develop.

Secondly, renewable energy helps to save clean water. Many people in Africa are dying because of a lack of water. By using renewable energy which is also a limitless source, people save natural resources for higher purposes.

Thirdly, using renewable energy enhances human consciousness of protecting the Earth. In recent years, our earth has been destroyed by overexploiting and it causes global warming. Global warming causes melting ice and tsunamis which cause natural disasters killing millions of people.

Lastly, using renewable energy prevents wars between countries. Countries compete for mineral sources and it causes instability among countries in the region. Therefore, replacing fossil energy by renewable energy can prevent this issue because they can share a limitless source.

In conclusion, we need to create a revolution in our era that using renewable energy is a big step to save our Earth. It is very essential to create a sustainable environment in which we last our lives longer and healthier.

STUDENT 18

I totally agree that renewable energy should replace fossil fuels. For obvious environmental reasons, people refer to renewable energy as "clean energy": it's a better option for Mother Earth.

Fossil fuels have always been the main energy source for its manufacturing capability but the harmful emission it causes is also undeniable. The resources for fossil fuels also put a heavy burden on poor countries and raised disputes between countries in the past. Countries with fewer resources are given a weaker negotiating position and develop more slowly.

With better sustainability, clean energy solutions have been highly invested over the years by the governments but there are still many problems to be solved.

Hydroelectric energy or wind energy is highly affected by weather conditions. Solar energy seems to be more dependable, especially with the Sahara desert being a giant field for "solar harvesting". Storage capacity and transportability become a problem. We haven't found a way to efficiently change solar energy into usable energy, let alone storing and transporting it to countries around the world. The ocean covers more than 70% of the surface of the Earth, making it a potential source of energy and more accessible when not many countries are landbound. There's technology developed to convert it into usable electricity but it is not industrialized and still under further development.

Renewable energy cannot replace fossil fuels for now but most countries are trying to be less dependent on fossil fuels by utilizing both solutions. Hopefully, in the near future, we can put an end to the fossil fuels era.

STUDENT 19

With the world changing at a fast pace, we do care a lot more about our environment, hence, leading to some thoughts that fossil fuels should be replaced by renewable

energy. I partly agree with this idea due to some advantages and disadvantages that renewable energy can bring to us.

One of the major issues rising nowadays that fossil fuels directly contribute to is city pollution. Utilizing renewable energy in daily life would benefit the environment by reducing city pollution levels since renewable energy uses natural resources such as sunlight, and water strength to produce energy. Therefore, the government is likewise particularly focused on these issues and is assisting the development of renewable energy companies by promulgating supportive regulations. Environmentalists also support the ideas and motivate people to use renewable energy rather than fossil fuels. In that case, to adapt to the changing world, more and more products use renewable fuels.

On the other hand, renewable energy is still too new for people to get used to, to be completely honest. Despite the incorporation of renewable energy into daily life, fossil fuels still have a significant influence on many of our activities. Besides that, the installation costs for renewable resources would be expensive and demanding because of the fact that it is very difficult to find a suitable place to install a solar power station.

In conclusion, I support the idea that we should apply the utilization of renewable energy in our lives for a brighter future.

STUDENT 20

Since the Covid epidemic in 2021 and 2022, online education has grown in popularity. Since medical isolation is required and there is still a need to communicate work, knowledge, and personal status, people have turned to online software such as Zoom, Google Meet, Skype, Discord, etc. After the Covid epidemic ended, many social structures have changed and many people do not feel the need to have a physical presence to absorb knowledge anymore, and due to the habit of accepting online learning. In addition, more programs and classes are being offered due to the practice of adopting online learning. To assess the matter fully and properly, let's first list the strengths and weaknesses of online learning so that we can evaluate the situation fairly and accurately.

Online learning has many advantages, including the ability to study any subject at any time and anywhere without being constrained by location, as well as low costs (since there is no need to purchase classrooms, tables, chairs, learning environments, such as air conditioners, TVs, lights, etc.). Additionally, since the number of students in an online class is not constrained, online education can help spread knowledge to a wider audience. However, there are a lot of drawbacks to online learning as well. First off, since it is signal-dependent, undeveloped areas cannot use this kind. Second, because there are no longer any physical restrictions on the class, such as size or number, anyone can put together the curriculum, making it challenging to maintain a high standard of instruction. Additionally, students in a class have little control over their learning.

Overall, online learning is an inexorable advancement of technology and time, therefore we must learn to accept it and find solutions to the remaining issues.

Table 5a. Detailed scores of 40 essays marked by Rater 1

Rater 1										
Student number	TR1	CC1	LR1	GA1	TOTAL E1	TR2	CC2	LR2	GA2	TOTAL E2
1	8	8	8	7	7.5	8	8	8	7	7.5
2	7	6	6	5	6	7	7	7	6	6.5
3	8	7	8	7	7.5	7	7	7	7	7
4	7	7	6	6	6.5	7	7	6	7	6.5
5	8	7	8	7	7.5	9	8	9	8	8.5
6	7	7	6	5	6	5	6	7	6	6
7	6	7	8	7	7	7	7	7	7	7
8	6	5	6	5	5.5	7	6	7	6	6.5
9	7	7	7	6	6.5	8	8	8	7	7.5
10	9	6	7	7	7	9	8	9	8	8.5
11	7	7	7	7	7	8	8	7	7	7.5
12	8	7	8	7	7.5	6	6	6	6	6
13	7	6	6	6	6	6	7	6	6	6
14	5	5	5	5	5	7	7	7	7	7
15	5	6	6	6	5.5	6	7	8	7	7
16	4	5	5	5	4.5	5	6	7	6	6
17	7	7	7	6	6.5	6	6	6	5	5.5
18	5	5	7	5	5.5	7	7	7	7	7

19	6	6	6	5	5.5	6	7	7	6	6.5
20	7	7	6	6	6.5	7	6	7	7	6.5

Table 6a. Detailed scores of 40 essays marked by Rater 2

Rater 2										
Student number	TR1	CC1	LR1	GA1	TOTAL E1	TR2	CC2	LR2	GA2	TOTAL E2
1	8	8	8	7	7.5	8	8	8	7	7.5
2	7	6	6	6	6	7	7	7	7	7
3	7	7	8	7	7	7	7	8	7	7
4	7	6	6	6	6	7	7	6	6	6.5
5	8	8	7	7	7.5	8	8	8	8	8
6	6	7	7	6	6.5	7	6	7	7	6.5
7	7	7	7	7	7	7	6	7	7	6.5
8	6	7	6	6	6	6	7	7	7	6.5
9	7	6	6	6	6	7	7	7	7	7
10	8	8	8	7	7.5	8	8	8	8	8
11	6	7	7	6	6.5	7	7	7	7	7
12	8	7	7	7	7	8	7	7	7	7
13	6	7	6	6	6	6	6	6	7	6

14	5	6	6	6	5.5	6	6	6	7	6
15	7	7	6	7	6.5	7	7	6	7	6.5
16	4	4	5	5	4.5	5	5	6	6	5.5
17	7	7	6	6	6.5	7	6	7	6	6.5
18	6	5	6	6	5.5	6	6	7	7	6.5
19	6	6	6	6	6	6	6	7	7	6.5
20	7	6	6	6	6	7	7	7	7	7

Table 7a. Detailed scores of 40 essays marked by Rater 3

Rater 3										
Student number	TR1	CC1	LR1	GA1	TOTAL E1	TR2	CC2	LR2	GA2	TOTAL E2
1	8	8	8	7	7.5	8	8	7	7	7.5
2	6	6	6	6	6	7	6	7	7	6.5
3	7	7	7	7	7	6	7	7	7	6.5
4	6	6	6	6	6	6	6	6	6	6
5	8	8	8	9	8	8	8	8	8	8
6	7	6	5	6	6	7	7	7	7	7
7	7	7	7	7	7	7	7	8	8	7.5

8	7	7	6	6	6.5	7	7	7	7	7
9	6	7	7	6	6.5	6	7	8	8	7
10	7	7	7	7	7	8	8	8	8	8
11	8	7	7	7	7	8	7	7	7	7
12	6	6	7	7	6.5	6	6	6	6	6
13	6	6	6	6	6	6	7	7	7	6.5
14	5	5	5	5	5	5	6	6	6	5.5
15	7	7	7	7	7	6	7	7	7	6.5
16	5	5	5	6	5	6	5	7	7	6
17	6	7	6	6	6	6	6	6	6	6
18	8	7	7	8	7.5	5	6	8	8	6.5
19	6	6	6	6	6	6	6	6	6	6
20	7	7	6	7	6.5	7	7	7	7	7

Table 17a. Detailed mean scores of 40 essays from three raters

Mean scores from three raters										
Student number	Mean TR1	Mean CC1	Mean LR1	Mean GA1	Mean Overall E1	Mean TR2	Mean CC2	Mean LR2	Mean GA2	Mean Overall E2
1	8	8	8	7	8	8	8	8	7	8

2	7	6	6	6	6	7	7	7	7	7
3	7	7	8	7	7	7	7	7	7	7
4	7	6	6	6	6	7	7	6	6	6
5	8	8	8	8	8	8	8	8	8	8
6	7	7	6	6	6	6	6	7	7	7
7	7	7	7	7	7	7	7	7	7	7
8	6	6	6	6	6	7	7	7	7	7
9	7	7	7	6	6	7	7	8	7	7
10	8	7	7	7	7	8	8	8	8	8
11	7	7	7	7	7	8	7	7	7	7
12	7	7	7	7	7	7	6	6	6	6
13	6	6	6	6	6	6	7	6	7	6
14	5	5	5	5	5	6	6	6	7	6
15	6	7	6	7	6	6	7	7	7	7
16	4	5	5	5	5	5	5	7	6	6
17	7	7	6	6	6	6	6	6	6	6
18	6	6	7	6	6	6	6	7	7	7
19	6	6	6	6	6	6	6	7	6	6

20	7	7	6	6	6	7	7	7	7	7

Table 24a. Full thematic coding for the interview transcripts of 20 students

Thematic coding from 20 students' interview transcript

High-scoring students (n=13) IDs: 1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 15, 18, 20

Low-scoring students (n=7) IDs: 4, 12, 13, 14, 16, 17, 19

STUDENT 1

1/ In general, how do you feel about the usage of the four AI writing tools for writing this IELTS essay? Do you find them helpful?

I find Grammarly and Google Translate helpful for checking synonyms, grammar, and spelling.

2/ Overall, do you think you produced a higher quality essay with the tools compared to the first essay that you wrote without the tools? Which part or sentence do you think you wrote better with the tools?

Yes, I believe I produced a higher quality essay with the tools, particularly in terms of grammar and spelling, which were improved using Grammarly. Additionally, I replaced synonyms using SketchEngine and Google Translate.

3/ In the brainstorming and planning step (before writing), did you find the tools useful for you in this step? Which tool helped you the most? How did it help you?

I did not need any tools during the brainstorming and planning step because the topic was quite common and simple. I can write in English with simple words first. Then I need to replace the words that are too simple.

4/ In the composing step, did you find the tools useful? Which tool helped you the most? How did it help you?

Quillbot was not working during my composing time so I had to find synonyms suggested by Google Translate. I can write in English with simple words first. Then I need to replace the words that are too simple. For example, I didn't feel that the phrase "increased relaxation time" was correct, so I checked the meaning in Vietnamese, and I found more synonyms of this word suggested by Google Translate. Among them, I picked the word "extend" and modified my phrase to "extended relaxation time". I also used Sketch Engine to help me find more words related to "top-notch", and then among the collocations I chose "world-class" to complete the phrase "world-class education". I

think this word sounds more suitable compared to other high-frequency words like “matchless”, “triple-A”, or “first-rate”.

5/ In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?

In the revising and editing step, Grammarly was particularly useful for checking overall spelling and grammar. Additionally, I used Sketch Engine to affirm the usage of terms like “real-life interactions”. Another word was the word “re-interact” from the phrase “re-interacting with their peers”, which was my term, but I am not confident in using it. The Vietnamese meaning of this word seemed accurate when it was translated by Google Translate, so I used this word. Grammarly is also helpful in correcting my errors in spelling, prepositions, and articles, such as “with affordable prices” to “at affordable prices”, and “choose the suitable time” should be “choose a suitable time”.

6/ Can you give me some examples of challenges that the tools could not help you in this writing task?

The tools could not help me find results for “re-interact” in SketchEngine, Google Translate, and QuillBot. I wanted to paraphrase the sentence, “Moreover, online learning offers students flexibility as they can choose a suitable time for their education, boosting their productivity and learning effectiveness” using a different, more formal structure but I did not succeed. And as I said, Quillbot was not working during my composing time.

7/ What do you want more from these four tools: SketchEngine, Google Translate, QuillBot, and Grammarly?

I would like Grammarly to evaluate coherence and cohesion. Sketch Engine is a bit difficult to use and does not always generate correct synonyms. Some collocations suggested by the tools did not fit the context. Additionally, I wanted the corpus tool to evaluate the range of vocabulary (level of vocabulary based on CEFR) and suggest some collocations that fit the context and topic of education.

8/ Do you recommend or would use any other tools for this IELTS essay writing practice? Why?

I recommend using ChatGPT to brainstorm more arguments and an AI-based IELTS essay scoring tool for evaluating cohesion and coherence.

STUDENT 2

1/ In general, how do you feel about the usage of the four AI writing tools for writing this IELTS essay? Do you find them helpful?

Yes, I find the four tools helpful for writing the IELTS essay. I would rate Google Translate to be the most important tool for me, followed by Grammarly, Quillbot and Sketch Engine.

2/ Overall, do you think you produced a higher quality essay with the tools compared to the first essay that you wrote without the tools? Which part or sentence do you think you wrote better with the tools?

One of the challenges I faced was the inability to recall some English words. I needed Google Translate to translate from Vietnamese to English for individual words and from English to English for synonyms, such as finding alternatives for the word "avoid". I am not sure about the quality of my arguments in the second essay, but I felt more confident in my spelling and grammar, especially with grammar. However, I did not have enough time to check all paraphrased sentences on Quillbot.

3/ In the brainstorming and planning step (before writing), did you find the tools useful for you in this step? Which tool helped you the most?

During the brainstorming and planning step, I did not find the tools necessary because

I generated ideas from my first language.

4/ In the composing step, did you find the tools useful? Which tool helped you the most?

Yes, in the composing step, I found the tools useful. I needed Google Translate to check the thesaurus for synonyms and paraphrases. I also used Sketch Engine to find concordances of "short-sighted", but I couldn't find it, so I switched to Google Translate.

I know one of the words to describe "cận thị" is "being short-sighted", but from Google Translate I learned another word that is "myopic". Grammarly also helped me recognise various spelling mistakes and errors with articles in my writing.

5/ In the revising and editing step, did you find the tools useful? Which tool helped you the most?

During the revising and editing step, I found the tools useful. I used Grammarly one time to check for spellings and grammar. Grammarly was particularly helpful in fixing articles (the, a), plurals, singulars, verb, and my spelling errors. For example, Grammarly helped me learned that I need to add "the" before "problems" in the phrase "one of the many problems", add "s" to make plurals with "students", "classmates", add "are" to the phrase "applications that are available", "distracted by" rather than "distracted into". I also learned that the past tense of "set up" is "set up" and not "setted up".

6/ What do you want more from these four tools: Sketch Engine, Google Translate, QuillBot, and Grammarly?

I found Sketch Engine difficult to use and struggled to find words. I did not have additional needs for Google Translate. For Grammarly, I don't think that it has enough correcting features

7/ Can you give me some examples of challenges that the tools could not help you in this writing task?

The tools could not suggest more argument ideas, expand the ideas and examples, or correct the cohesion between sentences, particularly in transitions.

8/ Do you recommend any other tools for the IELTS essay writing practice?

Not really. I do not like to rely on technology for writing. Instead, I think I would read to have more ideas from websites and sample essays, such as those found on IELTS-blog.

STUDENT 3

1/ In general, how do you feel about the usage of the four AI writing tools for writing this IELTS essay? Do you find them helpful?

I think these tools are useful for improving my writing in terms of language.

2/ Overall, do you think you produced a higher quality essay with the tools compared to the first essay that you wrote without the tools? Which part or sentence do you think you wrote better with the tools?

Yes, I believe I produced a higher quality essay in terms of lexical resource and grammar, especially because topic 2 was less familiar to me than topic 1. For instance, I used SketchEngine to find keywords and phrases like "harness energy". I used Google Translate to translate parts of this sentence: "Meanwhile, solar power generates energy directly from the heat and light of the sun without any by-products; and an electric car is far more environmentally friendly than a petrol-powered car due to the lack of exhaust fumes.", and I was satisfied with its output. Grammarly also helped me with the usage

of articles (a/an/the). However, these four tools cannot help me with more ideas, and the “Environment” topic is more difficult for me compared to the first topic about “Distance learning”.

3/ In the brainstorming and planning step (before writing), did you find the tools useful for you in this step? Which tool helped you the most? How did it help you?

For brainstorming, I like Sketch Engine because it helped me to find the verb “harness” in “harness energy” when I searched for the word “energy” to look for more ideas. I did not use any tools during the brainstorming and planning step.

4/ In the composing step, did you find the tools useful? Which tool helped you the most? How did it help you?

Yes, in the composing step, I found the tools useful. For composing, I mostly needed Google Translate to look for some phrases in English, such as the phrase “petrol-powered car”, “coal”, and “burning coal releases CO2” and I learned these new words. Quillbot was not very useful for me when I used it to look for synonyms, as the phrase “harmful by-products” that I tried with Quillbot did not result in better words.

5/ In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?

For editing, Grammarly is useful in correcting errors. Thanks to it, I can identify some article errors that I made, such as missing “a” before “electric car” and “petrol-powered car”. However, it did not provide good suggestions for vocabulary and writing style.

6/ Can you give me some examples of challenges that the tools could not help you in this writing task?

Quillbot did not provide good suggestions for the phrase "harmful by-products"

7/ What do you want more from these four tools: SketchEngine, Google Translate, QuillBot, and Grammarly?

I would like SketchEngine to suggest collocations classified by topics to help me with more ideas.

8/ Do you recommend or would use any other tools for this IELTS essay writing practice? Why?

I recommend brainstorming ideas with ChatGPT. Additionally, tools that can organise the arguments in a paragraph with cohesion and coherence would be beneficial.

STUDENT 4

1/ In general, how do you feel about the usage of the four AI writing tools for writing this IELTS essay? Do you find them helpful?

I think Grammarly is useful. I didn't use Sketch Engine, Google Translate and Quillbot.

2/ Overall, do you think you produced a higher quality essay with the tools compared to the first essay that you wrote without the tools? Which part or sentence do you think you wrote better with the tools?

Yes, I believe I produced a higher quality essay with Grammarly compared to the first essay I wrote without them. I only used Grammarly because I didn't have enough time to use the other three tools. Grammarly corrected sentences like "The equipment that we need for renewable sources is highly expensive, so everyone cannot afford them.",

"For instance, in Vietnam, climate conditions remain extremely challenging approximately every year, so the government would not rely on these resources."

3/ In the brainstorming and planning step (before writing), did you find the tools useful for you in this step? Which tool helped you the most? How did it help you?

During the brainstorming and planning step, I used some memorised phrases to write the outline. I didn't use any tools

4/ In the composing step, did you find the tools useful? Which tool helped you the most? How did it help you?

In the composing step, I didn't use any tools.

5/ In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?

During the revising and editing step, I used Grammarly to check errors. Grammarly was helpful, for example, it helped me correct from "the mesmerizing alternatives of other fuels" to "mesmerizing alternative to other fuels", adding "the" before "equipment".

6/ Can you give me some examples of challenges that the tools could not help you in this writing task?

The tools could not help with correcting the logic of the sentences (cohesion and coherence).

7/ What do you want more from these four tools: SketchEngine, Google Translate, QuillBot, and Grammarly?

I didn't use the other three tools. For Grammarly, I want it to help me correct the word choice. For example, there are three words that I am not sure whether they are used correctly in the topic, such as "equipment", "sunlight sources", "unpolluted", "pollution-

free”, as this tool did not detect any error.

8/ Do you recommend or would use any other tools for this IELTS essay writing practice? Why?

I think I would use ChatGPT for brainstorming ideas and check for errors, I think it can correct my essay better than Grammarly.

STUDENT 5

1/ In general, how do you feel about the usage of the four AI writing tools for writing this IELTS essay? Do you find them helpful?

I find the thesaurus feature from SketchEngine helpful for writing this IELTS essay.

2/ Overall, do you think you produced a higher quality essay with the tools compared to the first essay that you wrote without the tools? Which part or sentence do you think you wrote better with the tools?

Yes, I believe I produced a higher quality essay with the tools compared to the first essay I wrote without them. The tools helped improve vocabulary variety and provided more arguments.

3/ In the brainstorming and planning step (before writing), did you find the tools useful for you in this step? Which tool helped you the most? How did it help you?

I did not use any tools to help with ideas during the brainstorming and planning step.

4/ In the composing step, did you find the tools useful? Which tool helped you the most? How did it help you?

Yes, in the composing step, I found the tools useful. I used Quillbot to find synonyms of “distance learning”, and I learned that there are more words that can replace my phrase, those are “online learning”, “remote learning” and “distance education”. Additionally, I used Sketch Engine's Word Sketch to find more verbs related to “attention” but I did not find a relevant verb combination among its suggested, such as “turn”, “shift”, “focus” or “flatter”, so I composed my own phrase “hold the attention”.

5/ In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?

In the revising and editing step, Grammarly was helpful in correcting one error preposition. It corrected “the thirst of knowledge” to “the thirst for knowledge”. It also suggested adding some commas.

6/ Can you give me some examples of challenges that the tools could not help you in this writing task?

Sketch Engine was not very helpful with word collocations because the output for each word was based solely on frequency. I did not find a satisfying word that goes well with “attention”.

7/ What do you want more from these four tools: SketchEngine, Google Translate, QuillBot, and Grammarly?

I would like to see more suggestions from these tools that go beyond basic grammar correction, particularly in terms of improving clarity and enhancing writing style. For instance, when writing IELTS Task 2 essays, I want to find suggestions on how to vary sentence structures or incorporate more formal linking phrases, rather than just

simplifying sentences or removing complex grammar. But none of these tools can do it.

8/ Do you recommend or would use any other tools for this IELTS essay writing practice? Why?

Yes, I recommend using a Thesaurus dictionary to enhance vocabulary variety and refine word choice, particularly when revising topic sentences or paraphrasing the question. I also find ChatGPT helpful for generating ideas and developing arguments, but I use it carefully by drafting specific prompts. For example, instead of asking a general question like "Write an essay for this topic" I might ask, "What are three reasons why governments should invest in renewable energy?" or "Can you provide examples to support the idea that public transport reduces environmental pollution?" These prompts help me gather relevant points without copying full responses. I then select useful ideas and rephrase them in my own words to fit the IELTS essay structure.

STUDENT 6

1/ In general, how do you feel about the usage of the four AI writing tools for writing this IELTS essay? Do you find them helpful?

I rate the usefulness of the AI writing tools as follows: 1) Quillbot, 2) Sketch Engine, and 3) Grammarly and Google Translate.

2/ Overall, do you think you produced a higher quality essay with the tools compared to the first essay that you wrote without the tools? Which part or sentence do you think you wrote better with the tools?

I think I produced an essay with more word count and more advanced vocabulary with these tools. However, this also reduced my writing speed because I depended too much on the ideas provided by the tools. I did not have enough time to use Grammarly.

3/ In the brainstorming and planning step (before writing), did you find the tools

useful for you in this step? Which tool helped you the most? How did it help you?

In the brainstorming and planning step, I found the “Concordance function” of Sketch Engine and Quillbot quite useful. At first, I used Quillbot to paraphrase the phrase “distance learning” from the topic and I got the synonym phrase “remote learning”. Then I put both “distance learning” and “remote learning” in Sketch Engine to search for relevant concordances. From that, I can access other sentences about distance learning. I picked out some other good phrases that I can use, which are “social distancing”, “faculty and staff”, and “academic facilities”. From that, it helped me to think about what to write more regarding the merit of “distance learning” and helped me to write this sentence: “The faculty and staff of academic facilities uploaded the digital content to the cloud services therefore students can get an education without the interruption of the social distancing period.”. I think I could apply this way for more words and ideas if I had more time.

4/ In the composing step, did you find the tools useful? Which tool helped you the most? How did it help you?

Google Translate played a major role in the composing step. It helped me look for certain words I can't remember in English. I think I only need this tool at this step. Google Translate helped me wrote this sentence in English: “Students began to show symptoms of mental illness, dependence, and cheating during online examination time due to lack of face-to-face supervision and administration from the faculty and staff of academic facilities”. I know all of these English words but I just couldn't remember them, such as “symptoms”, “face-to-face supervision”, “administration from the faculty”, “academic facilities”, and Google Translate helped me retrieve them.

5/ In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?

In the revising and editing step, Quillbot was helpful. For Quillbot, I used it to paraphrase all of my paragraphs. I think it did a good job in rewriting. I replaced my original sentence “In the context of this essay, the argument is made that distance learning has more limites than benefit.” with “This essay makes the case that distance learning has more drawbacks than advantages.” because it sounds more concise, as it helped me learn

that one way to rewrite is: "This essay makes the case that...". For other parts, I did not change anything as I thought my original writing was better.

6/ Can you give me some examples of challenges that the tools could not help you in this writing task?

The tools did not help me with generating ideas or adding more examples and case studies relevant to the arguments. However, I did not need much help with language use.

7/ What do you want more from these four tools: Sketch Engine, Google Translate, QuillBot, and Grammarly?

I would like Sketch Engine's "Concordance" function to provide more academic articles relevant to the keyword. For Google Translate, I find it about 40% effective for checking my translations for editing, but I still had to edit the sentences myself. An example of a sentence improved with these tools is: "The rationales would be given in detail in the following paragraphs in the given Vietnam context" which I edited after using Google Translate.

8/ Do you recommend or would use any other tools for this IELTS essay writing practice? Why?

I recommend using ChatGPT for generating ideas and getting more structures and ideas for the essay. I would also use it to assess the band score of my essay, though I trust its evaluation about 50%.

STUDENT 7

1/ In general, how do you feel about the usage of the four AI writing tools for writing this IELTS essay? Do you find them helpful?

I find Quillbot and Grammarly helpful for writing this IELTS essay. I did not have time to use Sketch Engine.

2/ Overall, do you think you produced a higher quality essay with the tools compared to the first essay that you wrote without the tools? Which part or sentence do you think you wrote better with the tools?

I composed most of my writing without any tools. I only used Google Translate to find the English equivalent of the Vietnamese word “ủng hộ”, then I got the word “support” and “favor”, and I chose “support” because it fits more with the context of expressing an opinion. Then, I used Grammarly to check for errors of my writing. Then I paraphrased some parts with Quillbot. I felt more confident in terms of language when using the tools, especially Grammarly, but I wrote fewer words compared to the first essay (202 words). I was not familiar with the environmental topic. Grammarly helped me with reviewing grammar tense, but most of the errors were identified by myself.

3/ In the brainstorming and planning step (before writing), did you find the tools useful for you in this step? Which tool helped you the most? How did it help you?

I did not use any specific tools during the brainstorming and planning step.

4/ In the composing step, did you find the tools useful? Which tool helped you the most? How did it help you?

In the composing step, I composed most of my writing without any tools. I only used Google Translate to find the English equivalent of the Vietnamese word “ủng hộ”, then I got the word “support”.

5/ In the revising and editing step, did you find the tools useful? Which tool

helped you the most? How did it help you?

After completing my essay, I used Grammarly to check for errors. The tool helped me identify several grammatical mistakes, such as changing “reduce the exhaust” to “reduce exhaust”, correcting “Another factor causes” to “Another factor that causes” modifying “the earth surface” to “the earth’s surface” and adjusting “clean energy can be produce” to “clean energy can be produced” and “clean energy can be use” to “clean energy can be used”. Additionally, it corrected the spelling of “sea water” to “seawater”.

After checking errors with Grammarly, I used Quillbot to paraphrase my original writing. I changed this sentence: “Another reason I prefer renewable energy is that it is significantly less expensive than fossil fuels.”. Quillbot also helped me paraphrase lots of my sentences, for example, from “I totally agree with this standpoint for the following reasons.” to “ I completely concur with this viewpoint for the reasons listed below.”, and I learned that “concur” is a synonym of “agree”. Quillbot also helped me rewrite the order and structure of this sentence: “First, because it does not emit as much hazardous exhaust as fossil fuels do, recyclable energy can help to prevent the environment from becoming increasingly polluted.”. I also learn more words such as: “entirely”, “conventional energy sources”.

6/ Can you give me some examples of challenges that the tools could not help you in this writing task?

I have never used Sketch Engine and do not want to use it. Time limitation was a challenge, and I needed more arguments and ideas. I prefer using the Oxford and Cambridge dictionaries as they feel more trusted and reliable compared to Google Translate. Quillbot can only paraphrase one sentence at a time and loads slowly.

7/ What do you want more from these four tools: Sketch Engine, Google Translate, QuillBot, and Grammarly?

I would like more features from Quillbot to improve the writing process.

8/ Do you recommend or would use any other tools for this IELTS essay writing practice? Why?

I recommend using ChatGPT to ask for arguments and extract ideas. Additionally, any tools that can evaluate the text would be helpful. It is important to check evaluations from different sources.

STUDENT 8

1/ In general, how do you feel about the usage of the four AI writing tools for writing this IELTS essay? Do you find them helpful?

I find the AI writing tools helpful for writing this IELTS essay.

2/ Overall, do you think you produced a higher quality essay with the tools compared to the first essay that you wrote without the tools? Which part or sentence do you think you wrote better with the tools?

Yes, I believe I produced a higher quality essay with the tools. For instance, Quillbot helped me to paraphrase this sentence: "Distance learning appears to be gaining popularity. Despite some obvious disadvantages of this trend, I believe its benefits outweigh the drawbacks.". Sketch Engine and Quillbot assisted me in finding synonyms for words like "popular" and "distance learning". Grammarly is helpful for correcting articles and prepositions, such as from "from a distance" to "at a distance," changed "on tuition fees" to "for tuition fees" and adjusted "participate on" to "participate in". Additionally, for the sentence: "For example, young learners find it hard to concentrate fully on the lesson because they can stop themselves from using social media or playing games while studying at a distance", Quillbot helped me learn that "feel it hard" can be

paraphrased as "find it hard". Another example is, "In addition, the tuition fee of distance learning may be much more affordable than the regular fee, so people can easily access a variety of fields", where Quillbot paraphrased "cheaper" to "more affordable" and I selected this option.

3/ In the brainstorming and planning step (before writing), did you find the tools useful for you in this step? Which tool helped you the most? How did it help you?

For brainstorming, I found Sketch Engine useful. I used its concordance function to find more ideas for "distance learning". I found some terms that appeared in the concordance examples, such as "home study" and "open learning courses", which helped me with more ideas about its advantages.

4/ In the composing step, did you find the tools useful? Which tool helped you the most? How did it help you?

Yes, in the composing step, I really need Quillbot. I wrote short sentences and then paraphrased them with Quillbot before continuing to write. Quillbot helped me turn my simple words into more formal ones. For example, I had written the word "cheaper" and it suggested changing to the phrase "more affordable", and I chose this word because I think it sounds more formal than "cheaper".

5/ In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?

In the revising and editing step, Grammarly was the most helpful tool. This tool helped me identify my grammar errors in article usage, such as lacking "the" in "the surrounding environment", adding "for" in the phrase "especially for the poor".

6/ Can you give me some examples of challenges that the tools could not help you in this writing task?

Although Sketch Engine provided some key words, it wasn't sufficient. The other three tools didn't help generate ideas, but I really need a tool that can.

7/ What do you want more from these four tools: Sketch Engine, Google Translate, QuillBot, and Grammarly?

I would like Grammarly to evaluate the text based on the IELTS score and provide more explanations for using incorrect articles. Sketch Engine could be improved. Google Translate is good for translating under four words, but some words are not suitable for the context.

8/ Do you recommend or would use any other tools for this IELTS essay writing practice? Why?

I want to try Bing AI to help correct my writing, especially for coherence and word choice. However, I would still review its suggestions carefully, since not all AI-generated corrections are contextually accurate.

STUDENT 9

1/ In general, how do you feel about the usage of the four AI writing tools for writing this IELTS essay? Do you find them helpful?

Yes, I find the writing tools helpful for ensuring grammar accuracy. The tools I used include Google Translate, Quillbot, and Sketch Engine. But I didn't have time to use Grammarly.

2/ Overall, do you think you produced a higher quality essay with the tools compared to the first essay that you wrote without the tools? Which part or sentence do you think you wrote better with the tools?

Yes, I believe I produced a higher-quality essay with better-developed content and ideas. I used Google Translate for the sentence, "Moreover, some wind farms are built in the wrong places which threaten the migratory flow of native birds as the turbines are a deadly danger to birds if hit." Additionally, I used Quillbot for the phrase, "mitigate the negative environmental effects of fossil fuels" and I really like this phrase.

3/ In the brainstorming and planning step (before writing), did you find the tools useful for you in this step? Which tool helped you the most? How did it help you?

I relied on my own ideas and did not use any tools.

4/ In the composing step, did you find the tools useful? Which tool helped you the most? How did it help you?

I used Google Translate and Quillbot. Google Translate helped me to remember words in this context, such as "ecological risk", "wind propeller", "indestructible", and "the most optimal energy source".

5/ In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?

In the revising and editing step, Quillbot was the most helpful tool for checking my work. I was taught that the conclusion sentence of the essay should be a paraphrased version of the introduction. I usually have a difficult time rewriting my own sentences, but Quillbot outputs helped me to do it faster and explore more synonyms. After a few trials, I find that some phrases are even better than my original words, such as "mitigate", "alleviate the negative environmental implications" and "pose ecological challenges", which seem to be more formal than my original phrase "reduce the adverse environmental impacts" and "have potential ecological risks". I also tried the word "mitigate" in Sketch Engine to check for its collocations. After checking with Sketch

Engine, I found the noun “risk” was the first collocation that can combine with the verb “mitigate”, so I chose the phrase “mitigate risk”.

6/ Can you give me some examples of challenges that the tools could not help you in this writing task?

Sketch Engine did not provide good collocations.

7/ What do you want more from these four tools: Sketch Engine, Google Translate, QuillBot, and Grammarly?

These four tools could not suggest ideas. I need a tool that can help me with more ideas.

8/ Do you recommend or would use any other tools for this IELTS essay writing practice? Why?

I recommend using ChatGPT to ask for arguments, not for language. It helps with ideas to express in English and makes me confident that I will improve.

STUDENT 10

1/ In general, how do you feel about the usage of the four AI writing tools for writing this IELTS essay? Do you find them helpful? Rate the order of importance.

I find the AI writing tools helpful for writing this IELTS essay. I rate their importance as follows:

1. Quillbot

2. Sketch Engine

3. Grammarly and Google Translate (equally important)

2/ Overall, do you think you produced a higher quality essay with the tools compared to the first essay that you wrote without the tools? Which part or sentence do you think you wrote better with the tools?

Yes, I believe I produced a higher-quality essay with the tools. For instance, Quillbot helped me to improve my vocabulary in many phrases and sentences: "If our dependence on these sources is not reduced and no action is taken to replace them, their exhaustion will soon lead to a global energy crisis and irreversible environmental degradation", "Some may object to this shift, citing the enormous expenses required to develop the necessary infrastructure and technology to implement it globally" and "necessitating prompt action and preparation from governments worldwide". The phrase "the transition to renewable energy sources would be unavoidable" and the idea that it "would be well worth the investment" also benefited from using Quillbot.

3/ In the brainstorming and planning step (before writing), did you find the tools useful for you in this step? Which tool helped you the most? How did it help you?

No, I did not find the tools useful during the brainstorming and planning step.

4/ In the composing step, did you find the tools useful? Which tool helped you the most? How did it help you?

Yes, I found the tools useful in the composing step. Quillbot was particularly helpful. I tried Sketch Engine to find collocations and it suggested a collocation "tacitly" for "agree". But after checking it with Google Translate, I found that this word doesn't fit in my sentence, as I want to have a strong adverb with "agree", such as "wholeheartedly" rather than "tacitly" so I skipped it and did not use this tool anymore.

I used Quillbot to paraphrase specific sections of my essay. For instance, Quillbot suggested changing "led to debates" to "sparked discussion", which I found to be a good suggestion. I also adopted Quillbot's version of the first sentence: "Some may object to

this shift, citing the enormous expenses required to develop the necessary infrastructure and technology to implement it globally.”. I retained the following sentence as it was. Additionally, I changed “unavoidable” to “inevitable” and “making it essential” to “necessitating”. However, instead of using Quillbot’s suggestion of “immediate action,” I kept my own phrase, “prompt action”.

5/ In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?

I used Grammarly for checking, but since I did not have access to the premium version, it did not provide any other useful corrections for my essay.

6/ Can you give me some examples of challenges that the tools could not help you in this writing task?

Grammarly was not useful for revising my essay because I can identify most errors myself. There was one sentence: "In conclusion, I completely agree with the suggestion that fossil fuels be replaced by their renewable counterparts", which was flagged as erroneous but it did not offer any better suggestion. So I don't think Grammarly can help me at all. Additionally, Sketch Engine suggested "tacitly" for "agree" but this was not a suitable meaning, as translated by Google Translate, so I rejected it.

7/ What do you want more from these four tools: Sketch Engine, Google Translate, QuillBot, and Grammarly?

I would like tools that can suggest ideas. Grammarly sometimes provides informal suggestions for IELTS writing, so it would be helpful if it could choose a more formal writing style.

8/ Do you recommend or would use any other tools for this IELTS essay writing practice? Why?

I recommend using ChatGPT to suggest ideas and provide more wording options based on experience. Additionally, Google Bard can be used to generate more ideas.

STUDENT 11

1/ In general, how do you feel about the usage of the four AI writing tools for writing this IELTS essay? Do you find them helpful?

I find Quillbot and Google Translate helpful. Sketch Engine is also useful, while Grammarly is not very useful because it required subscription.

2/ Overall, do you think you produced a higher quality essay with the tools compared to the first essay that you wrote without the tools? Which part or sentence do you think you wrote better with the tools?

Not really, because of the time demand and the topic. I only have time to use Quillbot and Google Translate. I used Quillbot to paraphrase my introduction paragraph: "Although fossil fuels have been a significant part of human life since the beginning of modern civilization, many individuals believe that renewable energies should take their place. I personally quite agree with the viewpoint since renewable energy is more environmentally friendly and because it may be used indefinitely by humans in the future when fossil fuels run out."

3/ In the brainstorming and planning step (before writing), did you find the tools useful for you in this step? Which tool helped you the most? How did it help you?

I did not use any tools during the brainstorming and planning step.

4/ In the composing step, did you find the tools useful? Which tool helped you the most? How did it help you?

In the composing step, I did not use any tools. I composed in my simple English first.

5/ In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?

During the revising and editing stage, I used Quillbot and Google Translate. I used Quillbot to paraphrase the introduction, as it provided useful synonyms to enhance my writing. For example, it suggested changing "people's lives" to "human's life", "many

people" to "many individuals", and "people can take advantage of it" to "it may be used indefinitely by humans", and I agree with these suggestions, I think they are better than my own writing, they sounds more formal and suit with the context of the environment. Google Translate also helped me learn two new words. From using it, I learned that the translation of the phrase "vô tận" is "inexhaustibility", "trữ lượng" is "reserves".

6/ Can you give me some examples of challenges that the tools could not help you in this writing task?

I could not find synonyms for "harmless" using Quillbot. Grammarly also required me to buy subscription and it did not give me any useful correction.

7/ What do you want more from these four tools: Sketch Engine, Google Translate, QuillBot, and Grammarly?

I would like Quillbot to provide more words to paraphrase. Sometimes I couldn't find suitable alternatives for my original writing.

8/ Do you recommend or would use any other tools for this IELTS essay writing practice? Why?

I recommend using the Cambridge dictionary and thesaurus because they are more trusted. I use the closest suggestions in bold but mostly depend on my knowledge.

STUDENT 12

1/ In general, how do you feel about the usage of the four AI writing tools for writing this IELTS essay? Do you find them helpful?

I find the AI writing tools quite helpful. My order of preference is:

1. Google Translate (for checking word meaning of some phrases)
2. Quillbot
3. Sketch Engine
4. Grammarly

2/ Overall, do you think you produced a higher quality essay with the tools compared to the first essay that you wrote without the tools? Which part or sentence do you think you wrote better with the tools?

No, I did not produce a higher-quality essay with the tools due to time constraints. I don't think these tools helped me write any better in 45 minutes. I am not good with ideas for the environment topic and I truly need something that can help me search for more ideas. In terms of language, I used Sketch Engine to check whether the phrase "exhausted fumes" is used in a corpus, and it provided me with some results so I kept this phrase. I didn't use Google Translate because I did not have any problem with writing my ideas in English. I did not have time to use Quillbot as well. For Grammarly, I did not offer any corrections because it required me to pay for premium access. Overall, I feel like I didn't do as well on this second essay compared to the first one. I am not familiar with the topic of environment, so I cannot think of good examples. I need a tool that helps me with more ideas.

3/ In the brainstorming and planning step (before writing), did you find the tools useful for you in this step? Which tool helped you the most? How did it help you?

I generated ideas by myself and I used Google Translate and Sketch Engine to check words.

4/ In the composing step, did you find the tools useful? Which tool helped you the most? How did it help you?

In the composing step, Quillbot was useful. I also used Sketch Engine (concordance) to check words in context, such as "exhausted fumes".

5/ In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?

As I mentioned earlier, I tried using Sketch Engine to check whether the phrase "exhausted fumes" is used in a corpus, and it provided me with some results, so I kept this phrase. For the remaining parts, I used Grammarly. However, it did not offer specific corrections because it required me to pay for premium access and I only see some highlighted phrases. I still think that Grammarly was useful in the revising and editing step. For example, I edited the sentence from "They can create" to "They can be created by unlimited sources in nature such as the wind heat or sunlight" based on its highlighted suggestions and the "Passive voice misuse".

6/ Can you give me some examples of challenges that the tools could not help you in this writing task?

Quillbot could not generate synonyms for "fossil fuels", so I didn't find this tool useful. Grammarly usually identifies the use of passive voice as incorrect, yet I was taught that the passive voice is important for an IELTS essay.

7/ What do you want more from these four tools: Sketch Engine, Google Translate, QuillBot, and Grammarly?

I think Quillbot should offer more synonyms.

8/ Do you recommend or would use any other tools for this IELTS essay writing practice? Why?

I recommend using a collocation dictionary (OZdict), thesaurus, and Google Search to

search for sample readings to generate ideas.

STUDENT 13

1/ In general, how do you feel about the usage of the four AI writing tools for writing this IELTS essay? Do you find them helpful?

I find the AI writing tools helpful and rank them in the following order of usefulness:

1. SketchEngine
2. Grammarly
3. Quillbot
4. Google Translate

2/ Overall, do you think you produced a higher quality essay with the tools compared to the first essay that you wrote without the tools? Which part or sentence do you think you wrote better with the tools?

Yes, I believe I produced a higher-quality essay with the tools because I could check word usage and use more relevant vocabulary.

3/ In the brainstorming and planning step (before writing), did you find the tools useful for you in this step? Which tool helped you the most? How did it help you?

I did not use any tools during the brainstorming and planning step.

4/ In the composing step, did you find the tools useful? Which tool helped you the most? How did it help you?

In the composing step, I found Quillbot useful for paraphrasing the "renewable sources" to "alternative energy" and I chose its suggestions. I also used Google Translate to find words that I don't know in English. For example, I needed to use Google Translate to translate the word: "bệnh hô hấp" and I learned that it is "respiratory disease". I used

Sketch Engine to check if my phrases “CO2 emissions” and “green energy” were correct. I saw that both phrases appeared in concordance examples with other technical words like “generating power” and “greenhouse emissions”. That showed me they fit the topic, so I decided to use both in my essay.

5/ In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?

In the revising and editing step, Grammarly was the most helpful tool for checking grammar and spelling. It helped me notice mistakes, especially with plural nouns. For example, I wrote "I partly agree with this ideas" but Grammarly corrected it to "I partly agree with this idea" It also fixed "too much traditional sources" to "too many traditional sources" when talking about fossil fuels. I also made some mistakes with verbs. For example, I wrote "it completely disappear" but Grammarly changed it to "it completely disappears" Another mistake was with prepositions. I wrote "put a financial burden for some nations" but it corrected it to "put a financial burden on some nations" These corrections helped me improve my writing a lot.

6/ Can you give me some examples of challenges that the tools could not help you in this writing task?

Sketch Engine took a lot of time and its thesaurus is based on frequency, providing antonyms like "expensive" and "cheap" which were not always useful. However, it did help me confirm the usage of my two phrase: “CO2 emission” and “green energy”.

7/ What do you want more from these four tools: Sketch Engine, Google Translate, QuillBot, and Grammarly?

I would like Sketch Engine to provide more collocations

8/ Do you recommend or would use any other tools for this IELTS essay writing practice? Why?

I recommend using the Oxford Dictionary (Eng-Eng) for checking word meanings and ChatGPT for exploring and generating more ideas. ChatGPT can be compared with Google Translate for similar quality suggestions.

STUDENT 14

1/ In general, how do you feel about the usage of the four AI writing tools for writing this IELTS essay? Do you find them helpful? I find the AI writing tools helpful to varying degrees. I rank their usefulness as follows:

1. Grammarly
2. Quillbot
3. Google Translate
4. SketchEngine

2/ Overall, do you think you produced a higher quality essay with the tools compared to the first essay that you wrote without the tools? Which part or sentence do you think you wrote better with the tools?

No, the second essay took more time and I felt a lack of vocabulary, perhaps due to the topic familiarity. I did not have time to fully try Sketch Engine as it provides too much information but not exactly what I am looking for. I only used Google Translate and Grammarly. I also did not have time to try Quillbot.

3/ In the brainstorming and planning step (before writing), did you find the tools useful for you in this step? Which tool helped you the most? How did it help you?

No, I did not find the tools useful during the brainstorming and planning step. I tried the concordance “remote learning” but did not get any more ideas.

4/ In the composing step, did you find the tools useful? Which tool helped you the most? How did it help you?

For this step, I only used Google Translate to check for English words and my spelling. I translated the word “thúc đẩy” and I learned that it is “promote”, and “chênh vảnh” is “negligence”.

5/ In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?

Grammarly was the most useful tool for revising and editing. It helped me fix mistakes with plural nouns, like changing “so many course” to “so many courses”, “online course” to “online courses”, and “material” to “materials”. It also reminded me to add “the” before “internet”. Another mistake I made was with “is” and “are”. I wrote “there are no waste”, but Grammarly corrected it to “there is no waste”. It also helped me with verb tenses. I wrote “should followed”, but Grammarly pointed out that it should be “should follow”. Another error was “talent pay”, and Grammarly corrected it to “talent pays”.

6/ Can you give me some examples of challenges that the tools could not help you in this writing task?

The tools were not helpful in generating more ideas and improving the flow of ideas.

7/ What do you want more from these four tools: SketchEngine, Google Translate, QuillBot, and Grammarly?

Sketch Engine has a data-overloaded interface. I would like premium access to Quillbot and Grammarly. I think Google Translate could also be improved with more options to choose from its output, like Quillbot.

8/ Do you recommend or would use any other tools for this IELTS essay writing practice? Why?

I would opt for ChatGPT to explore more writing ideas.

STUDENT 15

1/ In general, how do you feel about the usage of the four AI writing tools for writing this IELTS essay? Do you find them helpful?

I find Grammarly and Quillbot helpful. However, Quillbot requires a certain level of language skill to recognise the vocabulary and context. I did not use Google Translate and did not have enough time to explore the usage of Sketch Engine. I might use Sketch Engine in the future to find more ideas and relevant keywords.

2/ Overall, do you think you produced a higher quality essay with the tools compared to the first essay that you wrote without the tools? Which part or sentence do you think you wrote better with the tools?

In terms of language, my Essay 2 can be better with the tools. However, in terms of language, my ideas in Essay 1 were better developed.

3/ In the brainstorming and planning step (before writing), did you find the tools useful for you in this step? Which tool helped you the most? How did it help you?

I did not use any tools during the brainstorming and planning step.

4/ In the composing step, did you find the tools useful? Which tool helped you the most? How did it help you?

I did not use any tools during the composing step.

5/ In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?

Grammarly helped me the most during the editing step. It helped me recognize my mistakes with plural nouns, such as correcting "hydros power" to "hydro power" and "this ideas" to "this idea." It also reminded me to add a comma after "For that reason," and corrected my preposition mistake by changing "replace fossil fuels to other energy" to "replace fossil fuels with other energy."

I also used Quillbot to improve my body paragraphs after completing my essay. It allowed me to modify words I thought were not suitable and see more options. For example, I learned that "successfully implement the usage of these energies into real life" can be paraphrased as "properly execute the use of these energies in real life" and I learned the new phrase "properly execute" I also discovered that "alter" is a synonym for "change" and that "much cheaper" can be rewritten as "significantly cheaper".

6/ Can you give me some examples of challenges that the tools could not help you in this writing task?

Not a challenge, but I found that Grammarly sometimes suggests removing extra words, which can be confusing. The tools cannot help much with generating ideas. For example, changing "to" to "in order to" was a suggestion from Grammarly. Moreover, I don't really trust the quality of outputs from Google Translate.

7/ What do you want more from these four tools: Sketch Engine, Google Translate, QuillBot, and Grammarly?

I would like premium access to Grammarly.

8/ Do you recommend or would use any other tools for this IELTS essay writing practice? Why?

I recommend using ChatGPT because it is an all-in-one tool. I paste my own ideas into ChatGPT and ask it to "make this idea more developed in language". The more specific the input, the better the output. I do not use it to evaluate the whole essay and do not trust it completely for that purpose. Grammarly is good for checking grammar, while ChatGPT paraphrases better than Quillbot and is better with contexts.

STUDENT 16

1/ In general, how do you feel about the usage of the four AI writing tools for writing this IELTS essay? Do you find them helpful?

These tools are so useful for fast writing within 45 minutes. I can write in Vietnamese, and then use Google Translate, check with Grammarly, and use Quillbot to produce an essay. But I didn't have time to use Quillbot. With Google Translate helping me with my whole essay I don't even have to think in English and I think its translation quality is better than my English level.

2/ Overall, do you think you produced a higher quality essay with the tools compared to the first essay that you wrote without the tools? Which part or sentence do you think you wrote better with the tools?

Yes. Google Translate and Grammarly helped me to produce an essay that is faster and has no grammar or spelling errors.

3/ In the brainstorming and planning step (before writing), did you find the tools useful for you in this step? Which tool helped you the most? How did it help you?

I didn't use any tools for brainstorming ideas.

4/ In the composing step, did you find the tools useful? Which tool helped you the most? How did it help you?

Google Translate

5/ In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?

Grammarly

6/ Can you give me some examples of challenges that the tools could not help you in this writing task?

No, I think Google Translate helped me well.

7/ What do you want more from these four tools: Sketch Engine, Google Translate, QuillBot, and Grammarly?

I can't understand Sketch Engine. This tool should offer a simpler platform.

8/ Do you recommend or would use any other tools for this IELTS essay writing practice? Why?

If I had more time, I would use a dictionary such as Thesaurus for better word choices.

STUDENT 17

1/ In general, how do you feel about the usage of the four AI writing tools for

writing this IELTS essay? Do you find them helpful?

I find Google Translate the most helpful, followed by Grammarly and Quillbot.

2/ Overall, do you think you produced a higher quality essay with the tools compared to the first essay that you wrote without the tools? Which part or sentence do you think you wrote better with the tools?

No, I did not produce a higher-quality essay with the tools. Among these four tools, I only need Google Translate the most because I need a dictionary to search for words that I do not know in English, such as “tsunami”, “compete”, “mineral” and “instability”. Quillbot is also good at searching more synonyms but I did not have time to try out other paraphrase options with Quillbot or try Sketch Engine to check collocations. But these tools cannot help me when I run out of ideas. I am not familiar with any environmental issues, so it is more difficult for me to write the second essay. I think my vocabulary for this topic is not good enough compared to the first topic of distance learning.

3/ In the brainstorming and planning step (before writing), did you find the tools useful for you in this step? Which tool helped you the most? How did it help you?

I did not use any tools during the brainstorming and planning step.

4/ In the composing step, did you find the tools useful? Which tool helped you the most? How did it help you?

Google Translate was the most helpful tool during the composing step. Among these four tools, I only need Google Translate the most because I need a dictionary to search for words that I do not know in English, such as “tsunami”, “compete”, “mineral” and “instability”.

5/ In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?

I tried using Quillbot one time. It helped me paraphrase the sentence from "People on earth develop once they are healthy" to "On earth, once a person is healthy, they develop", I think it sounded better than my previous writing, but I didn't learn any synonyms.

Grammarly were the most helpful tools during the revising and editing step. Grammarly helped with corrections like "The Earth" and pointing out errors that I need to write in plural form, such as from "human" to "humans", correct article usage from "the lack of water" to "a lack of water", and I learned that I made a spelling mistake with "breathe".

6/ Can you give me some examples of challenges that the tools could not help you in this writing task?

The tools did not help me generate more ideas.

7/ What do you want more from these four tools: Sketch Engine, Google Translate, QuillBot, and Grammarly?

I ran out of time for the second essay and I did not have enough time to fully exploit the potential of Sketch Engine and Quillbot.

8/ Do you recommend or would use any other tools for this IELTS essay writing practice? Why?

I recommend using Google Search for more ideas and searching for more sample essays to learn from them.

STUDENT 18

1/ In general, how do you feel about the usage of the four AI writing tools for writing this IELTS essay? Do you find them helpful?

I find Quillbot the most helpful, followed by Google Translate, Grammarly, and SketchEngine.

2/ Overall, do you think you produced a higher quality essay with the tools compared to the first essay that you wrote without the tools? Which part or sentence do you think you wrote better with the tools?

Yes, I believe I produced a higher quality essay with the tools. While I am not confident in terms of ideas, I feel more confident about my vocabulary and grammar after using these tools. Google Translate and Quillbot are the two tools that helped me the most with my second essay. I remember the word “landbound” while I was composing, but I wasn’t sure if it was the correct meaning, so I needed Google Translate. It also helped me realise that my word “hydro power” was wrong, and it was “hydroelectric energy”. Quillbot is also quite useful. For example, it helps me to paraphrase from “Countries with less resources are put in the lower hand in negotiation and slower in development” to “Countries with fewer resources are given a weaker negotiating position and develop more slowly. This tool helped me realise that “resources” should be written with “fewer” rather than “less”. Also, the phrase “develop more slowly” sounds more formal than my original phrase “and slower in development”. I had limited time during the task, if given more time, I would use Google Translate, Quillbot, and Sketch Engine together, to double-check and better understand Quillbot’s suggestions, and to read more example sentences for clarity.

3/ In the brainstorming and planning step (before writing), did you find the tools useful for you in this step? Which tool helped you the most? How did it help you?

I tried Sketch Engine to search for the term "renewable" but did not get a result. So I had to plan all the ideas myself.

4/ In the composing step, did you find the tools useful? Which tool helped you the most? How did it help you?

Google Translate and Quillbot were the most useful in the composing step. As I said, I wrote in simple English and then used Google Translate as a dictionary. There were two words: "hydro power" and "landbound" that I was not sure and I needed to confirm their Vietnamese meaning with Google Translate. Grammarly was also helpful for checking spelling and grammar at this stage.

5/In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?

Quillbot helped me the most. It was especially useful in improving the formality of my writing. For example, it helped me change a sentence like "Countries with less resources are put in the lower hand in negotiation" to a more formal version: "Countries with fewer resources are given a weaker negotiating position."

6/ Can you give me some examples of challenges that the tools could not help you in this writing task?

Grammarly doesn't provide useful suggestions in the free version. I can already revise my own spelling and grammar mistakes. I need a tool to correct complicated word choice errors.

7/ What do you want more from these four tools: Sketch Engine, Google Translate, QuillBot, and Grammarly?

Quillbot should offer more sentence structure options because during the composition

of Essay 2 sometimes I couldn't find suitable alternatives to modify my sentences.

8/ Do you recommend or would use any other tools for this IELTS essay writing practice? Why?

ChatGPT would be more helpful for generating more ideas for writing.

STUDENT 19

1/ In general, how do you feel about the usage of the four AI writing tools for writing this IELTS essay? Do you find them helpful?

I find the AI writing tools helpful and rank them in the following order:

1. Quillbot
2. Google Translate
3. Grammarly

I tried Sketch Engine one time with the word "renewable energy" but did not get a result, so I did not use this tool.

2/ Overall, do you think you produced a higher quality essay with the tools compared to the first essay that you wrote without the tools? Which part or sentence do you think you wrote better with the tools?

Yes, I believe I produced a higher quality essay with the tools. While I am not confident in terms of ideas, I feel that I have more language structures.

3/ In the brainstorming and planning step (before writing), did you find the tools useful for you in this step? Which tool helped you the most? How did it help you?

I used Sketch Engine to search for the term "renewable energy" but did not get a result

so I planned the ideas independently.

4/ In the composing step, did you find the tools useful? Which tool helped you the most? How did it help you?

During the composing step, I found Google Translate to be the most useful. I wrote in simple English and then used Google Translate as a dictionary. I was struggling to find the English word for “ban hành luật”, and Google Translate helped me solve this problem, now I know it is “promulgate”.

5/ In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?

Grammarly was the most helpful tool in the revising and editing step. For example, it corrected my phrase “into daily life” to “in life”. It assisted with correcting prepositions, articles, and spellings. For Quillbot, I used it to paraphrase the last paragraph, but I only changed these sentences: “On the other hand, renewable energy is still too new for people to get used to, to be completely honest. Despite the incorporation of renewable energy into daily life, fossil fuels still have a significant influence on many of our activities.” because I think this output contains better word choice and phrases than my original sentences, particularly the phrase “the incorporation of renewable energy into daily life” is better than my previous sentence: “the application of renewable energy into everyday life”, and this also contains a better structure: “fossil fuels still have a significant influence on many of our activities” compared to “many of our activities depend on fossil fuels”.

6/ Can you give me some examples of challenges that the tools could not help you in this writing task?

I don't like the interface of Sketch Engine, it looks too complicated and I couldn't use it. It also could not give me any ideas.

7/ What do you want more from these four tools: Sketch Engine, Google Translate, QuillBot, and Grammarly?

I think QuillBot could be improved by offering more varied sentence structures.

8/ Do you recommend or would use any other tools for this IELTS essay writing practice? Why?

I recommend using Google search for articles and sample essays to gather more ideas and vocabulary on the topic of renewable energy. The four tools are helpful for self-learning and self-practice.

STUDENT 20

1/ In general, how do you feel about the usage of the four AI writing tools for writing this IELTS essay? Do you find them helpful?

I find the AI writing tools helpful and rank them in the following order:

1. Google Translate
2. Quillbot
3. Grammarly
4. Sketch Engine

2/ Overall, do you think you produced a higher quality essay with the tools compared to the first essay that you wrote without the tools? Which part or sentence do you think you wrote better with the tools?

Within 45 minutes, I could only focus on completing my essay as fast as possible. I do not entirely trust the outputs of Google Translate, but I think it is faster and more

productive than trying to write in English. If there are any mistranslated sentences of Google Translate, I am confident in my reading skill that I can find out and edit. I did not have time to try Sketch Engine.

3/ In the brainstorming and planning step (before writing), did you find the tools useful for you in this step? Which tool helped you the most? How did it help you?

During the brainstorming and planning step, I wrote in Vietnamese and did not use any specific tools.

4/ In the composing step, did you find the tools useful? Which tool helped you the most? How did it help you?

Google Translate helped me translate my essay. All I needed to do was read its output and edited the paragraph spacing.

5/ In the revising and editing step, did you find the tools useful? Which tool helped you the most? How did it help you?

Grammarly was the most helpful tool in the revising and editing step, especially for checking punctuation and articles. For instance, it helped me recognise punctuation errors, one spelling error (Google meet should be GoogleMeet), remove the preposition "out" in "list out", and turning "times" to "time". After checking all errors with Grammarly, I used Quillbot to paraphrase the whole essay. I think the texts generated by Quillbot sound more formal to me in general, compared to those of Google Translate. I accepted all of its output, except for the last sentence of the conclusion: "Overall, online learning is an inexorable advancement of technology and time, therefore we must learn to accept it and find solutions to the remaining issues. It still has a lot of strengths and disadvantages that need to be analyzed.". The last sentence: "It still has a lot of strengths and disadvantages that need to be analyzed." I think is a bit redundant, so I did not include it. I can finish an essay with just Google Translate and Quillbot in 20 minutes and I just need to review.

6/ Can you give me some examples of challenges that the tools could not help you in this writing task?

I did not have enough time to use SketchEngine. If I had enough time, I would use this tool to check after Grammarly and Quillbot.

7/ What do you want more from these four tools: SketchEngine, Google Translate, QuillBot, and Grammarly?

I do not have specific additional needs for these tools, but I believe a certified human rater for correcting my essay is still important. Nothing can guarantee the output of all the mentioned tools.

8/ Do you recommend or would use any other tools for this IELTS essay writing practice? Why?

I recommend using ChatGPT to generate ideas and then selecting relevant arguments to put into Quillbot for rewriting. However, one should not rely on tools for long-term practice.

Table 35a. Full lesson plans of ten teacher participants

Teacher	Tool	Step	Activity
Teacher 1	Sketch Engine	Planning	The teacher selects some keywords: “speak” vs “write”, “nonverbal” and “verbal”, “communication”, “and powerful”. Nouns go with “speak”: dialect, accent, tone, voice,
	Google Translate		The teacher suggests that students write separate sentences as their initial thoughts about this topic and use Google Translate to assist in translating words. Example: “Spoken communication is less powerful than writing because it can be affected by dialect and accent and it is difficult to reach many people.”
	No tool	Composing	Students work individually for 40 minutes and write a full essay.
	Quillbot	Revising	This is a session for individual revision. The teacher suggests that students use Quillbot to get suggestions for their writing. Students will then selectively choose the sentences they want to paraphrase. This will help improve the quality and variety of their writing.
	Grammarly		The teacher allows students to use Grammarly to run the final check.
Teacher 2	Google Translate Sketch Engine	Planning	- T will divide the class into two groups: A and B. - T will show two pictures on the slide (or write two phrases on the board) about ‘spoken communication’ and ‘written communication’. - In 3 minutes, group A will list situations in which we should use spoken communication; Group B will list situations in which we should use written communication (must use English, using Google Translate to translate ideas if necessary). - T will regroup to ensure that one student in group A will be paired with one student in

			group B and allow them to exchange information (information-gap activity). - T will introduce the topic of writing task 2 "Spoken communication is more powerful than written communication. To what extent do you agree or disagree?" Content input: - T will introduce the structure to Ss: Intro, body 1 (spoken), body 2 (written), and conclusion. - Ss will work in pairs to find ideas for two body paragraphs (using Google Translate if necessary). Ss will brainstorm ideas of two aspects (i.e., why spoken communication is more powerful in some cases, and why written communication is more powerful in some cases) - T will summarise good ideas of body 1 and body 2 and write on the board to form an outline Language input: - Ss will use Sketch Engine to search for synonyms of keywords: 'spoken communication', 'powerful', and 'written communication', and could use Sketchengine to rephrase ideas in the previous step (content input). T will provide the phrases and Ss will use Sketch Engine to look for more words T will summarise and present some good vocabulary (synonyms and better phrases for the content) on the board. For example: <i>+spoken communication/ speaking/ verbal communication. T will not interrupt. Ss will share their items</i> <i>+ powerful/ effective/ important</i> <i>+written communication/ writing / non-verbal communication</i>
	No tool	Composing	Students will begin by using Grammarly to check their work and take note of any errors. Following this, they will exchange their writing with a peer to receive feedback. Additionally, students can use Quillbot to paraphrase sentences or phrases they find repetitive or too simple. The teacher will then ask for 1-2 volunteers to share the feedback they received with the class. Finally, the

			teacher will offer delayed feedback based on the students' overall performance.
	Grammarly Quillbot	Revising	Students will begin by using Grammarly to check their work and take note of any errors. Following this, they will exchange their writing with a peer to receive feedback. Additionally, students can use Quillbot to paraphrase sentences or phrases they find repetitive or too simple. The teacher will then ask for 1-2 volunteers to share the feedback they received with the class. Finally, the teacher will offer delayed feedback based on the students' overall performance.
Teacher 3	Sketch Engine	Planning	In this brainstorming (planning) stage, the teacher uses SketchEngine to prepare a list of needed vocabulary, phrases, and collocations that students may need for their writing, for example: <i>effective communication, two-way communication, verbal communication, communication channels, facilitate communication, enhance communication, maintain communication</i>, etc. Note: The brainstorming stage of my lesson is only about providing keywords and narrowing ideas, and no sentences should be formed at this stage.
	Google Translate Quillbot	Composing	In the composing stage, students can use tools like Google Translate to overcome language barriers when expressing their ideas in English. For instance, they can translate Vietnamese phrases such as "Trong một số tình huống thì giao tiếp bằng lời nói sẽ hiệu quả hơn là bằng văn bản" to "In some situations, verbal communication is more effective than written communication." However, it's important to note that the accuracy of translations depends on the clarity of the original ideas.

			Additionally, students can use Quillbot to paraphrase their thoughts, transforming simple statements into more scholarly language. For example, "In some situations, verbal communication is more effective than written communication" can be rephrased as "In certain circumstances, vocal communication outperforms written communication." These two tools can serve as valuable resources for students to enhance their English writing skills and academic expression.
	Grammarly	Revising	Student individually revise and proofread their essay before submitting them. Teachers will give the total marks and comments afterwards.
Teacher 4	Google Translate Sketch Engine	Planning	T asks Ss: "Why do you prefer texting/talking?" Then T asks "Which one is more powerful, spoken communication or written communication" T shows the topic and asks them to brainstorm ideas. T instructs them how to use Google Translate to translate their ideas. T asks S to show their ideas. Then T corrects their words with appropriate words. T uses Sketch Engine and the "Concordance" search to show examples of words and then lets Ss guess the meaning of "cue" and "powerful".
	No tool	Composing	Ss write a complete essay on Google Doc
	Grammarly	Revising	Use Grammarly to check grammar errors
	Quillbot Sketch Engine		Use Quillbot to paraphrase, use Sketch Engine to check collocations and then peer-review
Teacher 5	Sketch Engine Google Translate	Planning	Students read the topic carefully, identify keywords, and use Google Translate to thoroughly understand the topic. They then work in groups, employing Sketch Engine and Google Translate to find more ideas for

			the essay based on keywords like "spoken communication," "written communication," and "powerful." Next, students compare their ideas with their classmates to exchange good expressions and examples. They proceed to make drafts for the essay, working individually on their outlines using phrases and short sentences. Finally, the teacher provides comments on the students' ideas and drafts.
	Quillbot	Composing	Students use Quillbot to paraphrase sentences with different structures, phrases, and words to discover more effective expressions. For example, instead of saying, "It is said that spoken communication has more power than written one," they might say, "I agree with the statement saying.../I am in favour of the opinion...." In groups of four, students then use Quillbot to modify and customise their sentences, focusing on the two body paragraphs of their essays. The teacher observes and assists the students in effectively employing these tools when necessary.
	Grammarly	Revising	Students utilise Grammarly for self-correction. The teacher then gives comments on some essays to highlight common mistakes. If time permits, the teacher can have students read aloud one or two well-written essays, gather feedback on the use of AI tools, or demonstrate effective usage of these tools within five minutes.
Teacher 6	Google Translate Sketch Engine	Planning	• Introduce the topic of the writing test • Elicit keywords from the question • Ask students to think about the synonyms and paraphrases of the keywords "communication", "spoken", "written", "powerful" • Ask students to brainstorm about their usages of spoken and written forms in daily life, in studying and in the workplace, students can write in Vietnamese if they can't translate into English yet)

			• Do pair work to generate more ideas • Ask students to use their digital devices to go online and try out new tools for writing individually • Ask them to use Google Translate to translate their keywords or ideas into English (if their ideas generated in Activity 1 are in Vietnamese) • Ask students to use Sketch Engine to explore and take note of collocations and phrases with high frequency that would be useful for their writing (based on the words and ideas they generate in activity 1, they should explore different words of their choice) • Revise the outline of a writing task agree-disagree essay • Revise the terms use for expressing their agreement or disagreement • Teachers will show the terms that students already learned
	No tools	Composing	• Ask students to write a full task 2 essay based on the ideas generated and vocabulary noted from Sketch Engine
	Grammarly Quillbot	Revising	• Ask students to input their writing into Grammarly to check for grammar mistakes and make notes of these errors so the whole class can become aware of common mistakes. Then, have them apply the topic question and their topic sentences from their body paragraphs into Quillbot to see how they are phrased and evaluate which sentences are better, theirs or Quillbot's. Teachers should also ask students to paraphrase the topic sentence of the introduction, body paragraph 1, and body paragraph 2.
Teacher 7		Planning	Activity 1: Introducing the topic + Brainstorming Introducing (6 minutes)

			<i>crowd out the voice and arouse emotions, so it is more powerful than written communication."</i>
	No tools	Composing	Activity 3: Writing - Editing Compile arguments to complete an essay (30 minutes) Write individually, no tool support
	Quillbot Grammarly	Revising	Activity 4: Peer Feedback - Revising the Essay (40 minutes) Pair up students to comment on two different essays. They should use Grammarly for assistance, preferably with Google Docs. If a word is repetitive, they can use Quillbot to suggest paraphrasing, allowing students to decide on appropriate synonyms. After receiving feedback, students will edit their essays based on the comments and then submit them to the teacher. Activity 5: Teacher's Comments and Scoring The teacher will provide comments and assign scores to the essays, which might require additional time.
Teacher 8	Google Translate Quillbot	Planning	Engage students with the topic "Written communication or Spoken communication: Which is more important?" by asking open-ended questions such as: What do you use to communicate? How often do you use language to communicate? Which form do you use more often, writing or speaking? Which one do you consider more important, written communication or spoken communication? Encourage students to brainstorm ideas using various materials and Google Translate. Then, the teacher allows students to independently brainstorm using Google Translate, and afterwards, they can use Quillbot to paraphrase words like "powerful," "spoken communication," and "written communication."
	Google Translate Quillbot	Composing	In the composing process, students can use Google Translate to help express their

	Grammarly		opinions in paragraph form. To enhance the quality of their writing, students employ Quillbot to paraphrase the paragraph. This is to improve writing style and clarity. After that, students should use Grammarly to review and correct any grammatical errors, ensuring their work is polished and error-free. I hope after Grammarly, they can notice their errors and remember the corrections while composing.
	No tools	Revising	The revising process should be conducted by peers, completely free of tool usage. Finally, the teacher will give feedback and note down common errors from Grammarly and other errors that the tools cannot point out, such as errors in task response, coherence and word choice.
Teacher 9	Google Translate Grammarly	Planning	Students brainstorm ideas in Vietnamese and then use Google Translate to answer the topic question. The teacher will help students with some hints, such as “In what situation writing is more powerful than speaking?” (E.g. formal email in business, formal announcement) “In what situation speaking is more powerful than writing?” (E.g. Speeches of a leader) Then, students use Grammarly to check for errors in the translated texts.
	Google Translate Quillbot Grammarly	Composing	Use Google Translate as a dictionary when writing. Next, to avoid repetition and improve their writing style, students were guided to use Quillbot to rephrase these words in new ways, such as changing “write” to “compose” or “draft,” and “speak” to “discuss” or “articulate”
	Grammarly	Revising	After spelling errors and grammar errors are removed by Grammarly. The teacher will mark the essays, looking for errors in Task Response (TR) and Coherence and Cohesion (CC).

Teacher 10	Google Translate Grammarly	Planning	Students read the topic and choose a side between spoken communication and written communication. Then the teacher will ask students to brainstorm in Vietnamese and then use Google Translate to translate their ideas into short sentences. For example, their sentence may look like this: "Writing can be more powerful than speaking when you need to explain complex ideas clearly. For example, writing lets you choose your words carefully and make changes, which is helpful for legal papers and contracts". Students then use Grammarly to check for spelling and grammar errors in the draft.
	Quillbot Grammarly	Composing	Students will continue to compose their essays until the time limit of 45 minutes. Afterwards, they should use Quillbot to paraphrase sections of their writing that have repeated or informal words.
	Sketch Engine Grammarly	Revising	If students are unsure which output to choose from Quillbot, the teacher will guide them in checking Quillbot's suggested words using Sketch Engine and selecting words based on the frequency of collocations. Finally, the teacher will ask students to use Grammarly for a final check on spelling and grammar. Once Grammarly has reduced all spelling and grammar errors, the teacher will mark the essays, focusing on errors in Task Response (TR) and Coherence and Cohesion (CC). This might take extra time to correct.

Table 37a. Interview transcripts of ten teacher participants

Teachers' interview transcript
Teacher 1
1/ Based on your experience of participating in this study, how do you feel about using Sketch Engine for teaching IELTS essay writing? The tool Sketch Engine can be used in the planning stage, but it is not suitable for other writing steps. 2/ If you must use Sketch Engine to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology do you think would be the most suitable for Sketch Engine? The most suitable steps of instruction or teaching methodology for this tool would be planning. 3/ Do you have any concerns about using Sketch Engine as a writing tool? What are they? No, I do not have any concerns about using Sketch Engine as a writing tool. 4/ Can you think of anything you would want more from Sketch Engine to help teachers with writing lessons? Experts who are experienced with corpus tools like Sketch Engine could offer more training workshops and opportunities to enhance vocabulary learning. 5/ Based on your experience of participating in this study, how do you feel about using Google Translate for teaching IELTS essay writing? I find that Google Translate is easy to use for teaching IELTS essay writing. 6/ If you must use Google Translate to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology would be the most suitable? It should be used after planning and before composing, only for sentences from Vietnamese to English, to express ideas more correctly and check vocabulary. 7/ What are your concerns about using Google Translate as a writing tool? My concern is that Google Translate translation cannot express pragmatics or idioms. 8/ Can you think of anything you would want more from Google Translate to help teachers with writing lessons? I would like Google Translate to improve translations in terms of semantics and pragmatics, making them more suitable for Vietnamese idioms. 9/ Based on your experience of participating in this study, how do you feel about using Quillbot for teaching IELTS essay writing? I find Quillbot good for selecting synonyms based on semantics, better than a Thesaurus dictionary. 10/ If you must use Quillbot to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology would be the most suitable? The most suitable steps of instruction or teaching methodology for Quillbot would be revising.

11/ What are your concerns about using Quillbot as a writing tool?

My concerns about Quillbot are that it has a quite 'robotic' structure and highlights words that students can learn from, suitable with context.

12/ Can you think of anything you would want more from Quillbot to help teachers with writing lessons?

I would like Quillbot to filter and prioritise more words with high-frequency words in authentic texts.

13/ Based on your experience of participating in this study, how do you feel about using Grammarly for teaching IELTS essay writing?

I would encourage using Grammarly for revising tasks, as it provides good suggestions.

14/ If you must use Grammarly to teach an IELTS Task 2 essay, what step of instructions or teaching methodology would be the most suitable for this tool?

After students finished composing their essay, I would tell my students, 'Open Grammarly and run a final check before you hit submit.' I want that moment to feel like a self-audit, not a teacher-imposed correction. I've noticed that when students do this final pass themselves, they come to class already aware of recurring mistakes and are prepared to ask deeper questions about how to avoid them next time.

15/ What are your concerns about Grammarly?

My concerns about Grammarly are that it sometimes too strictly restricts prepositions and suggests linking words like 'However' to be separated into a new sentence.

16/ Can you think of anything you would want more from Grammarly to help teachers with writing lessons?

I would like Grammarly to improve how it evaluates texts, especially regarding passive voice.

17/ In general, do you have any other reflections on the lesson plan activity?

These tools are easy to use so my students would be familiar with using Sketch Engine, Google Translate, Quillbot, and Grammarly. Students have sufficient prior language to be able to engage in the tasks. Anticipated difficulties: Students may have difficulty in developing ideas through Sketch Engine. However, in practice, my students may rely more heavily on machine translation and use robotic or inappropriate lexical choices.

18/ Do you recommend any other tools for teaching the IELTS Task 2 essay?

"Yes, I would recommend using ChatGPT as a support tool for teaching the IELTS Task 2 essay, especially during the brainstorming phase. Many students struggle with generating ideas quickly under time pressure, and this tool can really help them overcome that initial block.

I also encourage students to use it to explore how arguments can be developed. For instance, they can ask, 'What are some reasons people support investment in public transport?' or 'Can you give examples to support the view that private car use should be reduced?' ChatGPT can offer a range of ideas that students may not think of themselves, and they can then select, adapt, and personalize those points in their own writing. Of course, I always emphasise the importance of critical thinking. They shouldn't copy full paragraphs or sentences, but instead treat the tool as a brainstorming partner—to help generate ideas, outline arguments, and even practice

paraphrasing. It's all about training them to engage with the tool actively and thoughtfully, not passively.

19/ Do you recommend any other tools for students to practice the IELTS Task 2 essay at home?

Also ChatGPT for all stages of practicing IELTS Task 2 essay at home, except composing.

Teacher 2

1/ Based on your experience of participating in this study, how do you feel about using Sketch Engine for teaching IELTS essay writing?

The combination of features allows for the context of authentic language, so I have high trust in its output.

2/ If you must use Sketch Engine to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology do you think would be the most suitable for Sketch Engine?

Pre-writing is the most suitable step. Additionally, corrections can be checked with COCA, a dictionary, or Google Scholar.

3/ Do you have any concerns about using Sketch Engine as a writing tool? What are they?

The interface is complicated and not very effective for planning. Discussing ideas would generate more ideas.

4/ Can you think of anything you would want more from Sketch Engine to help teachers with writing lessons?

It should be similar to ChatGPT in generating full ideas and sentences. Additionally, the concordance needs to be filtered based on contexts. Teachers might also create a corpus that limits the content and controls the academic language, similar to IELTS topics.

5/ Based on your experience of participating in this study, how do you feel about using Google Translate for teaching IELTS essay writing?

The tool is easy to use and fast. It can also be used for exchanging ideas during discussions.

6/ If you must use Google Translate to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology would be the most suitable?

Planning is the most suitable step.

7/ What are your concerns about using Google Translate as a writing tool?

It is not applicable to translating the whole essay. The output accuracy for content and language is around 50-60%, which may not be suitable for students aiming for above a 6.0 score in IELTS.

8/ Can you think of anything you would want more from Google Translate to help teachers with writing lessons?

It should suggest more synonyms with the translated output.

9/ Based on your experience of participating in this study, how do you feel about

using Quillbot for teaching IELTS essay writing?

The tool can suggest words, correct grammar errors, and even replace Grammarly. It offers a variety of text and style choices, which students can use to learn paraphrasing across different contexts, especially in the premium version.

10/ If you must use Quillbot to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology would be the most suitable?

Revising and editing are the most suitable steps.

11/ What are your concerns about using Quillbot as a writing tool?

The suggested list will only be suitable for high-level students. Low-level students (around 5.0 IELTS) need help to evaluate effectively and may find it applicable for only 1-2 words.

12/ Can you think of anything you would want more from Quillbot to help teachers with writing lessons?

The incorporation of a dictionary function would be helpful for ESL students.

13/ Based on your experience of participating in this study, how do you feel about using Grammarly for teaching IELTS essay writing?

The tool provides good grammar correction and distinguishes errors with colours, which helps students focus on certain types of errors first, such as spelling and grammar. The number of errors is also indicated, allowing students to know if they have improved. It has good accuracy, which helps teachers with marking.

14/ If you must use Grammarly to teach an IELTS Task 2 essay, what step of instructions or teaching methodology would be the most suitable for this tool?

Revising and editing are the most suitable steps.

15/ What are your concerns about Grammarly?

There are unnecessary corrections. Grammarly tends to break down small sentences and creates a robotic style by removing relative clauses and changing word order.

Of course it cannot offer suggestions for word choice errors and coherence. I don't think any existing tool can do this.

16/ Can you think of anything you would want more from Grammarly to help teachers with writing lessons?

Adding more examples for its suggestions, incorporating more authentic texts to illustrate word usage, and offering more suggestion options would be beneficial. It should also record learners' progress provide feedback based on long-term usage, and calculate errors.

17/ In general, do you have any other reflections on the lesson plan activity?

I like it because it promotes autonomy. I will try it in a real lesson.

18/ Do you recommend any other tools for teaching the IELTS Task 2 essay?

No, I don't know any better tools.

19/ Do you recommend any other tools for students to practice the IELTS Task 2 essay at home?

Students can focus on the skills that tools cannot perform, such as planning ideas. Students can use Google search for planning, and collocation dictionaries like Flax,

COCA, and Ozdic can use corpora from other countries. Students can also use ChatGPT to explore more ideas, but I would not encourage them to just ask the AI to “write the essay” and copy the answer.

Teacher 3

1/ Based on your experience of participating in this study, how do you feel about using Sketch Engine for teaching IELTS essay writing?

Sketch Engine is helpful for brainstorming because it shows the most frequent collocations of a word, which gives students natural language patterns to work with. For example, if they're writing about 'education,' they can quickly see high-frequency collocations like 'higher education,' 'education system,' or 'access to education.' This helps spark ideas and gives them direction for developing their arguments or structuring their paragraphs.

2/ If you must use Sketch Engine to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology do you think would be the most suitable for Sketch Engine?

Planning would be the most suitable step of instruction for using Sketch Engine.

3/ Do you have any concerns about using Sketch Engine as a writing tool? What are they?

Sketch Engine is not really suitable for students in band 6.0 in this study.

4/ Can you think of anything you would want more from Sketch Engine to help teachers with writing lessons?

It would be helpful if Sketch Engine could search with levels of English (CEFR), providing more academic concordances for beginners.

5/ Based on your experience of participating in this study, how do you feel about using Google Translate for teaching IELTS essay writing?

In the composing stage, I guide students to use Google Translate as a starting point to express their ideas in English, especially when those ideas were originally developed in Vietnamese. Many students struggle to write directly in English, so translating full Vietnamese phrases helps them produce a basic draft and overcome initial language barriers. However, I emphasize that Google Translate is not a shortcut or the final product, because it often produces literal or overly simple sentences. To address this, I encourage students to use Quillbot immediately afterward to paraphrase and improve the sophistication of their writing. For example, a simple translated sentence like “In some situations, speaking is more effective than writing” can be rephrased through Quillbot as “In certain circumstances, verbal communication outperforms written communication.” This process helps students not only express their ideas more clearly but also refine vocabulary, sentence structure, and academic tone. I see this combination of tools as a scaffolded approach that supports reflective learning and enables students to produce clearer, more polished writing.

6/ If you must use Google Translate to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology would be the most suitable?

Composing

7/ What are your concerns about using Google Translate as a writing tool?

The English-to-Vietnamese academic texts are not accurate when translated by Google

Translate.

8/ Can you think of anything you would want more from Google Translate to help teachers with writing lessons?

I would want Google Translate to have a filter of translated text that can be adjusted according to complexity level, maybe according to CEFR. If it can provide examples of each level of writing, it would be beneficial.

9/ Based on your experience of participating in this study, how do you feel about using Quillbot for teaching IELTS essay writing?

Quillbot provides more vocabulary for teachers and has a filter based on styles that can be used to give different-level examples.

10/ If you must use Quillbot to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology would be the most suitable?

Planning and composing (using the output of Google Translate and paraphrasing by Quillbot). Quillbot can also correct mistakes from Google Translate.

11/ What are your concerns about using Quillbot as a writing tool?

There are no concerns about using Quillbot as a writing tool.

12/ Can you think of anything you would want more from Quillbot to help teachers with writing lessons?

There are no additional needs yet identified for Quillbot.

13/ Based on your experience of participating in this study, how do you feel about using Grammarly for teaching IELTS essay writing?

Based on my experience in this study, I find Grammarly to be a very helpful tool for teaching IELTS essay writing. I highly recommend the function of highlighting each mistake and suggestions in Grammarly. This feature encourages students not just to correct errors quickly, but to actually learn from them. I always remind my students to read their work carefully, double-check the mistakes with Grammarly suggestions, instead of accepting it blindly, because understanding why an error occurred is what helps them improve in the long term.

14/ If you must use Grammarly to teach an IELTS Task 2 essay, what step of instructions or teaching methodology would be the most suitable for this tool?

Revising and editing would be the most suitable instructions for using Grammarly.

15/ What are your concerns about Grammarly?

Grammarly cannot correct errors that affects the TR and CC criteria in IELTS, including linking phrases, word choice errors, and even some word form errors. I really need to check them carefully by myself.

16/ Can you think of anything you would want more from Grammarly to help teachers with writing lessons?

The overall score is useful for students to know how much they have accomplished their set goals. However, it would be great if Grammarly could indicate the level of writing (elementary, intermediate, or advanced). Additionally, a filter function to select the level of output language would be helpful.

17/ In general, do you have any other reflections on the lesson plan activity?

Students may not be able to use a wide range of vocabulary, paraphrase sentences, or brainstorm and express ideas effectively.

18/ Do you recommend any other tools for teaching the IELTS Task 2 essay?

I would recommend OneLook Thesaurus. It can find all parts of speech of a word, its synonyms, words that go with it, as well as slang or idioms used with this word.

19/ Do you recommend any other tools for students to practice the IELTS Task 2 essay at home?

No, these four tools are fine for my students.

Teacher 4

1/ Based on your experience of participating in this study, how do you feel about using Sketch Engine for teaching IELTS essay writing?

I like the “Concordance” function. It is good to revise a phrase in different contexts. It’s especially useful for revising phrases, as it allows students to see how the same phrase is used in a variety of real contexts.

2/ If you must use Sketch Engine to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology do you think would be the most suitable for Sketch Engine?

Revising and editing. In my revision activities, I encouraged students to take a structured approach. First, they paraphrase their original sentences using Quillbot to improve lexical variety. Then, they use Sketch Engine to verify whether the new expressions make sense by checking collocations and concordance lines in authentic contexts. This step helps them evaluate the appropriateness of vocabulary usage. Finally, they exchange drafts with peers for feedback. I believe this sequence not only improves language accuracy but also builds their confidence in using academic English more independently.

3/ Do you have any concerns about using Sketch Engine as a writing tool? What are they?

It is difficult to use, particularly with its unfriendly interface and overloaded data. Also, picking concordances takes time, some concordances are too difficult and irrelevant.

4/ Can you think of anything you would want more from Sketch Engine to help teachers with writing lessons?

I want it to provide the level of vocabulary based on the CEFR for its collocations. For example, when I search for the collocation of “speech”, and I have the word “intelligible”, I would be curious to see if it’s a word at the C2 or C1 level.

5/ Based on your experience of participating in this study, how do you feel about using Google Translate for teaching IELTS essay writing?

Would not recommend students to use this tool to translate everything. They can use it as a dictionary.

6/ If you must use Google Translate to teach an IELTS Task 2 essay, what steps

of instructions or teaching methodology would be the most suitable?

Planning

7/ What are your concerns about using Google Translate as a writing tool?

Some words are translated too close to one meaning

8/ Can you think of anything you would want more from Google Translate to help teachers with writing lessons?

No

9/ Based on your experience of participating in this study, how do you feel about using Quillbot for teaching IELTS essay writing?

Quillbot can introduce various grammar structures and this can teach students in grammar. Furthermore, students can also learn more about synonyms with Quillbot.

10/ If you must use Quillbot to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology would be the most suitable?

Revising and editing

11/ What are your concerns about using Quillbot as a writing tool?

This tool is only good for advanced students to recognise irrelevant outputs.

12/ Can you think of anything you would want more from Quillbot to help teachers with writing lessons?

No

13/ Based on your experience of participating in this study, how do you feel about using Grammarly for teaching IELTS essay writing?

Identify grammatical mistakes (e.g. singular, plural nouns; S-V agreement)

14/ If you must use Grammarly to teach an IELTS Task 2 essay, what step of instructions or teaching methodology would be the most suitable for this tool?

I would use Grammarly for revising and editing. During the revision stage, I always tell my students to use Grammarly, not just to fix the mistakes, but to really look at what kinds of errors they're making. I've noticed a lot of them have trouble with things like subject-verb agreement and singular versus plural nouns, so I specifically ask them to pay attention to those. I usually tell my students in class: "Don't just click accept and try to understand why these sentences or words are wrong". I also encourage them to keep a list of common errors they make, so over time, they can become more aware and hopefully avoid repeating them.

15/ What are your concerns about Grammarly?

It cannot fix errors in logical thinking and can only optimise correcting with the "premium version". It also prioritises turning long sentences into shorter ones and I don't like it.

16/ Can you think of anything you would want more from Grammarly to help teachers with writing lessons?

I want it to help correct word choice errors.

17/ In general, do you have any other reflections on the lesson plan activity?

No. I think the tools are well made and will fit my lesson.

18/ Do you recommend any other tools for teaching the IELTS Task 2 essay?

No

19/ Do you recommend any other tools for students to practice the IELTS Task 2 essay at home?

No

Teacher 5

1/ Based on your experience of participating in this study, how do you feel about using Sketch Engine for teaching IELTS essay writing?

Sketch Engine is still quite new for both teachers and students, but it provides detailed lexical data that's useful for comparing word choices. It offers a wide range of collocations and phrase patterns, which can help users generate ideas and plan their writing in a more natural, native-like way.

2/ If you must use Sketch Engine to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology do you think would be the most suitable for Sketch Engine?

Planning

3/ Do you have any concerns about using Sketch Engine as a writing tool? What are they?

Low-level students might be confused because of too much data in Sketch Engine.

4/ Can you think of anything you would want more from Sketch Engine to help teachers with writing lessons?

No

5/ Based on your experience of participating in this study, how do you feel about using Google Translate for teaching IELTS essay writing?

It's fast and convenient. It can be used as a dictionary.

6/ If you must use Google Translate to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology would be the most suitable?

I'll let students use it for planning.

7/ What are your concerns about using Google Translate as a writing tool?

It still has many errors and I would not recommend it for students to practice English writing.

8/ Can you think of anything you would want more from Google Translate to help teachers with writing lessons?

I want more context-based translations, especially idioms and some features of language that are related to each culture.

9/ Based on your experience of participating in this study, how do you feel about using Quillbot for teaching IELTS essay writing?

This tool can offer various sentence structures and give students more confidence with lots of ideas in L1 and not very proficient in L2. However, low-level students might not be able to recognise the differences in the substitutions of words. It is suitable for students above 6.0 IELTS.

10/ If you must use Quillbot to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology would be the most suitable?

Mainly for composing, and maybe revising. I want students to see paraphrasing as part of a learning process. By working together and trying out different sentence versions in Quillbot, they not only improve their writing but also learn from each other's choices. It's important to let them discuss and reflect with each other about which version sounds more formal or appropriate, and why; that's how they can learn.

11/ What are your concerns about using Quillbot as a writing tool?

It requires users to have a certain proficiency (such as IELTS 6.0 or above), or they might choose some synonyms and structures that are not formal and appropriate

12/ Can you think of anything you would want more from Quillbot to help teachers with writing lessons?

I want it to offer more writing styles, focusing more on the coherence of its output.

13/ Based on your experience of participating in this study, how do you feel about using Grammarly for teaching IELTS essay writing?

Grammarly is a powerful tool to teach IELTS writing since it can help me correct my students' essays faster. This tool knows how to suggest other structures or expressions, based on the original ones. This makes students vary their ways of expressing ideas.

14/ If you must use Grammarly to teach an IELTS Task 2 essay, what step of instructions or teaching methodology would be the most suitable for this tool?

Revising and editing

15/ What are your concerns about Grammarly?

One of my concerns is that Grammarly doesn't consistently address issues related to word choice or incorrect word forms. These are crucial for students preparing for academic writing tasks like IELTS.

16/ Can you think of anything you would want more from Grammarly to help teachers with writing lessons?

Grammarly should offer more corrections in terms of writing styles (formal/informal) to access more kinds of learners/users

17/ In general, do you have any other reflections on the lesson plan activity?

It might take some extra time for the teacher to introduce the tools to students.

18/ Do you recommend any other tools for teaching the IELTS Task 2 essay?

I would also recommend students use a tool named "Mindmeister" for presenting their outline and improving critical thinking.

19/ Do you recommend any other tools for students to practice the IELTS Task 2 essay at home?

MindMeister

Teacher 6

1/ Based on your experience of participating in this study, how do you feel about using Sketch Engine for teaching IELTS essay writing?

It is good for reminding useful phrases

2/ If you must use Sketch Engine to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology do you think would be the most suitable for Sketch Engine?

Planning

3/ Do you have any concerns about using Sketch Engine as a writing tool? What are they?

It offers overloaded data, which might not be suitable for beginner students

4/ Can you think of anything you would want more from Sketch Engine to help teachers with writing lessons?

I want Sketch Engine to filter collocations into formal or informal usage, and add in the CEFR level of word and its collocations.

5/ Based on your experience of participating in this study, how do you feel about using Google Translate for teaching IELTS essay writing?

Google Translate is very easy to use and is familiar to students. It is good with short sentence translation,

6/ If you must use Google Translate to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology would be the most suitable?

Planning. Google Translate is very easy to use and is familiar to students. It is good with short sentence translation and helps them get started when they don't know how to say something in English.

7/ What are your concerns about using Google Translate as a writing tool?

This tool usually translates word-by-word and sometimes does not translate a word in its context correctly. I ask students to use Google Translate to translate their keywords or ideas into English if their ideas are still in Vietnamese. This gives them something concrete to work with before we move on to the next stage.

8/ Can you think of anything you would want more from Google Translate to help teachers with writing lessons?

I hope it can improve the context-based translation capacity

9/ Based on your experience of participating in this study, how do you feel about using Quillbot for teaching IELTS essay writing?

It offers flexible paraphrases. This is a good tool for students to learn different ways of paraphrasing

10/ If you must use Quillbot to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology would be the most suitable?

Revising

11/ What are your concerns about using Quillbot as a writing tool?

Sometimes the output is incorrect with irrelevant synonyms

12/ Can you think of anything you would want more from Quillbot to help teachers

with writing lessons?

I want it to have more relevant suggestions.

13/ Based on your experience of participating in this study, how do you feel about using Grammarly for teaching IELTS essay writing?

Grammarly functions better than the grammar-checker of other tools (Word, Google Docs). The suggestions can give more options and help teachers consider other sentences

14/ If you must use Grammarly to teach an IELTS Task 2 essay, what step of instructions or teaching methodology would be the most suitable for this tool?

Revising

15/ What are your concerns about Grammarly?

It can sometimes suggest removing the inclusion of a relative clause. I think this is not beneficial for language learning since it reduces the grammar variety. The tool also offers limited support for improving cohesion. It rarely provides useful feedback on linking phrases or how students can better connect ideas between sentences and paragraphs.

16/ Can you think of anything you would want more from Grammarly to help teachers with writing lessons?

Not yet.

17/ In general, do you have any other reflections on the lesson plan activity?

It takes a long time to prepare the lesson from the teacher. And I think this activity would only suit small classrooms, and be suitable for students at B2 level or above.

18/ Do you recommend any other tools for teaching the IELTS Task 2 essay?

I would stay close to the traditional methods: a dictionary (Cambridge dictionary or Oxford dictionary) because they can give me the level of each word. And I would use the Google search engine to look for concordances.

19/ Do you recommend any other tools for students to practice the IELTS Task 2 essay at home?

I would recommend my students use these four tools in combination with an online dictionary.

Teacher 7

1/ Based on your experience of participating in this study, how do you feel about using Sketch Engine for teaching IELTS essay writing?

The function of finding concordance and collocations aids students in discovering natural expressions, providing numerous collocations, and verifying meanings before usage. For example, when you search for the word "message," you might find the adverb collocation "effortlessly," allowing you to construct a sentence like, "Writing can effortlessly deliver the message." This process can also offer examples and ideas for essays.

2/ If you must use Sketch Engine to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology do you think would be the most suitable

for Sketch Engine?

Brainstorming and planning, expressing arguments

3/ Do you have any concerns about using Sketch Engine as a writing tool? What are they?

There is not enough data to brainstorm arguments for the given topic, so prior knowledge or experience is still needed, and teachers must prepare a word cloud beforehand.

4/ Can you think of anything you would want more from Sketch Engine to help teachers with writing lessons?

It is necessary to narrow down the search of Sketch Engine using relevant words, similar to how the application Netflix provides theme-based suggestions.

5/ Based on your experience of participating in this study, how do you feel about using Google Translate for teaching IELTS essay writing?

It's easy to use.

6/ If you must use Google Translate to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology would be the most suitable?

I would mainly use Google Translate for brainstorming and planning in my lesson. However, it can also be utilised for revising by translating from English to Vietnamese to verify meanings, and it can be used in conjunction with Grammarly.

7/ What are your concerns about using Google Translate as a writing tool?

I'm afraid that my students will become too lazy to think in English with Google Translate.

8/ Can you think of anything you would want more from Google Translate to help teachers with writing lessons?

Not yet

9/ Based on your experience of participating in this study, how do you feel about using Quillbot for teaching IELTS essay writing?

Quillbot can offer good paraphrases.

10/ If you must use Quillbot to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology would be the most suitable?

Revising

11/ What are your concerns about using Quillbot as a writing tool?

It's good as a paraphrasing tool. I have no concern.

12/ Can you think of anything you would want more from Quillbot to help teachers with writing lessons?

Not yet

13/ Based on your experience of participating in this study, how do you feel about using Grammarly for teaching IELTS essay writing?

I want this tool to suggest more errors in word choice or word form aside from just punctuation and spelling errors.

14/ If you must use Grammarly to teach an IELTS Task 2 essay, what step of instructions or teaching methodology would be the most suitable for this tool?

Revising and editing. Grammarly is really helpful for catching surface-level issues like spelling, punctuation, and basic grammar. I've found that when students use it before submitting their work, it saves me a lot of time on minor corrections. But it doesn't always handle more complex aspects like cohesion or word choice very well. That's why I always tell my students to go over those areas with a classmate during peer review and to pay special attention to my feedback afterwards. Grammarly can assist, but it's not a substitute for real human input when it comes to improving flow and clarity

15/ What are your concerns about Grammarly?

There isn't much concern since teachers will make the final decision. As a teacher, it would help me reduce marking time on minor grammar or spelling errors. But I'm concerned that Grammarly gives students a false sense of confidence. They assume their writing is correct once there are no more red lines, even though deeper issues remain.

16/ Can you think of anything you would want more from Grammarly to help teachers with writing lessons?

I want it to help me improve the cohesion and coherence between sentences, as this aspect requires significant time and careful consideration from teachers during marking.

17/ In general, do you have any other reflections on the lesson plan activity?

These four tools are not helpful enough with difficult or strange topics in IELTS, such as Crime or Space travel. I think low-level students will have more difficulties in brainstorming even with the assistance of these tools. Teachers also need to be more prepared to offer arguments and explanations from the output of these tools.

18/ Do you recommend any other tools for teaching the IELTS Task 2 essay?

I would suggest using ChatGPT to support students with brainstorming ideas for IELTS Task 2 essays. Many students find it difficult to generate relevant arguments or examples, especially when they are unfamiliar with the topic. I often show them how to use ChatGPT to explore different perspectives. For example, if the question is about whether university education should be free, I guide students to ask the tool for reasons in support and against the idea. This helps them understand both sides of the argument and decide which one to develop in their essay.

19/ Do you recommend any other tools for students to practice the IELTS Task 2 essay at home?

No. I think these four tools are good enough.

Teacher 8

1/ Based on your experience of participating in this study, how do you feel about using Sketch Engine for teaching IELTS essay writing?

Sketch Engine can assist students in identifying the structure and context for word usage, providing examples, and using the concordance feature to check if Quillbot's paraphrase outputs fit the context.

2/ If you must use Sketch Engine to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology do you think would be the most suitable for Sketch Engine?

Planning

3/ Do you have any concerns about using Sketch Engine as a writing tool? What are they?

It's tricky to use at first because the interface is complex, and it doesn't suggest relevant ideas for this topic.

4/ Can you think of anything you would want more from Sketch Engine to help teachers with writing lessons?

Might improve its interface, and provide more data for brainstorming ideas.

5/ Based on your experience of participating in this study, how do you feel about using Google Translate for teaching IELTS essay writing?

I find that Google Translate can really assist beginners in moving from Vietnamese to English. It helps them express their ideas and also suggests more appropriate vocabulary as they type. I think it's a useful tool for getting started during the planning stage, especially for students who lack confidence in generating English sentences on their own

6/ If you must use Google Translate to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology would be the most suitable?

Planning. In my lesson, I started with open-ended questions to get students thinking and talking about the topic of communication—what it means, how they use it, and which form they consider more important. After this discussion, I encouraged them to brainstorm ideas freely. For students who found it difficult to express their thoughts in English, I asked them to write in Vietnamese first, then use Google Translate to explore how those ideas might be expressed in English.

7/ What are your concerns about using Google Translate as a writing tool?

This tool translates in a robotic style, producing word-by-word outputs that cannot be improved with a higher vocabulary level. Grammarly is not reliable for identifying subtle grammatical patterns, such as misplaced modifiers or improper use of conditionals, which are common in student essays.

8/ Can you think of anything you would want more from Google Translate to help teachers with writing lessons?

I believe it can be improved by suggesting more synonyms for the translated output. Additionally, I hope it can create new versions of translations from Vietnamese to English.

9/ Based on your experience of participating in this study, how do you feel about using Quillbot for teaching IELTS essay writing?

Quillbot can help students paraphrase their own sentences, to learn more synonyms, offering various options to paraphrase.

10/ If you must use Quillbot to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology would be the most suitable?

Students can first use Google Translate to express their ideas in paragraph form during the composing stage. Once the basic ideas are on screen, they paste that text into

Quillbot to experiment with synonyms and more academic phrasing. This back-and-forth helps them flesh out ideas they hadn't considered, expand their vocabulary right from the outset, and walk into the composing stage with richer, better-worded points to develop.

11/ What are your concerns about using Quillbot as a writing tool?

It is necessary to use Quillbot in combination with Sketch Engine to determine if the suggested synonyms are contextually appropriate.

12/ Can you think of anything you would want more from Quillbot to help teachers with writing lessons?

Need to highlight the most used paraphrased sentences, and rate the frequency of usage similar to corpus search

13/ Based on your experience of participating in this study, how do you feel about using Grammarly for teaching IELTS essay writing?

Grammarly assists teachers by identifying spelling and grammar mistakes, suggesting words, and providing plagiarism checks. Overall, it offers effective corrections.

14/ If you must use Grammarly to teach an IELTS Task 2 essay, what step of instructions or teaching methodology would be the most suitable for this tool?

I would use Grammarly for brainstorming and composing. I think this tool would help with drafting sentences. However, to check the overall quality, I think my lesson should involve peer corrections and teachers' corrections. I hope that after Grammarly, they can notice their errors and remember the corrections after composing.

15/ What are your concerns about Grammarly?

Grammarly does not adequately address issues with advanced verb tense usage, such as past perfect or conditional forms, which are important for clarity and accuracy in formal writing.

16/ Can you think of anything you would want more from Grammarly to help teachers with writing lessons?

I want it to offer word choice corrections.

17/ In general, do you have any other reflections on the lesson plan activity?

I like this idea of using only the four tools as language input. But I'm not sure I can try it in my large-sized classrooms. I will try with individual students I think.

18/ Do you recommend any other tools for teaching the IELTS Task 2 essay?

I recommend students use Menti.com to create word clouds for idea generation. For classroom engagement, I believe more fun and interactive tools are necessary, such as Padlets, to share ideas with other groups.

19/ Do you recommend any other tools for students to practice the IELTS Task 2 essay at home?

I think ChatGPT can be a valuable tool to help students come up with more arguments on a wider range of IELTS Task 2 topics. Some students tend to rely on memorising vocabulary, and when they face unfamiliar questions. So I would guide them to use ChatGPT to explore the topic more deeply. For example, if the essay asks about the benefits of investing in space research, students can ask ChatGPT for reasons why

space programs might be important for society, and it could suggest ideas related to scientific innovation, environmental monitoring, or international cooperation. I would instruct students to ask the right kinds of prompts, like 'What are the advantages and disadvantages of [topic]?' or 'What are three reasons to support this opinion?' This helps students think more critically about how to structure their argument. Students should summarise those points in their own words and using the tool only as a starting point. (Teacher 8)

Teacher 9

1/ Based on your experience of participating in this study, how do you feel about using Sketch Engine for teaching IELTS essay writing?

It is a bit too complicated to use.

2/ If you must use Sketch Engine to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology do you think would be the most suitable for Sketch Engine?

I wouldn't use it because I think it would take up a lot of time in class. Instead, I would encourage group discussion.

3/ Do you have any concerns about using Sketch Engine as a writing tool? What are they?

As I said before, it is a bit too complicated to use. Besides, the context of concordances suggested by Sketch Engine I think is not appropriate for IELTS writing

4/ Can you think of anything you would want more from Sketch Engine to help teachers with writing lessons?

No

5/ Based on your experience of participating in this study, how do you feel about using Google Translate for teaching IELTS essay writing?

It's easy to use, but I would not recommend it for composing.

6/ If you must use Google Translate to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology would be the most suitable?

I would recommend this tool for the planning stage only, and only use it when students need help with translating specific words or phrases.

7/ What are your concerns about using Google Translate as a writing tool?

I do not trust its vocab context. I would suggest my students write directly in English independently, even though it's bad English. They can learn more from their mistakes in that way.

8/ Can you think of anything you would want more from Google Translate to help teachers with writing lessons?

No

9/ Based on your experience of participating in this study, how do you feel about using Quillbot for teaching IELTS essay writing?

It's a good option and it's easy to use.

10/ If you must use Quillbot to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology would be the most suitable?

Maybe at all stages: planning, composing, and also revising. This tool still requires input from the students, so I think it would not affect their original writing ideas. Besides, I think this tool helps with expanding vocabulary. If they notice repeated use of words like 'write' or 'speak,' I guide them to use Quillbot to paraphrase those sentences. This

helps them diversify their sentence structures and vocabulary, encouraging them to explore alternative ways to express their ideas more effectively.

11/ What are your concerns about using Quillbot as a writing tool?

To optimise the paraphrasing produced by this tool, I believe it is essential for teachers to guide word selection to ensure contextual accuracy. I don't think students randomly choosing synonyms would yield good results.

12/ Can you think of anything you would want more from Quillbot to help teachers with writing lessons?

I hope it can present the CEFR level of each synonym, which is kind of similar to a dictionary.

13/ Based on your experience of participating in this study, how do you feel about using Grammarly for teaching IELTS essay writing?

Grammarly is easy to use and quite beneficial.

14/ If you must use Grammarly to teach an IELTS Task 2 essay, what step of instructions or teaching methodology would be the most suitable for this tool?

I always begin with brainstorming in the students' first language because I believe well-written ideas need to be finished first before composing. Once they have their points in Vietnamese, I guide them to translate those into English, usually through Google Translate, and then immediately use Grammarly to check sentence structure and grammar. For instance, if we're discussing written versus spoken communication, I might ask, 'In what situation is writing more powerful than speaking?' This helps them think critically, and Grammarly ensures their ideas are expressed correctly from the beginning.

15/ What are your concerns about Grammarly?

Grammarly is unable to correct linking phrases. The tool also struggles with errors in parallel structure, particularly when students write more complex sentences with multiple verb forms or noun phrases. I have to look for those errors while marking.

16/ Can you think of anything you would want more from Grammarly to help teachers with writing lessons?

I hope it can help me correct the Task Achievement and Coherence and Cohesion criteria.

17/ In general, do you have any other reflections on the lesson plan activity?

Preparing a lesson with these four tools would require a significant amount of the teacher's time. Incorporating these tools in the revision phase, I estimate, would add at least another hour to the class

18/ Do you recommend any other tools for teaching the IELTS Task 2 essay?

I would recommend using ChatGPT for more idea generation, but not for writing the whole essay.

19/ Do you recommend any other tools for students to practice the IELTS Task 2 essay at home?

I would suggest using ChatGPT to support students with more advanced corrections, particularly in areas like vocabulary enhancement and the use of linking words. Many students tend to repeat basic vocabulary or rely on simple connectors like 'and,' 'but,' or 'because' in their essays. I show them how to use ChatGPT to suggest more precise or academic alternatives. For instance, if a student writes 'people like watching movies because they are fun,' I guide them to ask ChatGPT how to improve the sentence. It might suggest: 'people enjoy watching films due to their entertainment value and emotional appeal.' This helps expand their lexical range and increase formality.

Teacher 10

1/ Based on your experience of participating in this study, how do you feel about using Sketch Engine for teaching IELTS essay writing?

Sketch Engine is a helpful tool, but it may take students more time to fully understand and use it effectively.

2/ If you must use Sketch Engine to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology do you think would be the most suitable for Sketch Engine?

I would certainly use Sketch Engine for revising and editing. I believe using the corpus tool as an additional language reference for altering the original words would be the most optimal approach. For example, in terms of preposition errors, after students complete their essays, I can guide them to review the changes made by Grammarly and determine if those changes align with the most frequent collocations. Additionally, I can ask students to compare the synonyms suggested by Quillbot with the highest frequency words based on Sketch Engine, and then decide the final sentence based on the similarity of these two tools.

3/ Do you have any concerns about using Sketch Engine as a writing tool? What are they?

Sketch Engine provides an overwhelming amount of data with its "Collocation" and "Concordance" functions, and some words are entirely irrelevant to the topic keywords. Therefore, I believe students won't be able to find ideas during the planning stage by using this tool alone.

4/ Can you think of anything you would want more from Sketch Engine to help teachers with writing lessons?

I hope it will feature a simpler interface, reduce the amount of data displayed for collocations, and link noun, verb, and adjective collocations of the keyword with its synonyms.

5/ Based on your experience of participating in this study, how do you feel about using Google Translate for teaching IELTS essay writing?

It's a basic tool for language learners, I think. But I don't recommend it.

6/ If you must use Google Translate to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology would be the most suitable?

In my lesson, I asked students to choose a stance on the topic and brainstorm arguments in Vietnamese. Then, they used Google Translate to convert their ideas into short English sentences. This helped them get started with writing without feeling overwhelmed. I only allow Google Translate during the planning stage. It's useful for low-level students to express simple ideas, but it's not reliable for more advanced learners or complex sentences. After translation, students used Grammarly to correct grammar and spelling before moving on to full paragraphs. I see Google Translate as a bridge for early-stage writing, not a tool for the whole process.

7/ What are your concerns about using Google Translate as a writing tool?

This tool can translate simple sentences for low-level students, but it cannot guarantee the quality of translations for intermediate or advanced students.

8/ Can you think of anything you would want more from Google Translate to help teachers with writing lessons?

I think it would be better if Google Translate could offer more translated options to choose from in its output.

9/ Based on your experience of participating in this study, how do you feel about

using Quillbot for teaching IELTS essay writing?

A convenient tool to avoid repeating words and sentence structures, and also useful for learning vocabulary.

10/ If you must use Quillbot to teach an IELTS Task 2 essay, what steps of instructions or teaching methodology would be the most suitable?

Composing. I would ask students to use it for the whole essay. I would tell them to focus on changing the simple adjectives (not nouns because nouns are usually hard to replace), such as “good”, “bad”, “easy”, and “difficult”.

11/ What are your concerns about using Quillbot as a writing tool?

When students use this tool to paraphrase, some sentences can become overly complicated and may even alter their original meaning. Therefore, I believe teachers still need to carefully revise these sentences when marking essays.

12/ Can you think of anything you would want more from Quillbot to help teachers with writing lessons?

No

13/ Based on your experience of participating in this study, how do you feel about using Grammarly for teaching IELTS essay writing?

It is easy to use and suitable with learners at this intermediate level.

14/ If you must use Grammarly to teach an IELTS Task 2 essay, what step of instructions or teaching methodology would be the most suitable for this tool?

I would integrate Grammarly during brainstorming. My lesson plan includes a step where students brainstorm in Vietnamese and translate their position into English before they start composing. So if a student writes, ‘Writing can be more powerful than speaking when you need to explain complex ideas clearly,’ I tell them to run that through Grammarly right away. It helps correct issues that could cloud meaning, especially when they’re translating by themselves. I then encourage them to think of Grammarly not just as a finishing tool, but as part of their drafting toolkit, which is something they use right from the planning stage to refine both clarity and accuracy.

15/ What are your concerns about Grammarly?

Some of Grammarly’s corrections are a bit “dry” for the writing style that is expected in the IELTS test. For instance, it often flags long sentences with numerous noun phrases as issues. This could negatively impact the Grammatical Range and Accuracy criterion of the IELTS exam. Moreover, some suggestions of it are too “fixed” and might not be accurate all the time. For example, “go shopping” and “do shopping” are both correct in terms of grammar and semantics, but Grammarly would always identify the phrase “do shopping” as an error and change it to “go”.

16/ Can you think of anything you would want more from Grammarly to help teachers with writing lessons?

I believe that currently, it only corrects spelling and basic grammar. I would like it to help correct more lexical errors and, of course, improve the coherence and cohesion of students’ essays.

17/ In general, do you have any other reflections on the lesson plan activity?

I would like to try teaching a lesson using these tools instead of following a course book. However, it would take a lot of time to explain each correction pointed out by Grammarly, as well as compare the outputs of the four tools. I’m concerned this might lead to further discussion and confusion rather than student progress. It would take

longer than a usual writing lesson. More importantly, I don't feel prepared to challenge the AI results in all contexts as an ESL teacher.

18/ Do you recommend any other tools for teaching the IELTS Task 2 essay?

ChatGPT and Poe can offer advanced writing corrections.

19/ Do you recommend any other tools for students to practice the IELTS Task 2 essay at home?

I recommend that my students use ChatGPT to generate more ideas and arguments. It's an all-in-one AI tool that can help save time. The key factor will be crafting the right prompts, step-by-step, to compose the essay more effectively.