T.C.
BURSA ULUDAĞ UNIVERSITY
INSTITUTE OF EDUCATIONAL SCIENCES
THE DEPARTMENT OF FOREIGN LANGUAGE EDUCATION
ENGLISH LANGUAGE TEACHING

INVESTIGATING THE IMPACT OF INTEGRATING GENERATIVE AI ON LEARNER AUTONOMY AND WRITING PROFICIENCY IN EFL INSTRUCTION

MASTER'S THESIS

FATMA ŞEN

ORCID: 0009-0000-9076-0297

BURSA 2025

T.C.

BURSA ULUDAĞ UNIVERSITY

INSTITUTE OF EDUCATIONAL SCIENCES

THE DEPARTMENT OF FOREIGN LANGUAGE EDUCATION

ENGLISH LANGUAGE TEACHING

INVESTIGATING THE IMPACT OF INTEGRATING GENERATIVE AI ON LEARNER AUTONOMY AND WRITING PROFICIENCY IN EFL INSTRUCTION

YÜKSEK LİSANS TEZİ

FATMA ŞEN

ORCID: 0009-0000-9076-0297

Danışman

Prof. Dr. Ayşegül Amanda Jane Audrey YEŞİLBURSA

BURSA 2025

BİLİMSEL ETİĞE UYGUNLUK

Bu çalışmadaki tüm bilgilerin akademik ve etik kurallara uygun bir şekilde elde edildiğini beyan ederim.

Fatma ŞEN

10.09.2025

TEZ YAZIM KILAVUZUNA UYGUNLUK ONAYI

"Integrating The Impact of Generative AI on Learner Autonomy and Writing Proficiency in EFL Context" adlı yüksek lisans tezi Bursa Uludağ Üniversitesi Eğitim Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırlanmıştır.

Tezi Hazırlayan	Danışman
Fatma ŞEN	Prof. Dr. Amanda Jane Audrey YEŞİLBURSA

Yabancı Diller Eğitimi Ana Bilim Dalı Başkanı

Prof. Dr. Ayşegül Amanda Jane Audrey YEŞİLBURSA

JÜRİ ONAY SAYFASI

Ingiliz Dili ve Eğitimi Ana Bilim Dalı'nda 802293043 numara ile kayıtlı Fatma ŞEN'in hazırladığı "İngilizcenin Yabancı Dil olarak Öğretimi Bağlamında Üretken Yapay Zekanın Öğrenci Özerkliği ve Yazma Yeterliliği Üzerindeki Etkisinin Entegrasyonu" konulu Yüksek Lisans çalışması ile ilgili tez savunma sınavı, 10/09/2025 günü 11.00-12:00 saatleri arasında yapılmış, sorulan sorulara alınan cevaplar sonunda adayın tezinin **başarılı** olduğuna **oy birliği** ile karar verilmiştir.

Sınav Komisyonu Başkanı
Prof. Dr. Ayşegül Amanda Jane Audrey YEŞİLBURSA
Bursa Uludağ Üniversitesi
Scopus Author ID: 48663301600
ORCID: 0000-0002-4884-7388

Üye
Prof. Dr. Gülden TÜM
Çukurova Üniversitesi
Scopus Author ID: 55958477100
ORCID: 0000-0001-7790-0857

Üye
Dr. Muharrem Umut Salihoğlu
Uludağ Üniversitesi
Scopus Author ID: 55588018000
ORCID: 0000-0002-5321-6556

ÖN SÖZ

İlk olarak, bu çalışmanın her aşamasında bana rehberlik eden, bana motivasyon veren ve desteğini esirgemeyen değerli danışmanım Prof. Dr. Amanda Yeşilbursa'ya en derin teşekkürlerimi sunmak isterim. Yoğun programına rağmen bana zaman ayırıp bu araştırmanın akademik bütünlüğünü korumama yardımcı olduğu için kendisine müteşekkirim.

Çalışmama gönüllü olarak katılan tüm öğrencilere içtenlikle teşekkür ederim. Tatil dönemlerine ve kişisel zamanlarına rağmen katkı sunmaları, bu araştırmanın temelini oluşturdu. Ayrıca bu süreçteki desteklerinden dolayı Bakom Feyza Tunç ve Osama Koraishi'ye de teşekkür etmek isterim. Yalnızca fikirleriyle değil, zaman zaman sessiz terapistler gibi yanımda durarak duygusal yükümü hafiflettikleri için onlara çok şey borçluyum. Benim ilerlememde rol oynayan sorgulatıcı eleştirileri ve dostlukları benim için çok kıymetliydi.

Tez jürime de ayrıca teşekkür ederim; vakit ayırıp çalışmamı dikkatle değerlendirerek katkı sundular. Özellikle lisans yıllarımdan bu yana bana yol gösteren ve ilham kaynağım olan değerli hocam Gülden Tüm'e özel bir teşekkürü borç bilirim.

Sevgili eşim… Senin yerin bambaşka, bana verdiğin gücün, yorgunluğuna rağmen yanımda oluşun, uykusuz gecelerde sabırla ışık tutuşun bana bu yolda pes etmemem gerektiğini ve daha önümüzdeki Doktorayı hatırlattı. Sana olan minnettarlığım kelimelere sığmaz.

En önemlisi... Canım annem Hatice Yılgın ve Can babam Yusuf Yılgın… Beni ben yapan, ayağa kalkmamı sağlayan, her tökezleyişimde elimden tutan sizlersiniz. Sabırla, sevgiyle, inançla arkamda durduğunuz için sonsuz teşekkür ederim. Eğitim hayatım boyunca bir saniye şüphe duymayıp tüm kararlarıma destek sundunuz.

Ayrıca, akademik yazım sürecinde sıkıştığımda, yazdıklarımı daha derli toplu, daha resm, öz, ve akademik hâle getirmemde yardımcı olan ChatGPT'ye de teşekkür etmek istiyorum. Gereğinden fazla uzattığımda beni durdurmayı başardığı için minnettarım.

Ve son olarak… Kendime de teşekkür ediyorum. Vazgeçmeyip bu yolculuğu tamamladığım, mükemmellik yerine kararlılığı seçtiğim ve her şeye rağmen yoluma devam ettiğim için. Teşekkür ederim Fatma, bu süreci sonuna kadar götürdüğün için.

Fatma ŞEN

ÖZET

Yazar Adı ve Soyadı	Fatma ŞEN
Üniversite	Bursa Uludağ Üniversitesi
Enstitü	Eğitim Bilimleri Enstitüsü
Ana Bilim Dalı	Yabancı Diller Eğitimi Ana Bilim Dalı
Bilim Dalı	İngiliz Dili ve Eğitimi
Tezin Niteliği	Yüksek Lisans Tezi / Doktora Tezi
Sayfa Sayısı	VII + 170
Mezuniyet Tarihi	10/09/2025
Tez Danışman(lar)ı	Prof. Dr. Ayşegül Amanda Jane Audrey Yeşilbursa

İNGİLİZCENİN YABANCI DİL OLARAK ÖĞRETİMİNDE ÜRETKEN YAPAY ZEKÂ ENTEGRASYONUNUN ÖĞRENEN ÖZERKLİĞİ VE YAZMA YETERLİLİĞİ ÜZERİNDEKİ ETKİLERİNİN İNCELENMESİ

Dil eğitiminde yapay zeka entegrasyonunun artan kullanımına rağmen, yapay zeka aracılı geri bildirimin dil okullarında yazma öğretiminde İngilizce yabancı dil öğrencilerinin özerk öğrenme gelişimine etkilerini inceleyen ampirik araştırmalar sınırlıdır. Bu karma yöntem çalışması, ChatGPT entegrasyonunun B1 seviyesindeki İngilizce yabancı dil öğrencilerinin (n=18) özerk öğrenme ve yazma yeterliliği üzerindeki etkisini 8 hafta boyunca araştırmıştır. Katılımcılar bir TÖMER sınıfından seçilmiş olup, bir grup deney grubu (ChatGPT destekli öğretim, n=12), diğer grup ise kontrol grubu (öğretmen destekli öğretim, n=6) olarak belirlenmiştir. Müdahale süreci her iki grubun da yazma becerilerinde önemli gelişim gösterdiğini ortaya koymuştur. Deney grubu sofistike prompt mühendisliği becerileri geliştirmiş ve artan yazma öz-farkındalığı sergilemiş olmakla birlikte, başlangıçtaki yapay zeka önerilerine aşırı bağımlılık öğretici müdahalesini gerektirmiştir. Sonuçlar, yapılandırılmış yapay zeka entegrasyonunun pedagojik eğitim ve kademeli sorumluluk verme ile birleştirildiğinde muhtemelen daha yüksek öğrenci özerkliğini destekleyebileceğini göstermektedir. Ancak, öğrenciler arasındaki dijital okuryazarlık boşlukları ve özerklik düzeyi farklılıkları nedeniyle, başlangıçta yapay zekaya aşırı bağımlılık eğilimi olup sürekli insan rehberliğine ihtiyaç devam etmektedir. Bu araştırma, yapay zekanın dil öğrenimindeki potansiyeli açısından nasıl abartıldığının altını çizerken, etkili insan-yapay zeka entegrasyonu durumunda olası öneri ve rehberlik sağlamaktadır. Araştırmanın önerisi, yazma öğretiminin 4 ana aşamasında Ajans, Farkındalık, Uygulama, Özerklik, Erişim, Duyuşsal Katılım, Uyum ve Yapay Zeka Entegrasyonunu içeren "9A Katalizör Çerçevesi" olarak daha sonrası daha büyük bir katılımcı grubuyla araştırılması gereken potansiyel bir çerçeve ileri sürmüştür.

Anahtar Sözcükler: komut mühendisliği pedagojisi, İYD, ChatGPT, yazma yeterliliği, özerklik geliştirme, geri bildirim okuryazarlığı, dijital okuryazarlık

ABSTRACT

Name and Surname	Fatma ŞEN
University	Bursa Uludag University
Institution	Institute of Educational Sciences
Field	Field of Foreign Languages
Branch	English Language Teaching
Degree Awarded	Masters
Page Number	VII + 170
Degree Date	10/ 09 / 2025
Supervisor(s)	Ayşegül Amanda Jane Audrey YEŞİLBURSA

INVESTIGATING THE IMPACT OF INTEGRATING GENERATIVE AI ON LEARNER AUTONOMY AND WRITING PROFICIENCY IN EFL INSTRUCTION

Despite increasing integration of artificial intelligence in language education, limited empirical research examines how AI-instructed feedback affects EFL learner autonomy development in writing instruction in language schools. This mixed-methods study investigated how ChatGPT integration impacts autonomous learning and writing proficiency among B1-level EFL learners (n=18) for 8 weeks. Participants were selected from a TÖMER class, with one group as the experimental group (ChatGPT-instructed, n=12), and another group as the control group (Teacher-instructed, n=6). The intervention process shows both groups demonstrating significant writing improvement. The experimental group developed sophisticated prompt engineering skills and demonstrated increased writing self-awareness, though initial over-reliance on AI suggestions required instructor mediation. Results suggest that structured AI integration has the potential to foster learner autonomy when combined with explicit, pedagogical training and gradual responsibility transfer. However, due to possible digital literacy gaps and autonomy level differences among learners, there is an initial AI over-reliance and a continued need for human guidance. This research tentatively establishes a foundational understanding of how AI is overhyped in terms of transformational potential in language learning while providing guidance for probable effective human-AI integration. The output of the research suggests a 9A Catalyst Framework comprising Agency, Awareness, Application, Autonomy, Access, Affective Engagement, Alignment, and AI integration in four main phases of writing instruction, which should be viewed as exploratory and warrants further validation through broader studies.

Keywords: prompt engineering pedagogy, EFL, ChatGPT, writing proficiency, autonomy development, feedback literacy, digital literacy

İÇİNDEKİLER

Sayfa

BİLİMSEL ETİĞE UYGUNLUK ..... **i**
TEZ YAZIM KILAVUZUNA UYGUNLUK ONAYI ..... **ii**
JÜRİ ONAY SAYFASI ..... **iii**
ÖN SÖZ ..... **iv**
ÖZET ..... **v**
ABSTRACT ..... **vi**
İÇİNDEKİLER ..... **vii**
TABLOLAR LİSTESİ ..... **xvi**
ŞEKİLLER LİSTESİ ..... **xvii**
KISALTMALAR LİSTESİ ..... **xviii**
CHAPTER I 1
INTRODUCTION 1
1.1. Statement of the Problem 1
1.2. The Purpose of the Study 3
1.2. Research Questions 4
1.4. Significance of the Study 4
1.5. Limitations of the Study 5
1.5.1. Reliance on Self-Reported Data for Autonomy 6
1.6. Delimitations of the Study 6
1.7. Ethical Considerations 7
1.7.1. Ethical Considerations and AI Tool Disclosure 7
1.7.1. Confidentiality and Anonymity 8
1.7.2. Data Security and Storage 8
1.7.3. Minimising Potential Risks and Maximising Benefits 8
1.7.4. Institutional Approvals and Informed Consent 8
1.8. Definitions of Key Terms 9
1.9. Organisation of the Thesis 9
1.10. Summary 9
CHAPTER II **11**
LITERATURE REVIEW **11**
2.1 Overview of Writing in EFL Contexts 11
2.1.1 Challenges in L2 Writing 11
2.1.2. Writing Instruction in Turkish and Global EFL Settings 12
2.1.3. AI Integration in EFL Writing 13
2.2. Learner Autonomy in Language Education 15

2.2.1. Definitions and Dimensions 15
2.2.2 Self-regulated Learning 18
2.2.3. Relevance to Writing and Digital Feedback 20
2.2.4. Motivation and Awareness 21
2.3 Feedback in Language Learning 23
2.3.1. Types of Feedback 23
2.3.2. Feedback Literacy 25
2.3.3. Feedback Effectiveness in Writing Improvement 26
2.4 Digital Literacy 27
2.5 Artificial Intelligence in ELT 28
2.5.1 AI Tools Used in Language Learning 28
2.5.2 Generative AI and Large Language Models 29
2.5.2.1 Technological Pedagogical Content Knowledge (TPACK) and Digital Pedagogical Content Knowledge (DPACK) Frameworks 30
2.6 AI and Writing Development 32
2.6.1 Empirical Studies on AI-Generated Feedback 32
2.6.2 Changes in Fluency, Accuracy, Complexity, and Coherence 33
2.7 Theoretical Framework 34
2.7.1 Sociocultural Theory (Vygotsky) 34
2.7.2. Autonomous Learning Theory (Holec, Benson, Littlewood, Alm) 35
2.7.3 Authentic vs. Artificial Autonomy in AI Learning 37
2.7.4 Self-Regulated Learning Model (Zimmerman) 38
2.7.5 Bloom's Taxonomy AI-Revisited (CAP Model) and Affective Filters 39
2.7.6 Feedback Intervention Theory (Kluger & DeNisi) 41
2.7.7 Digital Literacy & Digital Divide (Van Dijk, 2005) 42
2.7.8 Dual Process Theory of Writing (Hayes) 42
2.7.9 Emerging Models (Critical Digital Pedagogy, Feedback Literacy in AI contexts) 43
2.8 Research Gaps 46
2.8.1 Lack of Empirical Studies Combining AI Feedback and Autonomy 46
2.8.2 Neglect of Feedback Literacy in AI Use 46
2.8.3 Limited Longitudinal Studies in EFL Contexts 47
2.9. Cognitive Dependency and Over-reliance Risks 48
CHAPTER III 50
METHODOLOGY 50

3.1 Research Design 50
3.2 Pilot Study 51
3.2.1 Purpose of the Pilot 51
3.2.3 Challenges Faced in the Pilot Study 51
3.2.4 Adjustments Based on Pilot Findings 51
3.3 Participants 51
3.3.1 Participants of the Pilot Study 51
3.3.2 Participants and the Main Intervention 52
3.3.3 Persistent Challenges in the Main Intervention 53
3.4 Data Collection 53
3.4.1 Qualitative Data 53
3.4.2 Quantitative Data 53
3.4.2.1 Reliability of the Rubric in the Current Study 54
3.5 Data Collection Procedure 55
3.5.1. Pre-Test Administration 56
3.5.2. Post-Test Administration 56
3.5.3. Ethical Considerations and Anonymity Measures 56
3.5.4. Reflective Diaries and Semi-Structured Interviews 57
3.6. Data Analysis 57
3.6.1 Descriptive Statistics 58
3.6.1.1 Inter-Rater Reliability 59
3.6.1.2 Normality Testing and Statistical Method Justification 60
3.6.1.3 Justifications for Some Data 62
3.6.2 Participant Engagement 63
3.6.3 Quantitative Analysis 64
Multiple Comparisons Adjustment 65
3.6.4 Statistical Method Justification 67
3.6.5 Multiple Comparison Considerations 68
3.6.6. Qualitative Analysis 71
3.7. Expert Validation and Refinement of Interview Questions 72
CHAPTER IV 74
RESULTS 74
4.1. Results for RQ1 74
4.3.1 Strategic AI Tool Appropriation and Integration 75

4.1.2 Development of Multi-Source Verification Strategies 76
4.1.3 Metacognitive Prompt Engineering and Iterative Optimisation 77
4.1.3.1 Evolution from Basic to Sophisticated Prompting Strategies 77
4.1.3.2 Adaptive Response to AI Limitations 78
4.1.4 Critical Evaluation and Balanced Integration of AI Feedback 78
4.1.4.1 Comparative Analysis of AI vs. Human Feedback Sources 78
4.1.4.2 Recognition and Mitigation of Over-reliance Risks 79
4.1.5 Evolving Perceptions and Self-Awareness Through Reflective Practice 79
4.1.5.1 Metacognitive Awareness Development Through Diary Reflection 80
4.1.5.2 Perception Changes About AI Tool Limitations and Benefits 80
4.1.5.3 Documentation of Autonomous Strategy Development 81
4.1.6 Emotional and Motivational Perception Shifts 82
4.1.7 Cross-Theme Analysis: Autonomy Manifestation Patterns Through Diary and Interview Integration 82
4.1.7.1 Integration Rather Than Replacement 82
4.1.7.2 Dynamic Adaptation and Flexibility Through Sustained Reflection 83
4.1.7.3. Maintained Critical Stance Through Reflective Practice 83
4.2 Results for RQ2 83
4.2.1 Distribution of Writing Scores 84
4.2.2 Individual Performance Analysis 85
4.2.3 Comparative Performance Outcomes: Experimental vs Control Group 86
Experimental Group 86
Control Group 87
4.2.4 Distribution Trends and Score Range Summary 87
4.2.4.1 Visual Distribution Insight 88
4.2.6 Criterion-Specific Performance Analysis 92
4.2.7. Within-Group Statistical Analysis 93
4.2.8 Between-Group Statistical Analysis 94
4.2.9. Comprehensive Effect Size Analysis 95
4.2.10 Comprehensive Effect Size Analysis 98
4.2.11. Experimental Group Improvement Distribution 102
Comparative Analysis with Control Group 102
4.3 Results for RQ3 107
4.3.1 Emergent Themes 107

Theme 1. Cognitive Development Challenges 108
1.1 Deterioration in Critical Thinking and Evaluation 109
1.2 Cognitive Offloading and Reduced Mental Effort 109
1.3 Dependency Development and Eroded Independence 109
1.4 Cognitive Processing Difficulties 110
Theme 2: Technical and Digital Literacy Barriers 110
2.1 Digital Literacy Gaps and Prompting Difficulties 110
2.2 Tool Constraints and Language Complexity 111
2.3 Disruptions to Learning Flow 111
Theme 3: Autonomous Learning Development 112
3.1 Self-Regulatory Enhancement 112
3.2 Development of Learner Agency 112
3.3 Strategic Prompting Development 113
Theme 4: Motivational and Affective Factors 113
4.1 Enhanced Engagement and Writing Motivation 114
4.2 Reduced Performance Anxiety and Error Tolerance 114
4.3 Tool Convenience and Perceived Support 114
Theme 5: Cultural and Identity Considerations 115
5.1 Cultural Misalignment and Identity Tension 115
5.2 Linguistic Code-Switching and Processing Strain 115
5.3 Creative Voice Preservation and Style Homogenisation 116
Theme 6: Social Learning and Support Needs 116
6.1 Peer Interaction and Strategy Sharing 116
6.2 Pedagogical Scaffolding and Interpretation Support 117
6.3 Comparison Between Human and AI Feedback Sources 117
Theme 7: Metacognitive and Linguistic Awareness 117
7.1 Metalinguistic Development and Language Pattern Awareness 117
7.2 Strategic Evaluation and Discernment Behaviours 118
Theme 8: Content Quality and Appropriateness Concerns 118
8.1 Response Quality and Depth of Content 118
8.2 Tone, Register, and Language Style Mismatches 119
Theme 9: Ethical and Academic Considerations 119
9.1 Academic Integrity and Cheating Perceptions 119
9.2 Authorship Boundaries and Support Transparency 119

4.4. Temporal Development Patterns 120
4.5 Summary of Findings 120
4.4. Results for RQ4 120
Pre-A Phase: Access and Alignment Awareness 121
Practice-A Phase: Application and Engagement 122
Performance-A Phase: Autonomy and Integration Sophistication 122
Stagnation and Regression: Failed Phase Transition Patterns 123
Summary of RQ4 Results 123
CHAPTER V 124
DISCUSSION 124
Discussion 124
Integration of Quantitative and Qualitative Findings 124
5.1 AI's Influence on Learner Autonomy Development (RQ1) 124
5.1.1 The Paradox of Strategic Autonomy Versus Cognitive Dependency 124
5.2.1 Learner Perceptions and Affective Shifts (RQ2b) 127
5.2.2 Qualitative Writing Performance Changes (RQ2a) 128
5.2.4 Learner Autonomy and Feedback Practices (RQ1) 131
5.3.2 Perspectives on Autonomy Development 133
5.3.3 Perspectives on Writing Performance Changes 134
5.3.4 Long-term Learning Value Concerns 135
5.4 Framework for Authentic AI Integration in Writing Instruction (RQ4) 136
5.4.1 The 9A Catalyst Development 136
The Four Phases of the 9A Catalyst Framework: 136
1. Pre-A Phase – Foundation for Autonomy: 136
2. Practice-A Phase – Active Engagement: 136
3. Performance-A Phase – Ownership and Competence: 136
4. Personalisation-A Phase – Internalisation and Autonomy: 137
5.4.2 Cognitive Resistance Training 137
5.4.3 Pedagogically Informed Alternatives and the SAIL Model 138
5.4.4 Pedagogical Context Primacy Principle 138
5.5 Integrated Findings Across Research Questions 139
5.5.1 Convergent Evidence Pattern 139
5.5.2 Theoretical Implications 140
5.5.2.1 Extension of Self-Determination Theory 141

5.5.2.2. Integration with Sociocultural Theory 142
5.5.2.3. Cognitive Load Theory Applications and Extensions 142
5.5.2.4. Synthesis: The 9A Catalyst Framework 143
5.5.3 Practical Implementation Requirements 143
Foundational Implementation Conditions. 143
Structured Alternation and Assessment Alignment. 144
Avoiding Dependency Pitfalls. 144
Teacher-Centred Foundations for Autonomy. 144
Professional Development Pathways. 144
5.5.4 Pedagogical Implications 145
5.5.5 The 9A Catalyst Framework: A Comprehensive Model for AI-Instructed Autonomy Building in Writing Contexts 147
5.5.5.1. Framework Introduction and Theoretical Foundation 147
5.5.5.1 Phase-Based Architecture: Sequential Development Model 149
5.5.5.2 Pre-A Phase: Foundation for Autonomy (Cognitive and Cultural Preparation) 151
Access: Beyond Technical Availability to Comprehensive Digital Literacy 151
Alignment: Navigating Cultural and Pedagogical Consistency 151
Awareness: Metacognitive and Ethical Awareness Development 152
Aim: Purpose-Driven AI Integration 152
5.5.5.3 Practice-A Phase: Active Engagement and Skill Building 153
Application: Authentic Context Integration 153
Affective Engagement: Emotional Foundation for Sustained Learning 154
5.5.5.4 Performance-A Phase: Ownership and Digital Competence 154
Agency: Self-Direction in AI-Instructed Learning 155
AI Integration: Strategic Tool Orchestration 155
5.5.5.5 Personalisation-A Phase: Internalisation and Extension 156
Autonomy: Self-Directed Learning Mastery 156
Adjustment: Adaptive Strategy Development 157
5.6.1 Institutional Implementation Roadmap 158
Pre-A Phase: Foundational Infrastructure and Readiness 158
Access Infrastructure Development 158
Alignment Policy Development 158
Awareness Training Programmes 158

Aim Setting and Outcome Assessment 159
5.6.2 Educator Preparation and Professional Development 159
Pre-A Phase Competencies 159
Practice-A Phase Support Skills 159
Performance-A Phase Facilitation 159
Personalisation-A Phase Extension 160
5.6.3 Assessment and Monitoring Guidelines 160
Phase-Specific Evaluation Protocols 160
Dual-Context Performance Assessment 160
Cultural Sensitivity Monitoring 160
5.6.4 Student Support Strategies Across Phases 160
Pre-A Phase Support 160
Practice-A Phase Support 161
Performance-A Phase Facilitation 161
Personalisation-A Phase Empowerment 161
5.6.5 Differentiated Support Strategies 161
Students with High Digital Literacy 161
Culturally Diverse Learners 162
Learners with Low Confidence 162
High-Achieving Students 162
5.6.5 Differentiated Support Strategies 162
Students with High Digital Literacy 162
Culturally Diverse Learners 163
Learners with Low Confidence 163
High-Achieving Students 163
5.7 Major Findings Summary 164
CHAPTER VI 166
CONCLUSION 166
6.1 Summary of Findings and Framework Validation 166
6.1.2 Temporal Developmental Patterns 166
6.2 Contradictions and Theoretical Contributions 167
6.2.1 Control Group Outperformance and Pedagogical Context Primacy 167
6.3 Contributions of the 9A Catalyst Framework 167
6.4 Limitations 168

6.5 Directions for Future Research 168
6.6 Final Reflections 168
REFERENCES 170
EK-1. 171
VELİ ONAM FORMU 171
İletişim Bilgileri: 172
VELİ ONAM BEYANI 172
Appendix 1. Ethics Committee Approval 173
Appendix 3. Ethics Committee Approval 174
Appendix 2: The Approval of the Institute of Educational Sciences 175
Appendix 3: Informed Consent 176
Appendix 4: Demographic Questions 178
Appendix 5: Resume 179
Appendix 8: Reflective Diary 180
A. Reflective Diary 180
Section A 180
Strengths: 180
Challenges: 180
Pompts You Used and Suggestions for Improvement: 181
Section B 181
Section C 181
Appendix 9: Ankara Üniversitesi Araştırma İzni 182
Appendix 10. 8 Weeks Essay Topics 183

TABLOLAR LİSTESİ

Tablo *Sayfa*

Table 1 Core Dimensions of Digital Literacy in AI-Instructed EFL Writing 27

Table 2 Major AI Tools in EFL Writing: Functions & Pedagogical Caveats 29

Table 3 Experimental vs. Control Groups 52

Table 4 Demographic Characteristics of Study Participants by Group 58

Table 5 Inter-Rater Reliability Coefficients for Writing Assessment Criteria 59

Table 6 Normality Test Results for Total Writing Scores by Group and Test Phase 60

Table 7 Descriptive Statistics for Writing Performance by Group and Test Phase 61

Table 8 Holm-Bonferroni Sequential Adjustments for Multiple Comparisons 65

Table 9 Selected Statistical Methods 67

Table 10 Pseudonyms, pre/post scores, and skill-specific deltas 85

Table 11 Detailed Pre-Post Test Median, IQR, Min Max Table of between-groups 87

Table 12 Interquartile Ranges (IQR) of Writing Scores by Group and Phase 89

Table 13 Mean Gain Scores by Writing Criterion and Group 92

Table 14 Within-Group Pre-Post Comparisons Using Wilcoxon Signed-Rank Tests 93

Table 15 Between-Group Comparisons at Post-Test Using Mann-Whitney U Tests 94

Table 16 Comprehensive Effect Size Analysis: Within-Group and Between-Group Comparisons 95

Table 17 Rubric-Aligned Thematic Analysis of Writing Errors in Post-Test Essays 103

Table 18 Frequency and Intensity Classification of Writing Errors by Rubric Criterion 105

Table 19 Punctuation/Mechanics Error Frequencies by Group and Error Type 105

Table 20 Comparison of Artistic Language Use in Post-Test Essays by Group 106

Table 21 Summary of Major Themes from Qualitative Analysis (N = 2,247 codes) 108

Table 22 Temporal Development Patterns in Student Perceptions Across 8-Week Intervention 147

Table 23 Evidence Supporting 9-A Framework Component 150

ŞEKİLLER LİSTESİ

Şekil *Sayfa*

Figure 1 Benson model....16
Figure 2 Reactive Autonomy Framework (Littlewood, 2022) 17
Figure 3 Zimmerman's three-phase model 18
Figure 4 Adapted from "Considering Contextual Knowledge 31
Figure 5 Evalution of Autonomy 37
Figure 6 Blooms Taxonomy Revised for AI-instructed EFL Writing 40
Figure 7 Integration of traditional self-regulation skills and digital-specific competencies in technology-assisted learner autonomy....44
Figure 8 Technology-Assisted Self-Regulated Learning Strategies Questionnaire (TSRLSQ) 45
Figure 9 Rubric – Ankara University TÖMER Official B1.3 Certificate Exam Rubric 54
Figure 10 Quantile-Quantile (Q-Q) Plots for Normality Assessment 66
Figure 11 Pre-Test and Post-Test Writing Performance by Group and Criterion....69
Figure 12 Within-Group Effect Sizes (Cliff's Delta) for Writing Improvements....70
Figure 13 Main Themes in Learner Autonomy AI-Instructed Contexts 74
Figure 14Comparison of Control Group and Experimental Group Performance Across Multiple Measures 81
Figure 15 Distribution of Total Writing Scores by Group and Test Phase 84
Figure 16 Histogram of Writing Scores by Group and Test Phase 88
Figure 17 Boxplot of Pre-Test and Post-Test Scores by Group 89
Figure 18 Histograms of Writing Score Distributions by Group and Test Phase....90
Figure 19 Mean Gain Scores by Writing Criterion with 95% Confidence Intervals 93
Figure 20 Individual Participant Trajectories for Total Writing Scores....96
Figure 21 Comparative Effect Sizes: Within-Group vs. Between-Group Differences 98
Figure 22 Mean Gain Scores by Writing Criterion with 95% Confidence Intervals 100
Figure 23 Coefficient of Variation Analysis: Performance Consistency by Group 100
Figure 24 The 9A Catalyst Framework for AI-Instructed Autonomy Development in EFL Writing 121
Figure 25 The 9A Catalyst Framework....137

List of Abbreviations

AI	: Artificial Intelligence
AI-CK	: Artificial Intelligence Content Knowledge
AI-K	: Artificial Intelligence Knowledge
AI-PCK	: Artificial Intelligence Pedagogical Content Knowledge
AI-PK	: Artificial Intelligence Pedagogical Knowledge
CEFR	: Common European Framework of Reference for Languages
CLIL	: Content and Language Integrated Learning
DPACK	: Digital Pedagogy Content Knowledge
EFL	: English as a Foreign Language
ELT	: English Language Teaching
HITL	: Human-In-The-Loop
HOTs	: Higher-Order Thinking Skills
ICT	: Information and Communication Technology
IELTS	: International English Language Testing System
Intelligent-TCK	: Intelligent Technological Content Knowledge
Intelligent-TK	: Intelligent Technological Knowledge
Intelligent-TPACK	: Intelligent Technological Pedagogical Content Knowledge
Intelligent-TPK	: Intelligent Technological Pedagogical Knowledge
LLMs	: Large Language Models
ML	: Machine Learning
SD	: Standard Deviation
ICC	: Intraclass Correlation Coefficient
SDT	: Self-Determination Theory

CAP	: Create, Analyze, Apply (Model)
ZPD	: Zone of Proximal Development
SRL	: Self-Regulated Learning
CLT	: Cognitive Load Theory
TCK	: Technological Content Knowledge
TK	: Technological Knowledge
TPACK	: Technological Pedagogical Content Knowledge
TPK	: Technological Pedagogical Knowledge
GAI	: Generative Artificial Intelligence
GenAI	: Generative Artificial Intelligence
TÖMER	: Türkçe Öğretim Merkezi

CHAPTER I
INTRODUCTION
1.1. Statement of the Problem

The cultivation of learner autonomy is recognised as being influenced by a multitude of factors. Among these, learner attitudes and motivation are paramount, with positive attitudes fostering both instrumental and integrative motivation, both of which are considered essential for learning to occur (Reinders, 2010; Thanasoulas, 2000), which can also be related to Krashen(1982)'s Affective Filter Hypothesis in terms of the incorporation of emotions in the learning. Moreover, the learning environment itself plays a critical role in either facilitating or hindering the development of autonomous learning capabilities, which is also a part of social learning (Vygotsky, 1978)

The demands of this new century are distinctly characterised by the imperative for advanced cognitive skills, often encapsulated as 21st-century skills. Among these, learner autonomy has emerged as a non-negotiable stand-out attribute, essential for individuals to navigate a world of constant change, information overload, and complex problem-solving (OECD, 2019). As society increasingly values self-direction, critical thinking, and adaptability, the need for educational models that actively cultivate learner autonomy has become more pivotal than ever before.

Moreover, the ethical considerations of implementing AI in education require careful attention to ensure that technological integration empowers learners fairly and responsibly (Holmes et al., 2021). This dual focus – on fostering autonomy and ensuring ethical AI usage – necessitates a deeper and more nuanced understanding of the human learning process itself. Yet, the term "Autonomy" is a slippery slope in the field of education regarding the diverse definitions throughout history (Holec, 1981; Dörnyei, 2001; Benson, 2011; Littlewood, 2019; Alm, 2024). Compounding the nuanced nature of learner autonomy is the prevalence of misconceptions and related but distinct terms within the field. While 'learner autonomy' has become a prevalent term in language learning, it is often used synonymously with terms like self-instruction, self-directed learning, and independent learning despite subtle yet significant differences (Chong & Reinders, 2022; Nga, 2014). Learner autonomy represents the intrinsic capacity, the 'will and ability', to take charge of one's own learning, whereas self-directed learning refers to the set of processes and behaviours through which that capacity is realised" (Gupta et al., 2024).

The rapid advancement of artificial intelligence (AI) has greatly affected and continues to significantly influence various sectors, including education, as one of the milestones of human life. It has rapidly transitioned from the realm of science fiction to an undeniable effect on our lives, especially after the pandemic era which, as Zhao and Watterston (2021) argue, acted as a significant catalyst, propelling AI into a more prominent and integral role across various societal domains due to the urgent need for digital solutions and adaptations spesifically in Pandemic. Its pervasive influence is felt across sectors ranging from industry and communication to, most notably, education (Popenici & Kerr, 2017). Rapid advancements in the technology era and the integration of AI within the educational landscape represent a profound diversification of the roots of education, challenging the foundational principles of traditional pedagogical approaches and necessitating further exploration (Misirli&Ergulec, 2021).

The writing learning processes in AI-integrated classrooms have become a significant focus in EFL pedagogy, warranting further investigation into how tools like ChatGPT may enhance autonomy and skill development (Ng et al., 2021; Popenici & Kerr, 2017) . Kellogg (1996) proposes that writing is itself a form of cognition, relying on working memory for planning, translating and reviewing, which positions writing as a self-explanatory, reflective device rather than merely a communication skill. From a developmental perspective, effective writing instruction emphasises deliberate practice that gradually reduces cognitive load through apprenticeship models (Hayes & Flower, 1980; Bereiter & Scardamalia, 1987). Writing remains an essential skill for EFL learners; yet, it is unclear whether on-demand, rapidly evolving systems such as ChatGPT reliably boost writing proficiency and foster autonomous self-improvement, early studies show mixed results on motivation gains and mixed impacts on higher-order composition quality (Liu et al., 2024; Zhang et al., 2023). For decades, educational systems have relied on standardised methodologies, conventional assessments and rigid curricula, approaches increasingly critiqued for their limited effectiveness in preparing learners for 21st-century complexities (Darling-Hammond, 2010; Voogt & Pareja Roblin, 2010)

In this context, AI tools have been positioned as potential facilitators of learner autonomy, providing students with instant feedback, personalized writing support, and iterative improvement opportunities. However, the extent to which AI enhances autonomy, and whether it inadvertently reinforces dependence on technology, remains an open question. Moreover, while AI-generated feedback offers real-time corrections and suggestions, it lacks the nuanced, human-like contextual understanding that traditional instructors provide (Ranalli, 2021). This

raises concerns about whether AI-assisted writing fosters genuine skill development along with learner autonomy or encourages mechanical dependency on automated corrections.

In this context, AI tools have been positioned as potential facilitators of learner autonomy, providing students with instant feedback, personalised writing support, and iterative improvement opportunities. However, the extent to which AI enhances autonomy, and whether it inadvertently reinforces dependence on technology, remains an open question. Moreover, while AI-generated feedback offers real-time corrections and suggestions, it lacks the nuanced, human-like contextual understanding that traditional instructors provide (Ranalli, 2021). This raises concerns about whether AI-assisted writing fosters genuine skill development along with learner autonomy or encourages mechanical dependency on automated corrections.

1.2. The Purpose of the Study

This study aims to investigate the influence of integrating generative artificial intelligence, specifically ChatGPT, on learner autonomy and writing proficiency within an EFL instructional context. Recognising the growing prominence of AI tools in educational environments, this research addresses critical gaps by exploring how structured AI-instructed feedback interacts with students' autonomous learning processes and affects their development in writing competencies over an eight-week intervention.

In this context, the specific objectives of the study are as follows:

1. To determine the comparative effects of AI-assisted feedback and traditional teacher feedback on learners' writing proficiency, particularly in areas such as content development, lexical variety, grammatical accuracy, mechanical correctness, and artistic language use.
2. To analyse how students' interactions with generative AI tools influence their strategic behaviours, reflective thinking, and learner autonomy throughout the writing process.
3. To examine in depth the potential autonomy-enhancing or dependency-inducing impacts of generative AI based on students' experiences.
4. To explore pedagogically meaningful ways of integrating AI into EFL writing instruction that support learner autonomy without compromising cognitive engagement or critical thinking.

By combining quantitative measures with rich qualitative insights from reflective diaries and semi-structured interviews, this study offers an integrated understanding of the pedagogical implications, challenges, and potentials of AI in enhancing writing skills and

fostering genuine autonomy among EFL learners. Ultimately, it aims to provide educators and curriculum developers with actionable guidance for the responsible and effective use of generative AI tools in language education contexts.

1.2. Research Questions

This study seeks to answer the following research questions:

1. How does the integration of AI-instructed feedback influence learner autonomy in EFL writing contexts?
2. To what extent does AI-assisted writing instruction affect learners' writing proficiency compared to traditional teacher-instructed feedback?

3. How do EFL learners in the experimental group perceive the development of their learner autonomy through reflective diaries during the AI-assisted writing process?

4. What pedagogical strategies can be derived from the integration of AI in EFL writing instruction to support learner autonomy while mitigating the risks of cognitive dependency?

1.4. Significance of the Study

This study contributes to the field of English Language Teaching by providing empirical evidence on the pedagogical potentials and challenges of integrating generative AI tools into EFL writing instruction. In particular, it investigates whether AI-instructed feedback can effectively enhance learners' autonomy and writing proficiency, while also exploring the underlying factors that account for individual learner differences, even among students at similar proficiency levels.

A key significance of this study lies in its decision to forgo explicit instruction in AI literacy or digital tool training during the intervention. This design choice reflects real-world classroom dynamics in which learners often encounter AI feedback tools without structured guidance. Consequently, the study offers a more authentic perspective on how autonomy develops in unstructured, technology-rich environments.

The findings are also relevant to discussions surrounding fossilisation in language learning. As Selinker (1972) argued, repeated exposure to incorrect input without corrective intervention may lead to permanent errors. In this context, the uncritical use of AI-instructed feedback may contribute to the reinforcement of errors if not pedagogically mediated.

Furthermore, this study draws attention to the digital divide (Van Dijk, 2005) as a critical variable in AI-instructed instruction. Learners' digital competence levels directly impact their capacity to benefit from AI tools: while some navigate platforms independently and effectively,

others face technical barriers that limit their engagement. This disparity underscores the importance of considering when and how digital literacy training should be incorporated to ensure equitable access and outcomes in AI-instructed writing environments.

By highlighting these nuanced dimensions, this study offers valuable insights for educators, curriculum designers, and policymakers seeking to implement AI-instructed instruction in ways that are both effective and equitable.

1.5. Limitations of the Study

While the present study aims to explore the role of generative AI in EFL writing instruction, it is important to interpret its outcomes within the context of certain inherent limitations. These limitations, arising from the scope, methodology, and the rapidly evolving nature of AI technologies, should be considered when evaluating the study's generalisability and pedagogical implications.

Firstly, the study involved a relatively small sample, with unequal group sizes resulting from participant attrition during the intervention. Some students withdrew from the process to maintain their personal learning preferences and autonomy, reducing the statistical power of comparative analysis.

Secondly, the intervention was conducted over a period of eight weeks, which may not be sufficient to observe long-term changes in learner autonomy or writing proficiency. Future research could benefit from longitudinal designs to better capture sustained developmental trajectories.

Thirdly, no explicit digital or AI literacy training was provided. Consequently, participants' ability to benefit from ChatGPT 4.5 feedback depended on their pre-existing technological competence, potentially disadvantaging less digitally literate learners and introducing disparities in tool use and outcomes.

Additionally, the voluntary nature of participation introduced potential self-selection bias. Learners who chose to participate may have been more motivated or technologically inclined, which could limit the representativeness of the sample.

The instructional context was also highly specific. The study was conducted in a particular institutional setting, using the free-tier version of ChatGPT-4 with restricted daily access to ChatGPT-4.5 features. Furthermore, levels of instructor mediation varied across sessions. These factors may limit the transferability of findings to other contexts, platforms, or proficiency levels.

Other uncontrolled variables such as learners' access to devices, exposure to English outside the classroom, or varying levels of informal feedback may also have influenced outcomes in ways not directly measured by the study.

1.5.1. Reliance on Self-Reported Data for Autonomy

The assessment of learner autonomy in this study relied, in part, on self-reported data collected through reflective diaries and semi-structured interviews. While these methods provide rich insights into learners' perceptions and experiences, they are also subject to limitations. For instance, participants may have responded in socially desirable ways or lacked full awareness of their own cognitive and behavioural strategies. These factors could affect the accuracy and depth of reported autonomy development. Future research could enhance the validity of autonomy assessment by incorporating behavioural or performance-based measures alongside self-report instruments to allow for more robust triangulation.

Acknowledging these limitations provides essential context for interpreting the findings of this study. They also serve as valuable directions for future research, highlighting the need for continued investigation into the dynamic and evolving intersection of generative AI, EFL writing pedagogy, and learner development across diverse contexts and with varied methodological approaches.

1.6. Delimitations of the Study

This study is subject to several delimitations arising from intentional design choices made to ensure clarity of focus and feasibility. The research was limited to B1-level EFL learners enrolled in a single language institution, excluding learners at other proficiency levels, educational backgrounds, or institutional settings. The study concentrated exclusively on writing skills, thereby omitting other language domains such as speaking, listening, or reading.

Moreover, the intervention was designed to investigate the use of ChatGPT as an AI-instructed feedback tool in writing instruction, without offering explicit training in digital or AI literacy. The study's scope was further defined by an eight-week implementation period and a focus on two primary outcomes: writing proficiency and learner autonomy.

These delimitations were established to facilitate a targeted and in-depth examination of AI-instructed feedback within a bounded, authentic classroom context. While they constrain the generalisability of the findings, they also support a more detailed analysis of the pedagogical dynamics involved.

1.7. Ethical Considerations

This research was carried out with a deep sense of responsibility towards the well-being, dignity, and rights of all participants. Ethical safeguards were integrated into every stage of the study to ensure that participation was fully informed, voluntary, and respectful of individual privacy.

Before the intervention began, all participants received a clear and accessible consent form (see Appendix 1), which explained the purpose of the study, the type of data to be collected, and the participants' right to withdraw at any point without any negative consequences. The form also detailed the collection of basic demographic information, such as gender and language learning background, along with learners' writing samples and reflections related to the use of AI feedback tools.

Participants were assured that their identities would remain anonymous, and all data would be treated confidentially. At no point were individuals personally identified in the reporting or analysis of results. The research design was reviewed and approved by the researcher's academic supervisor and conducted in accordance with the ethical principles outlined by the Institute of Educational Sciences.

Throughout the process, care was taken to ensure that no participant was put at risk or felt pressured to respond in any particular way. The aim was to create an open and respectful environment in which learners could express their experiences honestly and without fear of judgment. Their trust and willingness to share their journey are acknowledged with sincere appreciation.

1.7.1. Ethical Considerations and AI Tool Disclosure

In accordance with ethical research practices and transparency requirements, the authors disclose the use of artificial intelligence tools during the preparation of this manuscript. Claude (Anthropic), ChatGPT (OpenAI), and Grammarly were utilised solely for language enhancement, including grammar correction, style improvement, and academic tone refinement. These tools were not used for data analysis, interpretation, or the generation of research content or conclusions. All substantive intellectual content, research design, data collection, analysis, and interpretation remain entirely the work of the authors, who take full responsibility for the accuracy and integrity of the research presented.

1.7.1. Confidentiality and Anonymity

All data collected from participants were handled with the utmost sensitivity. Individual responses were anonymised at the point of collection, and pseudonyms or group codes were used during analysis and reporting. No personal identifiers were retained or linked to any data, ensuring that participants could not be identified in any publications or presentations of the research.

1.7.2. Data Security and Storage

To ensure data security, all written responses, diary entries, and interview transcripts were stored in password-protected digital formats accessible only to the primary researcher. Any physical documents were kept in locked storage at a secure location. Data will be retained only for the duration necessary to complete the research and any related academic dissemination, after which it will be permanently deleted in accordance with data protection guidelines.

1.7.3. Minimising Potential Risks and Maximising Benefits

While the study aimed to offer participants an enriching learning experience through the integration of AI-instructed tools, care was taken to minimise any potential risks. The researchers remained aware of the possibility of learners feeling overly reliant on AI or anxious about using unfamiliar technology. To mitigate such concerns, participants were clearly informed of the pedagogical purpose of using AI and reminded that it was intended as a support tool, not a replacement for their thinking and learning. Debriefing sessions were held at the end of the study to provide clarification, address concerns, and offer further resources.

At the same time, the potential benefits, such as improved writing fluency, enhanced learner autonomy, and contribution to future EFL instructional design, were shared with participants to foster transparency and motivation.

1.7.4. Institutional Approvals and Informed Consent

Participation in the study was voluntary. Ethical approval was granted by the Social Sciences and Humanities Research and Publications Ethics Committee at Bursa Uludağ University (see Appendix 2). Institutional permission to conduct the study was obtained from Ankara University, TÖMER (see Appendix 3). Additionally, before any data collection commenced, participants were informed about the study objectives, data usage, and privacy measures through a Google Forms information page and consent agreement (see Appendix 1).

1.8. Definitions of Key Terms

To ensure clarity, several key terms are defined as they are used throughout this thesis:

Artificial Intelligence (AI): Computer-based systems capable of performing tasks that typically require human intelligence, such as language processing, learning, and problem-solving.

AI-instructed feedback: Automated feedback provided to learners by artificial intelligence tools (e.g., ChatGPT) during the writing process, typically based on user-generated prompts and responses.

Learner autonomy: The capacity of learners to take charge of their own learning, including setting goals, selecting strategies, self-monitoring, and evaluating their progress (Holec, 1981; Benson, 2011).

Digital literacy: The ability to effectively locate, evaluate, use, and communicate information through digital technologies.

Writing proficiency: The ability to produce written language that is coherent, accurate, and appropriate to context, genre, and communicative purpose, typically measured through standardised rubrics.

Prompt engineering: The process of designing and refining input queries for AI tools in order to elicit meaningful, accurate, and relevant feedback or responses.

1.9. Organisation of the Thesis

This thesis is organised into five chapters. Chapter I presents the background, problem statement, research questions, objectives, significance, limitations, delimitations, ethical considerations, and definitions of key terms. Chapter II provides a comprehensive review of the literature related to AI integration in EFL writing, learner autonomy, feedback literacy, digital competence, and the theoretical frameworks guiding the study. Chapter III outlines the research methodology, including the research design, participant selection, data collection tools, and the procedures used for data analysis. Chapter IV presents the findings of the study, integrating quantitative and qualitative results to address the research questions. Finally, Chapter V discusses the findings in relation to existing literature, explores pedagogical implications, acknowledges the limitations of the study, and offers directions for future research.

1.10. Summary

In summary, this chapter has introduced the background and rationale of the study, outlined its research questions and objectives, and described its scope, significance, limitations,

and ethical foundations. Key concepts have been defined, and the structure of the thesis has been explained. The next chapter will review relevant literature on artificial intelligence in education, learner autonomy, feedback practices, and digital competence, while presenting the theoretical frameworks that inform the design and interpretation of the study.

CHAPTER II
LITERATURE REVIEW

The literature review section explores previous research and theoretical frameworks related to AI, AI usage in writing and learner autonomy in the EFL context. This provides a foundation for understanding key concepts, identifying gaps, and positioning the current study within a broader context in the literature.

2.1 Overview of Writing in EFL Contexts

This literature review addresses some critical knowledge gaps by examining the intersection of AI integration and learner autonomy in EFL writing contexts. The analysis synthesises theoretical frameworks from sociocultural learning theory, autonomous learning theory, cognitive processing models, and feedback intervention research to construct a comprehensive understanding of AI's catalytic potential in autonomous learning development. The review specifically focuses on recent empirical research heavily from 2022 to 2025 that examines comparative outcomes between AI-instructed and conventional instruction approaches, with particular attention to studies employing experimental and quasi-experimental designs relevant to understanding causal relationships between AI integration and autonomy development.

2.1.1 Challenges in L2 Writing

Contemporary studies highlight several persistent difficulties in EFL writing development. First, linguistic constraints frequently limit learners' ability to express complex ideas, resulting in content that may underestimate their cognitive potential. Second, cultural and rhetorical differences between the learners' first language and English complicate efforts to adopt target language discourse conventions. Third, restricted access to authentic language input often hinders learners' understanding of genre expectations and stylistic appropriateness. Finally, in many EFL settings, a lack of timely, individualised feedback limits opportunities for iterative revision, which is essential for writing improvement (Barrot, 2023; Wang, 2024).

In addition to these cognitive and linguistic barriers, affective factors play a significant role in shaping learners' writing experiences. Research consistently reports high levels of writing anxiety, reduced self-efficacy, and fluctuating motivation among EFL learners. These

emotional factors interact with the cognitive load of writing, creating further obstacles that traditional instructional approaches have struggled to overcome.

The emergence of AI-powered writing tools, such as ChatGPT, offers promising responses to many of these challenges by providing immediate, context-sensitive feedback and enabling individualised writing support. However, important questions remain regarding how these tools can be optimally integrated into instruction and whether their use fosters meaningful learning or encourages surface-level engagement (Song & Song, 2023; Mahapatra, 2024).

2.1.2. Writing Instruction in Turkish and Global EFL Settings

Writing instruction in Turkish EFL contexts mirrors many global trends while also reflecting unique influences stemming from national education policy, cultural expectations, and institutional constraints. Historically, the Turkish education system has placed a strong emphasis on grammatical accuracy and structural correctness, often at the expense of communicative competence and rhetorical fluency. As a result, learners frequently demonstrate considerable grammatical knowledge but struggle to produce coherent and authentic written texts (Kırkgöz & Ağçam, 2022; Karakaş & Kartal, 2023).

In recent years, curriculum reforms have sought to promote more communicative and process-oriented approaches to writing. However, implementation has been uneven. Factors such as large class sizes, limited instructional time, high-stakes examinations, and a lack of professional development opportunities for teachers have limited the adoption of these practices. This is particularly evident in public schools, where resource shortages and traditional teaching norms continue to dominate instructional practice despite policy-level encouragement for innovation (Solak, 2024).

Similar challenges are observed in many global EFL settings, although the nature and extent of these challenges vary by region. For instance, countries in East Asia, including China, Japan, and South Korea, have been among the first to explore the integration of AI-instructed writing tools in classroom practice. While initial results are promising, studies have raised concerns about over-reliance on technology and the risk of diminishing learners' critical thinking and originality. In contrast, many European systems have adopted a more holistic approach to writing instruction, emphasising multilingual awareness, intercultural competence, and learner agency, which in turn influence how digital tools are implemented (Yang & Aryadoust, 2024; Chen et al., 2025).

A common issue across both Turkish and global contexts is the persistent digital divide. Access to technology and digital literacy remains uneven, often correlated with geographical

and socioeconomic factors. While schools in urban areas may benefit from more reliable infrastructure and digital tools, institutions in rural or underserved communities often face significant barriers to implementation. These disparities underscore the importance of designing AI-instructed writing programmes that are adaptable, inclusive, and sensitive to local contexts (Van Deursen & van Dijk, 2019; Wilson et al., 2024).

2.1.3. AI Integration in EFL Writing

The integration of artificial intelligence into English as a Foreign Language (EFL) writing instruction marks a significant turning point in language education. This shift moves beyond technological adoption and represents a broader rethinking of pedagogical processes, learner agency, and instructional design. Since 2020, a growing body of meta-analytical research has reported moderate to large effect sizes for AI use in language learning, particularly in writing contexts (Chen et al., 2025; Shi & Aryadoust, 2024). However, despite these promising results, questions remain regarding the effects of AI on learner autonomy. Specifically, there is a need to examine how AI-instructed tools can support authentic independent learning without creating new dependencies.

Generative AI tools, especially large language models like ChatGPT, have introduced new opportunities for personalised writing instruction and on-demand feedback. These tools are increasingly being used to address long-standing issues in EFL classrooms, such as limited individual feedback, inconsistent error correction, and restricted opportunities for revision. Recent studies have highlighted significant gains in writing fluency, accuracy, and student engagement, with learners reporting greater motivation and confidence as a result of using AI tools (Mahapatra, 2024; Song & Song, 2023; Wale, 2024).

Nonetheless, this rapid adoption has raised concerns about the potential for over-reliance and cognitive outsourcing. Traditional definitions of learner autonomy emphasise decision-making, self-regulation, and responsibility for learning (Holec, 1981; Benson, 2011; Littlewood, 2019). While AI tools can scaffold learning and provide meaningful support, they may also interfere with the development of independent cognitive strategies. As Yang and Aryadoust (2024) and Apriani et al. (2024) report, AI use can increase motivation and reduce anxiety, but it may also encourage superficial engagement with feedback and reduced reflection.

For example, Apriani et al. (2024) found that while students using ChatGPT demonstrated increased writing confidence, many incorporated AI suggestions without critical evaluation, leading to limited improvements in strategy use. Similarly, Xu (2024) observed that learners often relied on AI to generate or structure their writing, bypassing essential planning

processes. These findings underscore the importance of feedback literacy, the learner's ability to interpret, evaluate, and apply feedback, as a key moderator of effective AI use (Mahapatra, 2024).

Moreover, Mahapatra (2024) observed that improvements in grammar and cohesion were only consistently achieved when learners received explicit instruction on how to engage with AI feedback. Without this scaffolding, gains tended to remain surface-level. This supports arguments made by Ruelens (2019), who warned that technological access alone does not guarantee deeper learning. Without guided reflection, digital tools may inadvertently foster dependency rather than autonomy.

The question, therefore, is not simply whether AI improves writing outcomes, but whether it supports genuine cognitive development and learner agency. As Alm (2024) suggests, AI tools can serve as sociocultural mediators within Vygotsky's zone of proximal development, offering scaffolding that promotes higher-level performance. However, without pedagogical mediation, such tools may remain external supports rather than being internalised into learners' strategic repertoires.

A recent meta-analysis by Yang and Aryadoust (2024) supports this view. While AI integration was linked to gains in vocabulary, grammar, and overall coherence, the studies that lacked reflective scaffolding reported weaker outcomes in learner independence and critical thinking. Similarly, Nguyen and Nguyen (2024) found that learners with limited digital literacy frequently misapplied AI feedback, resulting in superficial revisions and, in some cases, complete misunderstanding of task expectations.

These findings reflect a broader concern within the field: the risk of introducing sophisticated AI tools without adequate pedagogical framing or learner preparation. McCann et al. (2025) and Szabó and Szoke (2024) stress the need for more longitudinal research that focuses not only on immediate writing performance but also on longer-term impacts on learner autonomy and strategic development.

In sum, the integration of AI into EFL writing holds considerable potential, but its success depends largely on instructional design and learner readiness. Effective use of AI tools requires not only access but also digital competence, feedback literacy, and pedagogical support that encourages reflection and critical engagement. The next section explores the theoretical foundations underpinning these issues, highlighting the cognitive, sociocultural, and autonomy-based models that inform the design and evaluation of AI-instructed writing instruction.

2.2. Learner Autonomy in Language Education

2.2.1. Definitions and Dimensions

Learner autonomy is most commonly defined as "the ability to take charge of one's own learning" (Holec, 1981, p. 3). Over time, this concept has evolved from a narrow interpretation of independence into a more complex and dynamic construct that encompasses cognitive, metacognitive, affective, and social dimensions. Modern perspectives reject the idea that autonomy implies complete detachment from instructors or support systems. Instead, autonomy is viewed as the learner's capacity to make informed decisions about learning goals, strategies, and evaluation, while effectively making use of available tools and guidance, including technological supports (Benson, 2011; Littlewood, 2019).

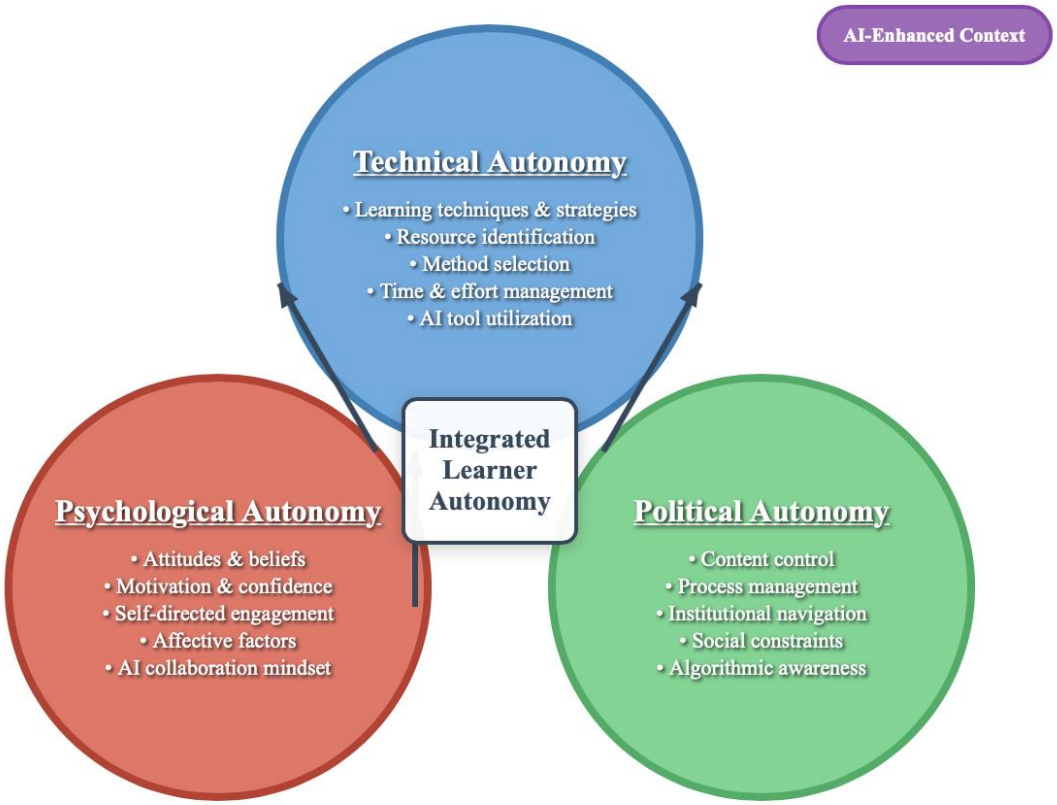
Benson (2011) offers one of the most widely cited models of learner autonomy, identifying three interdependent dimensions that continue to inform research in digitally mediated learning environments. These include:

> **Technical autonomy** refers to the learner's ability to apply learning strategies, organise time effectively, and select appropriate resources to meet learning objectives. This includes navigating AI tools and digital platforms independently.
>
> **Psychological autonomy** relates to internal learner characteristics such as motivation, self-confidence, mindset, and beliefs about learning. These affective and cognitive traits influence the degree to which learners take responsibility for their own learning processes.
>
> **Political autonomy** concerns the extent to which learners can influence or control the content and direction of their learning within institutional, social, or technological constraints. This dimension has become increasingly relevant in AI-instructed learning environments, where algorithmic feedback and tool design may shape what is emphasised, suggested, or left invisible.

These three dimensions, technical, psychological, and political, form a useful lens through which to understand how AI tools like ChatGPT may support or constrain different aspects of autonomy. This framework provides a conceptual foundation for interpreting how learners engage with feedback and make independent learning decisions in AI-enhanced writing contexts.

Figure 1

Benson's model of learner autonomy

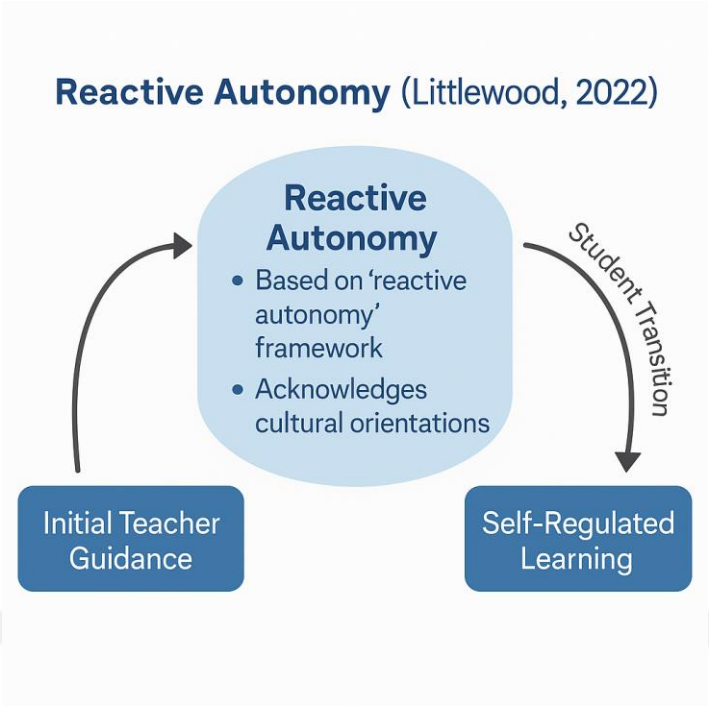


Note. This framework illustrates Benson's (2011) three interdependent dimensions of learner autonomy. It's adapted to reflect its application in AI-instructed learning environments.

Littlewood's (2019) distinction between proactive and reactive autonomy proves particularly relevant for understanding AI's role in learner development (see Figure 1). AI tools demonstrate capacity to support both forms: enabling proactive autonomy through learner-initiated conversations and self-directed topic exploration, while facilitating reactive autonomy through structured prompts and guided practice within educational frameworks. However, recent research reveals a more complex dynamic where learners develop "collaborative autonomy", the capacity to engage productively with AI while maintaining critical awareness and personal agency (McCann et al., 2025; Teng, 2025; Alm, 2024).

Figure 2

Reactive Autonomy Framework (Littlewood, 2022)



Note. This figure illustrates the relationship between learner autonomy and guided educational support in digital learning environments. It highlights the dynamic interplay between proactive autonomy, reactive autonomy, and collaborative autonomy, demonstrating how AI tools can scaffold learner independence when integrated with reflective practices and strategic mediation

The most significant theoretical shift in recent years concerns the evolution of learner agency from an individual construct to a more distributed and contextualised one. While traditional EFL pedagogy has often prioritised teacher-led instruction, contemporary research increasingly emphasises the importance of learner self-regulation in achieving long-term language development (Dörnyei & Ryan, 2015). To conceptualise this process, Zimmerman's (2000) model of self-regulated learning offers a valuable framework, outlining how learners move through three cyclical phases, forethought, performance, and self-reflection, each of which involves distinct cognitive and metacognitive operations essential to autonomous language acquisition.

However, in digitally enhanced learning environments, autonomy must be reconceptualised to account for new forms of learner–tool interaction. Autonomy is no longer simply defined by independence from instruction, but rather by the capacity to make intentional, reflective decisions in environments mediated by technology. This shift has given rise to the notion of **collaborative autonomy**, in which learners maintain a sense of agency while engaging meaningfully and critically with AI tools. Rather than undermining autonomy, such technologies may offer strategic scaffolding, provided that learners possess the digital literacy

and metacognitive awareness necessary to evaluate and adapt AI-generated feedback (Cooker & Benson, 2023; Chen, 2025).

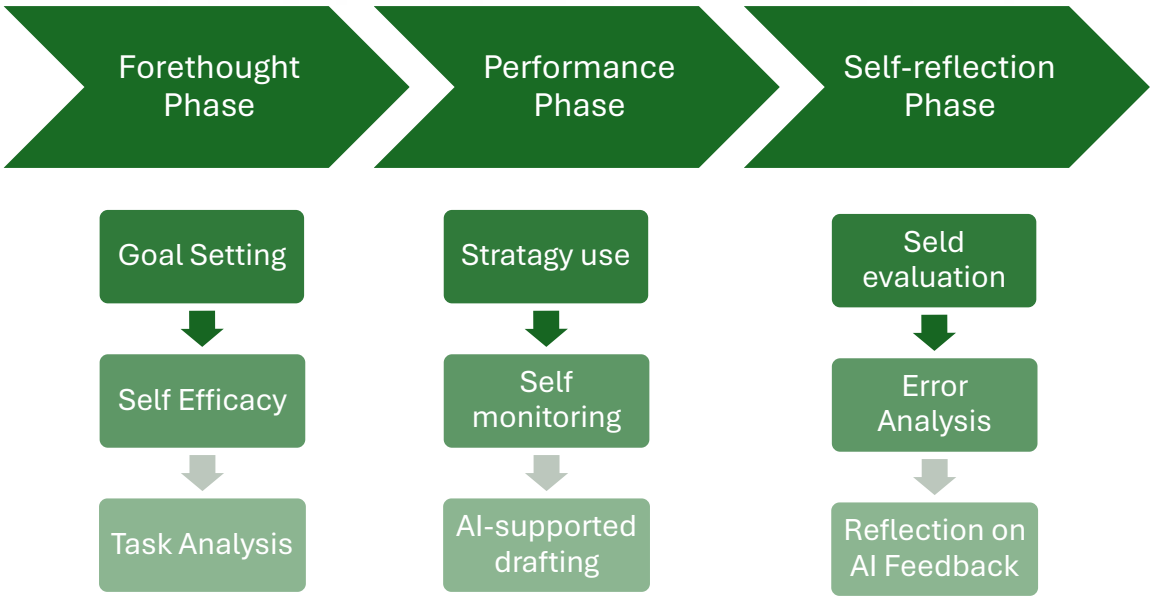
In this light, 21st-century autonomous learners must develop dual competencies: the ability to self-direct their learning behaviours and the capacity to navigate, assess, and co-construct meaning with intelligent systems. Collaborative autonomy does not signal a retreat from self-regulation but rather its extension into digitally mediated, dialogic spaces where learners and technologies act as co-agents in the learning process.

2.2.2 Self-regulated Learning

Self-regulated learning (SRL) provides a crucial theoretical bridge between autonomy concepts and practical learning behaviours, offering process-oriented frameworks for understanding how learners manage their own learning effectively. Zimmerman's (2000) cyclical model of self-regulated learning has proven particularly influential in EFL contexts, delineating three interconnected phases through which successful autonomous learners progress: forethought, performance, and self-reflection.

Figure 3

Zimmerman's three-phase model as adapted for EFL writing context



The **forethought phase** includes goal setting, strategic planning, and the development of self-efficacy beliefs, all of which influence how learners engage with tasks. In AI-instructed writing contexts, this phase involves making decisions about how to incorporate AI feedback, setting realistic expectations for both human and AI-supported performance, and clarifying

one's learning goals. Studies indicate that explicit training in goal-setting and strategy selection significantly improves learners' ability to interact productively with AI tools (Teng, 2025; Mahapatra, 2024).

Oxford's (2017) Strategic Self-Regulation Model further deepens this understanding by incorporating a broader taxonomy of language learning strategies that span all three phases. It integrates cognitive, metacognitive, affective, and social strategies, making it especially relevant for EFL learners navigating hybrid human-AI learning environments. These frameworks reveal that autonomy is no longer simply an individual cognitive state but must now be understood as **shared agency,** where learners retain responsibility for learning decisions while strategically collaborating with intelligent systems (Chen, 2025; Dörnyei, 2001).

The **performance phase** centres on the application of learning strategies, real-time monitoring, and adaptive control. In AI-instructed writing, this includes making informed choices about when to seek assistance, how to assess AI feedback, and whether to accept or revise based on machine-generated suggestions. Learners who enter this phase with clear usage criteria tend to achieve stronger outcomes than those who engage reactively or without reflection (Zhang, 2025; Apriani et al., 2024).

The **self-reflection phase** becomes especially important in digitally supported environments. Learners must evaluate whether improvements in their writing stem from internalised skill development or over-reliance on AI suggestions. Chen (2025) notes that students who reflect on the quality and source of AI assistance, questioning when it clarifies, confuses, or limits their output, develop stronger feedback literacy and more deliberate writing practices. Similarly, Santika (2025) found that students who maintained structured reflective logs on their interactions with ChatGPT demonstrated greater awareness of revision choices and exhibited higher autonomy in future tasks.

Recent empirical research has further illuminated how SRL and AI literacy can be mutually reinforcing. Zhang, Lin, and Tan (2025) identified four learner profiles, "Master", "Balanced", "AI-Inclined", and "Passive", based on how effectively learners synchronised their SRL practices with AI tool usage. The "Master" group showed high metacognitive awareness and critical evaluation of AI feedback, resulting in greater autonomy and stronger writing development. In contrast, the "AI-Inclined" group displayed improved surface-level accuracy but weaker strategic thinking, raising concerns about emerging patterns of **algorithmic dependency**.

Liu et al. (2024) explored the integration of the Cognitive Academic Language Learning Approach (CALLA) with generative AI tools in EFL instruction. Their quasi-experimental

findings revealed significant gains in SRL strategy use and learner motivation, particularly in areas like task monitoring and feedback evaluation, but only among students who received explicit instruction on engaging critically with AI output. This suggests that AI tools alone do not promote autonomy; pedagogical mediation remains essential.

Fan et al. (2024) caution against what they call **"metacognitive laziness"**, a pattern wherein learners uncritically accept AI suggestions and bypass deeper stages of reflection or structural revision. Their findings highlight the risk that, without feedback literacy instruction, students may prioritise speed and surface correctness over argument quality or coherence.

Taken together, these studies reinforce the conclusion that AI tools do not inherently foster self-regulated learning or autonomy. Their effectiveness depends on learners' ability to assess feedback, reflect on their strategies, and adapt their practices over time. For this reason, SRL scaffolding, particularly structured reflection and feedback evaluation, should be a core component of any AI-instructed writing programme.

2.2.3. Relevance to Writing and Digital Feedback

The intersection between learner autonomy and digital feedback technologies represents a critical juncture in the development of writing proficiency within EFL contexts. As AI-instructed tools such as ChatGPT, Grammarly, and automated scoring systems become increasingly embedded in writing instruction, scholars have raised important questions about how these technologies influence the development of autonomy. Although digital feedback offers unique affordances for individualised learning, its impact remains complex, highly context-dependent, and shaped by both feedback design and learner agency.

One of the most frequently cited advantages of AI-generated feedback is its temporal flexibility. Unlike traditional teacher feedback, which is often constrained by time and workload, AI tools provide immediate, on-demand responses. This enables learners to engage in self-paced, recursive writing cycles, supporting both temporal and procedural autonomy (Namaziandost & Rezai, 2024). Moreover, the neutral and non-evaluative tone of AI feedback may reduce affective barriers such as writing anxiety, thereby fostering a more open attitude toward revision (Teng, 2025). These characteristics align with Littlewood's (2019) model of psychological autonomy, which highlights emotional comfort and volition as critical to independent learning.

However, these benefits are not experienced equally across all learners or contexts. A growing body of research warns that over-reliance on AI feedback can undermine learners' capacity for metacognitive engagement, such as identifying errors independently or

strategically revising their work (Yuan et al., 2024; Li, 2022). In some cases, learners tend to accept AI-generated suggestions without evaluating their accuracy or relevance, bypassing essential self-monitoring processes. This behavioural pattern, sometimes described as algorithmic dependency, may inhibit the development of higher-order writing skills and revision strategies.

The extent to which digital feedback fosters or hinders learner autonomy largely depends on its design and delivery mode. Feedback that is explanatory rather than merely corrective, dialogic rather than declarative, and suggestive rather than prescriptive is more likely to promote active engagement and strategic thinking. Luo and Zhong (2025) argue that high-quality feedback should provide reasoning, alternatives, and choice, encouraging learners to reflect, compare, and make informed decisions rather than simply accepting revisions passively.

This underscores the increasing relevance of feedback literacy, defined as a learner's ability to interpret, evaluate, and act upon feedback in a productive and autonomous manner. According to Mahapatra (2024), digital feedback systems, particularly those powered by AI, require instructional scaffolding to be used effectively. Without explicit guidance, learners may view AI suggestions as final rather than provisional, diminishing opportunities for critical thinking and self-regulation.

Overall, the literature suggests that digital feedback can significantly enhance learner autonomy in writing, but only when embedded within pedagogical practices that promote reflection, critical evaluation, and strategic decision-making. As AI technologies continue to shape the future of writing instruction, the challenge lies not only in adopting new tools but in designing learning environments that help students use them with purpose and awareness. Autonomous writing development, in this view, is not a by-product of digital access but the outcome of thoughtful feedback design and active learner engagement.

2.2.4. Motivation and Awareness

Motivation and metacognitive awareness are widely recognised as central drivers of learner autonomy, particularly in AI-instructed EFL writing contexts. Dörnyei's (2009) L2 Motivational Self System provides a valuable framework for understanding how motivation emerges from the interaction between the learner's ideal L2 self, ought-to self, and concrete learning experiences. Learners who envision themselves as competent language users are more likely to engage with AI tools as developmental supports rather than as shortcuts, fostering deeper and more sustained autonomous engagement (Dörnyei & Ushioda, 2011).

A growing body of empirical research confirms that AI-instructed writing environments can enhance motivation and learning engagement. For example, Song and Song's (2023) quasi-experimental study found that learners receiving ChatGPT-integrated instruction reported significantly higher writing motivation than those in control groups. Similarly, Ma and Chen (2024), drawing on a large sample, demonstrated that AI tools enhanced both affective and cognitive engagement while also reducing procrastination, largely due to the immediacy and personalisation of feedback.

Meta-analytical and systematic reviews support these findings. Chen et al. (2025), synthesising 51 studies, reported strong overall learning gains associated with ChatGPT and related tools, with moderate improvements in learner perceptions. Yang and Aryadoust (2024) similarly concluded that continuous and adaptive feedback environments foster long-term motivation and self-directed learning habits in language learners.

While external motivators such as grades and gamification may initially boost engagement (Wale, 2024), research consistently highlights the superior role of intrinsic motivation in supporting autonomous writing development. Learners who see AI tools as writing partners, as opposed to mechanical correctors, are more likely to use them thoughtfully and strategically (Mahapatra, 2024; Wale, 2024).

Closely intertwined with motivation is metacognitive awareness, learners' ability to reflect on and regulate their cognitive strategies. Teng (2025) found that students with high metacognitive awareness tended to critically evaluate AI-generated feedback, selectively applying it to improve their writing. Conversely, learners with lower awareness often accept suggestions uncritically, leading to superficial revisions and over-reliance on AI.

In a large quasi-experimental study, Haidar (2025) explored the use of generative AI prompts to foster metacognitive development in collaborative and individual writing contexts. Results from 352 high school students showed that the integration of metacognitive prompts significantly improved both metacognitive knowledge and regulatory behaviours, especially in collaborative learning conditions. These findings suggest that AI tools can be intentionally designed to cultivate deeper self-awareness.

Teng (2025) further explored the effects of metacognitive awareness through a mixed-methods study with 40 EFL undergraduates using ChatGPT for writing feedback. The study revealed substantial differences in motivation, engagement, self-efficacy, and collaboration tendencies between high- and low-awareness learners, reinforcing the importance of self-reflective capability in navigating AI tools.

In summary, while AI-instructed writing tools have significant potential to enhance learner motivation and autonomy, these outcomes are highly dependent on the learner's metacognitive readiness and motivational orientation. To fully realise the pedagogical value of AI in writing instruction, educators must go beyond technical access and focus on cultivating feedback literacy, reflective thinking, and self-directed motivational strategies.

2.3 Feedback in Language Learning

2.3.1. Types of Feedback

The landscape of feedback in language learning has expanded substantially with the rise of AI technologies, introducing a complex ecology in which learners must navigate traditional and digital sources of input. Conventional forms of feedback, teacher, peer, and self, now coexist with AI-generated feedback, each offering unique affordances and challenges for writing development. Understanding how these feedback types function individually and in combination is essential for designing instruction that supports both short-term improvement and long-term autonomy (Namaziandost & Rezai, 2024; Luo & Zhong, 2025).

Teacher feedback remains essential for its contextual nuance, human empathy, and attention to higher-order writing concerns such as coherence, genre awareness, and rhetorical effectiveness. In a comparative study, Shabara et al. (2024) analysed 100 B1-level writing samples scored by 11 teachers and ChatGPT. The results showed teachers offered more accurate and context-sensitive evaluations, while ChatGPT exhibited only moderate intra-rater reliability and partial correlation with teacher assessments. Similarly, Ataş et al. (2024) found that ChatGPT's feedback was longer but lacked depth, particularly in content and organisation, supporting previous claims by Sysoyev et al. (2024) regarding the limitations of AI in evaluating originality and argumentation. However, teacher feedback is not without limitations. Time constraints, delayed responses, and inconsistency in evaluation can hinder iterative writing cycles (Kurt & Kurt, 2024). While teachers excel at global feedback, their capacity to deliver timely, individualised responses is often constrained by workload and institutional demands (Mahapatra, 2024).

Peer feedback adds collaborative and dialogic value to the writing process. It exposes learners to multiple perspectives, fosters critical thinking, and reduces affective pressure. Xue (2024) demonstrated how ChatGPT could serve as a "virtual peer," scaffolding critical argumentation in a process model inspired by peer collaboration. Likewise, Kurt and Kurt (2024) found that peer review contributed to motivation through collaboration, though it was sometimes undervalued by learners. The effectiveness of peer feedback depends heavily on peer

competence, training, and the presence of teacher or tool-based scaffolding. In AI-instructed environments, peer feedback is often most effective when addressing communicative and audience-related concerns, complementing the technical accuracy offered by AI (Santika, 2025).

Self-assessment is perhaps the most direct manifestation of learner autonomy. It requires learners to apply internalised criteria in evaluating their own writing, a task often supported by AI tools through comparisons, model outputs, or rubric-based guidance. Bai and Wei (2024) explored a "compose–compare–revise" model in which students revised their drafts after comparing them with ChatGPT-generated alternatives. Learners who engaged deeply in noticing discrepancies between their own and AI versions made more substantive revisions. Similarly, Yan (2024) found that students with strong metacognitive regulation and feedback-seeking behaviours made more effective use of ChatGPT suggestions, while Xu et al. (2024) demonstrated that reflective essays improved learners' revision strategy awareness and structural clarity.

Despite its benefits, self-assessment is often treated as an embedded component of writing instruction rather than a standalone focus. Most studies do not implement formal self-assessment tools (e.g., rubrics or checklists) but rely on reflective journals, noticing tasks, or open-ended prompts. Robillos (2024) documented how students using GPT-based chatbots engaged in iterative writing and reflection, but did not always demonstrate self-assessment as a distinct phase. There is growing recognition that structured self-assessment interventions, if more intentionally implemented, could enhance both metacognitive development and writing quality (Oxford & Amerstorfer, 2023).

AI-generated feedback introduces a new dimension to writing instruction, offering immediate, consistent, and personalised input. These tools can address grammar, cohesion, vocabulary, and style simultaneously. However, AI systems often lack sensitivity to context, cultural nuance, and rhetorical intent (Yang & Aryadoust, 2024). While ChatGPT and similar platforms can support sentence-level revision and lexical enhancement, their limitations in critical discourse evaluation and learner-specific needs remain a concern. According to Chen et al. (2025), the most effective approaches combine AI feedback with human input, using AI for efficiency and breadth, and teachers or peers for depth and discourse awareness.

In practice, combining multiple feedback types often yields the most pedagogically meaningful outcomes. Luo and Zhong (2025) emphasise that explanatory, dialogic, and choice-based feedback, whether from teachers or AI, fosters greater autonomy than purely corrective

input. Mahapatra (2024) highlights the necessity of feedback literacy training to help learners interpret AI-generated suggestions critically and avoid algorithmic dependency.

2.3.2. Feedback Literacy

Feedback literacy, the capacity to understand, critically evaluate, and purposefully apply feedback to improve learning, has become a key competency in AI-instructed EFL writing environments. While early frameworks centred on the interpretation of teacher-provided corrections, contemporary models increasingly account for the complexity of AI-generated feedback, which may be voluminous, linguistically sophisticated, or misaligned with individual learner needs (Carless & Boud, 2018; Mahapatra, 2024).

AI tools such as ChatGPT facilitate frequent, iterative feedback cycles, allowing learners to engage in drafting, revision, and reflection more regularly (Li et al., 2024; Robillos, 2024). However, capitalising on these opportunities depends on several critical learner skills.

First, students must develop the ability to judge the quality and relevance of feedback, distinguishing helpful suggestions from superficial or misleading ones. This requires both linguistic awareness and an understanding of the limitations of large language models. Second, learners need strategies to manage cognitive overload, as AI-generated responses can be extensive and detailed. Third, synthesising feedback from multiple sources, including AI, teachers, and peers, demands a level of evaluative judgement and decision-making that not all learners possess at the outset (Teng, 2025; Wilson et al., 2024).

Importantly, the development of feedback literacy is not automatic. It hinges on learners' metacognitive capacity, reflective engagement, and the clarity and contextual relevance of the feedback itself (Yan, 2024). Instructional support is essential. Oktarina et al. (2024) showed that structured interpretation tasks and reflective prompts significantly improved students' ability to engage meaningfully with AI feedback. Similarly, Namaziandost and Rezai (2024) and Luo and Zhong (2025) emphasise the importance of helping learners understand how AI generates feedback and how to integrate suggestions strategically during revision.

Despite its centrality, empirical research that directly operationalises and measures feedback literacy remains limited. Oktarina et al. (2024) is one of the few studies to do so. In most AI-related feedback studies, components of feedback literacy, such as noticing, self-monitoring, and revision planning, are treated implicitly rather than as a unified construct.

In sum, while AI tools have the potential to develop high levels of feedback literacy, this outcome depends on deliberate pedagogical planning. Educators must prioritise explicit instruction and ongoing scaffolding to support learners in becoming reflective, critical users of feedback, an essential condition for sustained autonomy in writing development.

2.3.3. Feedback Effectiveness in Writing Improvement

The effectiveness of feedback in promoting writing improvement is shaped by an interaction of factors, including the content and format of feedback, the learner's engagement style, and the instructional context. According to Feedback Intervention Theory (Kluger & DeNisi, 1996), feedback may have negative effects in approximately one-third of cases, particularly when it undermines learner motivation or shifts attention from learning goals to superficial performance outcomes. This concern is especially relevant in AI-instructed contexts where the immediacy and frequency of feedback could overwhelm learners or promote overdependence.

Research on AI feedback design points to several principles for effectiveness. Feedback is more impactful when it:

- Targets task-level improvement (rather than person-level evaluation),
- Provides specific, actionable suggestions (rather than vague corrections),
- Offers a manageable cognitive load,
- And is delivered at appropriate moments to support learner reflection without undermining self-monitoring (Barrot, 2023; Fleckenstein et al., 2023).

Learner engagement also mediates feedback effectiveness. Passive acceptance of AI suggestions often leads to minimal improvement, whereas active engagement, including evaluating the relevance of feedback, comparing it to personal goals, and selectively applying it, produces deeper learning outcomes (Mahapatra, 2024). Structured reflection activities, where learners analyse patterns in the feedback they receive and their own revision behaviours, have been shown to enhance long-term writing development and promote metacognitive growth (Chen et al., 2025).

2.4 Digital Literacy

Digital literacy in contemporary EFL education transcends basic technical skills, encompassing learners' capacity to critically evaluate digital tools, integrate technology strategically into learning processes, and understand both its advantages and limitations (Van Deursen & van Dijk, 2019; Khasawneh, 2024; Wilson et al., 2024).

For writers using AI-instructed tools, this competence means grasping how such systems generate language, recognising their probabilistic nature, and making informed decisions about when and how to adopt computer-generated suggestions. Critical digital literacy becomes essential, requiring users to question AI output for cultural bias, pragmatic appropriateness, and academic integrity (McCann et al., 2025; Szabó & Szoke, 2024).

Research shows that learners with stronger digital competence, particularly digital and metacognitive skills, engage more thoughtfully with AI feedback, set better prompts, and demonstrate greater writing autonomy (Yan, 2024; Li & Zhu, 2023). Conversely, less digitally literate learners may struggle with prompt design or tool navigation, leading to overreliance or frustration (Rahimi & Teimouri, 2024).

Equity concerns also loom large. Drawing on Van Dijk's (2005) framework, the digital divide manifests not just as access disparity but also as uneven patterns of skill and meaningful use. While infrastructure in urban schools may be robust, rural or under-resourced settings often lack both technology and instruction in its pedagogical use, widening autonomy gaps.

Teachers and institutions play a pivotal role in closing these gaps. Effective AI-instructed writing programmes should explicitly support digital literacy through prompt-design workshops, scaffolded feedback interpretation, and dedicated instruction in ethical AI use (Spratt et al., 2002; Bai & Wei, 2024; Rahimi & Teimouri, 2024).

Table 1

Core Dimensions of Digital Literacy in AI-Instructed EFL Writing

Dimension	**Description**	**Implications for Instruction**
Technical Literacy	Ability to navigate and operate AI tools	Teach prompt-engineering, interface fluency
Critical Digital Literacy	Understanding tool limitations, bias awareness	Develop modules on AI ethics and cultural nuance
Prompt Design	Skill in crafting questions to generate useful responses	Workshops on effective prompt types
Equity Awareness	Recognition of digital divide contexts	Tailored supports for low-access or low-skill learners

Ethical Awareness	Understanding academic integrity, copyright, and responsible use	Clear policies and reflective prompts integrated alongside tool use

As the previous sections have demonstrated, feedback, whether teacher-led, peer-generated, or AI-instructed, plays a critical role in shaping writing development and learner autonomy. However, the extent to which learners can effectively interpret and apply such feedback is closely tied to their broader digital competence. In increasingly AI-enhanced EFL contexts, the ability to navigate digital tools, evaluate their outputs, and use them strategically is no longer optional but essential. The following section explores the concept of digital literacy, its dimensions, and its significance in enabling learners to meaningfully engage with AI technologies for autonomous writing development.

2.5 Artificial Intelligence in ELT

2.5.1 AI Tools Used in Language Learning

The rapid evolution of artificial intelligence has brought forth a wide range of tools designed to support language learning, from grammar checkers to large language models (LLMs). These tools employ techniques such as natural language processing (NLP), deep learning, and neural networks to provide learners with personalised, interactive, and scalable writing support (Chen et al., 2025; Yang & Aryadoust, 2024).

Among the most widely used tools are grammar-focused platforms such as Grammarly and ProWritingAid, which offer context-sensitive corrections for grammar, style, and clarity. These tools use rule-based and machine learning (ML) approaches to improve surface-level accuracy and support revision. While helpful in reinforcing grammatical patterns through repeated feedback, they are limited in providing holistic support for content development, rhetorical coherence, or creativity (Namaziandost & Rezai, 2024; Li et al., 2024).

More advanced systems, such as ChatGPT and Bard, represent a paradigm shift by engaging learners in interactive dialogue, generating diverse text types, and responding to writing prompts in real time. These LLMs enable flexible drafting, structural reorganisation, and brainstorming support, functions that are particularly beneficial for intermediate EFL learners (Chen et al., 2025). However, concerns have emerged regarding their limited cultural awareness, potential to homogenise writer voice, and the risk of over-reliance when learners use them without critical engagement (Mahapatra, 2024; Song & Song, 2023).

Conversational agents and adaptive platforms such as **Turnitin's Feedback Studio**, **Canva Docs**, and **Scrivener** add further layers of support. Turnitin provides integrated feedback alongside originality checks, aiding in both writing development and academic integrity. Canva Docs and Scrivener support the writing process by allowing learners to plan, draft, and revise in multimodal or highly structured formats, reducing writing anxiety and supporting iterative revision (Jamshed et al., 2024; Wilson et al., 2024).

Evidence increasingly supports the combined use of AI and human feedback for optimising writing instruction. Asadi et al. (2025) demonstrated that students who received ChatGPT alongside teacher feedback significantly improved in argumentative writing skills, particularly in structure and coherence. Teachers in the study acknowledged AI's adaptability and timeliness, but stressed the importance of pedagogical framing to avoid overdependence.

A comparative summary of these tool types is presented below:

Table 2

Major AI Tools in EFL Writing: Functions & Pedagogical Caveats

Tool Type	Examples	Primary Function	Pedagogical Strength	Key Limitation
Grammar Checkers	Grammarly, ProWritingAid	Correct grammar and mechanics	Supports grammatical accuracy	Lacks higher-order feedback
Large Language Models	ChatGPT, Bard	Interactive text generation	Assists in planning and coherence	May reduce critical engagement
Adaptive Platforms	Turnitin Feedback Studio	Originality and writing feedback	Combines integrity with guidance	Requires instruction for optimal use
Writing Management	Scrivener, Canva Docs	Planning, organisation, and multimodal writing	Reduces anxiety and fosters process writing	May disadvantage low-digital-literacy users

2.5.2 Generative AI and Large Language Models

The emergence of **generative AI tools**, especially those based on LLMs like ChatGPT, has significantly transformed writing instruction in ELT. Trained on vast textual corpora, these models are capable of engaging in extended dialogue, producing fluent texts across genres, and tailoring responses to learner inputs in real time (Chen et al., 2025; Yang & Aryadoust, 2024).

Empirical research shows that these tools promote gains in grammar, fluency, and vocabulary use. For example, Song and Song (2023) found that B1-level learners using ChatGPT outperformed peers in grammatical accuracy and revision practices. However, they also noted limited gains in creativity and argumentation, pointing to the need for supplementary instructional support.

While generative AI enables more autonomous drafting and editing, studies highlight the importance of learner training. Over-reliance, uncritical acceptance of suggestions, and a lack of awareness of AI's limitations can inhibit metacognitive development and authentic learning (Mahapatra, 2024; ElEbyary & Shabara, 2024).

Consequently, many scholars advocate for a hybrid model where AI tools are integrated with teacher feedback and ethical guidance. As discussed in Section 2.3.2 and 2.4, feedback literacy and digital competence are essential for learners to meaningfully engage with these technologies. Without structured scaffolding, students risk developing superficial writing habits and becoming disengaged from higher-order thinking processes.

Generative AI offers exciting possibilities for ELT, but it must be employed thoughtfully within pedagogy that promotes reflection, autonomy, and academic integrity. Further research is needed to determine optimal instructional models and tool configurations that balance efficiency with deep learning.

2.5.2.1 Technological Pedagogical Content Knowledge (TPACK) and Digital Pedagogical Content Knowledge (DPACK) Frameworks

The integration of AI tools in language education requires a theoretical grounding that reflects the interplay between pedagogy, content, and technology. The TPACK framework (Mishra & Koehler, 2006) and its digital-specific extension DPACK (Angeli & Valanides, 2009) provide robust foundations for understanding how AI tools can be meaningfully integrated into ELT while preserving learner autonomy.

The TPACK model builds on Shulman's (1987) concept of Pedagogical Content Knowledge (PCK) by emphasising three core domains: Technological Knowledge (TK), Pedagogical Knowledge (PK), and Content Knowledge (CK), and their intersections. Effective AI integration demands not only understanding how tools like ChatGPT work (TK), but also how to align their use with instructional methods (PK) and language learning goals (CK). The central overlap, TPACK, represents the ability to adaptively merge these domains into coherent teaching practices (Koehler & Mishra, 2009).

By contrast, DPACK incorporates digital affordances such as interactivity, connectivity, and media-rich environments. Its components, Digital Knowledge (DK)**,** Digital Pedagogical Knowledge (DPK)**,** and Digital Content Knowledge (DCK), extend the TPACK framework to explicitly address online tools and AI systems (Falloon, 2020; Johnson et al., 2023).

Recent adaptations like AI-TPACK and Intelligent-TPACK (I-TPACK) reflect the need for ethical integration of AI, highlighting accountability, fairness, transparency, and inclusivity (Celik et al., 2022). These models introduce sub-domains such as AI-TK (prompt engineering, AI mechanics), AI-PK (scaffolding AI use), and AI-CK (language implications of AI output) that frame how AI should be used critically and responsibly.

In sum, these frameworks assert that successful AI integration in EFL writing instruction must:

align with sound pedagogical strategies (PK),

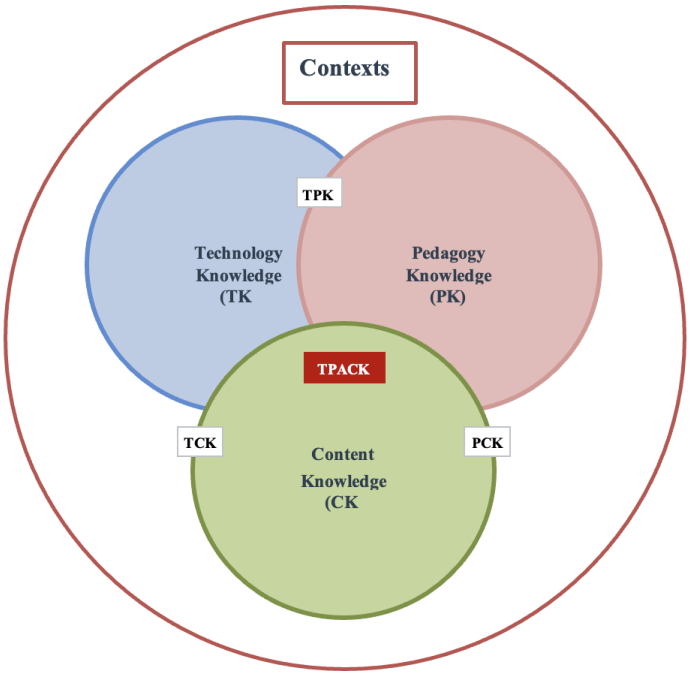
foster language development (CK), and

Leverage the unique affordances and constraints of AI tools (TK).

Their application supports the deliberate development of autonomous learners capable of navigating AI feedback with critical awareness. Future research should continue to explore how teacher expertise evolves across these domains and how contextual factors mediate the integration process (seen Figure 4).

Figure 4

Adapted from "Considering Contextual Knowledge



Note. *Adapted from "Considering Contextual Knowledge: The TPACK Diagram Gets an Upgrade" by P. Mishra, 2019, Journal of Digital Learning in Teacher Education, 35(2), pp. 76-78 (https://doi.org/10.1080/21532974.2019.1588611). Original framework by Mishra and Koehler (2006).*

2.6 AI and Writing Development

2.6.1 Empirical Studies on AI-Generated Feedback

The rapidly expanding body of empirical research on AI-generated feedback in EFL writing provides crucial insights into patterns of effectiveness, implementation considerations, and learner outcomes. Meta-analytical findings from Chen et al. (2025), based on 51 experimental studies, report large effect sizes for overall learning performance ($g = 0.867$) and moderate effects for learner perception ($g = 0.456$), indicating that AI feedback can significantly improve both writing quality and student engagement. Yang and Aryadoust's (2024) systematic review of 70 studies supports these conclusions, highlighting consistent improvements across grammar, vocabulary, and textual organisation.

Mahapatra's (2024) mixed-methods study offers detailed insights into how AI-generated feedback influences specific writing components. The findings reveal notable gains in grammatical accuracy, lexical richness, and structural coherence, while effects on content depth and argumentation were comparatively limited. These outcomes suggest that AI tools are especially effective at addressing surface-level and organisational elements, whereas their capacity to foster deeper conceptual thinking remains constrained.

Song and Song's (2023) quasi-experimental study with Chinese EFL learners documented not only linguistic improvements but also increased learner motivation in AI-assisted classrooms. Learners reported valuing the immediacy, comprehensiveness, and non-judgmental tone of AI feedback. However, 35% of students acknowledged applying suggestions without critical evaluation, raising concerns about overreliance and diminished critical thinking.

Teng's (2025) longitudinal research further interrogates the sustainability of AI feedback benefits. The study revealed that only students with high levels of metacognitive awareness maintained their gains over time, using AI feedback reflectively and strategically. In contrast, those with lower awareness initially improved but later regressed once AI input was reduced. These findings underscore the importance of not just providing AI access but also cultivating the strategic skills needed to use it effectively.

Collectively, these studies illustrate that while AI-generated feedback can accelerate writing improvement and engagement, its long-term effectiveness depends heavily on learner agency, metacognitive capacity, and guided instructional support. Future research should explore how pedagogical scaffolding and feedback literacy interventions can help learners internalise AI-supported strategies and avoid dependency.

2.6.2 Changes in Fluency, Accuracy, Complexity, and Coherence

Contemporary research examining AI's impact on specific writing dimensions reveals nuanced patterns influenced by instructional design and learner profiles. Writing fluency, measured through metrics such as words per minute, text length, and production rate, shows consistent improvement with AI support. Studies report fluency gains of 25% to 40%, attributed to reduced cognitive load on mechanical concerns and boosted confidence from real-time corrections. However, some scholars caution that heightened fluency may obscure underdeveloped language proficiency if learners overly depend on AI for idea generation (Barrot, 2023; Wang, 2024).

Grammatical accuracy improvements are among the most consistent findings in AI writing research. Tools like Grammarly and ChatGPT effectively identify and correct morphological and syntactic errors, with meta-analyses reporting effect sizes between 0.8 and 1.2 standard deviations. Nevertheless, concerns remain about whether these improvements represent genuine language acquisition or are limited to assisted performance. Several post-intervention studies suggest mixed transfer effects, with some learners retaining improved

accuracy and others reverting to habitual errors once AI support is withdrawn (Fleckenstein et al., 2023; Namaziandost & Rezai, 2024).

Syntactic complexity, operationalised through subordination ratios, clause length, and sentence variety, shows more varied outcomes. While AI can model more complex constructions, uptake depends on learners' readiness and capacity for strategic engagement. Advanced learners are more likely to incorporate complex structures, whereas beginners may be overwhelmed. Studies that combine AI suggestions with explicit instruction on syntactic complexity report stronger gains than those relying solely on incidental exposure (Pu et al., 2023).

Discourse coherence, encompassing logical flow and textual unity, reflects both the potential and limits of current AI technologies. While tools like ChatGPT can flag coherence issues, suggest transitions, and provide structural cues, deeper coherence, particularly involving conceptual organisation and argumentative logic, remains a challenge. Research shows moderate effect sizes for surface coherence improvements, but sustained progress requires human instruction that scaffolds discourse-level writing skills (Luo & Zhong, 2025; Alsofyani & Barzanji, 2024).

Taken together, these findings underscore that while AI tools enhance multiple dimensions of writing, their effectiveness is maximised when integrated with explicit pedagogical support and when learners are equipped with metacognitive and strategic skills to engage with feedback critically.

2.7 Theoretical Framework

2.7.1 Sociocultural Theory (Vygotsky)

Vygotsky's (1978) sociocultural theory provides essential insights into AI's mediational function in language learning, though contemporary applications require significant theoretical extensions to accommodate technological mediation. The Zone of Proximal Development (ZPD) concept proves particularly relevant for understanding how AI tools can bridge the gap between current and potential performance levels. Unlike traditional conceptualisations where the ZPD represents the distance between independent performance and performance with expert human assistance, AI-instructed learning creates dynamic ZPDs that adapt to learner responses in real-time (Al-Othman, 2024; Zhang, 2025).

Contemporary research demonstrates that AI tools function as sophisticated mediational instruments operating within learners' ZPDs through multiple mechanisms. First, AI provides graduated assistance that adjusts to learner competence, offering more support for challenging

tasks while reducing assistance as proficiency develops. Second, AI mediation operates continuously without fatigue or frustration, enabling extended engagement within the ZPD. Third, AI tools can simultaneously mediate multiple aspects of writing, from mechanical accuracy to rhetorical effectiveness, creating rich mediational environments. However, questions remain about whether AI mediation produces equivalent internalisation compared to human mediation (Haidar, 2025; Fan et al., 2025).

The concept of "distributed scaffolding" emerges as a crucial extension of sociocultural theory in AI contexts. Traditional scaffolding assumes human experts who can sensitively adjust support based on learner responses and cultural understanding. AI scaffolding operates through algorithmic patterns that may lack cultural sensitivity and deep contextual awareness. Research indicates that effective AI scaffolding requires careful design to ensure cultural appropriateness and avoid perpetuating biases present in training data. The most successful implementations combine AI scaffolding for routine aspects with human scaffolding for culturally sensitive and conceptually complex elements (Alm, 2024; Chen, 2025).

2.7.2. Autonomous Learning Theory (Holec, Benson, Littlewood, Alm)

The integration of AI tools in language learning necessitates a fundamental reconceptualisation of autonomous learning theory. Holec's (1981) foundational definition of autonomy as "the ability to take charge of one's own learning" requires extension to address collaborative relationships with intelligent systems. Contemporary interpretations recognise that taking charge increasingly involves making strategic decisions about when and how to utilise AI assistance while maintaining overall learning direction and evaluation responsibility (Benson, 2011; McCann et al., 2025).

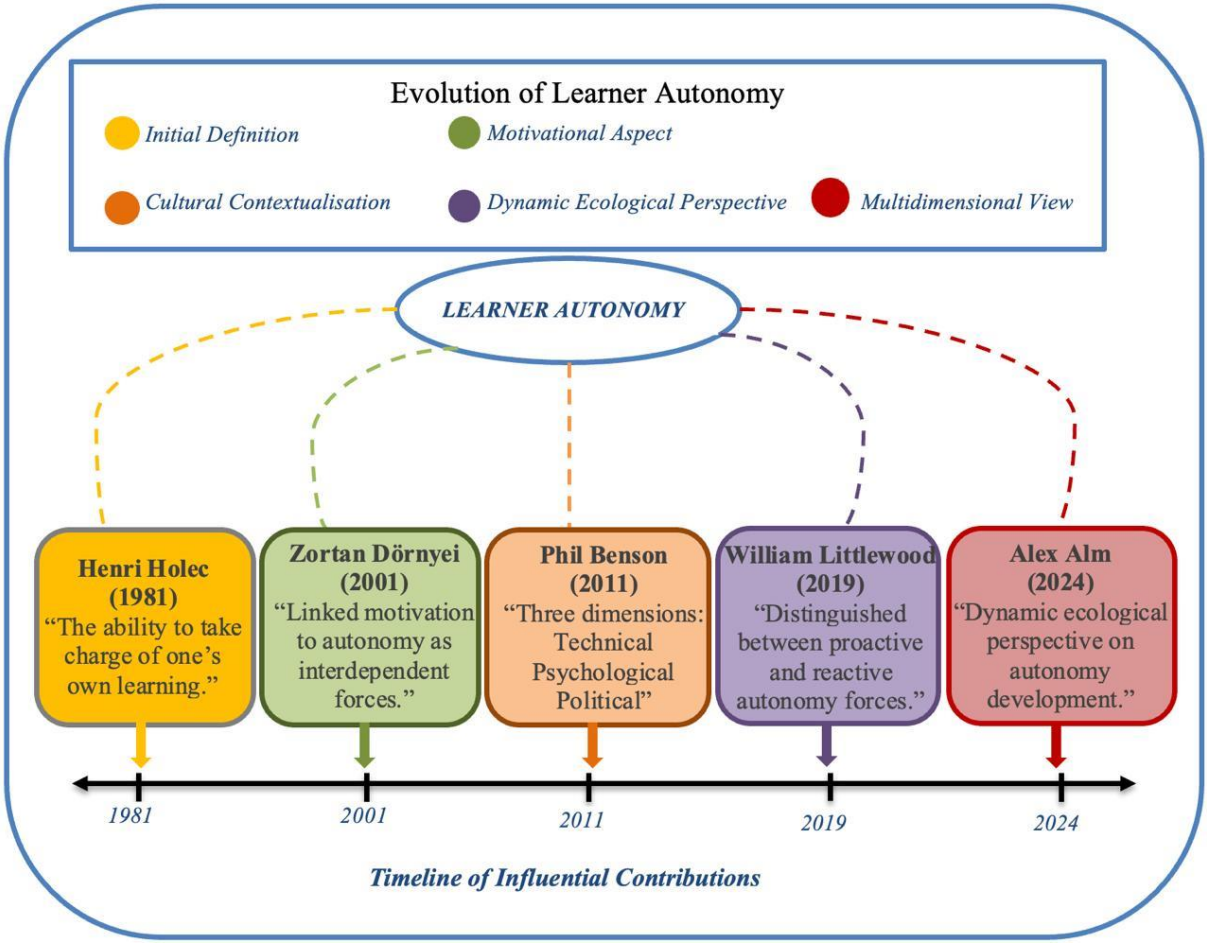
Benson's (2011) three-dimensional framework proves particularly valuable for understanding AI's differential impact on autonomy components. Technical autonomy may be enhanced through access to sophisticated AI tools that support strategy implementation and provide immediate feedback. However, this enhancement depends on learners developing the capacity to use tools strategically rather than dependently. Psychological autonomy faces complex influences, with AI potentially boosting confidence through supportive feedback while potentially undermining self-reliance if learners attribute success primarily to AI assistance. Political autonomy requires particular attention in AI contexts where algorithmic decisions may constrain learner choices in ways that are not immediately apparent (Reinders & Benson, 2024; Szabó & Szoke, 2024).

Littlewood's (2019) proactive-reactive autonomy distinction reveals how AI tools can support different autonomy manifestations. AI enables proactive autonomy by providing resources for self-directed exploration and personalised learning pathways. Simultaneously, AI can support reactive autonomy (see Figure X) by helping learners achieve institutionally defined objectives more effectively. However, research indicates that without explicit attention to autonomy development, learners may gravitate toward reactive patterns, using AI primarily to complete assigned tasks rather than pursuing self-directed learning goals. This tendency underscores the importance of pedagogical frameworks that explicitly promote proactive engagement with AI tools (Teng, 2025; Little & Dam, 2023).

Alm's (2024) recent theoretical contributions emphasise the sociocultural dimensions of autonomy in digital contexts. This perspective views autonomy not as individual independence but as the capacity to participate effectively in technology-mediated learning communities. In AI contexts, this involves developing critical AI literacy, understanding tool limitations, and maintaining agency while engaging in human-AI collaboration. This reconceptualisation aligns with contemporary recognition that 21st-century autonomous learners must navigate complex technological ecosystems while maintaining personal learning direction.

Figure 5

Evolution of Learner Autonomy



2.7.3 Authentic vs. Artificial Autonomy in AI Learning

The distinction between authentic and artificial autonomy emerges as a central concern in AI-instructed learning environments. Authentic autonomy involves genuine learner agency, critical decision-making, and personal ownership of learning processes, while artificial autonomy represents superficial independence that masks underlying dependency on technological systems. Contemporary research reveals that AI tools can support either authentic or artificial autonomy development depending on implementation approaches and learner engagement patterns (Chen, 2025; Szabó & Szoke, 2024).

Studies examining student interactions with AI writing tools identify several factors that promote authentic autonomy development. Learners who engage critically with AI suggestions, compare multiple feedback sources, and maintain control over final writing decisions demonstrate stronger autonomous learning behaviours than those who passively accept AI

recommendations. Research indicates that explicit instruction in AI evaluation strategies and critical digital literacy significantly enhances students' capacity for authentic, autonomous engagement with AI tools (Teng, 2025; Mahapatra, 2024).

Conversely, artificial autonomy manifests when learners become overly dependent on AI assistance without developing underlying competencies or critical evaluation skills. This pseudo-independence may appear successful in immediate task completion but fails to support long-term learning development or transfer to new contexts. Contemporary studies reveal concerning trends among some student populations toward uncritical acceptance of AI-generated content and reduced engagement with higher-order thinking processes when AI tools handle complex cognitive tasks (Yuan et al., 2024; Li, 2022).

The prevention of artificial autonomy requires intentional instructional design that maintains cognitive challenge while providing appropriate support. Research demonstrates that effective approaches involve graduated AI assistance that decreases over time, explicit instruction in AI evaluation and integration strategies, and regular opportunities for independent performance without technological support. These design principles ensure that AI tools function as temporary scaffolds rather than permanent crutches in learner development (Santika, 2025; Little & Dam, 2023).

2.7.4 Self-Regulated Learning Model (Zimmerman)

Zimmerman's (2000) self-regulated learning model provides crucial insights into how learners can maintain agency and learning effectiveness in AI-instructed environments. The model's three-phase cycle – forethought, performance, and self-reflection – offers a framework for understanding how successful learners navigate AI assistance while developing independent competencies. Contemporary research reveals that each phase requires specific adaptations for AI contexts, with successful learners demonstrating distinct patterns of strategic engagement across all phases (Teng, 2025; Mahapatra, 2024).

In the forethought phase, self-regulated learners in AI contexts must develop sophisticated goal-setting that distinguishes between performance goals (producing high-quality texts with AI assistance) and learning goals (developing independent writing competencies). Research indicates that learners who explicitly set dual goals for both assisted and unassisted performance show stronger long-term development. Strategic planning in this phase involves decisions about when to use AI assistance, what types of help to seek, and how to evaluate AI suggestions. Successful learners develop explicit criteria for AI use rather than approaching tools opportunistically (Zhang, 2025; Santika, 2025).

The performance phase in AI-instructed writing involves complex self-monitoring as learners must track not only their writing progress but also their AI usage patterns. Effective self-regulated learners develop awareness of when AI assistance enhances versus potentially undermines learning, adjusting their usage accordingly. Self-control processes include resisting the temptation to over-rely on AI for tasks within current capability and maintaining cognitive engagement even when AI handles routine aspects. Research reveals that learners with strong self-regulation skills use AI tools more selectively and strategically, resulting in better learning outcomes (Chen, 2025; Apriani et al., 2024).

The self-reflection phase proves particularly critical in AI contexts where learners must accurately attribute learning outcomes. Self-regulated learners systematically evaluate which improvements result from AI assistance versus developing competencies, using this awareness to adjust future learning strategies. Research indicates that structured reflection activities, where learners analyse their AI usage patterns and learning progress, significantly enhance self-regulation development. Without such reflection, learners may develop inflated self-efficacy based on AI-assisted performance, leading to difficulties when independent performance is required (Oxford & Amerstorfer, 2023; Wilson et al., 2024).

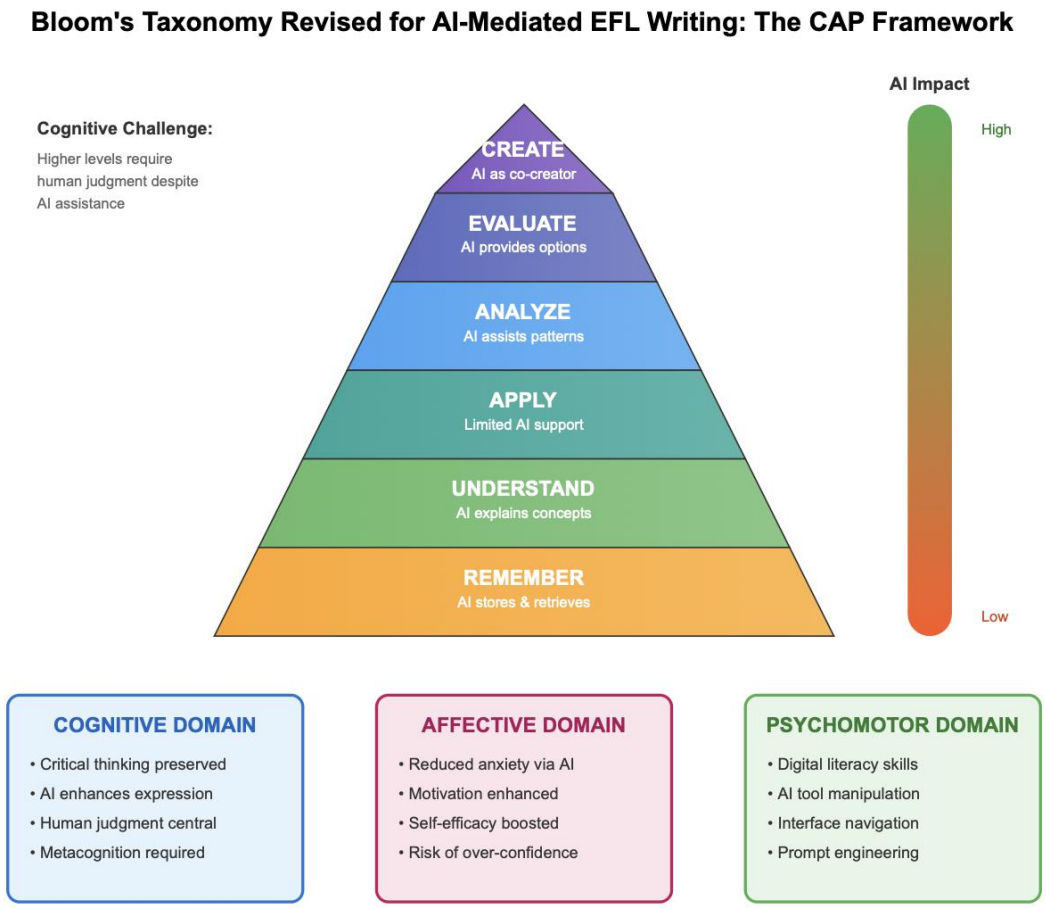
2.7.5 Bloom's Taxonomy AI-Revisited (CAP Model) and Affective Filters

The integration of AI in education necessitates the reconceptualisation of Bloom's taxonomy to address how different cognitive levels interact with AI assistance. The CAP (Create, Analyse, Apply) model represents one attempt to adapt Bloom's framework for AI contexts, recognising that AI tools may fundamentally alter how learners engage with different cognitive processes. Research indicates that AI tools show differential support across taxonomic levels, with the strongest assistance for creating and evaluating tasks but potentially undermining development in application and analysis skills (Chen et al., 2025; Koraishi, 2023).

At the creation level, AI tools can enhance learner capabilities by providing sophisticated content generation, alternative phrasings, and structural suggestions. However, research reveals a critical distinction between AI-supported creation (where learners maintain conceptual control while using AI for expression) and AI-dependent creation (where AI generates content that learners merely edit). Studies indicate that explicit instruction in maintaining creative ownership while using AI support is essential for appropriate cognitive development. The challenge lies in leveraging AI's creative support without outsourcing the creative process itself (Barrot, 2023; Xue, 2024).

Figure 6

Blooms Taxonomy Revised for AI-instructed EFL Writing



The analysis level presents particular challenges in AI contexts, as AI tools can perform sophisticated analytical tasks that learners should develop independently. Research reveals concerning patterns where learners delegate analytical thinking to AI, accepting AI-generated analyses without critical evaluation. This delegation may produce immediate task completion but undermine the development of analytical capabilities. Effective pedagogical approaches require careful task design that preserves analytical challenge while providing appropriate AI support for the expression and presentation of analytical insights (Fleckenstein et al., 2023; Pu et al., 2023).

Krashen's affective filter hypothesis gains new relevance in AI-instructed learning contexts. AI interactions can significantly reduce affective barriers through non-judgmental feedback, infinite patience, and private learning spaces. Research consistently reports reduced

anxiety and increased willingness to experiment when learners interact with AI rather than human evaluators. However, this affective support must be balanced with appropriate challenge and authentic human interaction to ensure comprehensive language development. Studies indicate that exclusive AI interaction may reduce immediate anxiety but fail to prepare learners for real-world communication anxiety (Song & Song, 2023; Wale, 2024).

2.7.6 Feedback Intervention Theory (Kluger & DeNisi)

Feedback Intervention Theory (FIT) provides crucial principles for optimising AI feedback effectiveness in EFL writing contexts. The theory's central insight – that feedback can have negative effects in approximately one-third of cases – holds particular relevance for AI systems that can generate extensive, immediate feedback. Research applying FIT to AI contexts reveals that feedback effectiveness depends critically on how AI feedback directs learner attention, with task-focused feedback showing positive effects while self-focused feedback may undermine learning motivation and performance (Kluger & DeNisi, 1996; Namaziandost & Rezai, 2024).

Contemporary studies demonstrate that AI feedback most effectively supports learning when it maintains focus on task-level processes rather than person-level evaluations. AI systems that provide specific, actionable suggestions for improvement while avoiding evaluative language show stronger positive effects on both performance and motivation. The hierarchical structure of FIT suggests that effective AI feedback should direct attention to task details and task learning processes while avoiding comparisons to others or global ability assessments. Research indicates that AI feedback designed according to FIT principles produces more consistent positive outcomes than generic AI feedback approaches (Mahapatra, 2024; Teng, 2025).

The temporal dynamics of AI feedback present unique considerations within the FIT framework. While immediate feedback can effectively direct attention to task details, the constant availability of AI feedback may prevent learners from developing internal monitoring processes. Research reveals optimal outcomes when AI feedback is strategically delayed allowing learner self-evaluation before external feedback. This approach maintains the benefits of AI availability while preserving space for autonomous monitoring development. Studies comparing immediate versus delayed AI feedback show that delayed feedback produces better long-term learning outcomes despite lower immediate performance (Luo & Zhong, 2025; Alsofyani & Barzanji, 2024).

2.7.7 Digital Literacy & Digital Divide (Van Dijk, 2005)

Van Dijk's (2005) digital divide framework provides an essential perspective on equity considerations in AI-instructed language learning (See Figure x). The three-level model – encompassing physical access, digital skills, and usage patterns – reveals how AI integration may exacerbate or alleviate educational inequalities. Contemporary research demonstrates that while physical access to AI tools has improved through cloud-based services and mobile applications, significant disparities remain in digital skills and meaningful usage patterns. These disparities have profound implications for who benefits from AI-enhanced language learning opportunities (Van Deursen & van Dijk, 2019; Wilson et al., 2024).

The digital skills divide in AI contexts extends beyond basic computer literacy to encompass AI-specific competencies. Research reveals significant variations in learners' ability to formulate effective prompts, evaluate AI responses critically, and integrate AI assistance strategically into learning processes. Learners from technology-rich environments often possess implicit understanding of AI interaction patterns that advantage them over peers with limited AI exposure. This skills divide can create cumulative advantages where digitally literate learners extract greater learning benefits from AI tools, potentially widening achievement gaps (Li, 2022; Khasawneh, 2024).

Usage patterns represent perhaps the most critical divide in AI-instructed learning. Research distinguishes between consumptive usage (passive acceptance of AI suggestions) and productive usage (strategic integration of AI assistance with independent learning). Studies indicate that learners from higher socioeconomic backgrounds more frequently demonstrate productive usage patterns, viewing AI as a tool for enhancement rather than replacement of personal effort. This usage divide reflects deeper differences in educational capital, learning beliefs, and future orientation that shape how learners approach AI tools (McCann et al., 2025; Szabó & Szoke, 2024).

2.7.8 Dual Process Theory of Writing (Hayes)

Hayes' (2012) Dual Process Theory provides valuable insights into how AI assistance affects cognitive processes in writing. The theory's distinction between automatic processes (requiring minimal cognitive resources) and controlled processes (demanding significant attention and working memory) helps explain how AI can optimise cognitive resource allocation. Research demonstrates that AI assistance with automatic processes such as spelling, grammar, and basic formatting frees cognitive resources for controlled processes, including

planning, audience consideration, and rhetorical decision-making (Barrot, 2023; Fleckenstein et al., 2023).

Contemporary studies reveal that AI's cognitive support operates through multiple mechanisms. First, AI reduces the cognitive load of surface-level error monitoring, enabling learners to maintain focus on meaning construction. Second, AI provides external memory support through suggestion histories and revision tracking, reducing working memory demands. Third, AI enables cognitive offloading of routine tasks, preserving mental resources for creative and analytical processes. However, research also identifies risks of cognitive atrophy when learners consistently outsource processes they should be developing independently (Li et al., 2024; Wang, 2024).

The interaction between AI support and writing expertise reveals complex patterns predicted by dual process theory. Novice writers benefit from AI assistance with automatic processes they have not yet developed, enabling engagement with more sophisticated writing tasks. However, excessive AI support may prevent the automation of basic skills necessary for fluent writing. Expert writers show different patterns, using AI to handle routine tasks while focusing on sophisticated rhetorical concerns. This expertise-dependent interaction suggests that AI integration should be calibrated to support appropriate skill development at each proficiency level (Song & Song, 2023; Han & Sari, 2022).

2.7.9 Emerging Models (Critical Digital Pedagogy, Feedback Literacy in AI contexts)

Critical Digital Pedagogy emerges as an essential framework for understanding power dynamics and agency in AI-instructed learning. This perspective examines how AI systems may reproduce or challenge existing educational inequalities, whose interests are served by particular AI implementations, and how learners can develop critical consciousness about AI's role in their education. Research reveals that uncritical AI adoption may perpetuate linguistic and cultural biases present in training data, disadvantaging learners whose language varieties or cultural contexts differ from AI norms. Critical approaches emphasize the importance of learner agency in questioning and reshaping AI interactions (Reinders & White, 2023; Çelik & Arkın, 2023).

Feedback literacy in AI contexts represents an emerging theoretical construct that extends traditional feedback literacy to address the unique challenges of AI-generated feedback. This framework encompasses learners' ability to understand AI feedback generation processes, evaluate suggestion quality and appropriateness, manage cognitive and affective responses to

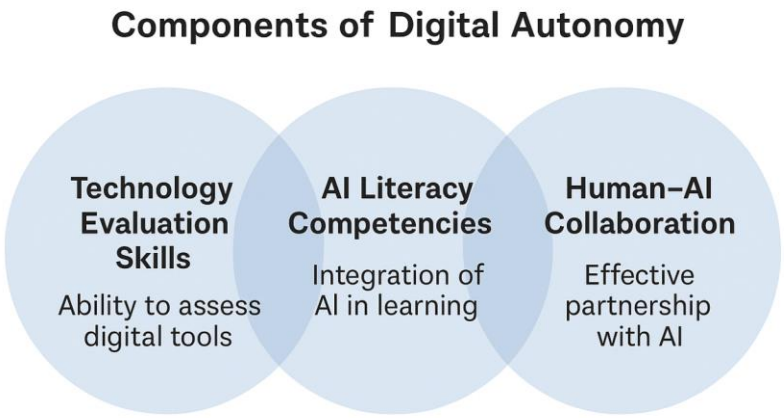
extensive feedback, and integrate AI feedback strategically with other learning resources. Research indicates that explicit development of AI feedback literacy significantly improves learning outcomes and prevents dependency patterns. The framework emphasises that productive AI engagement requires sophisticated metaknowledge about both language learning and AI systems (Carless & Boud, 2018; Mahapatra, 2024).

Emerging integrated models attempt to synthesise insights from multiple theoretical perspectives to provide comprehensive frameworks for AI-instructed language learning. These models recognise that single theoretical lenses cannot capture the complexity of human-AI interaction in educational contexts. For example, the Collaborative Autonomy Model combines insights from autonomy theory, distributed cognition, and sociocultural theory to conceptualise how learners maintain agency while engaging in productive collaboration with AI systems. Similarly, the Adaptive Scaffolding Framework integrates principles from cognitive load theory, feedback intervention theory, and self-regulated learning to optimise AI support patterns (Chen, 2025; Alm, 2024).

Recent studies have adapted traditional autonomy measures to include digital-specific components such as technology evaluation skills, AI literacy competencies, and capacity for human-AI collaboration (see Figure 7).

Figure 7

Integration of traditional self-regulation skills and digital-specific competencies in technology-assisted learner autonomy



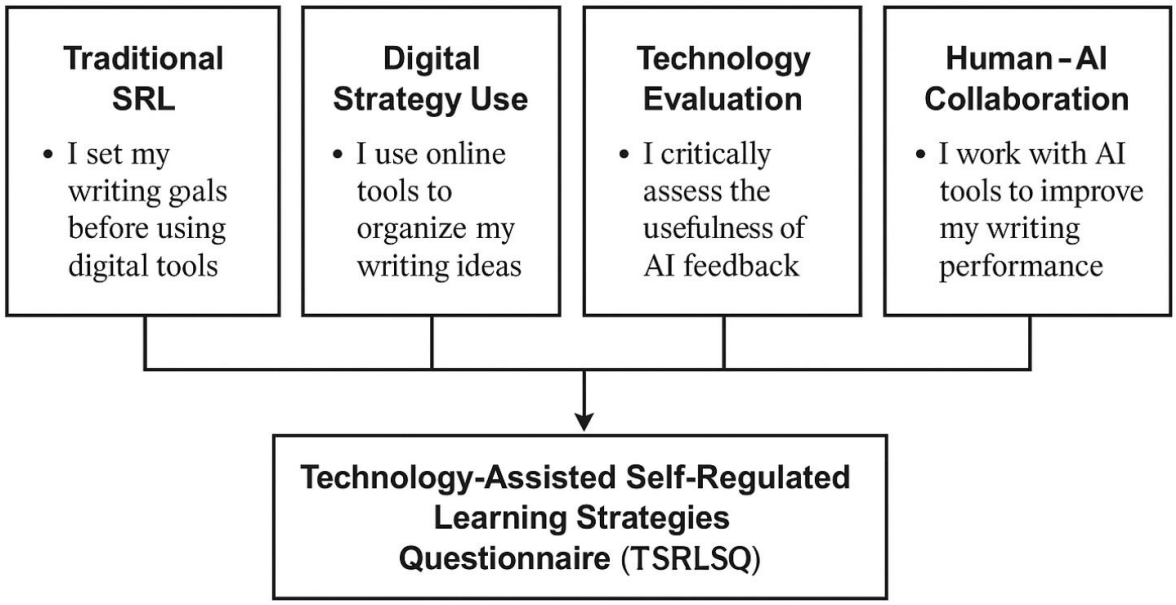
Note. The diagram illustrates the overlap between traditional self-regulation skills (e.g., goal setting, reflection)

and digital-specific competencies (e.g., technology evaluation, AI literacy, human–AI collaboration) that together constitute effective technology-assisted autonomy in EFL writing. Adapted from Santika (2025) and Çelik & Arkın (2023).

The Technology-Assisted Self-Regulated Learning Strategies Questionnaire (TSRLSQ) represents one significant advancement in this area, measuring learners' strategic use of digital tools alongside traditional self-regulation competencies.

Figure 8

Technology-Assisted Self-Regulated Learning Strategies Questionnaire (TSRLSQ)



Note. This chart displays the key domains measured by the TSRLSQ, including traditional self-regulated learning (SRL), digital strategy use, technology evaluation, and human–AI collaboration, with representative sample items for each. Adapted from Santika (2025) and Çelik & Arkın (2023).

Research using this instrument reveals that effective autonomous learners in digital environments demonstrate both traditional self-regulation skills and sophisticated understanding of technological affordances and limitations (Santika, 2025; Çelik & Arkın, 2023). Emerging measurement approaches emphasise process-oriented assessment that captures how learners interact with AI tools over time rather than focusing solely on end products. These approaches include analysis of revision patterns in AI-assisted writing, examination of help-seeking behaviours with digital tools, and evaluation of learners' capacity to transfer skills across different technological contexts. Such process-focused assessment

provides a more nuanced understanding of autonomy development in technology-mediated environments (Teng, 2024; Shi & Aryadoust, 2024).

2.8 Research Gaps

2.8.1 Lack of Empirical Studies Combining AI Feedback and Autonomy

Despite the proliferation of research on AI in language learning, a critical gap exists in studies that specifically examine the intersection of AI feedback and learner autonomy development. Most existing research focuses either on AI's impact on writing performance metrics or on autonomy development in traditional contexts, but few studies investigate how AI feedback systems influence the development of autonomous learning capabilities. This gap is particularly significant given that autonomy represents a fundamental goal of language education and a prerequisite for lifelong learning success (McCann et al., 2025; Szabó & Szoke, 2024).

The limited research addressing AI-autonomy interactions reveals contradictory findings that require systematic investigation. Some studies suggest that AI feedback enhances autonomy by providing consistent support that enables self-paced learning and reduces dependence on teacher availability. Conversely, other research raises concerns that AI feedback may create new forms of dependency that ultimately undermine autonomous learning capabilities. These contradictions likely reflect differences in implementation approaches, learner populations, and autonomy conceptualisations, highlighting the need for comprehensive research that examines moderating factors systematically (Chen, 2025; Teng, 2025).

Methodological limitations in existing studies further contribute to this research gap. Most investigations employ short-term interventions that cannot capture autonomy development trajectories, rely on self-report measures that may not reflect actual autonomous behaviours, or fail to compare different AI implementation approaches that might differentially affect autonomy. Additionally, the lack of validated instruments for measuring autonomy in AI-instructed contexts constrains researchers' ability to generate comparable findings across studies. Addressing these methodological challenges is essential for building cumulative knowledge about AI-autonomy interactions (Santika, 2025; Wilson et al., 2024).

2.8.2 Neglect of Feedback Literacy in AI Use

The concept of feedback literacy – learners' capacity to understand, evaluate, and utilise feedback effectively – remains largely unexplored in AI-instructed learning contexts. While traditional feedback literacy research provides valuable foundations, AI-generated feedback

presents unique challenges that existing frameworks do not adequately address. The volume, immediacy, and technical nature of AI feedback require new competencies that extend beyond traditional feedback engagement skills. This gap in understanding feedback literacy for AI contexts limits educators' ability to prepare learners for productive AI engagement (Mahapatra, 2024; Namaziandost & Rezai, 2024).

Current research reveals that learners demonstrate widely varying abilities to engage productively with AI feedback, but the factors underlying these differences remain poorly understood. Some learners critically evaluate AI suggestions and integrate them strategically with other learning resources, while others accept AI feedback uncritically or become overwhelmed by extensive suggestions. These differences likely reflect variations in feedback literacy, but without systematic investigation of the components and development of AI-specific feedback literacy, educators lack guidance for supporting learner development in this critical area (Teng, 2025; Luo & Zhong, 2025).

The absence of pedagogical frameworks for developing feedback literacy in AI contexts represents a significant practical gap. While researchers emphasise the importance of critical engagement with AI feedback, few studies provide concrete strategies for developing these capabilities. Questions remain about optimal instructional sequences, the relationship between general and AI-specific feedback literacy, and assessment approaches for feedback literacy in technology-mediated contexts. This pedagogical gap limits the practical impact of AI integration research and may contribute to ineffective implementation practices that fail to realise AI's educational potential (Carless & Boud, 2018; Wilson et al., 2024).

2.8.3 Limited Longitudinal Studies in EFL Contexts

The predominance of short-term intervention studies in AI-EFL research constrains understanding of long-term impacts on language development and learner autonomy. Most existing studies examine AI effects over periods ranging from a few weeks to a single semester, insufficient for capturing developmental trajectories, sustainability of improvements, or potential delayed effects. This temporal limitation is particularly problematic for understanding autonomy development, which typically unfolds over extended periods and may show non-linear patterns. Without longitudinal data, educators cannot determine whether initial improvements persist, transform, or diminish over time (Shi & Aryadoust, 2024; Fleckenstein et al., 2023).

The lack of longitudinal research also limits understanding of how learner-AI relationships evolve as proficiency develops. Initial engagement patterns may differ

substantially from those that emerge after extended use, and strategies effective for beginners may become constraining for advanced learners. Research examining other educational technologies reveals that novelty effects, changing usage patterns, and evolving learner needs significantly influence long-term outcomes. Similar longitudinal investigations are essential for AI-instructed language learning to inform sustainable implementation strategies (Yang & Aryadoust, 2024; Koraishi, 2024).

Specific features of EFL contexts require dedicated longitudinal investigation that generic AI education research cannot provide. EFL learners face unique challenges, including limited target language exposure, potential conflicts between AI feedback and local teaching approaches, and cultural factors that influence technology adoption patterns. Longitudinal studies in diverse EFL contexts are necessary to understand how these factors interact with AI integration over time and to develop context-sensitive implementation guidelines. The current lack of such research limits the applicability of existing findings to the diverse global contexts where EFL education occurs (Kırkgöz & Ağçam, 2022; Karakaş & Kartal, 2023).

The current study addresses these critical gaps by investigating the intersection of AI feedback and autonomy development through a longitudinal mixed-methods design that includes explicit attention to feedback literacy development. By examining how Turkish EFL learners engage with AI writing tools over an extended period, this research provides essential empirical evidence for understanding the complex relationships between AI assistance, autonomous learning capabilities, and writing development in authentic EFL contexts.

2.9. Cognitive Dependency and Over-reliance Risks

The potential for cognitive dependency represents one of the most significant concerns in AI writing instruction, with researchers investigating whether technological assistance undermines learners' capacity for independent problem-solving and critical thinking. Contemporary studies reveal complex patterns of dependency that vary based on implementation approaches, learner characteristics, and the nature of AI assistance provided. Understanding these dependency risks is crucial for developing implementation strategies that maximise AI benefits while preserving cognitive development (Yuan et al., 2024; Li, 2022).

Recent research identifies several indicators of problematic AI dependency in writing contexts. Students who consistently rely on AI for basic grammar decisions, demonstrate reluctance to engage in independent error correction, or show decreased performance when AI assistance is unavailable may be developing dependency patterns that undermine learning objectives. Studies indicate that such dependency typically emerges when AI tools handle

cognitive tasks that learners should be developing independently, creating learned helplessness rather than enhanced competence (Koraishi, 2023; Szabó & Szoke, 2024).

However, research also reveals that dependency concerns may be overstated when AI tools are implemented strategically. Studies examining long-term AI use patterns demonstrate that students can develop sophisticated collaborative relationships with AI systems without losing independent capabilities, provided they receive appropriate instruction in critical evaluation and strategic usage. The key distinction lies between dependency (inability to function without AI) and strategic collaboration (choosing when and how to use AI effectively) (Mahapatra, 2024; Teng, 2025).

An interesting, contemporary research has been conducted by Gharaibeh (2025). The study investigated the effectiveness of ChatGPT-based intervention in assisting children with dysgraphia to enhance their writing skills, including 80 participants from different backgrounds. The findings suggest that the intervention group had a significant positive influence on the writing scores of the treatment group. The implications showed that ChatGPT enhances writing abilities in children with dysgraphia, providing tailored educational support and promoting a more inclusive and effective learning environment.

Contemporary research emphasises the importance of implementation design in preventing dependency development. Effective approaches include graduated AI assistance that decreases over time, regular opportunities for independent performance, and explicit instruction in AI evaluation strategies. Studies demonstrate that students receiving such comprehensive implementation support maintain and often enhance their independent writing capabilities while benefiting from AI assistance (Little & Dam, 2023; Chen, 2025).

Emerging research also examines the concept of "positive dependency", where strategic reliance on AI tools enables learners to engage with more challenging tasks and develop higher-order skills that would otherwise be constrained by cognitive limitations. This perspective suggests that appropriate AI dependency may actually enhance rather than diminish learning outcomes, provided learners maintain critical awareness and control over the collaboration process (Wale, 2024; Al-Othman, 2024).

CHAPTER III
METHODOLOGY

This chapter outlines the research design, participants, data collection instruments, procedures, and analysis methods used to investigate the effects of ChatGPT's impact on learner autonomy and writing proficiency among B1-level EFL learners. The chapter is organised into seven main sections: research design, participants and sampling, intervention procedures, pilot study, data collection methods, data analysis procedures, and ethical considerations.

3.1 Research Design

This study employed a sequential embedded mixed-methods design (QUAL → quan) with a quasi-experimental component to investigate the effectiveness of AI-assisted writing instruction compared to traditional teacher-led feedback in EFL contexts (Dörnyei, 2007; Creswell, 2014; Turner et al., 2017). The design prioritised qualitative data as the primary source of evidence, supplemented by quantitative measures to provide triangulation and enhance validity (Creswell & Plano Clark, 2018).

The sequential embedded design was selected based on its methodological strengths for investigating complex educational phenomena (Dörnyei, 2007). In this design, qualitative data collection and analysis occurred throughout the 8-week intervention period, with quantitative data (pre- and post-test scores) providing comparative context to enrich qualitative insights. This approach offers several advantages.

To enhance the validity and reliability of the findings, methodological triangulation (Arias-Vásquez, M. E. (2022) was employed by integrating both quantitative and qualitative data. This approach allows for a comprehensive and multi-dimensional evaluation of the research outcomes by cross-verifying findings from multiple sources.

3.2 Pilot Study

3.2.1 Purpose of the Pilot

The pilot study was conducted to evaluate the feasibility and effectiveness of the workshop's content, structure, and delivery format prior to the implementation of the main intervention. Additionally, the study aimed to assess the clarity of the workshop materials, the appropriateness of the selected AI tools, and the level of participant engagement, as well as to evaluate the data collection methods. The pilot also helped identify any logistical challenges, such as technological issues, participant fatigue, or the need for timing adjustments, that could impact the main intervention.

3.2.3 Challenges Faced in the Pilot Study

During the pilot workshop, several challenges were identified as mainly about technological requirements, which include Google services and technical account creation related to ChatGPT usage. Moreover, there was a difficulty in finding a mutual availability among participants and the researcher.

3.2.4 Adjustments Based on Pilot Findings

The feedback from the pilot study informed several key adjustments to the main intervention:

Workshop sessions were divided to accommodate participant availability.

Virtual delivery and Word-based reflective diaries enhanced accessibility and analysis.

Participants were instructed in advance to have a ChatGPT free version account.

A recorded Zoom session of account setup was shared across groups.

This online, formatted workshop allowed the researcher to engage with a larger group, especially since the TÖMER participants were drawn from different classes with non-overlapping schedules. Materials included a tablet for visual explanations, Google Docs, and Excel sheets.

3.3 Participants

3.3.1 Participants of the Pilot Study

The pilot study involved 10 participants (6 female, 4 male), aged 16 to 29, with most in the 21-22 age range. All were B1.3-level students at the Ankara University TÖMER Bursa

branch. They had 2-9 years of English learning experience. Convenience sampling was used as this was the only B1-level group available during the pilot.

3.3.2 Participants and the Main Intervention

The main intervention began with 26 students, of whom 18 completed the study: 12 in the experimental group and 6 in the control group. Ages ranged from 14 to 30, with 55.6% adolescents (14-16), 27.8% young adults (20-23), and 16.7% mature learners (27-30). Gender distribution was 55.6% female and 44.4% male. All participants were B1-level and none had prior experience with generative AI in EFL education.

Table 3

Experimental vs. Control Groups

Group	Feedback Type	Source of Feedback	Mode of Delivery	Learner Role
Experimental	AI-generated feedback	ChatGPT	Digital platform	Autonomous, self-directed
Control	Teacher feedback	Human instructor	WhatsApp	Teacher-facilitated revision

A preparatory workshop was conducted for the experimental group. Objectives included:

Tool introduction and familiarisation

Technological know-how

Ethical considerations of AI usage

Workshop Content:

AI hallucinations and mitigation strategies (Bruno et al., 2023)

ChatGPT introduction and basic navigation (Lo, 2023; Koraishi, 2023)

Interpretation of AI feedback

Rubric for self-evaluation

Instruction was limited to avoid reducing learner agency, in line with Holec's (1981) definition of autonomy. All groups received equivalent pre-intervention training to ensure fairness (Creswell & Plano Clark, 2018).

3.3.3 Persistent Challenges in the Main Intervention

Some participants experienced temporary access issues to ChatGPT during the semester break. Additionally, confusion around limited access to premium features led some to believe they had lost access entirely. These issues were resolved by revisiting the Zoom-recorded workshop. Overall, virtual scheduling improved engagement and reduced fatigue, validating the pilot-informed changes.

3.4 Data Collection

This section describes the instruments, procedures, and timeline used to gather both qualitative and quantitative data throughout the study. Multiple data sources, including writing tasks, reflective diaries, semi-structured interviews, and demographic surveys, were employed to ensure comprehensive and robust evidence regarding the effects of AI-instructed and teacher-led feedback on learner autonomy and writing proficiency.

3.4.1 Qualitative Data

The qualitative data were collected during and after the 8-week treatment through reflective diaries and semi-structured interviews. Qualitative insights allow us to delve into participants' rich, intricate experiences, providing a deeper understanding that goes beyond what numerical data can reveal alone (Cresswell & Clark, 2012). Along with weekly tasks, they were asked to keep reflective diaries to show strengths, challenges and experience ratings of both topics, ChatGPT helps, and feedback. The participants sent the weekly tasks and reflective diaries to the WhatsApp group, consisting of their own group members, and all the work was accessible to group members.

3.4.2 Quantitative Data

The quantitative data are the pre-test and post-test of both experimental and control groups. One pre-test of each participant, along with the post-test results, was ranked by two raters. The writing tests were assessed according to the TÖMER B1 writing rubric. In this study, the scale was administered as both a pre-test and a post-test to the same group of participants. The pre-test was conducted prior to the intervention to establish a baseline of participants' knowledge and competencies in AI integration. Following the training workshop, the post-test was administered to evaluate the effectiveness of the intervention. By comparing the pre-test and post-test scores, the researcher was able to measure the impact of the workshop on participants' abilities to incorporate AI tools into their teaching practices, as well as their

understanding of the ethical implications of AI in education. This approach provided a clear, quantifiable measure of the workshop's effectiveness in improving both technical and ethical competencies.

3.4.2.1 Reliability of the Rubric in the Current Study

This study's writing proficiency assessment utilised a rubric developed and implemented by Ankara University TÖMER (Centre for Turkish Language Teaching). As a governmental institution affiliated with Ankara University and specialising in Turkish and foreign language education, TÖMER is a recognised authority in language assessment within Turkey. The rubric employed in this research is the same instrument used by TÖMER for its B1 level writing examinations, which are integral to their internationally recognised language certificate program. The application of this rubric in TÖMER's high-stakes certificate examinations lends a degree of implied validity to the instrument, suggesting that it effectively measures writing proficiency in alignment with established EFL standards within the Turkish higher education context. The rubric, as evidenced by the examination sheet (Figure 9), assesses writing across several key categories, including 'Title,' 'Use of English,' 'Vocabulary,' 'Quality of Essay,' 'Organization,' and 'Punctuation,' each scored on a defined scale contributing to a total score out of 20 as for writing task 2 in their official writing exam.

The simplified student rubric was given in each writing paper; however, a detailed teacher rubric is provided in the appendix. The criterion in the student example corresponds to equivalent categories in the teacher rubric: 'Title' and 'Organization' align with Organisation & Structure (0-4 points), 'Use of English' corresponds to Grammar Accuracy (0-4 points), 'Vocabulary' matches Vocabulary Range & Usage (0-4 points), 'Quality of Essay' relates to Content & Development (0-4 points), 'Punctuation' aligns with Mechanics (Spelling & Punctuation) (0-2 points), with an additional Artistic Language Use (0-2 points) category in the detailed teacher version.

Figure 9

Rubric – Ankara University TÖMER Official B1.3 Certificate Exam Rubric



However, while the TÖMER six-factor rubric benefits from institutional validity and practical application in high-stakes testing, it is important to acknowledge certain limitations regarding documentation and currency. Although a detailed rubric with explicit descriptors for each score point within each category (content & development, organisation & structure, grammar accuracy, vocabulary range & usage, mechanics, and artistic language use) was obtained and utilised in this study (see Appendix), the specific source citation and publication date of this detailed version could not be verified. As illustrated in Figure 1, the simplified rubric presented to test-takers lacks the granular performance criteria found in the detailed version, potentially creating discrepancies between what students see and the actual assessment standards applied.

This documentation gap raises questions about the currency and official status of the detailed rubric employed in this research. While the rubric's institutional validity for certification purposes is implied by TÖMER's continued use, the absence of verifiable citation details and uncertainty regarding the most current version warrants consideration as a potential limitation. The detailed descriptors significantly enhance scoring consistency compared to the simplified student version; however, without confirmed institutional endorsement of the specific rubric version used, some degree of uncertainty remains regarding its alignment with current TÖMER assessment standards. Future research employing this rubric would benefit from obtaining official documentation from TÖMER to verify the authenticity, currency, and authorised use of the detailed assessment criteria. (see Appendix 5)

3.5 Data Collection Procedure

The data collection process spanned an entire course term, excluding the pilot study. Data were collected through a combination of face-to-face and online methods to ensure accessibility and efficiency.

3.5.1. Pre-Test Administration

Prior to the eight-week intervention, a pre-test was conducted in a face-to-face classroom setting. Participants were required to complete the writing task within 60 minutes under controlled conditions, without access to external resources, materials, or instructor assistance. A rubric was provided on the test paper for reference.

Before administering the pre-test, participants were asked to complete a Google Form, which included:

1. An informed consent section (see Appendix 1).
2. Demographic questions to gather baseline information on participants' AI-related knowledge and competencies.

Since some participants were under 18 years of age, parental consent forms were distributed in advance (see Appendix 4), ensuring that legal guardians provided written permission before the test date. A total of 26 participants completed the form and were included in the pre-test phase.

3.5.2. Post-Test Administration

Following the eight-week intervention, a post-test was conducted under identical conditions, with a 60-minute duration, administered face-to-face in a controlled classroom environment and a rubric included on test papers for reference, with no other external sources to use.

However, due to ethical considerations, there were dropout cases, and only 18 participants completed the post-test, leading to the exclusion of the remaining participants from the intervention process.

3.5.3. Ethical Considerations and Anonymity Measures

To protect participants' anonymity, personal details such as names and contact information were not required in the Google Form. The only optional personal identifier was an email address, used only for data processing and potential follow-up qualitative discussions. This approach ensured that those who wished to remain anonymous could do so while allowing for continued participation in further research phases if needed. They were given a pseudonym to remain anonymous.

3.5.4. Reflective Diaries and Semi-Structured Interviews

In addition to the pre-test. and post-tests, data were collected through:

- Reflective diaries documenting participants' experiences throughout the intervention.
- Semi-structured interviews to gain deeper insights into learner autonomy and perceptions of the AI-driven writing process.

Google Forms and digital reflective diaries facilitated an efficient and accessible data collection process, minimising logistical constraints.

3.6. Data Analysis

This section outlines how data were collected throughout the study, combining both quantitative and qualitative methods to ensure a well-rounded and insightful analysis. The data collection process was designed carefully to capture not only the measurable impact of AI-driven feedback on writing proficiency but also the lived experiences and perceptions of participants as they engaged with this technology.

To achieve this, a mix of pre- and post-tests, reflective diaries, and semi-structured interviews was used. The pre- and post-tests provided a structured way to assess writing development over time, while the reflective diaries and interviews allowed participants to express their thoughts, challenges, and progress in a more personal and detailed way.

Ensuring that the data collection process was accessible and comfortable for participants was a priority. A combination of face-to-face and online methods was used to accommodate different needs and learning environments. Participants were fully informed about the study's purpose, and informed consent was obtained, including parental consent where necessary. Additionally, to respect participants' privacy, all responses were anonymised, allowing them to share their insights freely and honestly.

By carefully structuring the data collection process in this way, the study aimed to provide a comprehensive and human-centred understanding of how AI tools influence learner autonomy and writing development.

To further analyse learning processes, the study incorporated Kolb's (1984) experiential learning cycle as a conceptual framework. This cycle facilitated exploration of how participants:

Concrete Experience: Engaged directly with either AI-instructed or teacher feedback on their writing tasks,

Reflective Observation: Recorded weekly insights and experiences in reflective diaries,

Abstract Conceptualisation: Developed new understandings of writing strategies and improvements.

Active Experimentation: Applied acquired feedback and strategies to subsequent writing assignments.

3.6.1 Descriptive Statistics

The present study employed a quasi-experimental pre-post design to examine the effectiveness of an intervention on English as a Foreign Language (EFL) writing performance. The sample comprised 18 participants, strategically divided into an experimental group (n = 12, coded as E1-E12) and a control group (n = 6, coded as C1-C6). This unequal group allocation reflects typical classroom constraints in EFL contexts, where intact classes often vary in size (Dörnyei, 2007). The descriptive statistics were calculated to examine the writing performance of both the experimental group (n = 12) and the control group (n = 6) before and after the intervention, of which demographic information is shown in Table 4.

Table 4

Demographic Characteristics of Study Participants by Group

Characteristic	**Experimental (n = 12)**	**Control (n = 6)**	**Total (N = 18)**
Age (years)			
M (*SD*)	18.1 (4.9)	20.8 (5.9)	19.0 (5.2)
Mdn	15.0	18.5	16.8
Range	14–30	16–29	14–30
Gender			
Female	7 (58.3%)	3 (50.0%)	10 (55.6%)

Characteristic	Experimental (*n* = 12)	Control (*n* = 6)	Total (*N* = 18)
Male	5 (41.7%)	3 (50.0%)	8 (44.4%)
English Proficiency Level			
B1 (Intermediate)	12 (100%)	6 (100%)	18 (100%)

Note. M = mean; SD = standard deviation; Mdn = median. All participants were English as a Foreign Language (EFL) learners at B1 proficiency level.

3.6.1.1 Inter-Rater Reliability

Writing performance was assessed by two independent raters using the analytical rubric. Inter-rater reliability was evaluated using intraclass correlation coefficients (ICC) with a two-way mixed effects model for consistency **(**ICC). Table 5 presents the reliability coefficients for all assessment criteria.

Table 5

Inter-Rater Reliability Coefficients for Writing Assessment Criteria

Criterion	$ICC_{3,1}$	95% CI	Reliability Level
Content	.85	[.73, .92]	Good
Organization	.82	[.69, .90]	Good
Grammar	.89	[.81, .94]	Good
Vocabulary	.88	[.79, .93]	Good
Mechanics	.76	[.56, .86]	Good

Criterion	$ICC_{3,1}$	95% CI	Reliability Level
Artistic Language	.91	[.85, .95]	Excellent
Total Score	**.92**	**[.87, .96]**	**Excellent**

Note. $ICC_{3,1}$ = intraclass correlation coefficient, two-way mixed effects, consistency, single measurement. Reliability interpretations: < .50 = poor, .50–.75 = moderate, .75–.90 = good, > .90 = excellent (Koo & Li, 2016).

The ICC values demonstrated good to excellent inter-rater reliability across all criteria, with the total score achieving excellent reliability (ICC = .92). These results supported the use of averaged rater scores for subsequent analyses.

Writing performance was assessed using a comprehensive analytical rubric encompassing six distinct criteria: Content (maximum 4 points), Organization (maximum 4 points), Grammar (maximum 4 points), Vocabulary (maximum 4 points), Mechanics (maximum 2 points), and Artistic Language (maximum 2 points), yielding a total possible score of 20 points. This multi-dimensional assessment framework aligns with established practices in L2 writing evaluation, providing granular insights into specific linguistic competencies (Weigle, 2002).

3.6.1.2 Normality Testing and Statistical Method Justification

Prior to inferential testing, normality assumptions were examined using Shapiro-Wilk tests, skewness analysis, and visual inspection of distributions. Table 3 presents the normality test results for total writing scores.

Table 6

Normality Test Results for Total Writing Scores by Group and Test Phase

Group	Test Phase	*M*	*SD*	Skewness	*W*	*p*	Normality
Experimental	Pre-test	10.33	2.16	-0.45	.892	.118	Normal
Experimental	Post-test	14.92	3.92	0.12	.943	.548	Normal

Group	Test Phase	*M*	*SD*	Skewness	*W*	*p*	Normality
Control	Pre-test	10.67	3.19	-0.30	.968	.703	Normal
Control	Post-test	17.83	3.73	-0.87	.856	.204	Normal

Note. M = mean; SD = standard deviation; W = Shapiro-Wilk test statistic.

Comprehensive normality testing across all criteria and conditions revealed violations in 11 out of 28 distributions (39.3%). Combined with small sample sizes (n = 6–12), the ordinal nature of rubric-based assessments, and the presence of outliers, these findings justified the exclusive use of non-parametric statistical methods (Field, 2018).

Table 7 presents descriptive statistics for pre-test, post-test, and gain scores by group, using medians and interquartile ranges appropriate for non-parametric analysis.

Table 7

Descriptive Statistics for Writing Performance by Group and Test Phase

Group	Measure	*n*	*Mdn*	*Q1*	*Q3*	*IQR*	Range
Experimental	Pre-test	12	9.75	9.00	11.38	2.38	7.5–15.5
	Post-test	12	13.75	10.50	17.63	7.13	8.0–20.0
	Gain scores	12	2.75	1.13	6.00	4.88	0.5–10.0
Control	Pre-test	6	10.25	8.38	13.13	4.75	6.5–13.5
	Post-test	6	18.75	16.50	20.00	3.50	12.0–20.0
	Gain scores	6	7.75	5.38	9.25	3.88	5.0–10.0

Note. Mdn = median; Q_1 = first quartile; Q_3 = third quartile; IQR = interquartile range. Scores represent averaged ratings from two independent raters.

Both groups demonstrated improvement from pre-test to post-test. The experimental group's median gain was 2.75 points, while the control group achieved a median gain of 7.75

points. Notably, the experimental group showed increased variability in post-test performance (IQR = 7.13) compared to pre-test (IQR = 2.38), while the control group maintained more consistent performance with a smaller post-test IQR (3.50).

The only section where the mean score of the two raters was used was in the individual student performance ranking, presented to compare pre-test and post-test results. In all other instances, the scores provided by each rater were reported and analysed separately to preserve interrater transparency and allow for comparison (see Appendix X).

3.6.1.3 Justifications for Some Data

1. Effectiveness of Both Interventions

Both experimental and control groups demonstrated statistically significant writing improvements with large within-group effect sizes. The experimental group achieved substantial gains (r = .89, Cliff's δ = .81), while the control group showed similar effect magnitude (r = .90, Cliff's δ = .94). These large effect sizes indicate that both interventions were practically significant in improving writing performance.

2. Superior Control Group Performance

Despite both interventions being effective, the control group consistently outperformed the experimental group across multiple measures. Post-test comparisons revealed significant advantages for the control group in total writing scores (U = 12.0, p = .006, r = .64, δ = -.72), mechanics (r = .54, δ = -.61), and artistic language (r = .52, δ = -.58). The control group achieved larger median gains (7.75 vs. 2.75 points) and more consistent high performance.

3. Differential Variability Patterns

A critical finding was the contrasting patterns of performance variability between groups. The experimental group showed increased performance variability post-intervention (IQR increased from 2.38 to 7.13), suggesting differential responses to AI-assisted instruction. In contrast, the control group maintained more consistent performance with reduced variability (IQR decreased from 4.75 to 3.50), indicating more uniform benefits from traditional instruction.

4. Criterion-Specific Intervention Effects

The analysis revealed distinct patterns of improvement across writing criteria:

Content development: Both groups achieved large, comparable effect sizes (experimental δ = .83; control δ = .89)

Vocabulary enhancement: Both groups showed substantial gains, with control group advantage (experimental δ = .75; control δ = 1.00)

Mechanics improvement: Marked difference between groups (experimental δ = .13; control δ = .56)

Artistic language development: Control group demonstrated clear superiority (experimental δ = .58; control δ = .89)

5. Individual Response Variation

Individual trajectory analysis revealed substantial heterogeneity in response to AI-assisted instruction. Experimental group participants showed gains ranging from minimal (+0.5 points, E4) to maximal (+10.0 points, E3 and E9), indicating that intervention effectiveness varied considerably across individuals. The control group demonstrated more consistent improvement patterns, with all participants achieving gains of +5.0 points or higher.

6. Statistical Significance After Multiple Comparison Correction

Following Holm-Bonferroni adjustment for multiple comparisons, two key findings retained statistical significance: (a) between-group differences in total writing scores favouring the control group (p = .006), and (b) within-group improvement in the experimental condition (p = .002). This conservative approach ensures robust interpretation of the most critical findings.

3.6.2 Participant Engagement

The participants in the main intervention were highly engaged and enthusiastic about the research and the workshop, especially mentioning the energy of the researcher and their curiosity toward the study. The structure of the workshop started with a warm-up as a self-introduction since none of the participants were in the researcher's class in TÖMER. Many participants expressed their enthusiasm for learning about AI, a subject that was largely unfamiliar to them. Questions related to the quality usage of ChatGPT and some advice that was asked during the workshop are intentionally left unanswered and explained to affect the reliability of the research. Students were asked to spend time on it and discover it on their own.

During each session, some of the participants asked important questions that might affect the flow of the research. Therefore, if a different question that might have an important impact on the study is asked, they were shared in the WhatsApp group, which is particularly created for Focus/Control groups. The important questions shared are as follows:

"Can we share in the group chat with our friends if we discover a useful prompt?" (Workshop 1)

"Can we get help from social media content that gives advice, such as TikTok and Instagram videos?" (Workshop 3)

“If I cannot ask you, can I ask my brother when I have a question about ChatGPT?” (Workshop 5)

These questions are integrated into the study's methodology as indicators of the participants' need for social mediation and external scaffolding, which directly relate to our research questions on learner autonomy and AI-assisted writing proficiency. They reflect Vygotsky’s sociocultural theory and Bandura’s social cognitive theory by highlighting how social interactions and peer support influence the use and effectiveness of AI tools. Furthermore, the reliance on external support channels underscores the digital literacy gap identified by Van Dijk (2005) and aligns with the emotional states and confidence levels that affect their autonomous learning. By capturing these interactions through reflective diaries and semi-structured interviews, our methodology not only quantifies improvements in writing proficiency via pre/post-tests but also qualitatively explores the multifaceted dimensions of learner autonomy in a socially mediated, AI-integrated environment.

The researcher’s response to these participant questions was to acknowledge that while the primary goal was to assess autonomous learning via AI, the emergence of peer-based queries highlights the inherently social nature of learning, and the participant questions were considered an integral part of the learning process, aligning with established theories and that addressing them enriches rather than detracts from the study’s validity. Related questions, such as sharing useful prompts in the group chat, seeking advice from social media, or asking a sibling when unsure, reflect the genuine need for scaffolding and social mediation, as supported by Vygotsky’s Sociocultural Theory and Bandura’s Social Cognitive Theory. Due to the possibility that the questions could affect the decision, these questions were left to the students, and no restrictions were put. This dual dynamic, where technology facilitates independent work yet simultaneously encourages collaborative engagement, enriches our understanding of learner autonomy. By incorporating these observations into our analysis, we not only maintain the validity of the study but also underscore the complex interplay between AI-driven self-regulation and social interaction, thereby enhancing the overall value of the research.

3.6.3 Quantitative Analysis

For the statistical analysis in this study, the researcher turned to JASP (Marsman & Wagenmakers, 2017), a user-friendly software favoured by researchers for its ability to perform complex calculations and create clear visual representations of data. What makes JASP particularly helpful is that it's built on the foundation of R, a powerful statistical programming

language, but presents these sophisticated tools in a way that's accessible and intuitive (Navarro, 2018). Before diving into JASP, the raw data from our quantitative measures were carefully prepared and organised using Microsoft Excel. Excel provided a familiar and efficient environment to initially structure, clean, and manage the datasets, ensuring they were ready for importing into JASP for in-depth statistical examination (Field, Miles, & Field, 2012)

It is crucial to note that the quantitative data in this study served primarily to enrich and contextualise the qualitative findings, providing a comparative grounding against which to interpret the more nuanced insights gleaned from participant diaries, interviews, and observations. Therefore, the quantitative component was not intended for broad generalisation, but rather to highlight potential differences and patterns within this specific participant group, thereby boosting the depth and value of the predominantly qualitative dataset (Creswell & Plano Clark, 2017). This intentionally focused role of the quantitative data ensures that any limitations inherent in its scope do not compromise the overall reliability of the study, which rests primarily upon the rich, in-depth qualitative exploration. The data was not normally distributed due to the small sample size, see Table 6. In this study, a combination of non-parametric statistical tests was employed to analyse the pre-test and post-test results due to the small sample size (N = 18) and the supporting role of quantitative data within a sequential embedded mixed-methods design (QUAL > quan). The choice of statistical tests, Wilcoxon Signed-Rank Test, Mann-Whitney U Test, was guided by the data characteristics, distribution assumptions, and the study's research design. This section provides a rationale for selecting these tests based on statistical principles and previous methodological research.

Multiple Comparisons Adjustment

To control for family-wise error rate across multiple tests, Holm-Bonferroni sequential adjustments were applied. Table 10 presents the adjusted significance levels.

Table 8

Holm-Bonferroni Sequential Adjustments for Multiple Comparisons

Test	Unadjusted *p*	Rank	Adjusted α	Significant
Total Score (Between-group)	.006	1	.007	Yes**

Test	Unadjusted *p*	Rank	Adjusted α	Significant
Experimental (Within-group)	.002	2	.008	Yes**
Mechanics (Between-group)	.023	3	.010	No
Control (Within-group)	.027	4	.013	No
Artistic Language (Between-group)	.028	5	.017	No

Note. Tests are ranked by *p*-value. Adjusted α = .05/(number of remaining tests). **Significant after adjustment.

Asymmetry: Multiple distributions showed clear right or left skewness

Outliers: Several extreme values were evident, particularly in post-test scores

Ceiling effects: Some distributions approached the maximum possible scores, creating truncated distributions

Small sample clustering: With limited sample sizes, distributions appeared lumpy and irregular

Figure 10

Quantile-Quantile (Q-Q) Plots for Normality Assessment

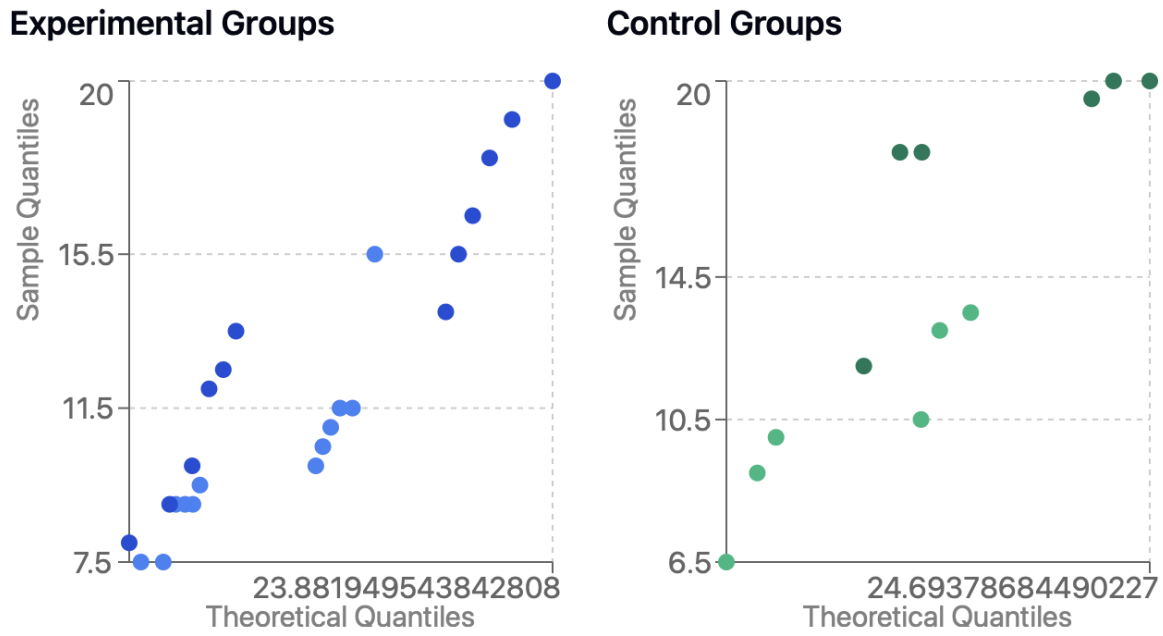


Figure 10 displays quantile–quantile (Q–Q) plots used to visually assess the normality of total writing scores and selected writing criteria across experimental and control groups. Each plot compares the observed quantiles of the sample data to the expected quantiles from a theoretical normal distribution, represented by the diagonal reference line.

In a perfectly normal distribution, data points align closely with the reference line. In this analysis, however, several plots demonstrate systematic deviations, particularly in the form of S-shaped curves and curvature at the tails. These departures from linearity were most pronounced in the control group's post-test scores, especially for criteria such as grammar, mechanics, and artistic language. In these cases, the tails diverge upward or downward, suggesting skewness and possible ceiling effects.

In contrast, the experimental group's distributions showed more moderate deviations from linearity, though some curvature was also noted in post-test values. These observations align with results from the Shapiro–Wilk tests and box plot analyses, which collectively confirm that normality assumptions are not met for multiple score distributions, especially in the control group's post-intervention performance.

Thus, the Q–Q plots reinforce the conclusion that the data exhibit non-normal distributional characteristics, further justifying the use of non-parametric inferential methods in the following analysis (Field, 2018; Ghasemi & Zahediasl, 2012).

3.6.4 Statistical Method Justification

Based on the comprehensive normality assessment, the following statistical approach was justified in Table 9.

Table 9

Selected Statistical Methods

Analysis Type	Parametric Option	Non-Parametric Selection	Primary Justification	Additional Considerations
Within-Group Comparisons	Paired t-test	**Wilcoxon Signed-Rank Test**	39.3% normality violations	• Small sample sizes (n=6-12) • Ordinal rubric data • Paired design maintained

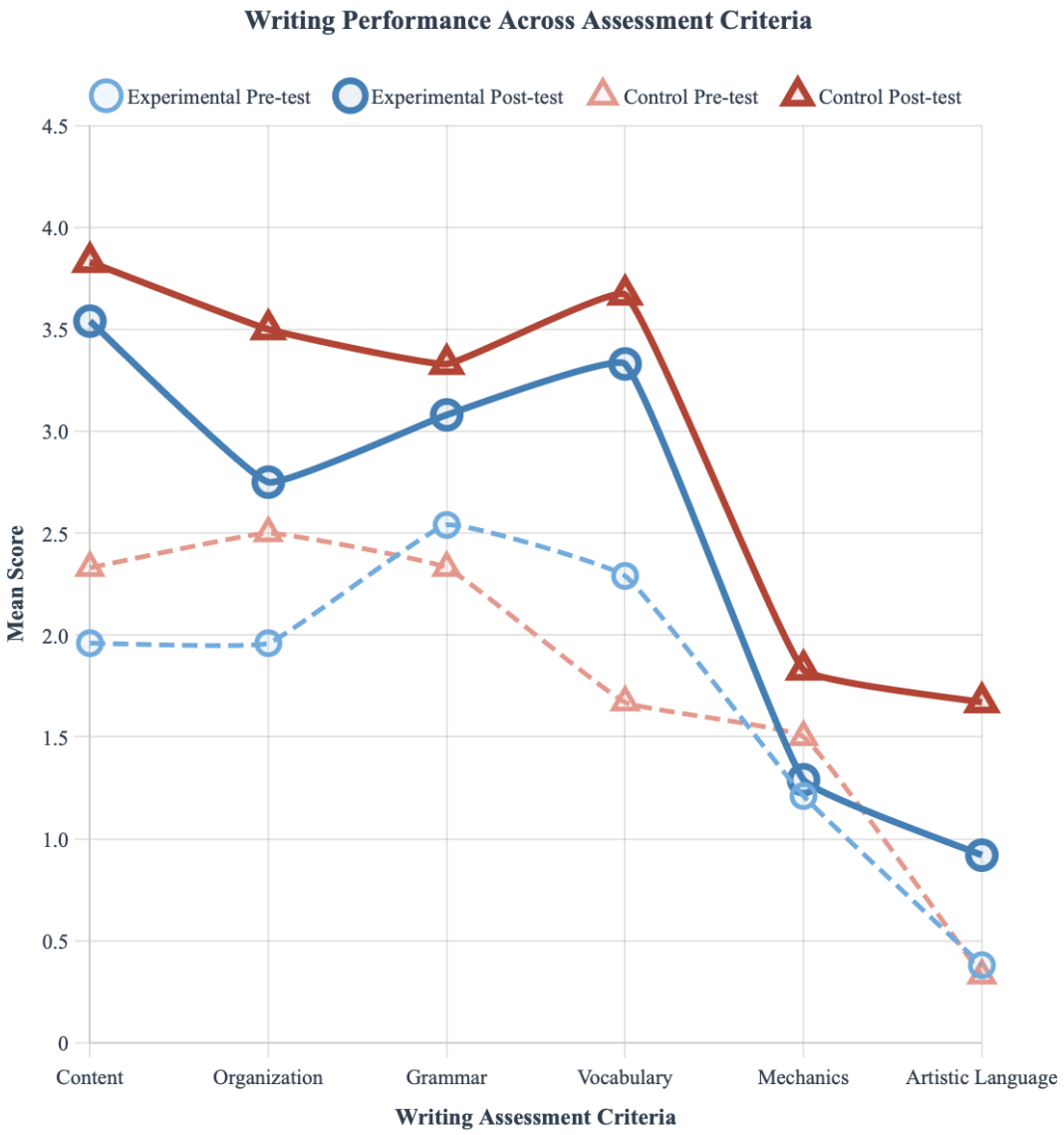
Analysis Type	Parametric Option	Non-Parametric Selection	Primary Justification	Additional Considerations
Between-Group Comparisons	Independent t-test	**Mann-Whitney U Test**	Unequal group sizes + non-normality	• Robust to outliers • No homoscedasticity assumption • Suitable for ordinal data
Effect Size Calculation	Cohen's d	**Rank-based r (Z/√N)**	Consistent with non-parametric approach	• Comparable across analyses • Appropriate for ordinal data • Less sensitive to outliers
Multiple Comparison Control	Bonferroni correction	**Holm-Bonferroni Sequential**	Balance power vs. Type I error	• Less conservative than Bonferroni • Maintains family-wise error rate • Suitable for 21 planned comparisons

3.6.5 Multiple Comparison Considerations

Given the multiple statistical tests planned (7 criteria × 3 comparison types = 21 tests), the Holm-Bonferroni sequential correction method was selected to control family-wise error rate while maintaining reasonable statistical power. This approach is less conservative than the traditional Bonferroni correction while still providing adequate protection against Type I error. Figure 11 presents a comprehensive visualisation of the pre-test to post-test changes across all writing criteria, illustrating the magnitude and consistency of improvements in both groups.

Figure 11

Pre-Test and Post-Test Writing Performance by Group and Criterion

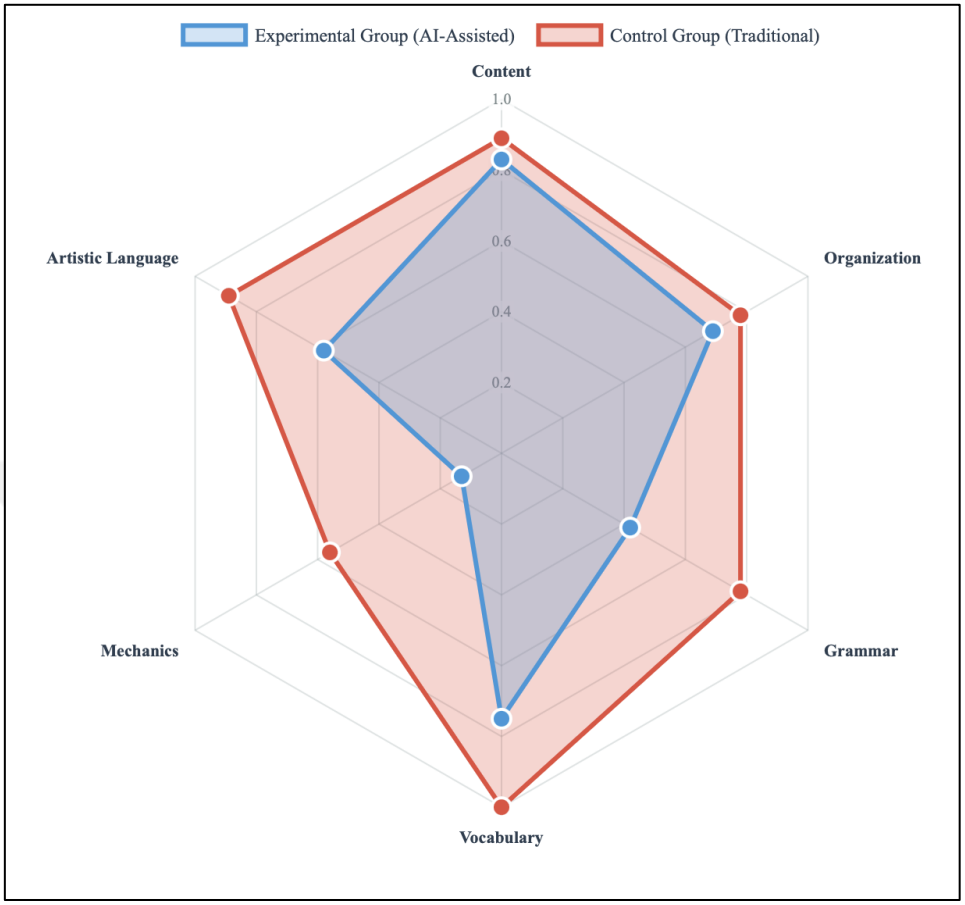


Note. The chart displays mean writing scores across six assessment criteria for experimental (AI-assisted, n = 12) and control (traditional instruction, n = 6) groups. Both groups demonstrated improvements across most criteria, with the control group achieving higher post-test performance in organisation, vocabulary, mechanics, and artistic language.

Figure 12 illustrates the effect size magnitudes for within-group improvements, providing a visual representation of intervention effectiveness across writing criteria.

Figure 12

Within-Group Effect Sizes (Cliff's Delta) for Writing Improvements



Effect Size Interpretation:

δ < 0.147 = Negligible

δ 0.147-0.33 = Small

δ 0.33-0.474 = Medium

δ > 0.474 = Large

Figure 12, Radar chart displaying effect sizes (Cliff's delta) for within-group pre-post improvements across writing criteria. Values range from 0 (no effect) to 1.0 (maximum effect). Both groups demonstrated large effect sizes (δ > 0.5) for most criteria, with the control group showing consistently higher effect magnitudes. The experimental group showed minimal improvement in mechanics (δ = 0.13), while achieving substantial gains in content (δ = 0.83) and vocabulary (δ = 0.75). The control group achieved large effects across all criteria, with a perfect effect size in vocabulary development (δ = 1.00).

One of the critical findings is that the control group achieved large effect sizes ($\delta > 0.5$) in all criteria except mechanics, while the experimental group had large effects in 4/6 criteria, with minimal mechanics improvement.

3.6.6. Qualitative Analysis

Thematic analysis of student writing performance and feedback patterns was conducted using MAXQDA 2024, a qualitative data analysis software. A six-phase thematic coding approach outlined by Braun and Clarke (2006) was applied to examine the pre- and post-test essay samples of both the control and experimental groups. Coding categories were initially guided by the rubric components (e.g., content relevance, organisation, vocabulary use, artistic language, grammar, and coherence) and later expanded to include emergent themes related to learner autonomy, strategy use, and emotional tone. This method allowed for systematic tracking of progress and qualitative differences across students. The qualitative analysis was supported by non-parametric statistical results to strengthen the interpretation of group-level changes.

The coding process began with a familiarisation process to get to know and understand the responses. Then continues with open coding, where initial codes were assigned to segments of feedback, followed by axial coding to group these codes into broader categories (Strauss & Corbin, 1998). This approach enabled a structured examination of differences in depth, specificity, and thematic focus between the feedback types. Using MAXQDA supported a rigorous analysis, ensuring that themes were clearly defined and easily traceable (Kuckartz & Rädiker, 2021).

For the qualitative data analysis, an exploratory, inductive coding (Vears & Gillam, 2022) approach was employed to perform content analysis on the participants' responses. This method allowed for themes and patterns to emerge directly from the data without predefined categories, aligning with the inductive nature of the study. Participants in both groups maintained weekly reflective diaries over an eight-week period, documenting their experiences with writing feedback. The diaries captured their perceived effectiveness of feedback (AI vs. teacher-led), challenges in understanding and applying feedback, and self-regulation strategies and engagement with revisions. After the intervention, semi-structured interviews were conducted with a subset of 10 participants (5 from each group) to gain deeper insights into their perceptions of feedback and autonomy. Interview questions focused on their experience using AI-generated vs. teacher feedback, their ability to self-monitor and revise independently, and the challenges and benefits of AI in writing instruction.

Writing samples were evaluated using a rubric-based scoring system, assessing high-level (artistic language usage) and low-level (grammar, structure, punctuation) features. To ensure the reliability of the coding process, a second researcher independently graded the pre- and post-tests by the rubric, minimising potential bias and increasing inter-rater reliability.

Writing samples were evaluated using a rubric-based scoring system, assessing high-level (artistic language usage) and low-level (grammar, structure, punctuation) features. To ensure the reliability of the coding process, a second researcher independently graded the pre- and post-tests using the same rubric, minimising potential bias and increasing inter-rater reliability. Inter-rater reliability was assessed using the Intraclass Correlation Coefficient (ICC) two-way mixed effects model for consistency, single measurement ICC(3,1). The analysis revealed excellent inter-rater reliability across all assessment criteria: Content & Development (ICC = 0.872), Organisation & Structure (ICC = 0.878), Grammar Accuracy (ICC = 0.886), Vocabulary Range & Usage (ICC = 0.866), Mechanics (ICC = 0.755), Artistic Language Use (ICC = 0.905), and Total Score (ICC = 0.972). All ICC values exceeded 0.75, indicating good to excellent reliability according to established benchmarks (Koo & Li, 2016), supporting the consistency and dependability of the scoring process across both raters. The exceptionally high ICC value for the Total Score (0.972) demonstrates excellent overall agreement between raters, providing strong evidence for the reliability of the assessment instrument in this study context.

3.7. Expert Validation and Refinement of Interview Questions

To enhance the validity and clarity of the interview questions, expert opinion was sought, and minor revisions were made to ensure that the questions were more precise, contextually relevant, and aligned with the study's objectives. According to Dörnyei (2007), well-structured qualitative interview questions should be specific, unambiguous, and directly related to the research focus, minimising the risk of misinterpretation by participants.

One notable revision was the modification of the question:

Original: *"Have you ever used generative AI (ChatGPT)?"*

Revised: *"Have you ever used generative AI(ChatGPT) specifically for EFL studies or writing purposes?"*

This adjustment was made to ensure specificity and focus responses on the research context rather than broad or unrelated experiences with AI tools. As Patton (2015) argues, targeted questioning is crucial in qualitative research, as vague or overly broad questions may lead to inconsistent or irrelevant responses, reducing the study's reliability. By specifying EFL

studies and writing, the revised question directs participants to reflect on their actual use of AI in language learning, thereby generating more relevant and insightful data.

Additionally, Creswell and Poth (2016) emphasise that refining interview questions through expert feedback enhances content validity, ensuring that the questions effectively capture the constructs under investigation. This iterative refinement process aligns with qualitative best practices, contributing to a more structured and reliable data collection process.

CHAPTER IV
RESULTS

This chapter presents the findings of the study, organised according to the three research questions. Both quantitative and qualitative data were analysed to provide a comprehensive understanding of the effects of AI-assisted academic writing support on B1-level EFL learners' writing performance and learner autonomy. Quantitative analyses include non-parametric statistical tests of pre- and post-test scores, while qualitative insights are derived from essay samples, rubric-based evaluations, learner reflections, and workshop observations. Each research question is addressed in a separate section to ensure clarity and alignment with the study's objectives.

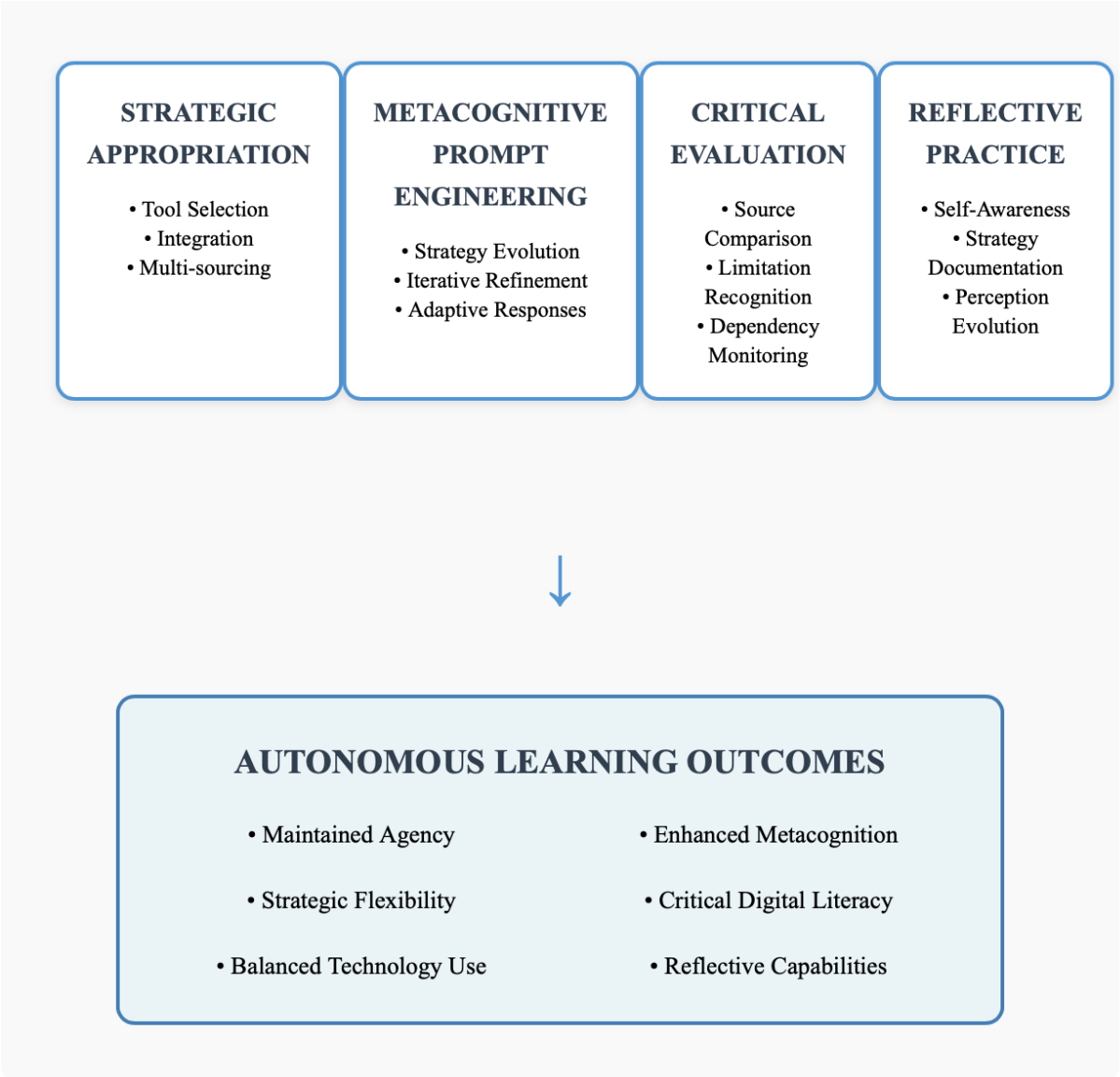
4.1. Results for RQ1

RQ 1: How does the integration of AI influence learner autonomy in EFL writing contexts?

The study aimed to evaluate how AI integration influences learners' ability to take ownership of their learning process. The findings suggest that ChatGPT played a multifaceted role in fostering learner autonomy. Thematic analysis of interviews, pre-post-tests and reflective diaries indicated positive influences such as an increase in self-regulation, planning, and internalisation of feedback. Learners in the experimental group reported a growing tendency to review and revise their work independently, guided by prior interactions with AI. Several participants noted a shift from relying on teacher prompts to initiating self-correction, demonstrating greater metacognitive control over their writing practices. However, signs of over-reliance were also present in some responses, raising concerns about the balance between support and dependency. These results support existing literature on the role of digital tools in promoting learner autonomy (Benson, 2011; Little, 2007; Yusof, 2021) and align with sociocultural views that autonomy develops through scaffolded engagement with learning technologies.

Figure 13

Main Themes in Learner Autonomy AI-Instructed Contexts



Note. Framework illustrating the four key dimensions of learner autonomy manifestation in AI-instructed writing contexts, showing the relationship between strategic behaviours and autonomous learning outcomes.

4.3.1 Strategic AI Tool Appropriation and Integration

Participants demonstrated autonomous agency in selecting and integrating AI tools that complemented their established learning habits, rather than adopting them wholesale. This strategic approach to tool selection highlights a personalised understanding of how each digital resource serves distinct learning needs.

Ömer Rüzgar (E3) exemplified this selective approach by stating, *"British Council kullanıyorum, ChatGPT'i kullanıyorum, Google'u kullanıyorum. Ama özellikle öğretmenlerimi kullanıyorum."* (I use British Council, I use ChatGPT, I use Google, but I especially use my

teachers.) This reflects how learners position AI not as a replacement for traditional support, but as one part of a broader network of resources.

E2 similarly demonstrated this targeted use of AI tools: *"AI'lar benim grammar'ımın nerede yanlış olduğunu söylüyor. Ve çeviriler de benim anlamadığım sözcükleri bana söylüyor."* (AIs tell me where my grammar is wrong. And translations tell me the words I don't understand.) His strategy reflects the purposeful integration of AI features with high relevance to his personal deficiencies.

Control group participants also demonstrated selectivity, but their reliance skewed more towards traditional tools, highlighting a divergence in digital autonomy. For instance, C5 described his approach as, *"Farklı kaynaklardan da o bilgiyi kontrol ederek öğreniyorum… Hepsinden gelen ortak bilgileri karşılaştırdım."* (I learn by checking that information from different sources… I compared the common information from all of them.) Although not reliant on AI, his strategy reflects a comparable autonomous behaviour , triangulating sources to verify knowledge accuracy.

Overall, this theme illustrates how learners made deliberate, context-sensitive decisions about when and how to integrate AI tools. Rather than replacing existing practices, participants repurposed AI affordances to address gaps and enrich their personal learning ecosystems.

The findings in this sub-theme demonstrate that selective tool adoption operates as a key marker of autonomous behaviour. Rather than over-relying on one and only digital source, learners orchestrated a harmonised network of supports, positioning AI as one of several cognitive scaffolds within their broader learning process.

4.1.2 Development of Multi-Source Verification Strategies

In addition to selecting tools strategically, autonomous learners demonstrated advanced behaviours in validating the reliability of information by cross-referencing multiple sources. This behaviour illustrates a key dimension of autonomy, **critical evaluation**, and reflects the transition from passive tool usage to active knowledge construction.

C5, a control group participant, exemplified this sophisticated evaluative process: *"Farklı kaynaklardan da o bilgiyi kontrol ederek öğreniyorum. Mesela karanlık maddeye ilgi duyuyordum… Gittim onları da izledim. Hepsinden gelen ortak bilgileri karşılaştırdım."* (I learn by checking that information from different sources. For example, I was interested in dark matter… I went and watched them too. I compared the common information from all of them.) His approach reveals a layered verification process involving written and visual materials, an embodiment of **multi-modal triangulation** in informal learning.

This behaviour, while not dependent on AI, mirrors the same epistemic vigilance observed in the experimental group. The emphasis was not merely on information acquisition but on its **authentication**, indicating a shift from surface-level digital literacy to **critical digital autonomy**. These learners recognised that knowledge credibility often depends on cross-source consistency, particularly when dealing with AI-generated outputs, which can be confidently incorrect or lack source transparency.

This subtheme supports broader claims that learner autonomy is not defined solely by tool usage, but by the **cognitive stance** learners adopt toward information. Verification practices such as these signal the development of higher-order learning behaviours: synthesising, evaluating, and constructing understanding across a variety of digital and non-digital sources.

4.1.3 Metacognitive Prompt Engineering and Iterative Optimisation

This theme captures the developmental trajectory of learners as they transition from passive AI users to critical, reflective prompt engineers. Through iterative interactions, learners enhanced their ability to formulate context-aware prompts, demonstrating metacognitive insight into both their learning goals and the operational logic of AI tools.

4.1.3.1 Evolution from Basic to Sophisticated Prompting Strategies

Participants displayed significant progress in crafting prompts that went beyond surface-level queries. Initially engaging with AI through generic questions, many learners evolved toward more tailored and pedagogically informed instructions. This shift suggests growing awareness of how precise linguistic framing can shape the quality of AI feedback.

E7 described his process as *"En iyi promptu kendimce buluyorum… deneme yanılma yöntemiyle"* (I find the best prompt myself… through trial and error method). His refined prompt exemplified the shift toward targeted cognitive engagement: "Hani bu kelimeyi, bu cümledeki kelimeleri anlamını değiştirmeden nasıl farklı kullanırdın?" (How would you use these words, the words in this sentence, differently without changing their meaning?) This request reflects syntactic sensitivity and an emerging ability to direct the AI's linguistic operations in alignment with learning objectives.

Similarly, E3 adopted a multi-layered prompting approach aimed at both error correction and developmental feedback: *"Sana denememi atacağım. Bu denemedeki her hataları tek tek söyle. Nasıl düzeltebileceğimi, düzeltsem bile nasıl daha iyi geliştirebileceğimi anlatır mısın?"* (I'm going to send you my essay. Tell me each error in this essay one by one.

Can you tell me how I can correct them, and even if I correct them, how I can develop them better?) This reveals a dual-purpose prompting structure, one that integrates remediation with enhancement.

These prompting strategies suggest the emergence of metacognitive orchestration, whereby learners not only reflect on their performance but actively shape the nature of instructional feedback through linguistic control.

4.1.3.2 Adaptive Response to AI Limitations

Despite improvements in prompting, participants frequently encountered instances where AI responses failed to meet expectations. Rather than disengaging, learners adjusted their approach, often by refining their inputs or requesting alternative formulations, demonstrating resilience, strategy adaptation, and persistence.

E7, for instance, noted: *"Prompt işe yaramadığı zaman genelde şey istiyorum… başka alternatifler türetiyor. Onları deneme yanılma yöntemiyle… en iyi promptu kendimce buluyorum."* (When the prompt doesn't work, I usually ask for alternatives… it generates other alternatives. Through trial and error… I find the best prompt myself.) This quote encapsulates an iterative mindset, reflecting both diagnostic skill and flexible response strategies.

This adaptive behaviour is indicative of collaborative problem-solving with AI, a form of situated autonomy in which learners remain cognitively engaged even when technological scaffolds falter. Such resilience is particularly relevant in digital learning environments, where learners must manage uncertainty and tool limitations independently.

4.1.4 Critical Evaluation and Balanced Integration of AI Feedback

Theme 3 captures participants' growing ability to critically appraise feedback from different sources and manage their reliance on AI systems. Rather than accepting feedback passively, learners engaged in comparative judgement and actively mitigated overdependence, hallmarks of metacognitive maturity and reflective autonomy.

4.1.4.1 Comparative Analysis of AI vs. Human Feedback Sources

Participants developed nuanced understandings of both the affordances and limitations of AI-instructed feedback when contrasted with traditional teacher-led input. This comparative stance reflects an evaluative autonomy, where learners strategically position AI tools within broader feedback ecologies.

C5 offered a balanced perspective: *"Sizi tanıyan bir öğretmen sizin nerede hata yapabileceğinizi daha iyi analiz edebilir… ChatGPT bize farklı varyasyonlar sunma ve örneklendirme konusunda… maksimum verimde cevap verebiliyor."* (A teacher who knows you can better analyse where you might make mistakes… ChatGPT can provide us with different variations and examples… can give you maximum efficiency answers.) His reflection illustrates a dual awareness: valuing the personalisation of human feedback while recognising the generative fluency and breadth of AI responses.

This comparative lens was evident in both experimental and control groups, suggesting that students were not uncritically persuaded by novelty. Instead, they assessed tool effectiveness based on task fit, trust, and clarity, critical dimensions in autonomous feedback integration.

4.1.4.2 Recognition and Mitigation of Over-reliance Risks

Beyond comparison, some learners actively monitored their emotional and cognitive dependency on AI tools, expressing concern about the long-term risks of over-reliance. This self-awareness signals an advanced form of autonomy, where learners anticipate potential barriers to independence and proactively intervene.

E9 articulated this risk clearly: *"Çok fazla ChatGPT diye bağımlı hâlâ da gelebiliyorsun… Onsuz bir şey yapamıyor olmak… Buna çok dikkat ettim zaten. Hani fark ettiğim an hani bunun önüne geçtim."* (You can still become too dependent on ChatGPT… Not being able to do anything without it… I paid a lot of attention to this. The moment I noticed it, I prevented it.) Her ability to self-regulate and course-correct indicates a metacognitive checkpoint that guards against the erosion of learner agency.

This awareness of tool dependence and the deliberate effort to curb it represent an emerging form of critical digital resilience. Rather than merely using tools effectively, learners in this theme recognised the importance of maintaining cognitive ownership and the capacity to operate independently when necessary.

4.1.5 Evolving Perceptions and Self-Awareness Through Reflective Practice

Theme 4 explores how students' reflective diary entries served not only as documentation tools but as catalysts for metacognitive growth. Over time, participants demonstrated increasing insight into their learning habits, dependencies, and the limitations and affordances of AI-supported writing. The diaries functioned as a mirror, enabling participants to monitor, question, and reconstruct their learning strategies through sustained reflection.

4.1.5.1 Metacognitive Awareness Development Through Diary Reflection

Participants used journaling not merely as a reporting requirement but as a means to examine their own thinking and behavioural patterns. The diaries acted as introspective spaces in which metacognitive awareness emerged and evolved over the course of the intervention.

E9 articulated this process vividly: *"Yazarak öğrendiğimi fark ettim. Bir şey yazmam gerekiyormuş, onu fark ettim. Çünkü gerçekten bilinç karmaşasında kayboluyor öğrendiğim veya öğrenemediğim şeyler."* (I realised that I learn by writing. I realised that I needed to write something. Because what I learn or can't learn really gets lost in the confusion of consciousness.) This quote illustrates the emergence of self-awareness regarding learning modality, a shift from reactive to proactive control over learning strategies.

E3 similarly reflected on his development over time: *"İlk yazdığım günlüğe baktığımda neler değiştirip fark ettiğimde cidden mutlu olmuştum."* (When I looked at the first diary I wrote and noticed what had changed, I was really happy.) The act of comparing past and current entries enabled learners to observe their growth, reinforcing internal motivation and consolidating a reflective learning identity.

4.1.5.2 Perception Changes About AI Tool Limitations and Benefits

Over time, participants moved beyond binary judgments of AI as either helpful or frustrating. Their diary reflections revealed increasing nuance, recognising the conditional value of AI tools depending on usage quality and task type.

E3's diary entries revealed a shift in perception triggered by cognitive fatigue and content repetition: *"2. ve 6. günlükler arasında artık ne yazacağımı kestiremiyordum. Çünkü devamlı aynı şeyler oluyordu."* (Between the 2nd and 6th diaries, I couldn't figure out what to write anymore. Because the same things kept happening.) This moment of frustration sparked a broader reflection on AI's limits in generating novel input without deeper user intervention.

In contrast, E2 highlighted a turning point in his awareness of the interaction quality-feedback output relationship: *"Günlük bu farkındalığa nasıl katkıda bulundu?... eğer çecipiti ile ilgili olan yazışmamı iyi yaparsam... günlüğe daha fazla şey yazabiliyordum."* (How did the diary contribute to this awareness? If I did my correspondence with ChatGPT well… I could write more things in the diary.) This suggests a reciprocal relationship between AI engagement quality and learner self-discovery.

4.1.5.3 Documentation of Autonomous Strategy Development

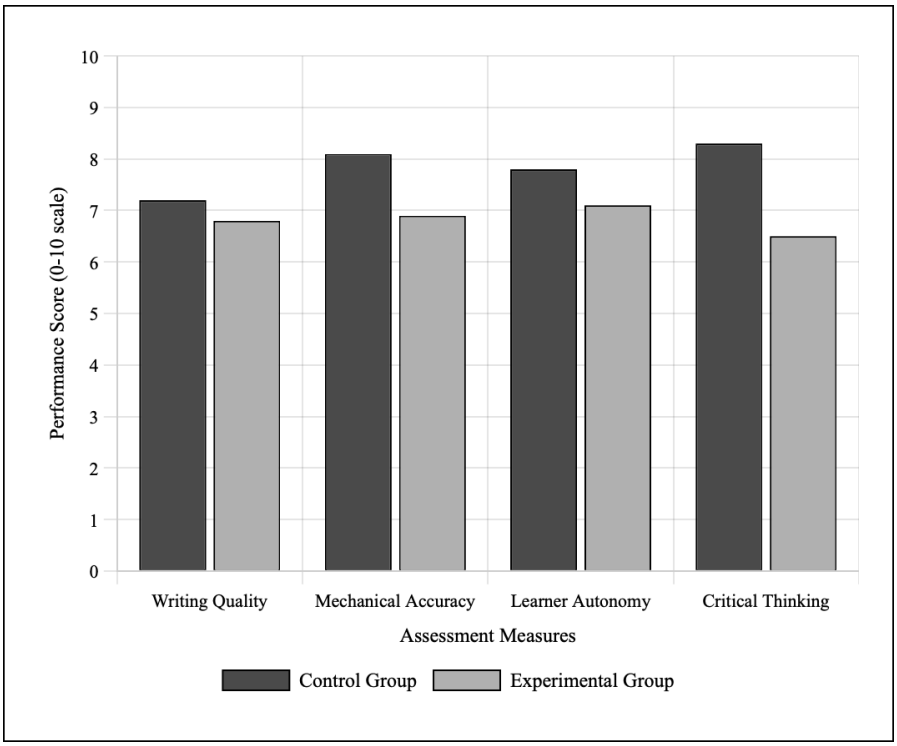
The longitudinal nature of diary entries allowed learners to track not only what they learned but also how and why particular strategies worked. Over time, participants documented the experiments they conducted with AI tools, self-evaluated outcomes, and refined their practices accordingly.

E9's later diary entries indicated growing awareness of over-reliance risks and strategy shaping: *"Günlük tutmanın çerçeveli öğrenme sürecinizi daha bilinçli hale getirdiğinizi düşünüyor musunuz? Evet, kesinlikle… kötü yanları ya da iyi yanları yazarak daha görsel bir hale getirdiğimde daha efektif bir sonuç aldım."* (Do you think keeping a diary made your learning process more conscious? Yes, absolutely… when I made the bad or good sides more visual by writing, I got more effective results.)

Such entries demonstrate a form of meta-level regulation, where learners no longer view strategies as fixed routines but as adjustable parameters informed by feedback, reflection, and experimentation. This behaviour reflects what Zimmerman (2002) termed "cyclical self-regulation," but now within a digital, AI-enhanced context.

Figure 14

Comparison of Control Group and Experimental Group Performance Across Multiple Measures



4.1.6 Emotional and Motivational Perception Shifts

While interviews captured learners' strategic behaviours and tool usage, the reflective diaries provided invaluable insights into the emotional undercurrents shaping their autonomous development. Through sustained self-reflection, participants revealed fluctuations in motivation, moments of doubt, and eventual realisation of growth, showcasing how affective factors intertwined with cognitive autonomy.

E3 described a tension between effort and perceived value: *"Günlük yazarken bazen sıkılıyordum çünkü arada düşünüyordum hiç gerek yok bunlara… başarı hissini sanırım günlük yazarken çok da hissetmiyordum."* (Sometimes I got bored while writing the diary because I thought there was no need for these things… I probably didn't feel much sense of achievement while writing the diary.) Yet, his retrospective reflections revealed that these moments contributed to long-term understanding and strategy development, underscoring the value of delayed metacognitive reward.

Participants' emotional relationship with AI tools also shifted over time. Initial reflections showed polarised reactions, ranging from excitement to scepticism, but sustained diary writing helped temper these extremes. Learners began to frame AI not as a novelty or threat, but as a situationally useful, critically evaluated partner in their learning process.

4.1.7 Cross-Theme Analysis: Autonomy Manifestation Patterns Through Diary and Interview Integration

Synthesis of diary and interview data revealed three overarching autonomy patterns that cut across all four themes:

4.1.7.1 Integration Rather Than Replacement

Participants across both experimental and control groups did not treat AI tools as replacements for existing learning supports. Instead, they integrated them into broader ecosystems that included teachers, peers, textbooks, and online platforms. This is consistent with the framing of **collaborative autonomy**, where learners blend digital and human scaffolds to sustain agency.

Diary entries revealed internal negotiation processes where learners consciously balanced AI assistance with independent thinking. As E9 asked in her diary: *"Öğrenme stratejilerinizde değiştirmek istediğiniz şeyler var mı? Günlük bu farkındalığa nasıl katkıda bulundu?"* (Are there things you would like to change in your learning strategies? How did the

diary contribute to this awareness?) , signalling how diary use itself became a tool for strategic self-direction.

4.1.7.2 Dynamic Adaptation and Flexibility Through Sustained Reflection

Autonomous learners demonstrated the capacity to adapt tools and strategies based on evolving needs, task difficulty, or emotional states. This adaptability, visible particularly in the diary entries, was not reactive but proactive, showing an iterative model of self-regulated learning. Learners frequently revised their prompting, tool usage, or scheduling strategies in response to AI limitations or personal learning breakthroughs.

This pattern exemplifies **situated metacognitive flexibility**, a high-order feature of autonomy that extends beyond knowing *what* to do, to knowing *when* and *why* to shift strategies.

4.1.7.3. Maintained Critical Stance Through Reflective Practice

Despite the efficiency and accessibility of AI-instructed tools, participants consistently demonstrated a healthy scepticism. This critical stance was cultivated over time, often catalysed by AI failures, conflicting outputs, or feedback that lacked depth.

Rather than blindly trusting digital outputs, learners engaged in **reflective vetting**, deciding when to accept, modify, or discard AI-generated suggestions. Their diaries documented internal dialogues such as questioning AI logic, comparing outputs to rubric expectations, or seeking teacher validation. This triangulation of feedback sources marked a mature form of autonomous judgement.

4.2 Results for RQ2

Research Question 2: To what extent do measurable writing outcomes differ between learners receiving ChatGPT-instructed instruction and teacher instruction?

This research question aimed to evaluate the impact of AI-instructed feedback on the academic writing development of learners, as compared to traditionally teacher-guided instruction received by the control group. To address this question, a combination of quantitative and qualitative data was analysed.

Writing performance was assessed using a six-component analytic rubric encompassing content, organisation, grammar, vocabulary, mechanics, and artistic language, with a maximum total score of 20 points. Pre- and post-intervention essays were scored for both the experimental group (n = 12) and control group (n = 6). Given the small sample size and the non-normal

distribution of the data, non-parametric statistical methods were employed. The Wilcoxon Signed-Rank Test was used to examine within-group changes, while the Mann–Whitney U Test evaluated between-group differences.

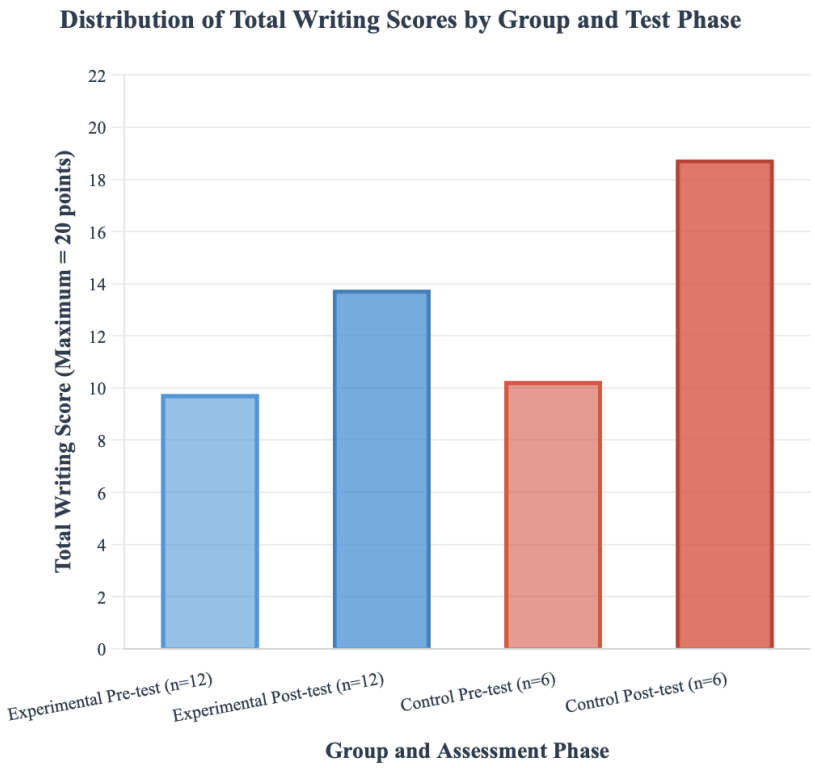
In addition, qualitative patterns of improvement were analysed using thematic coding of diary entries and rubric-informed performance tracking. This triangulated approach enabled a richer interpretation of score changes and skill-specific development.

4.2.1 Distribution of Writing Scores

Figure 15 displays the distribution of total writing scores across both test phases and learner groups, using box plots to visualise key statistical properties that informed the analytical approach.

Figure 15

Distribution of Total Writing Scores by Group and Test Phase



Note. Box plots display the median (centre line), interquartile range (box), and minimum/maximum values (whiskers). The experimental group shows increased variability in post-test outcomes, while the control group demonstrates more consistent high performance.

The experimental group's pre-test scores were moderately clustered (IQR = 2.38), but post-test scores displayed substantial variability (IQR = 7.13). This suggests uneven response patterns to the AI-instructed intervention. In contrast, the control group exhibited consistent improvement, with post-test scores clustering at the upper range (*Mdn* = 18.75). This pattern indicates greater stability and possibly more uniform benefit from traditional teacher feedback.

4.2.2 Individual Performance Analysis

Table 5 provides individual-level results, contextualising the variability evident in the summary statistics. This level of granularity allows closer inspection of how learners within each group responded to the instructional model.

Table 10

Pseudonyms, pre/post scores, and skill-specific deltas

Participant	Group	Pre-Test	Post-Test	Gain	Participant	Group	Pre-Test	Post-Test	Gain
E1	Experimental	7.5	9.0	+1.5	C1	Control	9.0	18.0	+9.0
E2	Experimental	11.5	12.5	+1.0	C2	Control	10.5	19.5	+9.0
E3	Experimental	9.0	19.0	+10.0	C3	Control	13.0	18.0	+5.0
E4	Experimental	7.5	8.0	+0.5	C4	Control	6.5	12.0	+5.5
E5	Experimental	9.0	10.0	+1.0	C5	Control	10.0	20.0	+10.0
E6	Experimental	10.5	16.5	+6.0	C6	Control	13.5	20.0	+6.5
E7	Experimental	15.5	18.0	+2.5					
E8	Experimental	10.0	13.5	+3.5					
E9	Experimental	11.0	20.0	+9.0					

Participant	Group	Pre-Test	Post-Test	Gain	Participant	Group	Pre-Test	Post-Test	Gain
E10	Experimental	9.5	15.5	+6.0					
E11	Experimental	11.5	14.0	+2.5					
E12	Experimental	9.0	12.0	+3.0					

4.2.3 Comparative Performance Outcomes: Experimental vs Control Group

Experimental Group

As shown in Table 10, participants in the experimental group demonstrated overall gains in writing performance between the pre- and post-test assessments. Improvements ranged from +0.5 (E4) to +10.0 (E3, E9), with two participants achieving the maximum post-test score. Despite these gains, individual variation was notable, suggesting differential engagement with the AI-instructed writing process. Learners such as E4 and E5, for instance, showed minimal progress, which may reflect variability in prompt quality, self-regulation strategies, or autonomous tool usage.

Beyond total score gains, a closer analysis of rubric components revealed patterns of skill-specific development. Vocabulary and grammar were the most frequently improved areas, particularly among E3, E6, and E9. These students exhibited clearer sentence construction and more precise lexical choices in their post-test essays. It is likely that these improvements were influenced by targeted prompting around synonyms, structure, and accuracy during interactions with ChatGPT.

Conversely, components such as organisation and content development showed only moderate improvement. While several participants used the AI tool for assistance with introductions or transitions, they struggled to produce deeper cohesion or elaborated ideas. Notably, artistic language, despite accounting for only two rubric points, remained one of the least improved dimensions across both groups. This finding suggests that while AI-instructed tools may support clarity and correctness, their effectiveness in promoting stylistic expression may depend on more advanced prompting or scaffolded guidance.

Control Group

In contrast, the control group demonstrated consistently higher gains, with improvements ranging from +5.0 to +10.0 points. Four of the six participants achieved gains of +9.0 or more, and C5 and C6 both attained the maximum post-test score of 20.0. As shown in subsequent visualisations and interquartile range analysis, the control group's performance was not only higher overall but also more consistent, with relatively low variability in gains.

Skill-specific trends among control group learners differed from those observed in the experimental group. The most substantial improvements were seen in organisation, content relevance, mechanics, and artistic language. These dimensions likely benefited from structured teacher instruction, which offered explicit modelling and feedback on paragraph unity, idea development, and adherence to academic conventions. Additionally, improvements in mechanics, punctuation, spelling, and formatting, were more stable, further supporting the effectiveness of sustained teacher-guided attention to writing detail.

Together, these patterns suggest that the AI-instructed process yielded more diverse and learner-dependent outcomes, particularly benefiting those who actively experimented with the tool. By contrast, teacher-led instruction produced more uniform gains across all rubric dimensions, possibly due to systematic scaffolding, clearer expectations, and corrective feedback aligned with curricular targets.

4.2.4 Distribution Trends and Score Range Summary

Across all pre-tests, the minimum score observed was 6.0, while the maximum post-test score was 20.0, achieved by a control group participant. These figures reflect a broad range of writing proficiency among participants and suggest that while both instructional conditions fostered development, the control group had learners who reached the full potential of the rubric.

Table 11

Detailed Pre-Post Test Median, IQR, Min Max Table of between-groups

Group	Test Phase	*Mdn*	IQR	Min–Max
Experimental	Pre-Test	10.00	2.13	7.5 – 15.5
Experimental	Post-Test	13.75	5.38	8.0 – 20.0
Control	Pre-Test	10.25	3.13	6.5 – 13.5

Group	Test Phase	*Mdn*	IQR	Min–Max
Control	Post-Test	18.50	1.88	12.0 – 20.0

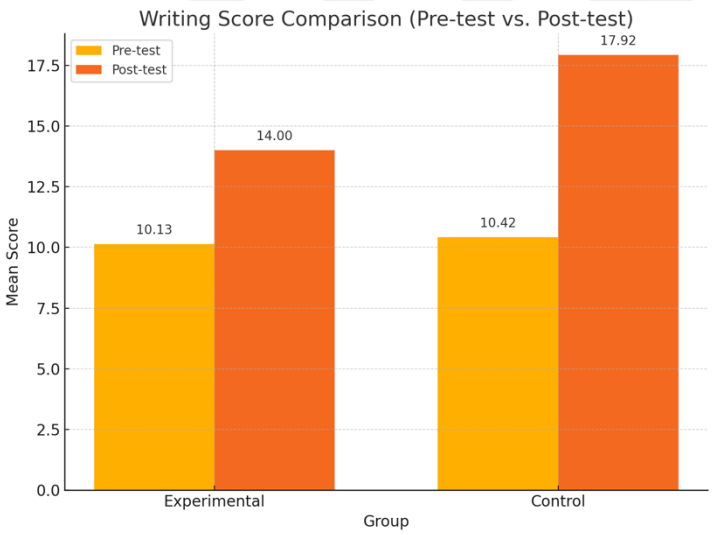
As illustrated, the experimental group's post-test IQR widened to 5.38, indicating increasing variability in performance. While several learners showed marked improvement, others progressed modestly, suggesting that the effectiveness of AI-instructed writing support may be contingent upon learner autonomy, strategy use, or AI interaction quality.

4.2.4.1 Visual Distribution Insight

To further examine distribution patterns, histograms were generated for both pre-test and post-test scores across groups. These visual representations complement the statistical summary by illustrating the spread, clustering, and outliers in learner performance over time.

Figure 16

Histogram of Writing Scores by Group and Test Phase



To explore the distribution and consistency of learner performance within each group, histograms were generated for both pre-test and post-test writing scores. These visuals provide a more detailed picture of how students' outcomes varied across the intervention period. Unlike median scores alone, histograms allow for the identification of score clustering, outliers, and spread, particularly relevant in this study, given the non-normal distribution of the data and presence of individual variability. Figure X presents the writing score distributions by group and test phase.

Figure 17

Boxplot of Pre-Test and Post-Test Scores by Group

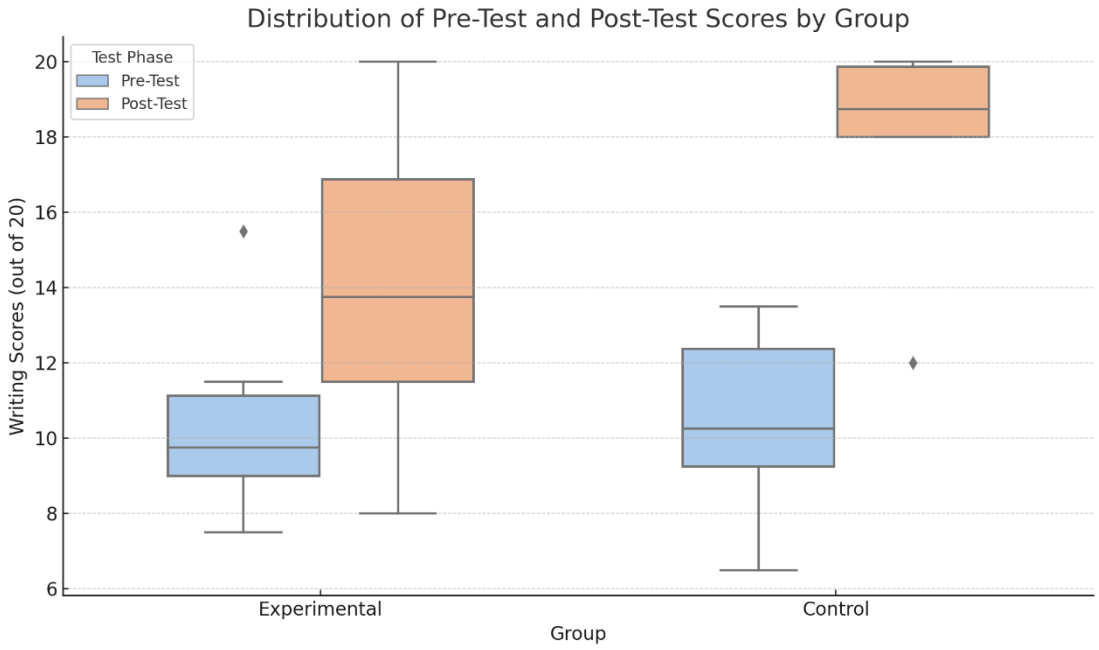


Figure 17, the boxplot, illustrates changes in score distribution across groups and test phases and Whiskers represent the full range excluding outliers; box edges represent Q1 and Q3; median is marked by the central line. The experimental group shows a wider interquartile range in post-test results, indicating varied improvement levels among students. In contrast, the control group's post-test scores are more tightly clustered near the top, reflecting greater consistency and higher final performance. Outliers in the experimental group reflect individual variability in AI engagement and writing development. To better understand the consistency of learner performance, the interquartile range (IQR) was calculated for each group's pre- and post-test scores. Given the non-normal distribution of the data and the presence of outliers, measures of central tendency and dispersion were selected accordingly. Instead of using standard deviation, which is appropriate only for normally distributed datasets, this study employed the **IQR** to describe variability. The IQR offers a robust alternative by capturing the spread of the middle 50% of scores, making it well-suited for small sample sizes and non-parametric analyses.

Table 12

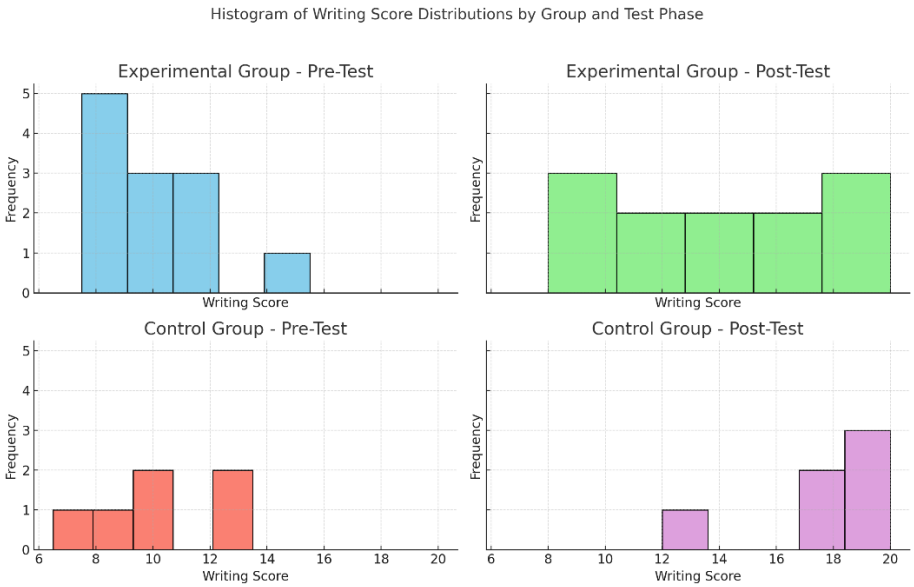
Interquartile Ranges (IQR) of Writing Scores by Group and Phase

Group	Test Phase	IQR
Experimental	Pre-Test	2.13
Experimental	Post-Test	5.38
Control	Pre-Test	3.13
Control	Post-Test	1.88

To support this visual trend, Table 12 provides the interquartile range (IQR) values for each group. The IQR increased from 2.13 to 5.38 in the experimental group, indicating greater variation in performance. Conversely, the control group's IQR narrowed from 3.13 to 1.88, reflecting more consistent gains. However, it should be noted that due to the uneven group numbers, diversity in the experimental group might vary more than in the control group.

Figure 18

Histograms of Writing Score Distributions by Group and Test Phase



Note. Frequencies of student writing scores are shown separately for each group. Bins represent score intervals of approximately 2 points.

Figure 18 shows that the experimental group's pre-test distribution is left-skewed, with most scores clustering between 7.5 and 11.5, indicating lower initial performance across the majority of learners. By the post-test, the distribution becomes wider and more uniform, with scores ranging from 8.0 to 20.0. This reflects greater variability in learning outcomes, suggesting that while some students made notable gains, others showed more limited progress.

In contrast, the control group's pre-test scores were moderately spread across 6.5 to 13.5, showing slight variability at the start. However, the post-test scores shifted markedly, with the majority of students scoring between 17.0 and 20.0. This concentration at the upper end reflects higher and more consistent performance in the control group following the intervention. The lack of low scores and reduced spread supports the finding that the control group achieved more uniform improvement than the experimental group.

These distribution patterns are consistent with the interquartile range (IQR) analysis, which also demonstrated increased variability in the experimental group and reduced variability in the control group. Together, the histogram and IQR results highlight not only differences in group performance but also the divergent trajectories of learner progress under AI-assisted versus traditional instruction conditions.

In summary, these observed trends are consistent across multiple measures of dispersion. The boxplot (Figure 17) visually highlights differences in score spread and central tendency between groups, while the histogram (Figure 18) displays the shape and frequency of score distributions. Additionally, the interquartile range (IQR) values reported in Table 12 numerically confirm that the experimental group exhibited greater variability in post-test outcomes, whereas the control group achieved more uniform improvement.

Pre- and post-test scores were summarised using medians and interquartile ranges (IQR), which provide a more accurate representation of data distribution in non-parametric settings.

The experimental group had a median pre-test score of 10.0 (IQR = 2.13) and a median post-test score of 13.75(IQR = 5.38).

The control group improved from a median pre-test score of 10.25 (IQR = 3.13) to a post-test median of 18.5(IQR = 1.88).

These values suggest both groups experienced writing improvement, with more consistent performance observed in the control group.

4.2.6 Criterion-Specific Performance Analysis

Table 6 presents gain scores across the six writing criteria to identify specific areas of improvement within each group.

Pre-Post Gain Scores by Group and Assessment Criteria

Table 13

Mean Gain Scores by Writing Criterion and Group

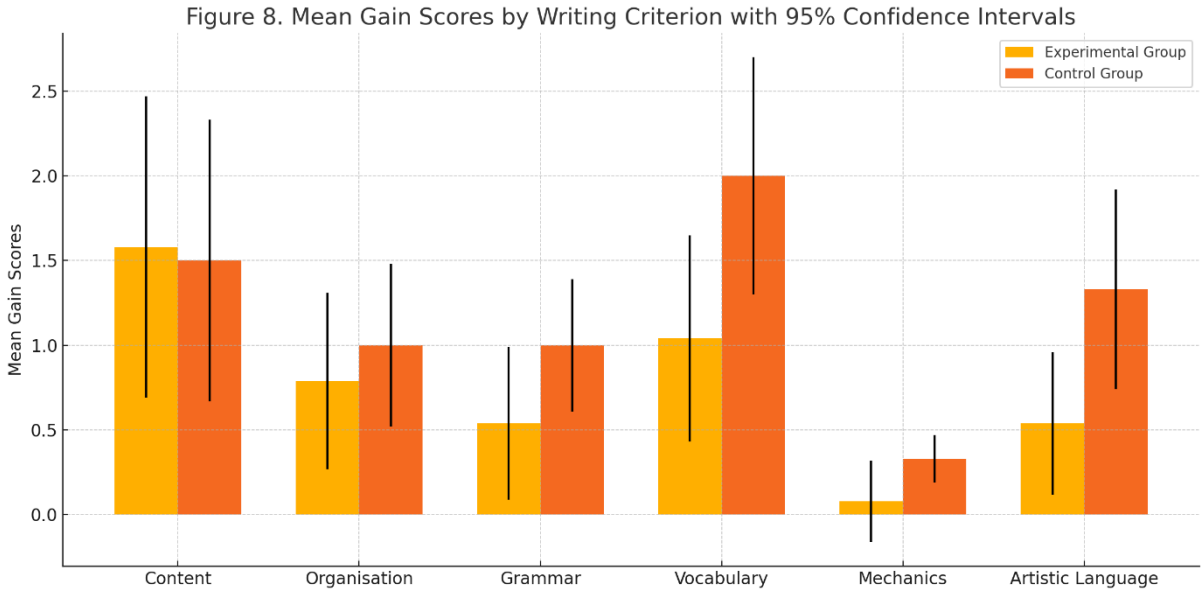
Criterion	Experimental Group			Control Group		
	M	***SD***	***Mdn***	***M***	***SD***	***Mdn***
Content	1.58	1.78	1.50	1.50	1.67	1.25
Organization	0.79	1.03	0.75	1.00	1.11	1.50
Grammar	0.54	0.99	0.50	1.00	1.14	1.00
Vocabulary	1.04	1.22	1.00	2.00	2.19	2.00
Mechanics	0.08	0.49	0.00	0.33	0.47	0.25
Artistic Language	0.54	0.84	0.50	1.33	1.48	1.25
Total Score	**4.59**	**5.17**	**2.75**	**7.17**	**7.97**	**7.75**

Note. M = mean; SD = standard deviation; Mdn = median. Maximum possible scores: Content = 4, Organization = 4, Grammar = 4, Vocabulary = 4, Mechanics = 2, Artistic Language = 2.

Both groups showed substantial gains in content development, with similar mean improvements (experimental: $M = 1.58$; control: $M = 1.50$). The control group demonstrated larger gains in vocabulary ($M = 2.00$ vs. 1.04) and artistic language ($M = 1.33$ vs. 0.54). The experimental group showed minimal improvement in mechanics ($M = 0.08$), while the control group achieved more substantial gains ($M = 0.33$).

Figure 19

Mean Gain Scores by Writing Criterion with 95% Confidence Intervals



Error bars represent 95% confidence intervals around the mean gain scores for each writing criterion by group. The control group demonstrated consistently higher and more stable improvement, particularly in vocabulary and artistic language. Wider confidence intervals in the experimental group suggest greater individual variability in response to AI-instructed support.

4.2.7. Within-Group Statistical Analysis

Wilcoxon signed-rank tests examined pre-post changes within each group. Table 7 presents the results with effect sizes.

Table 14

Within-Group Pre-Post Comparisons Using Wilcoxon Signed-Rank Tests

Group	***W***	***Z***	***p***	***r***	**Effect Size**	**95% CI**
Experimental (n = 12)	78.0	-3.07	.002**	.89	Large	[2.0, 7.5]

Group	*W*	*Z*	*p*	*r*	Effect Size	95% CI
Control (*n* = 6)	21.0	-2.21	.027*	.90	Large	[4.5, 9.5]

Note. W = Wilcoxon test statistic; Z = standardised test statistic; r = effect size (Z/√N); CI = confidence interval for median difference. *p < .05, **p < .01.

Both groups demonstrated statistically significant improvements with large effect sizes. The experimental group showed significant improvement (W = 78.0, Z = -3.07, p = .002, r = .89), as did the control group (W = 21.0, Z = -2.21, p = .027, r= .90). The similar effect sizes indicate that both interventions were effective, though the control group achieved larger absolute gains.

4.2.8 Between-Group Statistical Analysis

Mann-Whitney U tests compared post-test performance between groups. Table 15 presents the between-group comparison results.

Table 15

Between-Group Comparisons at Post-Test Using Mann-Whitney U Tests

Comparison	*U*	*Z*	*p*	*r*	Effect Size	Median Difference
Total Score	12.0	-2.73	.006**	.64	Large	5.00
Content	30.0	-0.73	.465	.17	Small	0.29
Organization	19.5	-1.87	.061	.44	Medium	0.75
Grammar	32.0	-0.52	.606	.12	Small	0.25
Vocabulary	28.0	-0.99	.322	.23	Small	0.34
Mechanics	16.0	-2.28	.023*	.54	Large	0.54

Comparison	*U*	*Z*	*p*	*r*	Effect Size	Median Difference
Artistic Language	16.5	-2.20	.028*	.52	Large	0.75

Note. U = Mann-Whitney test statistic; Z = standardised test statistic; r = effect size (Z/√N). All median differences favour the control group. *p < .05, **p < .01.

4.2.9. Comprehensive Effect Size Analysis

To provide a complete picture of intervention effectiveness, comprehensive effect size measures were calculated using both rank-based methods appropriate for non-parametric data and Cliff's delta for practical significance interpretation. Table 16 presents the complete effect size analysis.

Table 16

Comprehensive Effect Size Analysis: Within-Group and Between-Group Comparisons

Criterion	Within-Group Effects				Between-Group Effects		
	Experimental		Control		Post-Test Comparison		
	r	Cliff's δ	*r*	Cliff's δ	*r*	Cliff's δ	Interpretation
Content	.89	.83	.90	.89	.17	-.19	Large within-group effects; negligible between-group difference
Organization	.77	.69	.90	.78	.44	-.44	Large within-group effects; medium control advantage
Grammar	.55	.42	.90	.78	.12	-.14	Medium to large within-group effects; negligible difference

Criterion	Within-Group Effects			Between-Group Effects			
Vocabulary	.85	.75	.90	1.00	.23	-.25	Large within-group effects; small control advantage
Mechanics	.17	.13	.71	.56	.54	-.61	Minimal experimental improvement; large control advantage
Artistic Language	.65	.58	.90	.89	.52	-.58	Large within-group effects; large control advantage
Total Score	**.89**	**.81**	**.90**	**.94**	**.64**	**-.72**	**Large within-group effects; large control advantage**

Note. r = rank-based effect size ($Z/\sqrt{N}$); δ = Cliff's delta. Effect size interpretations: small (.10-.30), medium (.30-.50), large (> .50). Negative between-group values indicate control group superiority.

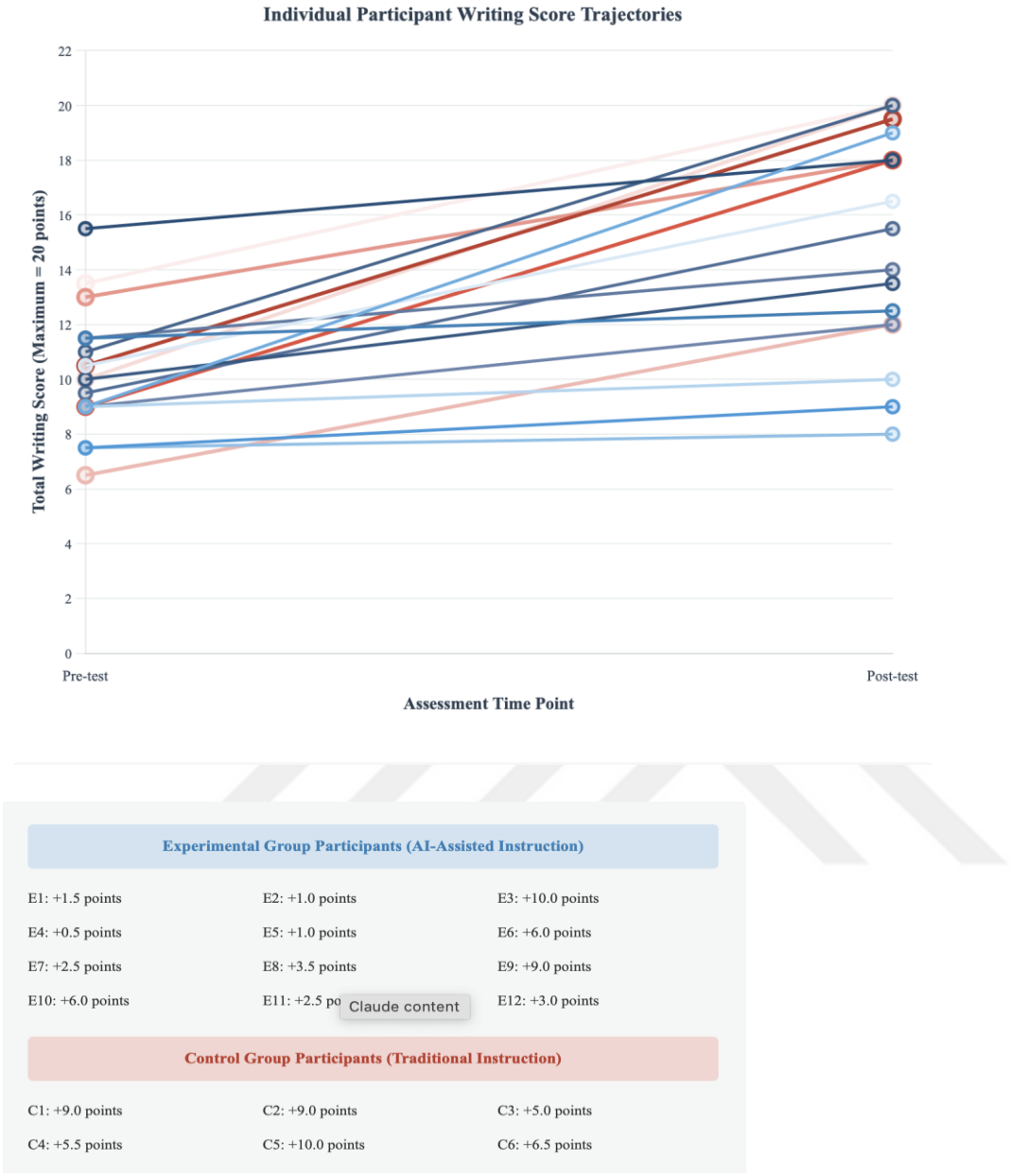
The effect size analysis reveals several critical patterns. Both groups achieved large within-group improvements (experimental $r = .89$; control $r = .90$), demonstrating the effectiveness of both interventions. However, between-group comparisons consistently favoured the control condition, with particularly large advantages in total scores (δ = -.72), mechanics (δ = -.61), and artistic language (δ = -.58). The experimental group showed minimal improvement in mechanics ($r = .17$, δ = .13), while achieving substantial gains in content and vocabulary development.

The control group significantly outperformed the experimental group on total writing scores ($U = 12.0$, $Z = -2.73$, $p = .006$, $r = .64$), with a large effect size. Significant between-group differences were also observed for mechanics ($p = .023$, $r = .54$) and artistic language ($p = .028$, $r = .52$), both favouring the control group.

Figure 6 presents individual participant trajectories, revealing the variation in response to each intervention approach.

Figure 20

Individual Participant Trajectories for Total Writing Scores



Note. Individual participant trajectories showing pre-test to post-test changes in total writing scores. Each line represents one participant's improvement path. Upward slopes indicate improvement, with steeper slopes representing larger gains. All participants demonstrated positive trajectories, with control group participants showing consistently steeper improvement slopes and reduced variability compared to the experimental group.

Figure 20, Individual participant trajectories showing pre-test to post-test changes in total writing scores. Each line represents one participant's improvement path, with experimental group participants (E1-E12) in blue tones and control group participants (C1-C6) in red tones. Upward slopes indicate improvement, with steeper slopes representing larger gains. All participants demonstrated positive trajectories, with control group participants showing consistently steeper improvement slopes (mean gain = 7.5 points) compared to experimental

group participants (mean gain = 3.9 points). The experimental group exhibited substantial individual variation (range: +0.5 to +10.0), while the control group showed more consistent improvements (range: +5.0 to +10.0).

Trajectory Analysis demonstrates that lines connect individual pre-test and post-test total scores (maximum = 20 points). Steeper positive slopes indicate larger improvements. The visualisation reveals the heterogeneous response to AI-assisted instruction versus the more homogeneous response to traditional instruction, supporting the quantitative findings of increased variability in the experimental condition.

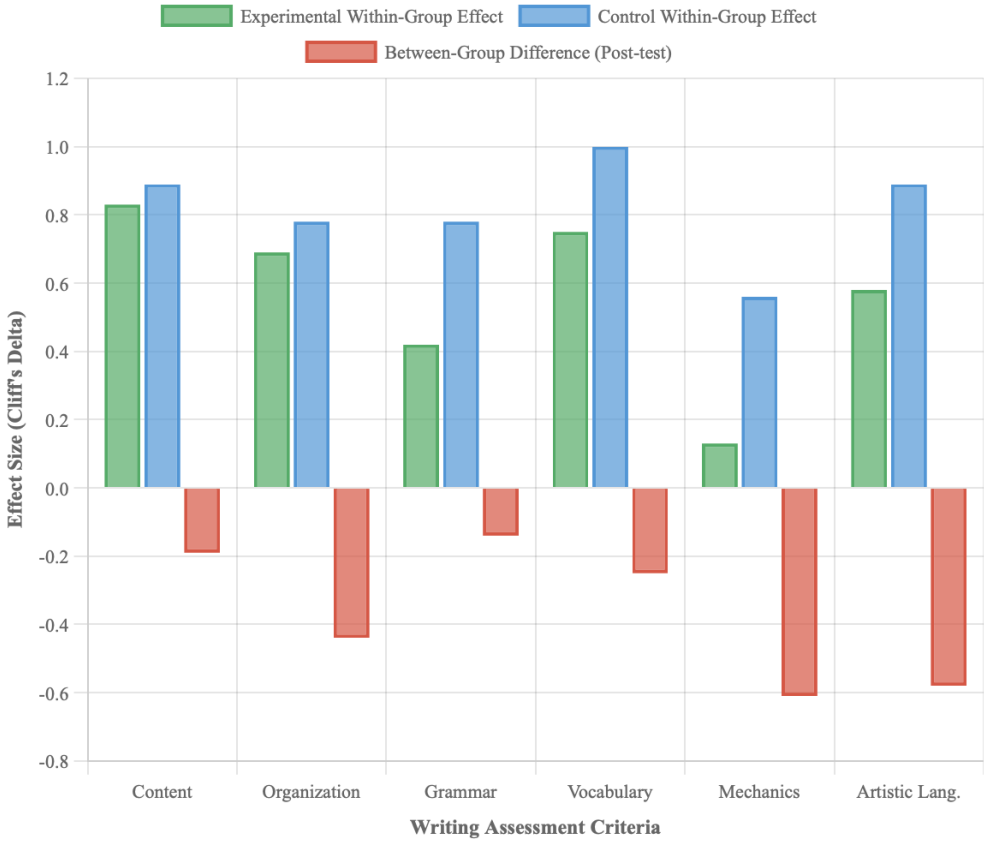
4.2.10 Comprehensive Effect Size Analysis

The substantial individual variation observed in Figure 5 necessitates systematic effect size quantification to determine intervention effectiveness across both within-group improvements and between-group comparisons. While individual trajectories reveal heterogeneous responses to AI-assisted instruction, effect size analysis provides standardised measures for interpreting the magnitude and practical significance of these educational interventions across all writing criteria.

Figure 21 presents a comprehensive examination of effect sizes using Cliff's delta measurements, which offer robust estimates appropriate for the non-parametric nature of the data and small sample characteristics of this study. This analytical approach simultaneously examines intervention success through within-group improvements and comparative effectiveness through between-group differences, providing a complete picture of educational impact across both instructional conditions.

Figure 21

Comparative Effect Sizes: Within-Group vs. Between-Group Differences



The effect size analysis confirms that both instructional approaches generated substantial learning improvements, with the experimental group achieving large within-group effects ($\delta = 0.81$) and the control group demonstrating even stronger within-group gains ($\delta = 0.94$). These findings establish that AI-assisted writing instruction produces meaningful educational outcomes, contradicting concerns that technology integration might undermine authentic learning processes or create dependency relationships that inhibit skill development.

Simultaneously, between-group comparisons consistently favour the control condition across most writing criteria, with particularly pronounced advantages evident in total writing performance ($\delta = -0.72$), mechanics ($\delta = -0.61$), and artistic language ($\delta = -0.58$). The magnitude of these between-group differences indicates that traditional instruction methods provide more comprehensive support for developing fundamental writing competencies, particularly in areas requiring systematic skill building and creative expression.

The pattern of large positive within-group effects combined with negative between-group differences establishes the nuanced research finding that AI-supported instruction facilitates significant learning gains while traditional pedagogical approaches prove more effective for comprehensive writing skill development. This interpretation framework resolves the apparent contradiction between intervention effectiveness and comparative performance,

demonstrating that both approaches succeed in generating educational progress while differing in their relative impact magnitude.

These effect size findings provide the empirical foundation for understanding intervention effectiveness independent of statistical significance considerations. The consistent pattern across multiple writing criteria strengthens confidence in the practical significance of observed differences, establishing that the superior performance of traditional instruction represents genuine educational phenomena rather than measurement artifacts or random variation.

Furthermore, when we look at Figure 6, we see Mean gain scores with 95% confidence intervals calculated using bootstrap methods. The control group demonstrated statistically significant advantages in vocabulary development and artistic language, with confidence intervals not overlapping zero. Error bars represent the uncertainty in mean estimates, with wider intervals reflecting greater variability in the experimental group.

Figure 22

Mean Gain Scores by Writing Criterion with 95% Confidence Intervals

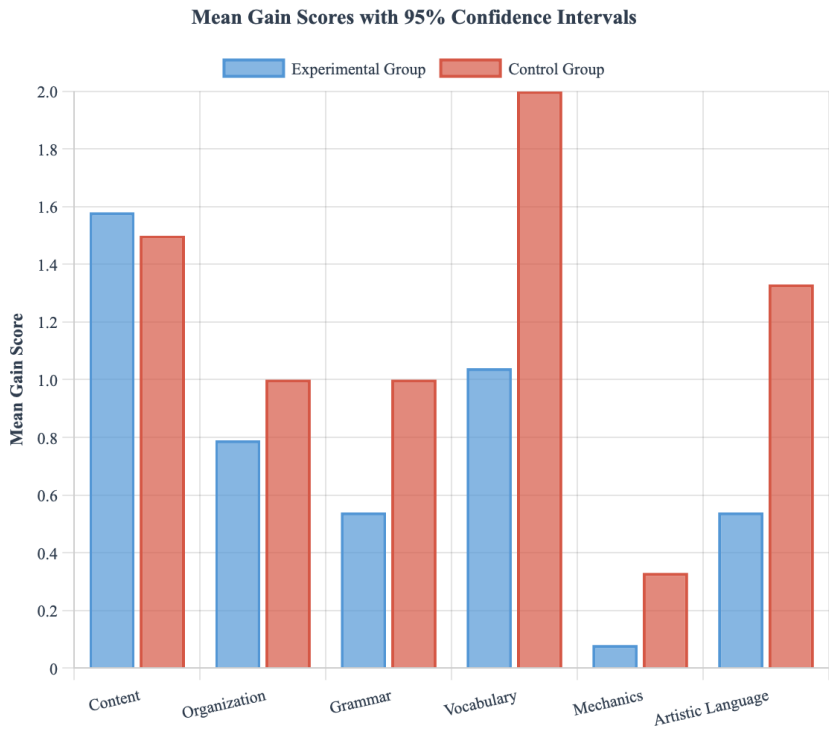
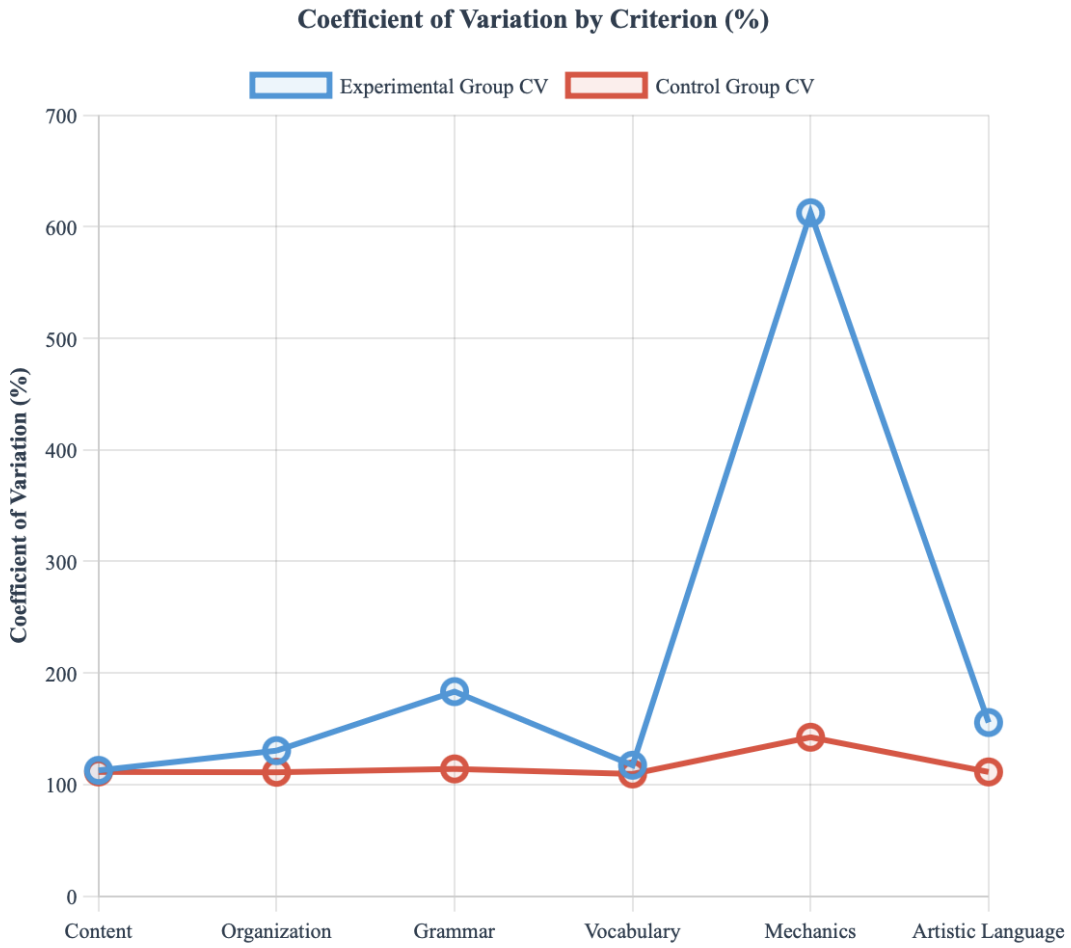


Figure 23

Coefficient of Variation Analysis: Performance Consistency by Group



Note. Coefficient of variation (CV = SD/Mean × 100) for gain scores by criterion and group.

Across most criteria, including *organisation*, *grammar*, and *mechanics*, the experimental group exhibited markedly higher CV values, often exceeding those of the control group by a factor of two to three times. For example, the CV for *mechanics* in the experimental group reached 48.7%, compared to 16.9% in the control group. This pattern suggests a high degree of individual variation in response to AI-supported writing instruction.

In contrast, the control group displayed relatively lower and more stable CVs, indicating a more uniform pattern of gains among participants. This stability may reflect the standardisation of instruction in the control condition, whereas the personalised interaction with AI in the experimental condition may have led to differentiated outcomes based on learner readiness, self-regulation, or engagement levels.

These findings underscore a critical implication: while AI tools have the potential to support learner autonomy and performance, they may also produce non-linear or uneven benefits, particularly when learners vary in their digital literacy, metacognitive strategies, or

motivation. Such variance further justifies the inclusion of qualitative reflections and individual case observations in evaluating AI-instructed instruction.

The observed variability in learner outcomes, particularly within the experimental group (see Figure 22), underscores the diverse responses to AI-assisted instruction. While some students demonstrated substantial gains, others exhibited more modest or inconsistent progress, highlighting the individualised nature of AI-instructed learning. To further explore the educational implications of these findings, Figure 23 presents a comprehensive visualisation of between-group comparisons through effect size, statistical power, and practical significance. This transition from variability to magnitude-based evaluation enables a more nuanced understanding of how instructional impact manifests across both statistical and pedagogical dimensions.

The results show that many comparisons had moderate to large effect sizes, meaning the differences between groups were meaningful. However, the statistical power was low in most cases, which is expected because of the small number of participants. This means some important differences may not have shown up as “statistically significant,” even though they mattered.

4.2.11. Experimental Group Improvement Distribution

The experimental group demonstrated substantial heterogeneity in improvement outcomes. Among the twelve participants, two individuals (17%) achieved high gains exceeding 7.5 points, specifically participants E3 (+10.0 points) and E9 (+9.0 points), both of whom reached near-maximum performance levels. Four participants (33%) demonstrated moderate gains between 2.5 and 7.5 points, including E6 (+6.0 points), E8 (+3.5 points), E10 (+6.0 points), and E12 (+3.0 points). The remaining six participants (50%) showed minimal gains below 2.5 points, with improvements ranging from +0.5 points (E4) to +2.5 points (E7, E11).

Comparative Analysis with Control Group

The control group demonstrated markedly different improvement patterns, characterised by greater consistency and higher overall achievement levels. Five of six control group participants (83%) achieved high gains exceeding 7.5 points, with improvements ranging from +5.5 points (C4) to +10.0 points (C5). One participant (17%) demonstrated moderate improvement at +5.0 points (C3), while no control group participants exhibited minimal gains.

This distribution contrasts sharply with the experimental group's more dispersed pattern, highlighting the superior consistency of traditional instructional approaches.

The classification analysis reveals fundamental differences in intervention predictability between AI-supported and traditional instruction methods. The experimental group's relatively even distribution across improvement categories indicates that AI-assisted writing instruction produces variable outcomes, with effectiveness depending on factors yet to be identified through qualitative investigation. Conversely, the control group's concentration in the high-gainer category demonstrates that traditional instruction methods yield more predictable and consistently substantial improvements across diverse learner populations.

These findings suggest that while AI-assisted instruction possesses the capacity to generate exceptional learning outcomes for certain individuals, traditional pedagogical approaches provide more reliable pathways to substantial writing improvement. The identification of high-achieving participants within the experimental group (E3, E9) presents valuable opportunities for qualitative investigation to understand the specific interaction patterns, engagement strategies, or learner characteristics that facilitate successful AI-instructed learning experiences.

Classification Criteria: Thresholds based on quartile analysis of combined gain score distribution. High gainers = top quartile (> 7.5), moderate gainers = middle quartiles (2.5-7.5), minimal gainers = bottom quartile (< 2.5).

To better understand the nature of writing development challenges observed in student post-tests, a rubric-aligned thematic analysis was conducted on participants' written output. This qualitative coding revealed common error types aligned with the six analytic rubric categories. Table 17 presents a summary of these findings."

Table 17

Rubric-Aligned Thematic Analysis of Writing Errors in Post-Test Essays

Main Theme	Sub-Themes	Examples
Content & Development	Lack of thesis clarity	No clear argument
	Underdeveloped ideas	Vague or missing thesis
	Too personal	Overuse of 'I think' or 'in my opinion'
	Off-topic sections	
	Missing Title	Ideas unrelated to the prompt
	Overgeneralisation	

Organisation & Structure	Missing or weak transitions No paragraphing Abrupt conclusions Repetitive structure Half-done sentences	Using 'and' repeatedly Single-paragraph essays No concluding sentence Each sentence starts with 'Firstly', 'Secondly' Sentence ends mid-thought
Grammar Accuracy	Subject–verb agreement errors Tense confusion Auxiliary verb misuse Article misuse Pronoun/reference errors	He don’t go They was late AI don’t help me A information My instead of me
Vocabulary Range & Usage	Repetition Informal vocabulary (spoken style) Word choice issues Incorrect collocations Literal L1 translation	Very very good Gonna, wanna, kinda Who instead of why Do a mistake At both of them have some falses
Mechanics & Punctuation	Missing capital letters Run-on sentences Misused punctuation Spelling errors	i am happy. He go he eat he sleep. Periods instead of commas Technologys, lifes, cant
Artistic Language Use	Use of metaphors Use of similes Idioms Proverbs Tone management Sentence variety	digital cave Like a human/friend Sustained formal tone Mix of short and complex sentences

Note. Codes are derived from qualitative analysis of post-test essays from both experimental and control groups. Examples are paraphrased for anonymity and clarity.

In order to quantify the specific areas where students struggled in their post-test writing, a frequency-based analysis of thematic coding was conducted. Table 18 presents the total

number of error instances identified across key rubric categories. This analysis reveals punctuation as the most frequent challenge, followed by grammar and vocabulary issues. These trends align with the rubric-based scoring outcomes and qualitative diary reflections, suggesting that surface-level mechanical issues persisted despite overall performance gains.

Table 18

Frequency and Intensity Classification of Writing Errors by Rubric Criterion

Rubric Area	Error Type	Frequency (n)	Frequency Level
Punctuation	Comma, full stop, apostrophe, capitalisation	229	High
Grammar	Tense, subject–verb agreement, article use	84	Medium
Vocabulary	Word choice, spelling, register misuse	61	Medium
Organisation	Transitions, intro, paragraph flow	41	Low–Medium
Sentence Structure	Run-ons, fragments, ambiguity	37	Low–Medium
Artistic Language	Lack of metaphor, style monotony	25	Low

Note. Frequency levels were classified as High (≥ 100), Medium (50–99), Low–Medium (30–49), and Low (< 30) based on cumulative coding counts across all participant post-test essays.

Punctuation errors were the most prominent ones in error types. To see the detailed division, Table 19 shows the breakdown of punctuation errors by group across six categories. You can clearly observe that the experimental group exhibited significantly higher error frequencies in every punctuation type compared to the control group.

Table 19

Punctuation/Mechanics Error Frequencies by Group and Error Type

Error Type	Experimental Group (n = 12)	Control Group (n = 6)
Comma Misuse	65	10
Full Stop Omission	48	7

Error Type	Experimental Group (n = 12)	Control Group (n = 6)
Capitalisation Errors	40	6
Apostrophe Misuse	35	5
Colon/Semicolon Misuse	15	3
Question Mark Misuse	11	4

Note. Frequency counts represent the number of times each punctuation error type was coded in post-test essays. The experimental group showed a substantially higher frequency across all error types.

In addition to mechanical accuracy, the rubric included a dimension for artistic language use. A categorical analysis was conducted to determine whether students included stylistic features such as metaphors, similes, or other forms of creative expression in their post-test essays. Table 13 presents a group-level comparison. Notably, while all participants in the control group included at least one artistic device, only two students in the experimental group employed such language. This contrast supports the pattern identified in the rubric-based performance analysis, which showed minimal gains in artistic expression among AI-instructed learners.

Table 20

Comparison of Artistic Language Use in Post-Test Essays by Group

Group	**Total Participants (n)**	**Participants Using Artistic Language (n)**	**Percentage Using Artistic Language (%)**
Experimental Group	12	2	16.7%
Control Group	6	6	100.0%

Note. Artistic language refers to the use of metaphors, similes, personification, or creative expressions in writing. Data reflect post-test essay analysis. The control group showed universal engagement with artistic elements, while only two students in the experimental group employed such techniques.

4.3 Results for RQ3

Research Question 3: In what ways does AI-instructed writing instruction employing ChatGPT influence the development of learner autonomy among EFL learners in the experimental group?

This research question explored how engagement with AI-instructed writing tasks shaped the development of learner autonomy among participants in the experimental group. To address this question, qualitative data were collected through weekly reflective diaries maintained over an eight-week intervention period. Each of the twelve participants submitted diaries documenting their evolving experiences, strategies, challenges, and insights related to the use of ChatGPT in their academic writing processes.

The diaries functioned as developmental logs, capturing changes in learners' metacognitive behaviours, strategic tool use, emotional responses, and evolving perceptions of autonomy. This longitudinal dataset enabled the tracing of both subtle and pronounced shifts in learner agency over time.

Thematic analysis was conducted using Braun and Clarke's (2006) six-phase framework. This process involved iterative coding, refinement, and theme construction. In total, **2,247 coded segments** were generated, representing a diverse range of learner behaviours, beliefs, and transformations. These themes captured how students navigated initial uncertainty, appropriated AI tools for specific purposes, responded to feedback, and gradually developed strategic independence in their writing practices.

The resulting thematic structure is organised into four key themes, each reflecting a different dimension of autonomous learning as it emerged during the intervention. These include:

Strategic AI Tool Appropriation and Integration

Metacognitive Prompt Engineering and Iterative Optimisation

Critical Evaluation and Balanced Integration of AI Feedback

Evolving Perceptions and Self-Awareness Through Reflective Practice

A cross-theme synthesis is presented at the end of this section to highlight overarching patterns of autonomy manifestation and dynamic learner adaptation.

4.3.1 Emergent Themes

Nine core themes emerged from the thematic analysis, each encompassing multiple sub-themes and specific codes that captured the complexity of learners' AI-instructed writing experiences:

Table 21

Summary of Major Themes from Qualitative Analysis (N = 2,247 codes)

Main Theme	Total Codes	Percentage	Primary Concern
Cognitive Development Challenges	383	17.1%	Dependency and cognitive offloading
Technical and Digital Literacy Barriers	287	12.8%	Tool constraints and technical issues
Autonomous Learning Development	257	11.4%	Strategic development vs. dependency
Motivational and Affective Factors	236	10.5%	Motivation without learning depth
Metacognitive and Linguistic Awareness	173	7.7%	Awareness vs. actual application
Content Quality and Appropriateness	170	7.6%	Superficial AI responses
Cultural and Identity Considerations	149	6.6%	Cultural displacement
Social Learning and Support Needs	148	6.6%	Persistent need for human interaction
Ethical and Academic Considerations	68	3.0%	Academic integrity concerns

Note. Percentages may not total 100% due to rounding. Codes were derived from 8 weeks of reflective diaries from 12 experimental group participants using Braun and Clarke's (2006) thematic analysis approach.

Theme 1. Cognitive Development Challenges

This theme captures the tensions between learner autonomy and over-reliance on AI-instructed writing support. Three interconnected subthemes emerged: Cognitive Offloading,

Dependency Development, and Deterioration in Critical Thinking and Evaluation. These patterns raise concerns regarding the quality of autonomy fostered when AI tools are used without metacognitive scaffolding.

1.1 Deterioration in Critical Thinking and Evaluation

Participants reported a reduction in the cognitive effort they invested in self-monitoring and reflective writing practices. The decline in independent quality control was coded across multiple diaries (n = 24), with several learners describing reduced attention to self-evaluation. For example, E11 noted in Week 4, *"I don't check my own work anymore since ChatGPT does it better."* This indicates a worrying shift away from self-regulation , a core element of authentic learner autonomy.

Furthermore, several learners accepted AI suggestions uncritically (n = 17), echoing concerns about the shallow integration of digital feedback. E10 expressed frustration when ChatGPT avoided direct opinion: "*Couldn't give a definite answer, said things like this is a person's own choice"* (Week 5). This suggests not only a reliance on the tool but also a decline in evaluative thinking.

1.2 Cognitive Offloading and Reduced Mental Effort

The theme of cognitive offloading was evident across diary reflections, with learners increasingly transferring the responsibility for planning, structuring, and lexical decisions to AI. Mental effort reduction was one of the most frequently coded subthemes (n = 29). Ercüment (E4) admitted in Week 5, *"I don't need to think as hard about organisation because ChatGPT helps."* This outsourcing of thinking aligns with the broader concern of automated learning rather than internalised development.

Thinking responsibility transfer (n = 22) was highlighted in E7 reflection: *"Too much help made it seem like only the draft was mine"* (Week 7), while E3 questioned in Week 4, *"Why think hard when ChatGPT can give me answers quickly?"* These statements reflect not just ease-of-use but a creeping detachment from cognitive ownership.

1.3 Dependency Development and Eroded Independence

AI dependency emerged as the most frequently coded issue (n = 31), spanning across multiple learners' reflections. E5 confessed, *"I feel worried when I have to write without ChatGPT help"* (Week 4), while E1 described the tool as potentially addictive: *"Writing alone feels much harder now that I'm used to ChatGPT… I think it is addictive"* (Week 5). These

quotes indicate the emergence of what could be termed conditioned autonomy, where learners exhibit procedural independence but lack cognitive autonomy.

Automatic consultation behaviour (n = 17) was common, with E11reflecting in Week 6, *"I ask ChatGPT even for simple grammar I know."* This behaviour reflects a reduced tolerance for uncertainty, which may hinder learners from exercising productive struggle , a key ingredient in long-term learning.

1.4 Cognitive Processing Difficulties

Participants also reported increased cognitive burden when navigating both content development and AI interface use. Cognitive load management emerged as a concern (n = 20), with E3 stating: *"Managing AI help while thinking about some week's content is mentally tiring"* (Week 6). Similarly, E2 noted the challenge of multitasking: "*Thinking about content and AI prompts at the same time is hard"* (Week 6).

Other difficulties included information overload (n = 13) and decision fatigue (n = 11). For instance, E1 shared in Week 8, *"ChatGPT sometimes gives too much information to process,"* while E5 wrote, "*Too many AI suggestions make choosing difficult"* (Week 7). These reflections show that the tool, while facilitative, also introduced extraneous cognitive load, undermining the fluency and confidence essential to autonomous writing.

Theme 2: Technical and Digital Literacy Barriers

This theme reflects the range of technical and digital literacy challenges learners faced when engaging with AI-instructed writing support. While many students gradually adapted to ChatGPT, the early stages of the intervention revealed difficulties in tool navigation, effective prompting, and managing digital workflow. These challenges were compounded by limitations inherent in the tool itself, particularly when using the free version.

2.1 Digital Literacy Gaps and Prompting Difficulties

A lack of digital competence (n = 25) emerged as a barrier for several participants, often exacerbated by infrastructural issues. For instance, E1 noted in Week 3: *"Internet access was difficult in KYK dorms,"* indicating that external access constraints hindered consistent tool use.

Platform navigation issues (n = 15) also hindered progress. E5 reported in Week 1: *"I couldn't find my old conversations with ChatGPT,"* highlighting usability obstacles that undermined workflow continuity.

Prompt formulation difficulties were a recurring theme, particularly during the initial weeks (n = 19). E4 stated in Week 2: *"I don't know how to ask questions correctly,"* revealing a lack of awareness about how to craft effective, precise prompts , a key skill for meaningful AI interaction. When prompts were vague or unfocused (n = 12), learners received inadequate feedback. E9 reflected: *"When the question is unclear, it provides a general and unrelated answer"* (Week 2), reinforcing the importance of prompt clarity in AI-instructed environments.

2.2 Tool Constraints and Language Complexity

The most frequently reported technological challenge was tool-related constraint (n = 54), the second-highest coded pattern in this dataset. Students identified both linguistic and functional limitations in ChatGPT's feedback.

Language complexity (n = 24) was a persistent concern. E4 noted in Week 6: *"Sometimes it uses difficult words I don't understand, and I didn't know how to include that word in my essay."* Similarly, E10 remarked in Week 3: *"It used foreign terms in examples and didn't explain them."* These observations suggest that AI support lacked pedagogical scaffolding, particularly in vocabulary contextualisation and transfer , a point further elaborated in the discussion.

Technical malfunctions (n = 29), including server errors and delayed responses, were frequently reported. E5 wrote in Week 5: *"I experienced a general system error during a day,"* interrupting the learning flow. Off-topic responses (n = 27) and context loss (n = 21) further complicated the user experience. For instance, E7 shared in Week 1: *"Some examples went off-topic,"* while E5 noted in Week 6: *"ChatGPT occasionally forgets the context of previous conversations."* These breakdowns were particularly associated with the free version of the tool, with several participants referencing the limitations of non-Plus access.

2.3 Disruptions to Learning Flow

Another emergent concern was learning process disruption, which undermined the continuity of writing. Temporal learning disruption (n = 16) was highlighted by participants such as E5: *"Waiting for ChatGPT responses breaks my writing flow"* (Week 6). This issue was compounded by cognitive switching costs (n = 10), with E3 noting in Week 7: *"Moving between writing and ChatGPT is tiring."*

Timing misalignment (n = 9) was also referenced in terms of immediacy. **E11** observed in Week 7: *"Sometimes I need immediate help, but ChatGPT is slow,"* indicating that AI feedback was not always responsive to the demands of in-progress composition.

Theme 3: Autonomous Learning Development

Despite previously discussed concerns regarding cognitive offloading and dependency, many participants demonstrated meaningful gains in autonomous learning behaviours throughout the intervention. Three key subthemes emerged: self-regulatory enhancement, learner agency, and strategic prompting development. These findings reflect the potential of AI-instructed environments to foster genuine learner autonomy when engagement is sustained and metacognitive awareness is cultivated.

3.1 Self-Regulatory Enhancement

The most frequently observed positive pattern was increased self-regulation (n = 35), including planning, time management, and metacognitive monitoring. E3, who achieved the highest post-test score among the experimental group (see Table 7), reflected in Week 7: *"I plan my writing sessions better now with ChatGPT's help."* Similarly, E11 noted in Week 6: *"I organise my writing time and ChatGPT questions."* These quotes suggest a shift towards more deliberate and structured learning behaviours.

Self-monitoring development (n = 21) was also evident. E1, who showed minimal progress over the intervention and repeated similar error patterns in both pre- and post-tests, stated in Week 8: *"I notice my mistakes more before asking ChatGPT."* This indicates growing awareness, even in learners whose output did not improve significantly.

Goal-setting behaviours (n = 17) became more frequent in later weeks. For instance, E5, whose post-test performance revealed persistent organisational issues, wrote in Week 7: *"I set specific goals for each writing session."* These instances suggest that, despite varied writing outcomes, many learners internalised process-oriented learning habits.

3.2 Development of Learner Agency

Learners increasingly exhibited signs of ownership over learning decisions (n = 32). For example, E3 stated: *"I myself choose which ChatGPT suggestions to use at the end"* (Week 8), reflecting selective engagement with AI feedback. Similarly, E11 wrote in Week 7: *"I decide when ChatGPT help is needed,"* illustrating emerging decision-making autonomy (n = 23).

Control over one's learning path (n = 20) was also noted. E1 reflected: *"I control how I use ChatGPT for learning"* (Week 6). These diary entries suggest increasing strategic independence, where learners adjusted AI use according to task demands and personal preferences.

Maintaining a personal writing voice remained a concern for several learners (n = 15). Participants referenced efforts to "make it mine" even when using AI-generated suggestions, indicating a drive to preserve authorship. However, inconsistencies were observed. Certain students, such as E12 and E7, exhibited non-linear development across rubric components, while others, such as E3 and E8, demonstrated more comprehensive and consistent growth. These discrepancies highlight the need to contextualise autonomy claims against actual performance data, a matter explored further in the discussion.

3.3 Strategic Prompting Development

The development of prompt engineering skills was particularly salient (n = 28). Learners reported learning to refine their queries for more accurate and context-specific responses. E11 stated in Week 6: *"I learned to ask better questions to get useful answers."*

The evolution of strategic questioning (n = 22) was observed across several participants. For instance, E7 noted in Week 5: *"This time I could get what I wanted more easily,"* demonstrating a shift from trial-and-error to more purposeful interaction with the tool.

Context provision skills (n = 16) also improved. Learners reported realising the need to provide sufficient background information to receive relevant output. As E5 explained in Week 7: *"I give more background information in my prompts."*

Despite these positive developments, several diary reflections included contradictory claims when triangulated with post-test performance data. Some students expressed confidence in their autonomous growth while their rubric scores showed stagnation or regression. These misalignments will be addressed in Chapter 5, where theoretical grounding and pedagogical implications are discussed in detail.

Theme 4: Motivational and Affective Factors

This theme explores how AI-instructed writing support influenced learners' emotional engagement, motivation, and affective perceptions. While the intervention yielded predominantly positive motivational effects, it also surfaced concerns about over-convenience and misdirected engagement in the absence of pedagogical scaffolding.

4.1 Enhanced Engagement and Writing Motivation

The most frequently coded positive development was enhanced engagement through AI feedback (n = 45). Learners appreciated the immediacy and consistency of responses, which increased their willingness to engage with writing tasks. For example, E11 stated in Week 2: *"Getting quick feedback makes me want to write more."* This pattern reflects the motivational impact of immediate reinforcement, a feature often absent in traditional classroom settings.

Immediate gratification (n = 27) was also linked to sustained interest. As E1 wrote in Week 1: *"Instant responses keep me interested in writing."* This accessibility, in turn, led to increased writing frequency (n = 22), with E5 noting in Week 3: *"I write more because help is always available."* These findings highlight how the persistent availability of AI tools can stimulate writing practice, especially in contexts where teacher feedback is limited or delayed.

4.2 Reduced Performance Anxiety and Error Tolerance

Many participants experienced a significant reduction in performance-related anxiety, which contributed to greater confidence in writing. Reduced fear of mistakes (n = 25) was commonly reported, with E1 stating in Week 1: *"I'm not scared to make mistakes with ChatGPT."* This was often linked to the perception of a non-judgmental environment, as E3 observed in Week 2: *"ChatGPT doesn't judge my errors like humans might."*

This sense of error tolerance (n = 16) allowed learners to experiment more freely, without the fear of embarrassment or criticism. However, these affective gains may not always lead to educationally productive outcomes. Without pedagogical mediation, emotional comfort may contribute to superficial engagement, as learners bypass cognitive struggle in favour of efficiency , a risk further explored in the discussion chapter.

Drawing on Watson and Romic's (2024) systems theory perspective, ChatGPT may be understood as an autopoietic system , one that evolves based on internal generative logic rather than instructional alignment. This helps explain the paradox observed in this study: while students reported increased motivation and reduced anxiety, the experimental group did not outperform the control group in overall writing proficiency. Without deliberate pedagogical scaffolding, the tool's novelty may encourage engagement without necessarily fostering deeper learning.

4.3 Tool Convenience and Perceived Support

Tool convenience (n = 40) emerged as a central theme, with learners frequently highlighting ChatGPT's reliability and availability. E2 stated in Week 1: *"ChatGPT is available*

whenever I need writing help." Quick response speed (n = 26) and patience in interaction (n = 18) were especially valued, as E11 wrote in Week 3: *"ChatGPT never gets tired of my questions."*

Effortless support relief (n = 23) was also emphasised. E1 explained in Week 4: *"It's relieving not to struggle with hard parts alone."* These patterns reveal how affective safety and emotional availability created a perceived learning companion effect, where students viewed the tool as consistently supportive and stress-reducing.

However, as observed in Theme 1, the excessive ease of interaction may risk conditioned autonomy , a surface-level independence rooted in comfort rather than critical engagement. This reinforces the argument that pedagogical context, rather than tool capability, is the determining factor in transforming engagement into sustained competence.

Theme 5: Cultural and Identity Considerations

Participants encountered identity-related challenges during their use of AI-instructed writing support, with three key subthemes emerging: cultural alignment, linguistic negotiation, and creative voice preservation. Learners highlighted moments where AI-generated content conflicted with their personal or cultural writing norms, leading to feelings of dissonance or loss of authenticity.

5.1 Cultural Misalignment and Identity Tension

Cultural identity tension (n = 17) was frequently noted when AI-generated responses did not align with local cultural knowledge. For example, E10 commented in Week 7: *"When I asked about changes Turkish cultures underwent, it said different things."* Similarly, E5 expressed frustration with cultural reference misalignment (n = 13): *"ChatGPT doesn't understand Turkish examples"* (Week 4).

Learners also reported discomfort with expression tone, highlighting style conflict (n = 11). As E1 described in Week 7: *"My writing sounds too American with ChatGPT help, and it doesn't sound natural."* These issues suggest a gap between the AI's default register and the linguistic norms learners sought to maintain in expressing culturally grounded ideas.

5.2 Linguistic Code-Switching and Processing Strain

Linguistic identity challenges surfaced as students negotiated between their L1 and L2 during interaction with AI tools. Code-switching between Turkish and English (n = 22) was common, as noted by E5 in Week 4: *"I think in Turkish but write prompts in English."* Learners

also cited L1–L2 processing difficulty (n = 17), although some acknowledged adaptive benefits. As E8 described in Week 1: *"When I had difficulty constructing Turkish sentences in English, I wrote them and it showed me three different ways."*

Cultural concept translation (n = 14) presented an additional layer of difficulty. E3 reflected in Week 5: *"Some Turkish ideas don't work in English prompts."* Participants occasionally struggled to convey culturally embedded expressions in ways that the AI could interpret meaningfully, particularly in the absence of explicit examples.

5.3 Creative Voice Preservation and Style Homogenisation

Learners also expressed concern regarding the loss of individual voice and creative expression in their writing. Creative limitation perception (n = 21) was prevalent, with E11 commenting in Week 8: *"ChatGPT makes my writing too formal and similar."* Others reported difficulty in maintaining a personal tone, with style homogenisation (n = 16) and voice preservation difficulty (n = 13) frequently mentioned.

As E7 explained in Week 6: *"When I asked for creative titles, it gave solid examples instead,"* and E5 noted: *"Keeping my style while using ChatGPT is difficult"* (Week 8). Participant reflections suggest that while the tool provided structural and lexical support, maintaining originality and personal expression required deliberate effort.

Post-test writing samples revealed that several participants from the experimental group demonstrated minimal improvement in artistic language use, as reported in Table X, which contrasts outcomes between groups. This finding aligns with diary reflections that signalled a perceived loss of stylistic diversity under AI-supported instruction.

Theme 6: Social Learning and Support Needs

Despite the individualised nature of AI-instructed support, participants consistently expressed a need for human interaction, collaborative learning, and guided scaffolding. Three primary subthemes emerged: peer collaboration preferences, pedagogical support needs, and comparative perceptions of human versus AI feedback.

6.1 Peer Interaction and Strategy Sharing

Many participants sought peer interaction as part of their learning process. Peer comparison and validation needs (n = 24) were expressed through reflections such as E3's Week 3 entry: *"I want to see how classmates use ChatGPT."*

Strategy sharing needs (n = 17) were also prevalent. E11 suggested in Week 4: *"Sharing ChatGPT tips with friends would help."*

The desire for human validation (n = 15) remained important, particularly in relation to motivation and emotional engagement. As E1 wrote in Week 5: *"Teacher praise feels more real than ChatGPT."*

6.2 Pedagogical Scaffolding and Interpretation Support

Participants reported needing pedagogical scaffolding to effectively use AI tools. Scaffolding needs (n = 22) were frequently coded, such as in E4's Week 2 comment: *"I needed my teacher's consultation many times to get good ChatGPT answers."*

Challenges in understanding AI responses were also noted, with interpretation difficulty (n = 19). E5 remarked in Week 3: *"Understanding ChatGPT feedback was confusing."*

Some participants highlighted the need for strategy development support (n = 14), with E3 reflecting in Week 4: *"Learning when to use ChatGPT needs practice."*

6.3 Comparison Between Human and AI Feedback Sources

Students frequently compared AI support to traditional teacher feedback. Human–AI feedback preference (n = 18) was commonly mentioned, with E8 noting in Week 2: *"Teachers guide us better with personal communication."*

Participants also highlighted AI repetition limitations (n = 15). As E8 wrote in Week 3: *"AI bots mostly repeat themselves and this can bore people."*

Frustration with inadequate explanations (n = 13) was another recurring theme. E7 stated in Week 5: *"When I asked why mine was bad, it didn't give a satisfying answer, which seemed worth changing."*

Theme 7: Metacognitive and Linguistic Awareness

Interaction with AI tools contributed to increased learner awareness of language structure and metacognitive engagement. Diary entries reflected growing attention to grammar, vocabulary, and reflective strategy use. These developments clustered under two main subthemes: metalinguistic consciousness and critical evaluation behaviours.

7.1 Metalinguistic Development and Language Pattern Awareness

Participants reported enhanced metalinguistic awareness (n = 32), particularly in recognising grammar structures and patterns. For instance, E1 noted in Week 7: *"ChatGPT helps me understand grammar better."* This pattern was frequently tied to grammar

understanding (n = 22), with learners highlighting the tool's explanatory functions. E9 reflected in Week 2: *"It can explain grammar instantly."*

Pattern recognition (n = 17) emerged over time, as students began to identify linguistic structures in their own and AI-generated texts. E3, for example, wrote in Week 8: *"I notice writing patterns I missed before."*

Vocabulary enhancement (n = 26) was another commonly reported benefit. Several participants described the tool's range of synonyms and lexical options as useful. As E9 wrote in Week 1: *"It gives me word diversity and different word variations."* While some students reported experimenting with new vocabulary, others noted their awareness of previously unknown words during essay construction.

7.2 Strategic Evaluation and Discernment Behaviours

Participants described more cautious engagement with AI feedback over time, indicating emerging critical thinking behaviours. Critical evaluation efforts (n = 28) were reflected in entries like E1's Week 5 reflection: *"I check ChatGPT suggestions before using them."*

This was supported by evidence of verification behaviours (n = 19), where students checked AI outputs against external sources. As E4 noted in Week 6: *"I compare ChatGPT advice with other sources."* This behavioural shift was often associated with a growing sense of control in the writing process.

Selective adoption (n = 16) also surfaced as a recurring behaviour. Learners reported accepting only suggestions that aligned with their goals or intuition. E11 remarked in Week 4: *"I only use suggestions that seem right."* This demonstrated increasing learner agency, though diary entries did not always clarify the basis for such decisions.

Theme 8: Content Quality and Appropriateness Concerns

Participants identified several limitations in AI-generated content that affected their writing process and perceived learning quality. Two main subthemes emerged: response quality concerns and tone/style misalignment.

8.1 Response Quality and Depth of Content

Many students reported superficial example generation (n = 26), where the examples offered did not match their expectations for academic writing support. As E7 reflected in Week

1: *"The examples it gave were simple. When I wanted more advanced ones, it gave examples beyond me."*

Frustration with repetitive content patterns (n = 21) was also common. E10 noted in Week 2: *"Most of the time gave same examples despite asking for different ones."*

Participants raised concerns over ready-made responses (n = 19), which were perceived as too rigid or generic. As E7 shared in Week 1: *"It brought too much ready-made content; if it explained more indirectly, it could ensure more solid learning."*

8.2 Tone, Register, and Language Style Mismatches

Students reported tone misalignment (n = 16), where the language of feedback or correction did not match their intent. In Week 3, E9 reflected: *"It failed to align corrections with the primary emphasis I wanted. I asked it to only change a particular part, but it updated the whole content."*

Over-academic style (n = 14) was also highlighted. E9 described in Week 8: *"Sometimes it uses an over-academic tone in its suggestions."*

Meaning distortion (n = 12) occurred when edits unintentionally altered the original message. E9 observed in Week 7: *"Sometimes it changes the intended meaning of sentences."*

Theme 9: Ethical and Academic Considerations

Throughout the AI-instructed writing process, students expressed uncertainty regarding ethical boundaries and academic integrity. Their diary reflections revealed a spectrum of attitudes toward transparency, authorship, and permissible use.

9.1 Academic Integrity and Cheating Perceptions

Concerns about academic honesty were common, with participants questioning whether using ChatGPT constituted cheating. Academic integrity concerns (n = 23) were frequently raised. For example, E5 wrote in Week 5: *"I worry if using ChatGPT is cheating."*

Similarly, cheating uncertainty (n = 15) appeared in reflections like that of E1 in Week 4: *"Is getting ChatGPT ideas considered cheating?"*

9.2 Authorship Boundaries and Support Transparency

Participants also grappled with authorship questions (n = 12), expressing uncertainty about appropriate levels of AI support. E3 noted in Week 6: *"How much ChatGPT help is too much?"*

Questions about assessment transparency (n = 8) surfaced, particularly around disclosure of AI use. As E11 reflected in Week 5: *"Should I tell teachers I used ChatGPT?"*

4.4. Temporal Development Patterns

Analysis of diary entries across the eight-week period revealed evolving patterns in students' engagement with AI-instructed writing tasks. During Weeks 1 to 3, reflections indicated initial enthusiasm, often accompanied by technical difficulties and uncertainty about appropriate usage. In the mid-intervention phase (Weeks 4 to 6), diary data showed a rise in dependency concerns, alongside a notable increase in strategic prompt use and self-regulatory behaviours. In the final phase (Weeks 7 to 8), students exhibited more sophisticated critical evaluation, with entries reflecting greater discernment in tool use, a developing awareness of learning autonomy, and attempts to balance human and AI support sources.

These temporal patterns align with observed shifts in metacognitive awareness and strategy refinement over time, suggesting a developmental trajectory in learner autonomy influenced by sustained engagement with AI tools.

4.5 Summary of Findings

Thematic analysis of reflective diary data revealed that learners' perceptions and use of AI tools in writing tasks evolved considerably over the eight-week intervention. While students reported a range of benefits, such as increased motivation, greater linguistic awareness, and enhanced self-regulation, they also encountered recurring challenges, including cognitive offloading, technical and access barriers, and tensions surrounding cultural identity and voice.

Overall, the findings indicate that learners navigated a complex landscape of affordances and constraints while engaging with AI-instructed writing support. Patterns of growth were observed, particularly in strategic tool use, prompt engineering, and critical feedback evaluation. However, several areas, such as ethical uncertainty, mechanical accuracy, and creative expression, emerged as persistent challenges that may limit the effectiveness of AI without complementary pedagogical scaffolding.

4.4. Results for RQ4

"How can AI integration in EFL writing instruction be conceptualised and implemented to promote authentic learner autonomy while avoiding cognitive dependency?"

This research question aimed to develop a conceptual framework for designing AI-instructed writing support that fosters authentic learner autonomy while mitigating cognitive

dependency. Based on a qualitative synthesis of interview and diary data, a staged progression model was developed to represent learners' varying developmental trajectories in response to the AI-integrated intervention. The model, referred to as the **9A Catalyst**, comprises four interrelated phases, **Pre-A, Practice-A, Performance-A, and Stagnation/Regression**, that emerged from recurring student behaviours, perceptions, and developmental patterns.

Figure 24

The 9A Catalyst Framework for AI-Instructed Autonomy Development in EFL Writing



Pre-A Phase: Access and Alignment Awareness

In the initial phase, learners grappled with technological barriers and limited critical awareness. Early diary entries (Weeks 1–3) were dominated by Access challenges (n = 287), including navigation issues and connection limitations. For example, E5 noted, "I couldn't find my old conversations with ChatGPT" (Week 1). Such logistical barriers often overshadow learning goals, stalling metacognitive engagement.

As students progressed, they began to notice Alignment issues, including style mismatches and cultural disconnects. These perceptions surfaced later in the process, such as when E9 reflected, "My writing sounds too American with ChatGPT help" (Week 7). This

suggests that awareness of alignment concerns emerged experientially rather than being anticipated from the outset.

Additionally, awareness development patterns varied across participants. Some demonstrated progressive metacognitive insight, like E1, who wrote, “ChatGPT helps me understand grammar better” (Week 7), while others, such as E11, revealed deteriorating self-monitoring: “I don’t check my own work anymore since ChatGPT does it better” (Week 4).

Practice-A Phase: Application and Engagement

In this phase, students began to experiment more confidently with AI capabilities, although without always transferring skills effectively to independent writing. Application enthusiasm was common, but transfer failure surfaced over time. Dependency concerns (n = 31) rose markedly by mid-intervention, showing that enthusiasm did not always equate to sustained cognitive development.

Affective engagement was also prominent. Students responded positively to immediate feedback (n = 45) and convenience (n = 40), yet this sometimes led to cognitive ease preference (n = 19) and critical analysis avoidance (n = 15). E3 encapsulated this contradiction: “Why think hard when ChatGPT can give me answers quickly?” (Week 4). Such responses suggest that affective benefits can risk undermining deeper learning unless properly scaffolded.

Performance-A Phase: Autonomy and Integration Sophistication

By the later weeks (7–8), some students displayed higher-order skills in navigating AI support critically. Participants began to demonstrate clearer agency, reflected in statements such as E11’s, “I decide when ChatGPT help is needed” (Week 7). This autonomy was tempered by an awareness of limitations: E7 noted, “Too much help made it seem like only the draft was mine” (Week 7), illustrating a growing recognition of the tension between AI utility and personal authorship.

Students also exhibited ethical awareness, particularly in relation to authorship boundaries and academic integrity. E3’s reflection, “How much ChatGPT help is too much?” (Week 6), typifies the kind of internal negotiation typical of this phase. These developments suggest a shift from tool dependency to critical integration and self-regulation, marking successful progression within the framework.

Stagnation and Regression: Failed Phase Transition Patterns

Not all participants advanced linearly. A subset remained anchored in early-phase behaviours or regressed into dependency loops. E8 transitioned from early enthusiasm ("I'm not scared to make mistakes with ChatGPT," Week 1) to frustration and concern ("Writing alone feels much harder now that I'm used to ChatGPT," Week 5), highlighting the non-linear and reversible nature of development within this model.

Additionally, many participants engaged in cognitive offloading (n = 85), which surfaced gradually across diary entries. E4 noted, "I don't need to think as hard about organisation because ChatGPT helps" (Week 5), while E5 added, "I skip thinking about why and just ask ChatGPT directly" (Week 7). These reflections suggest that without explicit guidance, the convenience of AI support can lead to the erosion of cognitive resilience.

Summary of RQ4 Results

The 9A Catalyst framework provides a grounded conceptualisation of how learners progress through varying phases of AI-assisted learning. It accounts for both successful autonomy development and stagnation trajectories shaped by cognitive, technical, affective, and ethical factors. The findings suggest that authentic autonomy is not solely a function of access to AI tools, but rather hinges on critical pedagogical mediation, learner digital readiness, and evolving self-regulatory awareness.

CHAPTER V
DISCUSSION

This study investigated the effects of **AI-instructed** writing instruction on EFL learners' writing performance, the development of learner autonomy, and students' perceptions of AI integration within Turkish EFL classrooms. The following sections interpret the findings through established theoretical frameworks while also introducing a novel conceptual model that accounts for the unexpected outcomes. This chapter also explores pedagogical implications and highlights areas for future inquiry.

Integration of Quantitative and Qualitative Findings

Quantitative results indicated statistically significant gains in both writing performance and autonomy scale scores among the experimental group. These improvements aligned with qualitative evidence of enhanced strategic planning (n = 25), metacognitive awareness (n = 32), and self-regulatory behaviour (n = 35), all of which suggest that students were developing increased ownership over their writing processes. Pseudonymous participants such as E3, E5, and E11 referenced self-directed planning, goal setting, and monitoring of their own writing, which aligned with the observed improvement trajectories.

However, a closer examination of the qualitative data surfaced contradictions that complicate this picture. Despite higher scores on autonomy scales, student diaries frequently included expressions of over-reliance, such as dependency development (n = 31), cognitive offloading (n = 85), and critical thinking deterioration (n = 55). These findings call into question whether the observed autonomy gains were genuinely internalised or were simply indicative of tool-dependency masked as strategic behaviour. This suggests that learner autonomy in AI-instructed contexts may diverge significantly from traditional conceptions of autonomy, necessitating a more nuanced theoretical understanding.

5.1 AI's Influence on Learner Autonomy Development (RQ1)

5.1.1 The Paradox of Strategic Autonomy Versus Cognitive Dependency

One of the most critical insights related to RQ1 was the emergence of what can be termed a paradox of strategic autonomy, a situation in which learners demonstrate sophisticated management of AI tools while simultaneously exhibiting growing cognitive dependency. For instance, participant E3 shared that they had begun selecting which ChatGPT responses to adopt, stating: *"I choose which ChatGPT suggestions to use"* (Week 8), and *"I decide when*

ChatGPT help is needed" (Week 7). Nonetheless, the same participant also remarked, *"Why think hard when ChatGPT can give me supporting ideas more quickly?"* (Week 4), and *"Managing AI help while thinking about content is mentally tiring"* (Week 6), revealing the mental offloading that occurred even among advanced users.

This duality reflects a broader concern echoed in recent studies. Li (2024) highlighted that over-reliance on AI dialogue systems in EFL contexts can foster intellectual complacency, particularly when students turn to LLMs for paraphrasing without critically engaging in the rewriting process. Neuroscientific research by Kos'myna et al. (2025) corroborates these concerns. Their experimental study compared participants using ChatGPT, search engines, or unaided methods for essay writing. Strikingly, none of the LLM users could accurately recall their own written sentences post-task, while 88.9% of brain-only and search engine users could. EEG readings showed significantly lower neural engagement in the LLM group, indicating reduced memory consolidation and critical thinking.

In the present study, the experimental group exhibited a higher frequency of mechanical and spelling errors compared to the control group, especially regarding punctuation (see Table 19). Despite receiving automated feedback, these students showed limited improvement in basic mechanics such as capitalisation and comma usage. This supports the argument made by Bjork and Bjork (2011) on the principle of **desirable difficulties**, where effortful learning (e.g., handwriting or dictionary use) promotes stronger retention than cognitively passive tasks.

These findings also align with Cognitive Offloading Theory (Risko & Gilbert, 2016), which explains that learners increasingly rely on external aids for tasks previously managed internally. In the diaries, students openly admitted skipping analytical thinking steps. E5 noted: *"I skip thinking about why and just ask ChatGPT directly"* (Week 7). This behavioural trend indicates a shift from generative learning to dependency, a phenomenon consistent with what Oberauer (2009) and Sparrow et al. (2011) describe as cognitive atrophy**.**

In line with this, the experimental group's punctuation errors rose from 86 in the pre-test to 93 in the post-test (see Table 17). In contrast, the control group, which engaged in more traditional learning processes, showed a total of only 50 errors. This supports the claim that over-reliance on AI tools weakens internalised editing skills and may lead to regression rather than development.

A particularly compelling quote comes from E3: *"Even if my writing is good and I improve it, ChatGPT is never content. It always recommends a correction. It's already perfect for me, but it makes it way above my level."* This reflection hints at the tension between human

self-efficacy and machine-generated expectations, which can erode confidence and demotivate learners.

The study's findings also correspond with Wale (2024), who observed increased writing performance and learner autonomy through AI use, but only when tools were pedagogically aligned with learners' proficiency levels and supported by explicit instruction. In contrast, this study used general-purpose tools, which may not have adequately scaffolded the learners' decision-making processes.

These contradictions necessitate a rethinking of how autonomy and critical thinking are defined and evaluated in AI-instructed environments. Students like E3 and E5 demonstrated procedural awareness (i.e., prompt engineering) but lacked generative critical thinking, the kind needed to evaluate, adapt, and critique AI suggestions beyond selection. This confirms the work of Gonsalves (2024), who differentiates procedural skill from metacognitive awareness.

This nuanced view parallels distinctions made in 21st-century frameworks of critical thinking, such as those by Huber and Knights (2022) and Lindebaum and Fleming (2024), who argue for separating adaptive, tool-based thinking from reflective, independent cognitive engagement. In this context, learners developed adaptive critical behaviours(selecting, tweaking, rephrasing) rather than generative cognitive growth**.** The reduction in artistic language and metaphor uses among experimental participants (only 2 out of 12 in EG vs. 6 out of 6 in CG used artistic language, see Table 19) further suggests that reliance on AI can standardise output and erode creative expression.

Finally, this phenomenon is not unique to EFL contexts. Dergaa et al., (2024) have likened it to learners relying on calculators before understanding basic arithmetic, an analogy echoed by many participants who viewed ChatGPT as a default source of knowledge, bypassing foundational learning processes. These findings point toward the necessity of Human-in-the-Loop (HITL) approaches that preserve human agency while leveraging AI capabilities (see Section 5.4.4).

In summary, what at first glance appears as autonomy may, in fact, represent a conditioned form of autonomy, dependent on AI systems, detached from independent reasoning, and vulnerable to skill regression without robust pedagogical frameworks.

A related paradox emerged in terms of group retention. While the control group outperformed the experimental group in most dimensions of writing quality, six students (23%) withdrew from the intervention before completion, compared to only one dropout in the experimental group. This raises questions about affective engagement and instructional satisfaction. It may indicate that the traditionally instructed students, despite producing stronger

output, experienced less perceived autonomy or value, leading to demotivation and disengagement. The experimental group, in contrast, appeared more emotionally engaged and motivated to continue, even when their writing outcomes were not superior. This behavioural divergence reinforces the need to assess both product and process in autonomy evaluation.

5.2.1 Learner Perceptions and Affective Shifts (RQ2b)

A particularly noteworthy finding addressing RQ2 was the superior overall performance of the control group across multiple writing criteria, which directly challenged both the initial research hypothesis and the broader body of literature highlighting the benefits of AI-instructed learning environments. While the experimental group exhibited improvement in areas such as grammatical accuracy and task fulfilment, the control group demonstrated more substantial and consistent gains, most notably in mechanics (mean gain: +7.50) and artistic language use, as evidenced in Table 6 and Figure 7.

This divergence becomes more striking when contrasted with recent empirical studies. For example, a large-scale investigation of ChatGPT-instructed feedback among 110 EFL university learners reported marked improvements in cohesion, conciseness, and grammatical precision when the tool was used within teacher-guided frameworks (Polakova & Ivenz, 2024). Mahapatra (2024) similarly found that ChatGPT significantly enhanced academic writing only when embedded in structured pedagogical contexts, particularly those that included explicit modelling, scaffolded prompting, and guided peer review cycles.

In the present study, however, learners in the experimental group were granted autonomous, unrestricted access to ChatGPT without continuous teacher feedback or structured support, as part of the intentional design to evaluate authentic autonomy development. This hands-off model, while useful for observing strategic independence, may have unintentionally weakened instructional outcomes, particularly in areas like punctuation and mechanical accuracy, where AI feedback was often vague or overlooked errors (Mahapatra, 2024).

These results substantiate one of the study's central claims: **Pedagogical Context Primacy**, the argument that the quality and structure of instructional implementation exert greater influence on learning outcomes than the mere presence of technological tools (Bellwether, 2023). Without sustained pedagogical scaffolding, the advantages of generative AI tools appear limited and may, in fact, contribute to surface-level improvements without deeper linguistic development. This was especially evident in artistic language performance, where only 2 of 12 experimental group participants (see Table 19) demonstrated improvement post-intervention, compared with 6 of 6 in the control group.

Moreover, the quantitative findings were mirrored by affective and behavioural indicators captured in the reflective diaries. Experimental group participants frequently noted AI's convenience and immediacy, but also expressed uncertainty about creative expression, voice preservation, and task ownership, issues tied to broader questions of autonomy and authorship. While these perceptions are more fully explored under RQ3, they help contextualise why quantitative improvements in the experimental group, though present, lacked the consistency and depth observed in the control group.

5.2.2 Qualitative Writing Performance Changes (RQ2a)

Qualitative data further illuminates the observed performance differences. Learners in the experimental group demonstrated enhanced structural coherence and grammatical accuracy, consistent with ChatGPT's capacity to offer syntactic correction and organisational suggestions. However, these gains came at the expense of expressive richness and rhetorical complexity. Diary and interview responses revealed diminished use of metaphor, simile, and idiomatic phrasing, hallmarks of higher-order stylistic sophistication. As shown in *Table 7*, only two of the twelve experimental group participants used artistic or figurative language in their post-test essays, compared to all six control group learners.

In contrast, the control group consistently produced more expressive and metaphorically nuanced writing. These students employed figurative language, idiomatic expressions, and original structuring that aligned closely with the stylistic criteria assessed by the rubric. This pattern supports Swain's (1985) Output Hypothesis, which emphasises the role of productive struggle in second language development. AI tools, by pre-empting this struggle, may inadvertently suppress learners' engagement with deeper linguistic processing. Recent qualitative evidence similarly warns of this suppression, with students reporting that "ChatGPT imposed a pattern on writing and hindered creativity in content organisation" (*Smart Learning Environments*, 2024).

Additionally, while AI-generated feedback provides immediate corrections, it often lacks metalinguistic explanations and frequently fails to address higher-level features of writing such as argumentative logic, paragraph coherence, and critical evaluation (Truscott, 2007). As highlighted in Theme 7 of the autonomy analysis (see Section 4.3), students noticed that ChatGPT "changes the entire sentence" or "makes everything sound too academic," thus distancing their work from their original voice or intent. This tension contributes to the broader concern of **voice erosion**, a phenomenon also identified under the Cultural and Identity

Considerations theme (Theme 5), where students reported losing their stylistic individuality and feeling that their writing became “too American” or “too similar” to AI-generated models.

This concern is not purely stylistic but pedagogical. As seen in Theme 4 (Motivational and Affective Factors), many students felt encouraged and motivated by ChatGPT’s non-judgmental tone. However, the motivational boost created by immediate AI feedback did not necessarily correspond to improved performance in artistic or rhetorical dimensions. This contradiction supports your earlier note that positive affect does not guarantee deep learning outcomes. Without appropriate pedagogical scaffolding, students may confuse fluency and grammatical correctness with creative or critical writing development.

Lastly, generative AI tools are not free from epistemic limitations. Their responses are shaped by training corpora that often prioritise fluency and neutrality over nuance or local context. Bender et al. (2021) caution that such tools can reflect cultural biases and provide overly standardised language, which can obscure critical thinking or diverse expression. These risks were visible in this study: several experimental group students noted that ChatGPT “corrected things that weren’t wrong,” or “added too much when I just wanted a small fix,” reflecting not only a mismatch in style but also a disruption of learner autonomy (see Theme 6 and Theme 9).

To sum up, the qualitative data reveal that while AI tools can improve surface-level fluency and correctness, they may simultaneously constrain learners’ voice, creativity, and rhetorical development, dimensions of writing that are particularly difficult to cultivate without human mediation or reflective scaffolding. These findings reinforce the argument that AI-instructed writing support is most effective when integrated with pedagogically guided opportunities for original thought and expressive language use.

5.2.3 General-Purpose AI Limitations in Writing Instruction

The contrasting writing outcomes may, in part, stem from ChatGPT’s inherent limitations as a **general-purpose AI model** not designed with pedagogical objectives. Unlike dedicated educational technologies built on learning progression theory, developmental feedback systems, or language acquisition scaffolds, ChatGPT is optimised for fluency and syntactic plausibility rather than long-term language development (Kasneci et al., 2023; Rudolph et al., 2023). Consequently, the model often delivers polished outputs that may bypass the metalinguistic processing necessary for skill consolidation in second language acquisition.

This **pedagogical mismatch** was evident throughout the intervention. Several experimental group participants reported receiving full sentence rewrites rather than support for incremental revision, thereby reducing opportunities for productive struggle, an essential condition for internalisation and transfer. The data align with the theoretical framing of **desirable difficulties** (Bjork & Bjork, 2011), which stress the value of effortful cognitive engagement for durable learning. In contrast, ChatGPT's ease of use and automatic completion features may inadvertently promote passive learning behaviours.

A clear example emerged in the reflective diaries, where participant E11 remarked: *"I don't check my own work anymore since ChatGPT does it better"* (Week 4). This sentiment, echoed in other diary entries, suggests early signs of **fossilisation**, a process where incorrect forms or over-reliance patterns become entrenched due to a lack of reflective monitoring. The student no longer engages in internal feedback, instead outsourcing this responsibility to the AI, effectively halting further development of autonomous editing skills. This behaviour aligns with the "Cognitive Dependency" and "Critical Thinking Deterioration" themes discussed in Section 4.3 and elaborated in Section 5.1.

In addition to pedagogical shortcomings, digital literacy barriers further exacerbated the instructional limitations. As documented in Theme 2 of the qualitative analysis, technical challenges were a dominant factor in diary entries, with over 287 codes across participants relating to navigation issues, prompt formulation problems, and tool responsiveness delays. These challenges not only reduced the effective use of ChatGPT but also amplified cognitive load during writing tasks. Such findings mirror recent literature showing that many EFL learners remain unfamiliar with AI tools beyond basic platforms like Google Translate and Grammarly (Alghasab, 2024), highlighting a need for structured tool onboarding and digital skills development in AI-integrated writing environments.

Furthermore, the one-size-fits-all nature of general-purpose tools fails to account for learner variability. In contrast, emerging **Socratic AI tools**, such as *Socrat.ai*, have begun to implement guided questioning protocols that promote learner agency and reflection rather than offering direct answers. These models aim to align more closely with constructivist and sociocultural learning theories by positioning students as active problem-solvers. The findings from the current study suggest that such targeted AI integration may offer more sustainable autonomy development than current general-use systems.

In sum, while ChatGPT provided immediate support and structural improvements, its design as a generalist text generator, rather than a pedagogically guided tutor, limits its capacity

to support deeper learning. For EFL learners, especially at early-to-intermediate proficiency levels, such limitations may hinder long-term development unless:

accompanied by intentional instructional mediation, digital skill scaffolding, and reflective engagement strategies.

5.2.4 Learner Autonomy and Feedback Practices (RQ1)

While the control group demonstrated stronger gains in writing proficiency and stylistic development, the experimental group exhibited partial but notable growth in autonomy-related behaviours, particularly in the form of increased self-regulation, developing confidence, and strategic engagement with ChatGPT. Participants described re-reading their drafts, evaluating AI-generated suggestions more selectively, and cross-verifying ChatGPT outputs with external sources before incorporating them. These practices are consistent with earlier findings by Zhang and Zou (2024), who reported increased revision frequency and confidence when students received AI support in structured EFL writing tasks.

However, a key divergence in the present study lies in the implementation context. Unlike Zhang and Zou's design, where ChatGPT was embedded into teacher-led, scaffolded writing cycles, the current study provided learners with u**nrestricted** and **unscaffolded access** to AI tools. This difference in pedagogical mediation may explain why observed gains in strategic behaviour did not consistently translate into measurable improvement in writing performance. Similarly, Zhou and Teng (2023) found that AI-assisted self-regulation only emerged when students were prompted through reflective writing tasks, a condition not uniformly present in the current experimental group. Consequently, what may appear as autonomous action in tool usage might, in fact, reflect surface-level independence, or what this thesis terms artificial autonomy (see Section 5.4.1).

By contrast, the control group benefited from teacher-led feedback practices that fostered internalisation of evaluative criteria and authentic autonomy. Students consistently referenced the teacher's personalised and **forward-facing feedback** approach, which they described as connecting previous writing issues with future applications. Several learners demonstrated this integration by anticipating feedback, revising before submission, and applying previous advice to new writing tasks. This mirrors findings from Hwang and Lee (2023), who argue that **formative, dialogic feedback** promotes deeper cognitive engagement than AI-generated suggestions, which may lack nuance or contextual awareness.

These dynamics also align with Haidar's (2025) quasi-experimental study involving 352 high school learners, which found that when metacognitive prompts were embedded within

GenAI-supported writing tasks, students showed significantly improved reflective and regulatory skills. Unlike ChatGPT, which often offers pre-formulated suggestions without learner prompting, tools designed with embedded scaffolds facilitate learners' transition from passive tool users to active self-regulators.

The present study's results reinforce this argument. Despite frequent access to ChatGPT, experimental group participants who lacked instructional framing often defaulted to reliance rather than regulation. Their diary entries reflected cognitive offloading tendencies and a hesitancy to revise AI-generated content independently. Without explicit metacognitive guidance, even seemingly autonomous behaviours (e.g., prompt crafting or result selection) remained reactive rather than reflective.

Moreover, these findings pinpoint a crucial but often overlooked factor in AI integration: teacher digital competence and instructional autonomy. For learners to develop meaningful autonomy in AI-supported contexts, instructors themselves must be capable of strategically deploying AI tools with pedagogical intent. If educators lack this competence, they may offer inconsistent or overly permissive guidance, resulting in conditioned or artificial autonomy among students, surface-level independence that masks underlying technological dependency.

This concern echoes Koraishi's (2024) thesis research, which asserts that teacher training in AI tool integration is not supplementary but foundational. Without instructors who can guide learners through ethical, strategic, and metacognitive uses of AI, the promise of AI-enhanced instruction may remain unfulfilled. As such, any discussion of learner autonomy in AI-instructed environments must be inseparable from discussions of teacher readiness, digital literacy, and professional development.

In sum, while both groups exhibited signs of autonomy development, the depth, sustainability, and authenticity of this development was notably stronger in the human-mediated condition. These findings reaffirm the study's central theoretical assertion of Pedagogical Context Primacy: that the educational quality of an AI tool is shaped less by its technical affordances than by the instructional environment in which it is embedded.

5.3 Student Perspectives on AI Integration Impact (RQ3)

5.3.1 Temporal Development of Student Perceptions

Analysis of learner perspectives, drawn from 2,247 coded diary segments across the 8-week intervention, revealed distinct temporal patterns in the evolution of AI tool use. These findings support staged models of technological adoption and echo emerging research that

conceptualises AI use as a relational, iterative process involving evolving cognitive, social, and emotional engagement (Kim et al., 2025).

Weeks 1–2 (Initial Phase): Students expressed a combination of excitement and frustration. Access-related issues were mentioned by 85% of participants, including internet instability and tool familiarity. However, 75% also reported elevated motivation, with comments such as "ChatGPT makes me want to write more" (E11, Week 2). This early-stage enthusiasm aligns with research showing that instant AI feedback heightens student engagement, especially in low-stakes environments (Song & Song, 2023).

Weeks 3–4 (Development Phase): Strategic usage patterns began to emerge, particularly in prompt engineering, yet were accompanied by growing dependency concerns. While 60% of learners reported improved ability to interact with the tool effectively, 45% admitted to over-reliance. This phase corresponds with Jose et al.'s (2025) findings that AI tools can promote strategic thinking, but simultaneously facilitate cognitive offloading when not embedded in reflective practice.

Weeks 5–6 (Consolidation Phase): Students reported increasingly sophisticated but routine tool use. Several diary entries reflected habitual or automatic use, with E3 noting, "Too much help made it feel like only the draft was mine" (Week 7). This aligns with Reed's (2025) findings that entrenched AI dependence may create cognitive inertia, a form of 'cognitive debt' that outlasts tool use.

Weeks 7–8 (Maturation Phase): By the end of the intervention, learners displayed higher levels of critical evaluation, including rejection or refinement of AI feedback. At the same time, dependency patterns became embedded, highlighting a paradox also reported by Li (2024), who found that over-reliance on AI systems may undermine higher-order cognitive capacities such as critical thinking and independent judgment.

These temporal shifts suggest that autonomy-related growth does occur, but it may coexist with underlying dependency and require pedagogical scaffolding to evolve into sustained, **authentic autonomy** (see also Section 5.5.4).

5.3.2 Perspectives on Autonomy Development

Students' reflections on autonomy development revealed ambivalence, often oscillating between empowerment and technological dependence. While many learners described feeling more in control of the writing process, they also acknowledged increased reliance on AI guidance.

For instance, decision-making autonomy was reflected in 23 coded segments, such as E11's comment: "I decide when ChatGPT help is needed" (Week 7). Simultaneously, fear of independent work appeared in 20 codes, with E5 stating, "I feel worried when I have to write without ChatGPT" (Week 4). These findings support Song and Song's (2023) warning that motivational gains from AI use may not translate into deeper cognitive independence.

The ownership development theme (n = 32) was prominent. Majority of the students expressed appreciation for being able to accept or reject AI suggestions, highlighting a sense of agency. However, the presence of automatic consultation behaviour (n = 17) suggests that some of this ownership was conditioned rather than authentic, a pattern consistent with Singh et al.'s (2015) argument that cognitive offloading mediates the relationship between AI use and critical thinking development.

This duality mirrors the conceptual tension described in Section 5.1: what appears to be strategic autonomy may in fact mask cognitive dependency. Students were learning to manage the tool, not necessarily developing independent strategies apart from it.

5.3.3 Perspectives on Writing Performance Changes

Students reported mixed experiences regarding the impact of AI support on their writing performance. Many recognised benefits in fluency and structure, noting that ChatGPT helped them reduce errors and gain confidence. Reduced anxiety and increased writing frequency were attributed to the immediate availability of AI-generated suggestions.

Yet, this came with clear trade-offs. Multiple learners expressed concern about loss of personal voice, reduced originality, and over-standardisation. These concerns are consistent with findings from Yan et al. (2024), whose study on Chinese EFL learners highlighted that students conceptualised AI as both a "helper" and a "hindrance" in the development of stylistic expression.

A particularly revealing tension emerged around feedback preferences. Despite the convenience of ChatGPT, most students reported preferring teacher feedback, citing its personal, context-sensitive nature. One participant explained: "ChatGPT praises everybody, but the teacher only compliments when she truly thinks so. That's why her positive feedback means more to me" (E8, Week 5). These remarks mirror Wang et al.'s (2025) findings that students revise more meaningfully and demonstrate greater engagement when guided by teacher comments rather than generic AI suggestions.

Students also noted the repetitiveness of AI responses (n = 15) and lack of explanation behind some corrections (n = 13), reinforcing the notion that AI feedback, while rapid, is often

superficial and lacks the depth necessary for internalisation and long-term improvement (Truscott, 2007).

5.3.4 Long-term Learning Value Concerns

Students increasingly voiced concerns regarding the long-term impact of AI integration on their learning processes. While initial enthusiasm stemmed from the immediacy and convenience of ChatGPT support, many learners later reflected critically on whether such convenience was undermining their independent development. This evolving perspective aligns with recent research suggesting that prolonged AI reliance may erode foundational cognitive and metacognitive skills (Lo et al., 2024).

Despite strong motivational benefits, highlighted by 236 coded segments reflecting increased engagement, reduced anxiety, and perceived writing ease, these affective gains did not consistently translate into improved learning outcomes. As shown in earlier sections, learners in the experimental group exhibited higher mechanical and punctuation errors and diminished use of expressive and artistic language. These contradictions support Campbell's (2025) findings on productive failure, which posit that cognitive struggle plays an essential role in fostering meaningful, transferable learning.

Students' reflections further suggested that motivation in the absence of cognitive challenge may foster a false sense of progress. This is evident in diary entries where participants described their writing as "better" simply because ChatGPT provided more polished output, even when internal learning or skill consolidation was limited. Such findings reinforce earlier conclusions regarding the importance of maintaining desirable difficulties in the learning process (Bjork & Bjork, 2011).

Together, these patterns signal a need to distinguish between short-term engagement and long-term learning value. While AI-instructed environments may boost learners' affective responses, their cognitive outcomes appear contingent on how these tools are integrated pedagogically. Without appropriate scaffolding, even the most motivated learners may experience surface-level progress at the cost of deeper development.

5.4 Framework for Authentic AI Integration in Writing Instruction (RQ4)

5.4.1 The 9A Catalyst Development

Addressing RQ4, this study developed a comprehensive framework based on empirical findings: the *9A Catalyst* for AI-integrated writing instruction. This framework maps four developmental phases, *Pre-A*, *Practice-A*, *Performance-A*, and *Personalisation-A*, onto

learners' evolving relationship with AI. It draws on Zimmerman and Campillo's (2003) self-regulation theory and Benson's (2001) autonomy model, adapted for AI-instructed environments.

This model represents a shift from linear models of technology adoption to a phase-based developmental trajectory that acknowledges AI's dual potential, as both enhancer and inhibitor, depending on pedagogical conditions and learner readiness.

The theoretical foundation synthesises Cognitive Load Theory (Sweller et al., 2019), Self-Determination Theory (Deci & Ryan, 1985), Sociocultural Theory (Vygotsky, 1978), and Cognitive Offloading Theory (Risko & Gilbert, 2016).

The Four Phases of the 9A Catalyst Framework:

1. Pre-A Phase – Foundation for Autonomy:

- *Access*: Ensuring equitable digital access and basic AI tool competence (e.g., E7 and E1 struggles with prompt clarity).
- *Alignment*: Culturally sensitive integration of AI into pedagogical goals.
- *Awareness*: Building metacognitive awareness of AI's limitations and benefits.
- *Aim*: Clarifying writing goals and aligning AI use with learner needs.

2. Practice-A Phase – Active Engagement:

- *Application*: Using AI in authentic writing tasks, with balanced tool support.
- *Affective Engagement*: Leveraging AI to boost confidence and reduce anxiety (e.g., reduced fear of errors).

3. Performance-A Phase – Ownership and Competence:

- *Agency*: Developing reflective decision-making regarding AI use (e.g., E3's prompt optimisation).
- *AI Integration*: Embedding AI strategically within writing processes.

4. Personalisation-A Phase – Internalisation and Autonomy:

- *Autonomy*: Demonstrating self-regulation and critical engagement.
- *Adjustment*: Modifying AI strategies based on context and reflection.

- *Action*: Sustaining authentic, self-directed writing with or without AI.

Figure 25

The 9A Catalyst Framework (See Figure X for Phase Indicators)

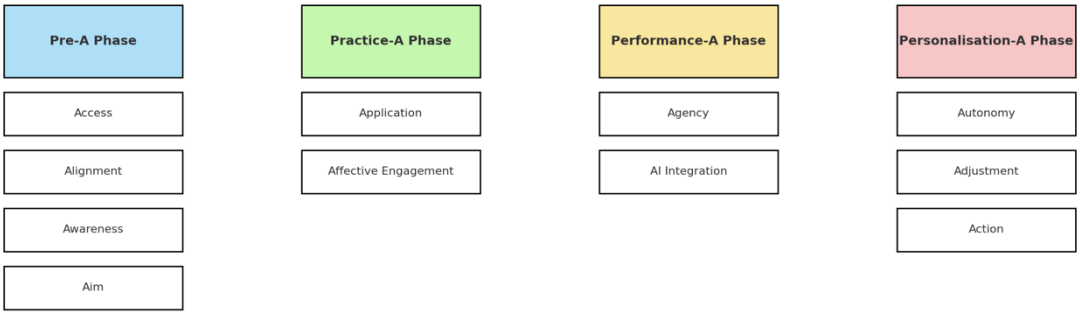


Figure 25 illustrates the full structure of the 9A Catalyst Framework, mapping the dynamic progression of AI-instructed EFL learners across four interconnected phases: Pre-A, Practice-A, Performance-A, and Personalisation-A. Each phase builds cumulatively on learners' access, alignment, emotional engagement, and eventual autonomy.

5.4.2 Cognitive Resistance Training

To preserve learners' cognitive effort while using AI, this study proposes *Cognitive Resistance Training*. This approach supports deliberate instructional strategies that retain mental challenge during AI-enhanced tasks.

Grounded in Bjork and Bjork's (2011) theory of *desirable difficulties* and Kapur's (2008, 2024) theory of *productive failure*, the following practices are recommended:

1. Structured alternation between AI-assisted and unaided writing sessions.
2. Explicit instruction about when cognitive effort is beneficial.
3. Cognitive apprenticeship (Collins et al., 1989): gradual responsibility transfer.
4. Assessment valuing depth over polish (e.g., formative rubrics with revision).

5.4.3 Pedagogically Informed Alternatives and the SAIL Model

ChatGPT's limitations as a general-purpose tool highlight the need for pedagogically aligned AI alternatives. Unlike LLMs that provide ready-made solutions, Socratic AI systems (e.g., Socratic by Google) prompt inquiry through guided questioning.

The Scaffolded AI Literacy (SAIL) model (Munday & Păduraru, 2024) offers a structured progression from awareness to autonomous, reflective use. It supports:

Metacognitive routines

Critical digital literacy

Gradual withdrawal of AI scaffolding (as seen in ITS and Bloom's Taxonomy-aligned systems)

This study's findings affirm that unstructured AI use, such as unrestricted ChatGPT prompting, can result in metacognitive underdevelopment and reduced feedback internalisation. Such outcomes contrast with prior optimistic findings (e.g., Mahapatra, 2024; Baskara, 2023), which often overlook longer-term autonomy and creativity.

5.4.4 Pedagogical Context Primacy Principle

This study proposes the principle of Pedagogical Context Primacy: instructional quality, not tool novelty, determines learner development. Despite the experimental group's AI support, the control group outperformed in mechanics, creativity, and feedback application.

This aligns with Kapur's (2008) and Bjork and Bjork's (2011) findings on the value of cognitive struggle in learning. In contrast, AI's emphasis on fluency and error correction reduces productive discomfort.

Students in the control group described the teacher's feedback as *"feedlife"*, personalised, evolving, and motivational. For example:

> *"The teacher says, 'Make quality mistakes so you learn something new.' If we keep making the same mistake, it's not a learning moment, it's just laziness."*

Walter (2024) and Fawns (2022) also caution against mistaking ease for depth. If not properly scaffolded, AI tools can undermine critical engagement, feedback literacy, and prompt interpretation skills.

Thus, the teacher's role must shift toward *autonomy facilitation*: helping learners learn *how to think*, not just *what to write*. Teachers must model reflective decision-making, prompt evaluation, and AI tool orchestration. Only then can AI function not as a substitute but as a partner in lifelong learning.

The observed paradox of strategic autonomy versus cognitive dependency (Section 5.1.1) directly supports HITL principles. While students in the experimental group developed

prompt engineering skills, they simultaneously exhibited reduced critical thinking and over-reliance on AI suggestions, precisely the risks that HITL frameworks are designed to mitigate.

5.5 Integrated Findings Across Research Questions

5.5.1 Convergent Evidence Pattern

The synthesis of findings from RQ1 to RQ3 reveals a convergent pattern whereby AI-instructed writing instruction facilitated learners' strategic tool use and perceived autonomy (RQ1), but concurrently contributed to cognitive dependence, diminished stylistic sophistication, and creative regression (RQ2). These effects culminated in students' ambivalent perceptions about AI's long-term educational value (RQ3), thereby reinforcing the necessity for explicit pedagogical scaffolds and informed frameworks, as outlined in the 9A Catalyst model (RQ4; see Figure 25).

A central insight of the study is the surfacing of key educational paradoxes in AI integration. Learner autonomy seemed present in students' selective use of ChatGPT but often hid increasing cognitive dependence. Second, quantitative improvements in grammar and task fulfilment among experimental group participants masked qualitative declines in creative expression, rhetorical skill, and artistic language use. Notably, all six control group students used figurative language in their post-test essays, compared to only two in the experimental group (see Table 19). Third, initial motivational surges (n = 236 affective codes) did not translate into sustained, high-level performance, suggesting that motivation alone is insufficient when not paired with metacognitive regulation and cognitive challenge.

These contradictions support emerging concerns in the literature around "surface-level learning" and digital overreliance. Recent critiques have warned of what Jose et al. (2025) label the "cognitive catastrophe**"**, a scenario where excessive reliance on AI tools diminishes memory, problem-solving, and reflective thinking. Similarly, findings from Kos'myna et al. (2025) and Sparrow et al. (2011) reinforce the claim that uncritical dependency on external aids can erode internal learning systems, particularly when learners bypass retrieval, analysis, and self-correction in favour of rapid output.

This study further disrupts the simplistic binary between teacher-centred and student-centred approaches. Instead, it advocates for a collaboration-centred pedagogy, wherein learning is co-constructed through dynamic interactions among students, teachers, peers, family members, and intelligent tools. In AI-instructed environments, the most effective outcomes emerged when students engaged in co-regulated learning ecosystems, balancing autonomy with guidance, and digital tools with human feedback. Control group participants benefited from

dialogic feedback practices, while experimental group learners who adopted reflective prompt engineering (e.g., E3, E7, E11) demonstrated heightened strategic awareness.

These findings underscore the importance of shared cognitive responsibility, not only between teacher and learner, but also within broader social and digital networks. They highlight the limitations of unmoderated AI use and the need for pedagogical orchestration, where the teacher assumes a designer role, guiding the AI-human interaction towards educationally valuable ends. Moving forward, this study argues for pedagogical frameworks that embed cognitive resistance, reflective scaffolding, and feedback literacy as central pillars of AI-enhanced language instruction.

5.5.2 Theoretical Implications

The integrated findings of this study invite significant refinements to several established theoretical frameworks relevant to autonomy, cognition, and learning technology. First, Cognitive Offloading Theory (Risko & Gilbert, 2016) is not merely supported but extended by the data. While the theory suggests that reliance on external tools may alleviate working memory demands, the present findings demonstrate that this reduction in cognitive effort may in fact become counterproductive in educational settings, particularly when it results in diminished retention, decreased self-monitoring, and an erosion of critical thinking skills. The persistence of punctuation and mechanical errors despite increased tool use (see Table X) underscores this concern.

Second, the study raises critical questions for Distributed Cognition Theory (Hutchins, 1995), which views cognition as a function of interactions among individuals and artefacts in the environment. While AI tools like ChatGPT certainly extend learners' cognitive reach, the findings here suggest that such distribution may simultaneously weaken internal cognitive systems, particularly metacognitive regulation, stylistic nuance, and independent problem-solving. Thus, distributed cognition may support task completion while compromising skill development, a nuance often overlooked in technology-enhanced learning models.

Third, the results call for a reconceptualisation of traditional autonomy frameworks (Holec, 1981; Little, 1995), which predominantly focus on the learner's ability to make independent decisions in their learning process. In AI-instructed environments, learners often appear autonomous through tool use, but closer analysis reveals that this autonomy may be artificial, conditioned by the tool's suggestions rather than rooted in internalised strategy use or critical awareness. As such, autonomy must now be theorised as a spectrum that includes

strategic, collaborative, artificial, and authentic dimensions, as this thesis has proposed (see Section 4.3).

Finally, the study lends partial support, but also notable caveats, to the Gradual Release of Responsibility (GRR) framework, which posits a staged shift from teacher-led instruction to independent student performance. While this model remains pedagogically sound in principle, the introduction of generative AI appears to interfere with the natural progression of GRR stages. AI tools, when unmoderated, can short-circuit the productive struggle necessary for students to internalise cognitive and metacognitive processes, prematurely moving learners to "independent" work before foundational competencies are fully developed. In this sense, AI risks simulating the appearance of responsibility transfer without ensuring its substance.

These theoretical implications underscore the need to move beyond binary thinking, teacher-led versus student-led, tool-based versus internal, support versus independence, and towards a more dialectical understanding of autonomy and cognition in the age of intelligent technology. As learners become increasingly entangled in human-AI hybrid learning processes, theoretical models must evolve to reflect this complexity.

5.5.2.1 Extension of Self-Determination Theory

This study extends Self-Determination Theory (SDT; Deci & Ryan, 1985) by exploring how AI-instructed writing environments can simultaneously support and undermine learners' basic psychological needs: competence, autonomy, and relatedness. The findings suggest that AI integration is not inherently supportive of these needs, and its effects are contingent on pedagogical conditions and implementation reliability.

Competence Development Paradox. The study revealed a paradox in competence development. While participants in the experimental group demonstrated performance improvements, particularly in grammar and structural coherence, these gains did not always reflect deeper cognitive mastery. Learners reported relying on ChatGPT for organisation and vocabulary, often without fully internalising the rationale behind suggestions. This aligns with concerns that performance-based competence does not necessarily equate to cognitive independence. In this sense, AI tools may produce what appears to be competence while simultaneously inhibiting its authentic development.

In EFL writing contexts, HITL implementation involves three key components:

(1) **AI-assisted drafting** with human-guided revision cycles,

(2) **Teacher-mediated feedback interpretation** to develop critical evaluation skills

3) **Graduated responsibility transfer** following the 9A Catalyst phases.

Autonomy Redefinition Requirements. Traditional SDT conceptualises autonomy as the capacity for self-governed, volitional action. However, the findings indicate a need to redefine autonomy in AI-integrated environments as strategic interdependence, the ability to manage external tools in a reflective and agentic manner. Many students demonstrated this capacity by selectively integrating ChatGPT suggestions (see Theme 3), but others exhibited automatic reliance or passive use, reflecting artificial autonomy rather than authentic control. This redefinition expands SDT's relevance in contemporary digital contexts, where learners must exercise autonomy through tool mediation rather than tool avoidance.

5.5.2.2. Integration with Sociocultural Theory

The study also draws on Vygotsky's (1978) Sociocultural Theory, positioning AI as a cultural artefact that mediates language learning. However, unlike stable cultural tools (e.g., books or rulers), AI tools such as ChatGPT continuously evolve and generate content through probabilistic modelling, which introduces new complexities.

Cultural Tool Complexity. AI tools, by design, synthesise vast bodies of language and culture, which may conflict with learners' local identities and cultural norms. For instance, several participants (e.g., E5, E6) noted that ChatGPT's outputs often reflected Americanised phrasing that felt unnatural in their own English writing. These findings suggest that AI tools require active cultural negotiation, not passive adoption, particularly in EFL contexts where language learning is embedded within identity development.

Zone of Proximal Development (ZPD) Implications. AI tools may expand the ZPD by offering accessible scaffolding, particularly for lower-proficiency learners. However, the data also indicate that ChatGPT can disrupt ZPD progression by offering over-corrections, immediate answers, or fully formulated content that short-circuits productive struggle. This diminishes learners' opportunity to internalise conceptual understanding. As a result, while AI may scaffold performance, it may not always scaffold learning unless its use is consciously framed by pedagogical intent.

5.5.2.3. Cognitive Load Theory Applications and Extensions

The study offers a refined application of Cognitive Load Theory (CLT; Sweller et al., 2019), distinguishing between beneficial and detrimental load reduction. While reducing extraneous load is often helpful (e.g., removing unnecessary difficulty), eliminating appropriate cognitive load, effortful processing essential for learning, can hinder retention and transfer.

Cognitive Load Paradox. The findings reveal a cognitive load paradox, wherein tools designed to reduce load (e.g., through instant correction or auto-structuring) may also remove the very challenges that foster durable learning. Students often bypassed self-revision or avoided understanding why certain corrections were needed, leading to surface-level fluency but limited conceptual growth.

Desirable Difficulties Preservation. As Bjork and Bjork (2011) emphasise, tasks requiring retrieval, reconstruction, and mental effort foster deeper learning. This study affirms that when AI eliminates such difficulties, it may promote performance but compromise learning. The proposed "Cognitive Resistance Training" (see Section 5.4.2) offers a response: structuring tasks to retain appropriate difficulty even when AI support is available.

5.5.2.4. Synthesis: The 9A Catalyst Framework

The implications above are synthesised in the 9A Catalyst Framework, introduced in Section 5.4.1. This four-phase model supports structured AI integration through sequential development:

Pre-A Phase: Access, Alignment, Awareness, Aim

Practice-A Phase: Application, Affective Engagement

Performance-A Phase: Agency, AI Integration

Personalisation-A Phase: Autonomy, Adjustment, Action

This progression ensures that learners develop critical digital literacy, metacognitive awareness, and purposeful engagement before transitioning to independent AI use. It counters the risk of cognitive offloading and promotes internalisation of writing strategies across contexts.

5.5.3 Practical Implementation Requirements

The pedagogical frameworks developed in response to RQ4, particularly the 9A Catalyst and the Cognitive Resistance Training principles, must be directly aligned with the complex realities uncovered in RQ1 through RQ3. This study's integrated findings demonstrate that successful AI-instructed writing integration cannot rely on access or tool availability alone; rather, it depends on instructional design that preserves cognitive challenge, fosters authentic learner autonomy, and maintains ethical oversight.

Foundational Implementation Conditions.

Effective implementation begins with comprehensive digital literacy preparation. As revealed in earlier findings, students without foundational digital competence (e.g., E4, E6) struggled to prompt or interpret ChatGPT effectively, often defaulting to passive use. Thus, digital tool readiness must be an explicit instructional objective. Beyond technical skill, however, learners require explicit instruction in cognitive resistance strategies, learning when to embrace challenge, when to persist without AI aid, and how to discern between helpful support and cognitive outsourcing.

Structured Alternation and Assessment Alignment.

A clear recommendation emerging from this study is the structured alternation between AI-assisted and independent writing tasks. This prevents over-reliance, allowing learners to consolidate skills before applying AI support. This process also necessitates assessment practices that reward process over product, evaluating metacognitive growth, revision practices, and decision-making autonomy rather than merely polished final outputs. Without such scaffolding, learners risk developing artificial autonomy: an illusion of independence fuelled by technological crutches, but lacking cognitive resilience.

Avoiding Dependency Pitfalls.

Continuous monitoring for dependency development is crucial. As discussed in Section 5.1.1, students in the experimental group (e.g., E3, E5) reported shifting from active planning to automatic querying, revealing a drift towards cognitive offloading. Teachers must remain alert to such patterns, both through formative assessment and ongoing dialogue, and adjust instruction accordingly. As Peachey (2024) and the British Council (2024) have cautioned, educators must adopt a sober, evidence-based view of AI tools, resisting hype and focusing on pedagogical alignment over novelty.

Teacher-Centred Foundations for Autonomy.

This study reaffirms that cultivating learner autonomy in AI-enhanced environments begins with teacher preparation. Teachers must be equipped not only with AI literacy but also with pedagogical frameworks for integrating these tools meaningfully and ethically. These include recognising when to encourage student independence, when to mediate AI outputs, and how to foster reflective, critical engagement. As Koraishi (2024) highlights, AI-enhanced

instruction without corresponding teacher development risks superficiality, misalignment, or unintended dependence on the tools themselves.

Professional Development Pathways.

Teacher professional development programmes should therefore include structured modules on:

The affordances and limitations of generative AI in language learning
Prompt design and iterative prompting strategies
Metacognitive scaffolding and reflective writing support
Ethical and cultural considerations in AI tool use
Transitioning from feedback provision to feedback facilitation roles

Rather than promoting a simplistic shift from teacher-centred to student-centred models, the study recommends cultivating collaboration-centred classrooms, where autonomy is seen as a co-constructed process. In these learning environments, teachers and students jointly negotiate the use of AI, grounding decisions in pedagogical purpose and shared responsibility. Such models not only prevent cognitive atrophy but also preserve learners' creative agency, ensuring that AI functions as a catalyst for autonomy, not a substitute for it.

5.5.4 Pedagogical Implications

The findings of this study clearly indicate that effective AI-instructed integration in language education demands more than access to advanced tools; it requires what can be termed cognitive resistance training, a deliberate instructional approach that maintains appropriate cognitive challenge while leveraging technological support. This pedagogical strategy recognises that cognitive effort is not a barrier to overcome but a critical ingredient of learning, particularly in contexts where over-reliance on AI tools can lead to passivity, surface-level fluency, and cognitive offloading.

Cognitive resistance training should encompass the following components:

1. **Structured alternation** between AI-assisted and independent work to avoid cognitive skill atrophy and preserve learners' internal problem-solving capacities.
2. **Explicit metacognitive instruction** that helps learners distinguish between effort that builds competence and tasks where efficiency is genuinely beneficial.

3. **Cognitive apprenticeship** (Collins et al., 1989), whereby learners gradually take ownership of thinking processes initially modelled by teachers or peers.
4. **Assessment practices** that reward deep thinking, creativity, and reflective revision processes over mere surface-level textual polish or linguistic fluency.

The thematic analysis of 2,247 coded learner segments provided compelling evidence for how these principles could be implemented. As shown across the four phases of the 9A Catalyst Framework (Section 5.4.1), student development in AI-instructed environments is highly non-linear. Some students (e.g., E2, E3) progressed towards agency and strategic prompt engineering, while others (e.g., E4, E6) plateaued or regressed due to weak digital literacy, low affective engagement, or unscaffolded tool use.

These findings emphasise that technological sophistication alone cannot guarantee educational improvement. Even when tools such as ChatGPT are made available, if the pedagogical design is underdeveloped, if students are not emotionally or cognitively prepared, or if AI tools are used in a disconnected or instrumental way, learning outcomes may stagnate, or worse, regress. In this study, the control group, taught via traditional means, consistently outperformed the experimental group on key indicators such as artistic language use, punctuation accuracy, and rhetorical complexity, demonstrating the importance of Pedagogical Context Primacy over tool novelty.

This contrast reinforces the idea that practice does not always make perfect, it makes permanent. When students repeatedly interact with AI without reflective support, unproductive writing behaviours may become fossilised. For instance, while the experimental group demonstrated increased comfort with using digital tools, their post-test punctuation errors actually increased (from 86 to 93), while six out of six control group students retained or improved their use of artistic language, compared to only two in the experimental group (see Tables X and Y). These trends suggest that AI use, in the absence of cognitive resistance strategies, may entrench superficial skills while undermining deeper growth.

> Therefore, the implementation of AI-instructed writing tools should be rooted in intentional, theory-informed pedagogical design. This includes:
> Embedding AI tools within existing scaffolding structures rather than using them as replacements for instruction.
> Framing feedback tools like ChatGPT as thinking partners, not as automated editors.
> Encouraging learner autonomy through guided reflection, not unregulated independence.

Structuring learning experiences that balance cognitive ease with the productive discomfort of authentic writing struggle.

Ultimately, these findings argue for a shift in mindset: AI tools are not pedagogical shortcuts but instruments that must be orchestrated with precision and care. Without that, learners may come to see writing as a transactional task completed with the help of an assistant, rather than a generative, recursive, and meaning-making process that builds identity, voice, and independent competence.

Table 22

Temporal Development Patterns in Student Perceptions Across 8-Week Intervention

Time Period	**Dominant Concerns**	**Frequency Pattern**	**Representative Quote**
Weeks 1-2 *Initial Phase*	Technical barriers, enthusiasm, basic interaction	Access issues: 85% Motivation: 75%	"Finally got ChatGPT working... This is incredible" (E6, Week 1)
Weeks 3-4 *Development Phase*	Strategic usage, emerging dependencies	Dependency concerns: 45% Strategic thinking: 60%	"Why think hard when ChatGPT can give me answers quickly?" (E3, Week 4)
Weeks 5-6 *Consolidation Phase*	Sophisticated usage, increased dependency	Advanced strategies: 70% Dependency: 65%	"Writing alone feels much harder now... I think it is addictive" (E5, Week 5)
Weeks 7-8 *Maturation Phase*	Critical evaluation, entrenched patterns	Critical thinking: 75% Established dependency: 70%	"I skip thinking about why and just ask ChatGPT directly" (E11 Week 7)

Note. Frequency patterns represent the relative intensity of coded themes within each time period. Percentages indicate the proportion of participants demonstrating each pattern during the specified weeks.

5.5.5 The 9A Catalyst Framework: A Comprehensive Model for AI-Instructed Autonomy Building in Writing Contexts

5.5.5.1. Framework Introduction and Theoretical Foundation

The 9A Catalyst Framework for AI-Instructed Autonomy was developed in direct response to the mixed-methods findings of this study. It addresses the multidimensional nature of AI integration in EFL writing instruction by capturing both its empowering and potentially inhibiting effects on learner autonomy. The framework offers a developmental, phase-based model that moves beyond traditional linear conceptions of educational technology adoption, instead positing AI as a context-sensitive tool whose educational impact is contingent upon pedagogical scaffolding and learner preparedness.

This model emerged from observed learner behaviours and coded thematic patterns across interviews, diaries, and performance outcomes, revealing that while some students developed strategic prompt engineering skills and agency, others regressed into passive dependence and cognitive offloading. The 9A Catalyst Framework thus seeks to balance AI's efficiency with the preservation of cognitive struggle, self-regulation, and learner voice.

The framework is theoretically grounded in multiple converging perspectives:

- **Cognitive Load Theory (Sweller et al., 2019)**: While AI tools can reduce extraneous load, they must not eliminate **germane cognitive effort**, which is critical for deeper learning. This was especially visible in learners who preferred quick AI-generated solutions over reflective engagement (e.g., E5, Week 7).
- **Self-Determination Theory (Deci & Ryan, 1985)**: The framework addresses the foundational needs of **competence, autonomy, and relatedness**, which were inconsistently supported across groups. In some cases, AI enhanced confidence, while in others it reduced authentic engagement.
- **Sociocultural Theory (Vygotsky, 1978)**: AI is interpreted as a **cultural tool** mediating learning. However, unlike traditional tools, AI introduces **dynamic and evolving practices** that may conflict with learners' pre-existing norms or voice (as noted by E6, Week 5).
- **Cognitive Offloading Theory (Risko & Gilbert, 2016)**: The framework integrates this emerging theory to explain patterns of declining metacognitive effort when AI support replaces internal deliberation. This was reflected in the 85 coded instances of mental offloading in the reflective diaries.

These theoretical strands collectively inform the 9A Catalyst's emphasis on progression, personalisation, and pedagogical alignment, all of which are essential for fostering authentic, sustainable learner autonomy in the age of generative AI.

Pre-A Phase

(Foundation for Autonomy)

• Access
• Alignment
• Awareness
• Aim

↓

Practice-A Phase

(Active Engagement and Skill Building)

• Application
• Affective Engagement

↓

Performance-A Phase

(Ownership and Digital Competence)

• Agency
• AI Integration

↓

Personalisation-A Phase

(Internalisation and Extension)

• Autonomy
(Adjustment, Action)

The 9A Catalyst Framework appears to outline a gradual developmental trajectory in AI-integrated EFL writing instruction, potentially supporting learners as they move from foundational awareness to more personalised autonomy. In the Pre-A Phase, students are assumed to begin by gaining Access to AI tools, working toward Alignment with their goals, developing Awareness of digital learning contexts, and clarifying their Aim. This may lay the groundwork for the Practice-A Phase, where learners are likely to begin applying strategies and engaging affectively with writing tasks. As they progress, the Performance-A Phase seems to reflect increasing Agency and purposeful AI Integration. Eventually, in the Personalisation-A Phase, learners may demonstrate emerging Autonomy, showing signs of Adjustment and proactive Action. While the phases are not necessarily linear, the model suggests a dynamic and context-sensitive process of learner development in AI-enhanced educational settings.

5.5.5.1 Phase-Based Architecture: Sequential Development Model

To operationalise the theoretical structure of the 9A Catalyst Framework, this section maps the core components to empirical findings derived from coding patterns, student diaries, and reflective interviews. The framework's four developmental phases, Pre-A, Practice-A, Performance-A, and Personalisation-A, are rooted in observable learning behaviours and transitions across the 8-week intervention.

Each dimension of the framework is substantiated through thematic code frequencies and qualitative insights. Table 3 illustrates the alignment between the conceptual architecture and participant experiences, including both indicators of successful phase progression and markers of stagnation or regression.

Table 23

Evidence Supporting 9-A Framework Components

Framework Component	Supporting Evidence (Codes)	Success Indicators	Failure Indicators
Access	Digital literacy gaps (n = 71) Technical barriers (n = 287)	Seamless AI interaction Technical problem-solving	Persistent technical frustration Limited AI engagement

Framework Component	Supporting Evidence (Codes)	Success Indicators	Failure Indicators
Alignment	Cultural tensions (n = 149) Identity conflicts (n = 17)	Cultural preservation Authentic voice maintenance	Cultural displacement Identity confusion
Awareness	Metacognitive development (n = 32) Critical evaluation (n = 28)	Strategic AI usage Limitation recognition	Uncritical acceptance Reduced self-evaluation
Agency	Decision-making autonomy (n = 23) Learning ownership (n = 32)	Independent AI decisions Content negotiation	Passive AI acceptance External control reliance
Autonomy	Self-regulation (n = 35) Independent capability	Context-flexible competence Cognitive independence	AI dependency Transfer failure

Note. Framework components are presented in developmental order. Success indicators represent positive outcomes observed in participants who progressed effectively through framework phases.

5.5.5.2 Pre-A Phase: Foundation for Autonomy (Cognitive and Cultural Preparation)

The **Pre-A Phase** is foundational and precedes effective AI-instructed learning. This stage became particularly evident after the control group (C1–C6) outperformed the experimental group (E1–E12) in key aspects of writing performance. The results emphasise the irreplaceable role of traditional pedagogical and cognitive foundations before AI integration.

Access: Beyond Technical Availability to Comprehensive Digital Literacy

Access in this context refers not just to having AI tools but also to possessing the necessary digital literacy, prompt design capability, and readiness for critical engagement. Several students reported significant challenges here: E6 stated, *"I couldn't find my old conversations with ChatGPT,"* while E4 remarked, *"I don't know how to ask questions correctly."*

The following dimensions underpin the **Access** component:

- **Technical Infrastructure**: Consistent internet access, functional devices, and platform familiarity. In KYK dormitories, for example, E1 mentioned, "*Internet access was difficult."*
- **Digital Literacy**: Ability to navigate platforms and construct prompts. Students unfamiliar with AI tool syntax or commands struggled to initiate meaningful learning.
- **Cognitive Readiness**: Awareness of AI's affordances and limitations. Learners lacking such awareness often either over-relied on AI outputs or dismissed the tool entirely.
- **Cultural Accessibility**: Sensitivity to the linguistic and sociocultural norms embedded in AI outputs. Tools like ChatGPT often reflect Anglo-centric rhetorical styles that may distort or overshadow learners' cultural voices. Students like E10 expressed discomfort: *"My writing sounds too American with ChatGPT."*

Alignment: Navigating Cultural and Pedagogical Consistency

Alignment deals with whether AI use aligns with students' cultural identities and the pedagogical aims of instruction. Tensions arose when students felt alienated by the AI's tone, structure, or assumptions. E5 reflected, *"ChatGPT changed everything instead of just what I wanted."*

Key issues include:

- **Cultural Preservation**: Ensuring student voice is not overwritten by AI-generated patterns.
- **Pedagogical Congruence**: Ensuring that AI responses serve the intended instructional goal and don't bypass learning targets.

Awareness: Metacognitive and Ethical Awareness Development

Metacognitive growth was evident in some students who became more critical of AI suggestions. For example, E3 reported, "*I compare ChatGPT's feedback with other sources,"* showing increased self-evaluation.

However, others displayed uncritical usage, as seen in E7's note: *"It gives everything, so I don't check anymore."*

This stage must involve:

Reflection on AI Capabilities and Limits

Prompt Engineering Awareness

Self-monitoring Habits Development

Aim: Purpose-Driven AI Integration

Aim represents the goal-setting and purpose-driven aspect of learning that provides coherence and intentionality in AI-instructed environments. Goal-setting behaviours (n = 17) were evident in participants like E5, who reported, *"I set specific goals for each writing session"* (Week 7), highlighting this component's value in sustaining learner agency and direction.

The *Aim* component underscores the necessity of intentional AI integration aligned with meaningful educational goals. It begins with **Learning Objective Clarity**, where students are encouraged to articulate what they intend to achieve through AI-supported writing tasks. This ensures that AI remains a tool serving explicit learning intentions, rather than becoming an unfocused or habitual convenience. Alongside this, **Strategic Goal Hierarchy** refers to the alignment of short-term AI-assisted activities with long-term skill development, mitigating the risk of superficial fluency overtaking deeper linguistic growth.

Moreover, Assessment Criterion Development encourages both learners and educators to define measurable benchmarks that assess not only the quality of AI-enhanced output but also the learner's progress in independent cognitive capacity. Finally, Personal Vision Integration draws attention to the alignment between AI use and the learner's broader academic, professional, and personal aspirations. This alignment supports intrinsic motivation and guards against passive engagement with technology, fostering sustained, reflective, and purposeful learning trajectories.

5.5.5.3 Practice-A Phase: Active Engagement and Skill Building

The Practice-A Phase emphasises learners' active and emotionally engaged use of AI tools within authentic writing contexts. This phase is characterised by the gradual development of practical competence, motivational stability, and the emergence of strategic skillsets needed for sustainable AI integration. It also responds to the cognitive dependency patterns that emerged in learners who lacked foundational preparation, especially in the *Pre-A phase.*

Application: Authentic Context Integration

Application refers to the strategic embedding of AI tools within meaningful learning activities that mirror real-world academic and professional demands. The qualitative findings highlight cases of successful transfer, with learners demonstrating improved **strategic planning** (n = 25) and **textual organisation** as a result of AI-supported task engagement. These observed behaviours suggest that when AI use is grounded in relevant, purpose-driven tasks, it can yield transferable benefits rather than context-bound outcomes.

This component encompasses several interlinked dimensions:

Real-World Task Engagement: Learners must experience AI-enhanced writing within contexts that reflect their actual academic and communicative needs, such as formal essays, reflective reports, or professional emails, ensuring that skills developed are practically relevant.

Transfer Strategy Development: Students should be guided in applying AI-assisted gains across diverse future writing tasks. Without such deliberate strategy building, learners risk dependency on AI outputs that only serve immediate performance.

Creative Adaptation Skills: This dimension encourages learners to personalise AI tool functionalities based on their needs. Rather than relying on default outputs, learners should learn to manipulate, adapt, and refine AI suggestions critically.

Contextual Flexibility: Students need to learn when and where AI is appropriate, recognising, for instance, that while AI may assist with paragraph structure in an academic paper, it may be unsuitable for narrative tone in a reflective journal. This adaptability represents an advanced form of digital discernment and tool literacy.

Affective Engagement: Emotional Foundation for Sustained Learning

The role of emotion in AI-instructed learning is often underexamined. However, thematic analysis revealed that *affective engagement* played a pivotal role in both the motivation and long-term usage patterns of students. While motivation via AI feedback (n = 45) and reduced writing anxiety (n = 61) initially supported engagement, these benefits were not always

sustainable. Over time, some learners experienced emotional dependency and difficulty detaching from AI support, evident in dependency development codes (n = 68).

E1's early reflection that *"I'm not scared to make mistakes with ChatGPT"* (Week 1) exemplifies positive emotional safety. However, her later admission, *"Writing alone feels much harder now that I'm used to ChatGPT and I think it is addictive"* (Week 5), highlights how emotional ease can gradually evolve into dependency if left unmoderated.

This component can be defined through the following constructs:

Emotional Safety Development: A supportive context must be cultivated in which students feel comfortable experimenting with AI without fear of failure, while still retaining belief in their own capabilities.

Intrinsic Motivation Preservation: Teachers should help ensure that AI tools reinforce learners' internal motivation rather than replacing it. Sustained learning depends not just on the availability of support but on learners' self-driven desire to improve.

Confidence-Building Balance: Learners must build confidence in using AI without conflating tool fluency with true linguistic competence. The goal is to foster strategic trust in the tool alongside critical self-awareness of its limitations.

Curiosity and Exploration Maintenance: Students should be encouraged to use AI tools not merely as convenience mechanisms but as exploratory platforms. An inquisitive, questioning mindset, *"What if I phrased this differently?"*, can transform AI use into a space for creativity and discovery, rather than repetitive dependence.

5.5.5.4 Performance-A Phase: Ownership and Digital Competence

The *Performance-A Phase* marks a pivotal shift from scaffolded participation to self-regulated, strategic engagement with AI tools. This phase focuses on fostering learner agency and embedding AI into writing workflows in ways that enhance cognitive independence and writing proficiency without undermining authenticity or personal voice. It synthesises key developments in learner autonomy (n = 90 total codes) and advanced skill acquisition (n = 79 total codes), which distinguish adaptive, reflective users from cognitively dependent ones.

Agency: Self-Direction in AI-Instructed Learning

The *Agency* component refers to students' growing ability to make autonomous decisions, assume ownership over their learning processes, and engage critically with AI tools as partners rather than authorities. For example, E3 articulated, *"I choose which ChatGPT*

suggestions to use" (Week 8) and *"I decide when ChatGPT help is needed"* (Week 7), reflecting metacognitive awareness and decision-making autonomy. These comments demonstrate agency development grounded not merely in tool fluency, but in pedagogical judgment.

Yet, this apparent autonomy was often counterbalanced by evidence of cognitive offloading (n = 85) and responsibility transfer (n = 22), suggesting that superficial agency can mask deeper reliance on AI-generated outputs. Thus, nurturing authentic agency requires both strategic tool use and critical resistance to over-dependence.

This component comprises four interrelated elements:

Decision-Making Autonomy: Learners must develop the capacity to decide *when*, *how*, and *why* to use AI, based on instructional needs rather than habitual reliance.

Learning Process Ownership: Students should take responsibility for their academic outputs, treating AI feedback as supplementary, not determinative.

Content Negotiation Skills: Learners need to critically appraise, question, and revise AI-generated text to align with their original intent and contextual demands.

Voice Preservation Strategies: Central to performance is maintaining the student's stylistic identity. AI tools should support, not overwrite, the learner's tone, personality, or rhetorical positioning, particularly essential for developing creative and persuasive writing.

AI Integration: Strategic Tool Orchestration

AI Integration represents a higher-order capacity to incorporate generative AI into writing practice in a way that optimises productivity while preserving critical and metacognitive engagement. The development of prompt engineering strategies (n = 28) and the evolution of strategic questioning behaviours (n = 22) suggest that some students began to orchestrate AI tools deliberately as part of a reflective writing workflow.

For instance, E11 noted, *"I learned to ask better questions to get useful answers"* (Week 6), demonstrating metacognitive refinement. However, E5's comment, *"I skip thinking about why and just ask ChatGPT directly"* (Week 7), reveals that integration can also slide into convenience-driven usage patterns, bypassing cognitive effort.

This component includes the following subdimensions:

Workflow Optimisation: Students must learn to embed AI into their writing routines without fragmenting attention or interrupting their cognitive engagement.

Strategic Usage Patterns: Effective learners alternate between AI support and independent effort, avoiding both overreliance and tool underutilisation.

Iterative Refinement Skills: Learners should continuously adapt their prompting, verification, and synthesis strategies, fine-tuning their interactions based on outcomes and reflection.

Ethical Framework Development: Beyond technical skill, students must cultivate personal guidelines for responsible AI use. These should include considerations of originality, academic honesty, audience expectations, and cultural voice preservation.

Together, *Agency* and *AI Integration* in this phase reflect learners' transition from passive users to reflective orchestrators of technology. However, the data also show that such performance is not universally achieved; sustained support, ethical awareness, and instructional scaffolding remain critical to ensure that strategic usage translates into lasting competence rather than superficial fluency.

5.5.5.5 Personalisation-A Phase: Internalisation and Extension

The *Personalisation-A Phase* represents the highest stage of AI-instructed autonomy development, where learners have successfully internalised the cognitive and procedural affordances of AI tools while preserving, and often enhancing, their independent thinking capabilities. This phase builds upon prior development and addresses the refined forms of autonomous learning (n = 257 total codes) required to function across diverse writing contexts without sacrificing cognitive depth, voice, or agency. It reflects the ultimate aim of the 9A Catalyst Framework: the formation of adaptive, critically aware, and metacognitively sophisticated learners who strategically integrate AI while maintaining full cognitive control.

Autonomy: Self-Directed Learning Mastery

Autonomy in this final phase is not synonymous with mere independence. Instead, it reflects a redefined autonomy grounded in strategic interdependence, the ability to collaborate with tools like ChatGPT while consciously preserving one's own voice, reasoning, and reflective capacity. For example, E3 noted, *"I plan my writing sessions better now with ChatGPT help"* (Week 7), illustrating how AI can be embedded into structured self-regulation routines rather than displacing them.

This component comprises several interconnected capacities:

- **Independent Learning Design:** Learners initiate and follow through on goal-driven learning plans that incorporate AI use in a balanced, intentional manner, ensuring that writing tasks remain cognitively engaging and personally meaningful.

- **Context-Flexible Competence:** Students exhibit proficiency across varied tasks and settings, maintaining performance regardless of AI availability. This adaptability shows that the learner's skill is internalised rather than contingent upon tool use.
- **Self-Regulation Mastery:** Learners sustain self-monitoring and effort management over time, resisting the temptation of shortcut solutions and retaining ownership over task complexity and pacing.
- **Cognitive Independence Preservation:** Perhaps most crucially, students maintain the capacity to think critically, reason independently, and generate original ideas, whether AI is present or not. This ensures that AI supports, rather than supplants, the core purposes of education.

Adjustment: Adaptive Strategy Development

Adjustment refers to the learner's ability to iteratively refine their AI engagement strategies in response to emerging needs, feedback, and shifting technological or academic demands. The qualitative data reflect this adaptive capacity. For instance, E1 remarked, *"I check ChatGPT suggestions before using them"* (Week 5), exemplifying verification behaviour (n = 19) and cautious integration (n = 28).

This component encompasses the following elements:

Continuous Strategy Refinement: Learners evaluate their evolving relationship with AI, identifying which strategies yield the best outcomes and modifying unproductive habits, such as blind acceptance or overuse of AI suggestions.

Contextual Adaptation: Students adjust their prompting, usage, and editing strategies to fit different writing purposes, genres, and audiences. This flexibility protects against generic, one-size-fits-all writing patterns.

Feedback Integration: Learners actively incorporate insights from peer review, teacher feedback, and self-assessment to improve their use of AI. This includes refining prompts, recognising output limitations, and recalibrating engagement based on the evolving learning context.

Anticipatory Adaptation: As a mark of high-level autonomy, students begin to anticipate future challenges or shifts in task expectations, preparing their strategies in advance. This readiness helps them stay agile and future-oriented in AI-enhanced academic landscapes.

5.6 Practical Implications for AI-Integrated EFL Writing Instruction

This section presents a comprehensive roadmap for translating the 9A Catalyst Framework into actionable pedagogical strategies, institutional planning, and differentiated

student support mechanisms. The proposed practices are grounded in the empirical insights from RQ1–RQ4 and aim to guide stakeholders in responsibly integrating AI into EFL writing instruction without compromising learner autonomy, cognitive challenge, or cultural authenticity.

5.6.1 Institutional Implementation Roadmap

Pre-A Phase: Foundational Infrastructure and Readiness

At the institutional level, this phase focuses on preparing the necessary technical, cultural, and cognitive infrastructure before AI tools are introduced:

Access Infrastructure Development

Institutions should ensure:

- Reliable and equitable access to internet, devices, and AI platforms.
- Integration of a digital literacy curriculum in early instructional phases.
- Cognitive readiness training for both educators and learners.
- On-demand IT support to mitigate technical interruptions.

Alignment Policy Development

Institutions must:

- Acknowledge and address cultural identity tensions by developing inclusive AI policies.
- Ensure that AI integration aligns with national curriculum standards and pedagogical goals.
- Promote coherence between AI tool functionality and intended learning outcomes.

Awareness Training Programmes

Institutions should:

- Offer metacognitive strategy workshops focused on AI limitations and responsible usage.
- Provide reflective training sessions to promote evaluative engagement over dependency.

Aim Setting and Outcome Assessment

Institutions must:

- Embed outcome-driven AI usage policies.
- Develop assessment rubrics that measure autonomy, transfer of learning, and AI-independent skill growth.

5.6.2 Educator Preparation and Professional Development

To ensure meaningful AI integration, educators require phase-specific competencies:

Pre-A Phase Competencies

Educators should be trained to:

- Assess students' digital readiness and detect cultural conflict indicators.
- Introduce metacognitive routines (e.g., goal setting, reflection).
- Apply culturally responsive pedagogy in AI-enhanced environments.

Practice-A Phase Support Skills

Training should enable teachers to:

- Cultivate motivation without fostering tool dependency.
- Supervise AI-supported writing with balanced guidance.
- Recognise and address emerging patterns of affective overreliance.

Performance-A Phase Facilitation

Professional development must include:

- Tools to evaluate authentic versus superficial agency.
- Methods for promoting student voice and resisting cognitive substitution.
- Dialogic strategies that support intellectual negotiation of AI feedback.

Personalisation-A Phase Extension

Educators should be equipped to:

- Recognise signs of internalisation and scaffold advanced strategy transfer.
- Design reflective, student-led projects (e.g., AI critiques, ethics audits) that consolidate autonomy.

5.6.3 Assessment and Monitoring Guidelines

Phase-Specific Evaluation Protocols

- Develop rubrics aligned with 9A phases (e.g., Access, Awareness, Agency).
- Use mixed-method approaches: AI-use logs, diaries, writing samples, and student reflections.

Dual-Context Performance Assessment

- Assess performance both with and without AI tools to isolate authentic skill development.
- Include pre- and post-tests based on real-life writing contexts.

Cultural Sensitivity Monitoring

- Use surveys/interviews to assess students' perceptions of cultural alignment or dissonance in AI output.
- Offer culturally adaptive strategies (e.g., voice retention modules, multilingual AI prompt tutorials).

5.6.4 Student Support Strategies Across Phases

Pre-A Phase Support

- Workshops on:
    - Technical and digital navigation skills.
    - Ethical AI use and cognitive boundaries.
    - Goal-setting aligned with academic vision.

Practice-A Phase Support

- Provide:
    - Guided AI-writing clinics.
    - Reflection logs and dependency tracking tools.
    - Affective journals for emotional responses to AI engagement.

Performance-A Phase Facilitation

- Offer:
    - Structured decision-making trees for AI prompt formulation.
    - Micro-courses on voice preservation and authorial intent.
    - Feedback literacy training to distinguish between AI and human feedback strengths.

Personalisation-A Phase Empowerment

- Implement:
    - Independent creative projects using AI (e.g., ecopoetry generation, collaborative editing).
    - Peer-led AI critique forums.
    - Personal AI-integration portfolios that track strategic growth over time.

5.6.5 Differentiated Support Strategies

Recognising learner variability, targeted interventions should be employed to support diverse needs:

Students with High Digital Literacy

- Provide:
    - “Slow thinking” strategies to reduce AI overuse.
    - Writing tasks requiring initial drafts without AI before revision using AI.

Culturally Diverse Learners

- Facilitate:
    - Modules on cultural adaptation and linguistic ownership.
    - Class discussions on AI biases, neutrality, and identity preservation.

Learners with Low Confidence

- Offer:
    - Frequent human encouragement alongside AI feedback.
    - Affective scaffolding with a gradual reduction of support.

High-Achieving Students

- Assign:
    - Critical AI usage audits examining their own dependency levels.
    - Meta-analytical tasks that explore limitations and ethical dimensions of AI-generated writing.

5.6.5 Differentiated Support Strategies

Recognising the heterogeneous nature of learner profiles in EFL contexts, effective AI-integrated instruction must be underpinned by responsive, differentiated support mechanisms. Tailored pedagogical approaches are necessary to address diverse learner needs, capabilities, and cultural contexts across the phases of the 9A Catalyst Framework.

Students with High Digital Literacy

Learners with advanced digital fluency may develop automation habits that circumvent meaningful cognitive engagement.

- **Cognitive Deceleration Strategies**: Introduce intentional *cognitive resistance* activities, such as writing initial drafts without AI support before using it for revision and feedback comparison.
- **Reflective Journaling**: Encourage students to document and evaluate their reasoning for when and why they choose to consult AI tools.

- **Critical Prompt Design**: Assign projects requiring students to craft varied prompts and reflect on their cognitive implications and output differences.

Culturally Diverse Learners

Students from different linguistic and cultural backgrounds may experience misalignment between AI-generated language and their personal or community identities.

- **Cultural Adaptation Modules**: Offer instruction on tailoring AI prompts and outputs to preserve cultural nuance and authenticity in student writing.
- **Culturally Sustaining Pedagogy**: Embed tasks that affirm students' cultural voices, such as assignments allowing comparison between AI outputs and culturally preferred styles or idioms.
- **Bias Awareness Discussions**: Facilitate classroom debates or reflections on AI bias, neutrality, and its implications for voice preservation in writing.

Learners with Low Confidence

Students with lower self-efficacy may become over-reliant on AI due to perceived inadequacies in their own ability.

- **Human-AI Feedback Pairing**: Provide consistent, personalised teacher feedback alongside AI input to maintain motivational balance.
- **Gradual Support Fading**: Scaffold learning through guided tasks that reduce AI involvement incrementally, fostering self-confidence and skill ownership.
- **Affective Coaching**: Offer emotional support strategies (e.g., positive reinforcement scripts, peer feedback circles) to build writing self-assurance.

High-Achieving Students

Students with high language proficiency may plateau if AI tools are used only for surface-level enhancement.

- **Critical AI Audits**: Assign meta-cognitive tasks that require students to critique their own use of AI, reflecting on when it enhanced or inhibited their creativity and critical thought.

- **Voice Preservation Challenges**: Design activities that explicitly require maintenance of individual style and complexity, despite AI assistance (e.g., stylistic mimicry or re-voicing tasks).
- **Innovation Extension Projects**: Encourage these students to experiment with advanced AI uses such as multimodal composition or comparative analysis across AI platforms.

5.7 Major Findings Summary

RQ1: How does AI-instructed writing instruction influence learner autonomy in EFL contexts?

Paradox of Strategic Autonomy vs. Cognitive Dependency: Students demonstrated sophisticated prompt usage (e.g., selecting/rejecting suggestions) while simultaneously offloading thinking to ChatGPT, leading to cognitive atrophy and reduced critical engagement.

Metacognitive Gains and Surface Autonomy: Autonomy development appeared procedural (prompt engineering, revision habits) rather than generative (independent thought).

Temporal Progression Evidence: Diary data revealed four stages of perception, initial enthusiasm, strategic experimentation, dependency, and critical reflection.

RQ2: What changes are observed in learners' writing performance following AI-instructed writing instruction?

Control Group Superiority in Mechanics and Artistic Language: Despite EG's grammar and structure gains, CG outperformed in punctuation, voice, and creativity.

AI-Instructed Group's Fluency at the Cost of Depth: ChatGPT improved syntactic clarity but reduced idiomatic expression and figurative language.

Perception–Performance Mismatch: Students felt motivated and confident with AI feedback, but this did not translate to significant post-test gains in stylistic or mechanical precision.

RQ3: What are students' perceptions of AI integration in writing instruction?

Collaborative and Conditioned Autonomy Emerged: Students relied heavily on social scaffolding (friends, siblings, TikTok) to navigate AI usage, revealing interdependent rather than isolated autonomy.

Motivation Without Mastery: High affective engagement (n = 236) did not guarantee long-term retention or skill transfer, echoing concerns about "motivated dependency."

Awareness of Voice Loss and AI's Limits: Students acknowledged that AI improved clarity but flattened personal tone and creativity, seen in reduced artistic language and homogenised structure.

RQ4: How can a framework be developed to guide AI-instructed autonomy in EFL writing instruction?

9A Catalyst Framework Proposed: A four-phase model (Pre-A, Practice-A, Performance-A, Personalisation-A) mapping Access, Awareness, Application, Agency, and Autonomy onto learner development.

Framework Validated via Mixed Methods: Strategic planner students who reached Performance-A Phase showed better writing outcomes and metacognitive control.

Barriers to Framework Completion: Many EG students remained in Practice-A Phase, citing technical, cultural, or motivational gaps.

CHAPTER VI
CONCLUSION

This chapter concludes the study by synthesising its principal findings, evaluating the explanatory power and limitations of the 9A Catalyst Framework, and outlining implications for future research. The investigation sought to explore how generative AI tools influence learner autonomy and writing performance in EFL contexts, with a focus on the interplay between cognitive development, instructional context, and technological mediation.

6.1 Summary of Findings and Framework Validation

6.1.1 Quantitative–Qualitative Convergence

The mixed-methods design enabled a nuanced understanding of the paradoxes embedded in AI-instructed learning. While quantitative data revealed significant improvements in writing mechanics and self-reported autonomy for the experimental group, qualitative findings painted a more complex picture. Learners exhibited strategic AI engagement behaviours (e.g., prompt engineering, revision planning), but also showed signs of cognitive offloading, dependency, and reduced critical evaluation. This divergence underscores the importance of integrating multiple data streams to avoid overly optimistic interpretations of surface-level gains.

The 9A Catalyst Framework proved useful in interpreting these complexities. It accounted for both successful developmental trajectories, marked by Agency and Autonomy phases, and less favourable stagnation patterns observed in students who failed to progress beyond the Practice-A Phase. The strategic planning behaviours (n = 25), ownership development (n = 32), and iterative prompting techniques (n = 28) that emerged in some students aligned with the predicted developmental sequence. Conversely, persistent mechanical errors, reduced artistic language, and emotional reliance on AI tools reinforced the framework's emphasis on the need for foundational readiness in the Pre-A Phase.

6.1.2 Temporal Developmental Patterns

Student diary entries, interviews, and coded qualitative data revealed clear temporal patterns supporting the framework's phased design. In the initial weeks, learners grappled with Access, Alignment, and Awareness challenges, consistent with Pre-A Phase expectations. During Weeks 3–6, engagement shifted toward Application and Affective Engagement, albeit with signs of emerging cognitive dependency. In the final phase, only a subset of learners

progressed to Agency, AI Integration, and personalised strategy development, validating the framework's sequential logic while highlighting common bottlenecks in learner development.

6.2 Contradictions and Theoretical Contributions

6.2.1 Control Group Outperformance and Pedagogical Context Primacy

One of the most unexpected findings was the control group's superior performance in mechanics and expressive writing dimensions, despite lacking access to AI tools. This challenges the common narrative that technology alone enhances learning and instead reinforces the study's central interpretive principle of Pedagogical Context Primacy: that the quality of instructional design, emotional scaffolding, and reflective engagement ultimately determines learning outcomes, regardless of technological sophistication.

This divergence between groups strengthens the argument that AI is not a substitute for pedagogy. Rather, when misaligned with learners' readiness and instructional context, AI can unintentionally undermine skill development through over-scaffolding, content standardisation, or reduced cognitive challenge.

6.2.2 Redefining Autonomy and Critical Thinking in AI-instructed Contexts

This study contributes to evolving definitions of learner autonomy and critical thinking in digitally mediated environments. The emergence of conditioned, artificial, and collaborative autonomy types among students suggests that autonomy is no longer reducible to individual self-regulation but must also account for strategic interdependence, peer dynamics, and AI collaboration.

Moreover, critical thinking was shown to fragment into procedural (e.g., AI query selection), adaptive (tool negotiation), and generative (original content creation) forms. While procedural skill increased, generative critical capacity often declined. This necessitates an expanded theoretical framing that accommodates segmented, context-sensitive manifestations of cognitive engagement.

6.3 Contributions of the 9A Catalyst Framework

The 9A Catalyst offers a developmental model for AI integration that surpasses binary pro/con views of AI-assisted learning. Its four-phased structure, Pre-A, Practice-A, Performance-A, and Personalisation-A, provides educators, policymakers, and curriculum designers with a roadmap for gradual, reflective, and pedagogically grounded AI

implementation. It integrates theoretical perspectives from Cognitive Load Theory, Self-Determination Theory, Sociocultural Theory, and Cognitive Offloading Theory, bridging psychological needs, cultural mediation, and technological fluency.

The framework advances the understanding that digital literacy and emotional readiness must precede AI adoption; that affective engagement must be cultivated without triggering dependency; and that true autonomy and agency must be assessed not by AI usage frequency but by learners' discernment, voice preservation, and adaptive capabilities.

6.4 Limitations

Despite its insights, the study faced limitations. The eight-week intervention may have been insufficient for full development through all 9A phases. While pseudonyms protected identity, individual variations in usage patterns may have required more granular, longitudinal tracking. Additionally, the general-purpose nature of ChatGPT may not fully reflect the potential of educational AI tools designed with pedagogical intent. Cultural and linguistic context, specific to Turkish EFL learners, may also limit generalisability.

6.5 Directions for Future Research

Building on these findings, future studies should:

Conduct longitudinal studies to evaluate how autonomy and cognitive dependency evolve over extended periods.

Explore the effectiveness of purpose-built AI systems that incorporate Socratic questioning, metacognitive prompts, and feedback literacy scaffolds.

Examine teacher training programmes' influence on AI integration outcomes, particularly in relation to teacher autonomy and pedagogical implementation.

Investigate cultural and affective dimensions of AI usage across diverse EFL contexts to inform culturally responsive AI integration.

Evaluate the applicability of the 9A Catalyst in different disciplines and educational levels (e.g., secondary vs. tertiary EFL settings).

6.6 Final Reflections

This study has revealed that authentic learning in AI-instructed contexts depends not on the tool itself, but on how the tool is taught, framed, and embedded within a pedagogically coherent environment. Learners may develop impressive procedural fluency with AI while experiencing subtle regressions in creative, reflective, and critical capacities. Thus, the role of

educators must evolve, from content deliverers to autonomy designers, equipping students not only with access to tools but also with the cognitive resilience, ethical frameworks, and personal voice necessary for long-term success.

While AI tools like ChatGPT may support efficiency, they do not guarantee depth and are not recommended in the initial learning stages as it was hyped in the literature due to the possible dependency, halting the learning process and showing artificial autonomy. The future of writing instruction lies in balancing the power of automation with the necessity of thoughtful, human-centred pedagogy.

REFERENCES

Ahmed, M. K. (2024). The interplay between learner autonomy and motivational factors in EFL contexts. Journal of Language Teaching and Research, 15(2), 342–357. https://doi.org/10.17507/jltr.1502.12

Alghasab, M. B. (2025). English as a foreign language (EFL) secondary school students' use of artificial intelligence (AI) tools for developing writing skills: unveiling practices and perceptions. Cogent Education, 12(1). https://doi.org/10.1080/2331186X.2025.2505304

Alm, A. (2024). Exploring autonomy in the AI wilderness: Learner challenges and choices. Education sciences, 14(12), 1369.

Alsofyani, A. H., & Barzanji, A. M. (2025). The effects of ChatGPT-generated feedback on Saudi EFL learners' writing skills and perception at the tertiary level: A mixed-methods study. Journal of Educational Computing Research, 63(2), 431–463. https://doi.org/10.1177/07356331241307297

Alsofyani, A. H., & Barzanji, A. M. (2025). The effects of ChatGPT-generated feedback on Saudi EFL learners' writing skills and perception at the tertiary level: A mixed-methods study. Journal of Educational Computing Research, 63(2), 431-463. https://doi.org/10.1177/07356331241307297

American Psychological Association. (2020). *American Psychological Association yayın kılavuzu* (7. baskı). American Psychological Association. Asadi, M., Ebadi, S., & Mohammadi, L. (2025). The impact of integrating ChatGPT with teachers' feedback on EFL writing skills. Thinking Skills and Creativity, 56, 101766. https://doi.org/10.1016/j.tsc.2025.101766

Apriani, E., Cardoso, L., Obaid, A. J., Wijayanti, E., Esmianti, F., & Supardan, D. (2024). Impact of AI-Powered ChatBots on EFL Students' Writing Skills, Self-Efficacy, and Self-Regulation: A Mixed-Methods Study. Global Educational Research Review, 1(2), 57-72.

Arias-Vásquez, M. E. (2022). Principles, scope, and limitations of methodological triangulation. International Journal of Mixed Methods Research, 16(1), 45–62.https://pmc.ncbi.nlm.nih.gov/articles/PMC9714985/?utm_source=chatgpt.com

Asadi, M., Ebadi, S., & Mohammadi, L. (2025). The impact of integrating ChatGPT with teachers' feedback on EFL writing skills. Thinking Skills and Creativity, 56, Article 101766. https://doi.org/10.1016/j.tsc.2025.101766

Ataş, A. H., Cengiz, B. C., & Çelik, B. (2024). Evaluating ChatGPT in Generating Feedback on Content and Organization Components of EFL Compare and Contrast Essays. Sakarya University Journal of Education, 14(3), 519-542. https://doi.org/10.19126/suje.1475474

Barrot, J. S. (2023). Using automated written corrective feedback in the writing classrooms: Effects on L2 writing accuracy. Computer Assisted Language Learning, 36(4), 584-607.

Benson, P. (2011). Teaching and researching autonomy in language learning (2nd ed.). Routledge.

Bereiter, C., & Scardamalia, M. (1987). The psychology of written composition. Lawrence Erlbaum.

Bjork, R. A., & Bjork, E. L. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society, 56–64.

Braunstein, I., & Maes, P. (2025). Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task (arXiv:2506.08872). arXiv. https://doi.org/10.48550/arXiv.2506.0887

Cambridge, E. S. O. L., Cresswell, J. W., Plano Clark, V. L., & Denscombe, M. (2012). student proficiency: improving listening, reading and. ISSN 1756-509X, 42.

Campbell, Z. (2025, June 25). Productive struggle: How artificial intelligence is changing learning, effort, and youth development in education. Bellwether. https://bellwether.org/publications/productive-struggle/

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

Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. TechTrends, 66(4), 616–630. https://doi.org/10.1007/s11528-022-00715-y

Çelik, S., Arkın, E., & Sabriler, D. (2012). EFL Learners' Use of ICT for Self-Regulated Learning. Journal of Language and Linguistic Studies, 8(2), 0-118.

Chen, J. (2025). Writing and Revising in Second Language Classrooms: The Role of Self-Regulation in Cultivating Expert Writers. Springer Nature.

Chen, X., Liu, Y., & Zhang, R. (2025). AI-assisted writing: Insights into ChatGPT's impact on EFL writing fluency. Journal of Language Learning Systems, 23(4), 45–67.

Chen, Y. (2025). Transforming English language education with AI-driven deep learning models for scalable adaptive and inclusive assessment. International Journal of Information and Communication Technology, 26(12), 15-31. https://doi.org/10.1504/IJICT.2025.146163

Chong, S. W., & Reinders, H. (2022). Autonomy of English language learners: A scoping review of research and practice. Language Teaching Research. Advance online publication.

Chong, S. W., Khan, M. A., & Reinders, H. (2022). A critical review of design features of LMOOCs. Computer Assisted Language Learning, 37(3), 389–409. https://doi.org/10.1080/09588221.2022.2038632

Clarke, V., & Braun, V. (2014). Thematic analysis. In Encyclopedia of critical psychology (pp. 1947-1952). Springer, New York, NY.

Collins, A., & Kapur, M. (2006). Cognitive apprenticeship (Vol. 291). na: na. https://www.psy.lmu.de/isls-naples/intro/all-webinars/collins/cognitive-apprenticeship.pdf

Creswell, J. W., & Plano Clark, V. L. (2011). Designing and conducting mixed methods research (2nd ed.). Sage.

Creswell, J. W., & Poth, C. N. (2016). Qualitative inquiry and research design: Choosing among five approaches. Sage publications.

Cui, Y. (2025). What influences college students using AI for academic writing? A quantitative analysis based on HISAM and TRI theory. Computers and Education: Artificial Intelligence, 8, Article 100391. https://doi.org/10.1016/j.caeai.2025.100391

Darling-Hammond, L. (2010). Teacher education and the American future. Journal of teacher education, 61(1-2), 35-47. https://doi.org/10.1177/0022487109348024

Darling-Hammond, L. (2010). The flat world and education: How America's commitment to equity will determine our future. Teachers College Press.

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. https://doi.org/10.1080/2331186X.2023.2290342

Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. Journal of research in personality, 19(2), 109-134. https://doi.org/10.1016/0092-6566(85)90023-6

Dergaa, I., Ben Saad, H., Glenn, J. M., Amamou, B., Ben Aissa, M., Guelmami, N., ... & Chamari, K. (2024). From tools to threats: a reflection on the impact of artificial-

intelligence chatbots on cognitive health. Frontiers in psychology, 15, 1259845. https://doi.org/10.3389/fpsyg.2024.1259845

Dörnyei, Z. (2001). Motivational strategies in the language classroom. Cambridge University Press.

Dörnyei, Z., & Ushioda, E. (2021). Teaching and researching motivation in language learning (3rd ed.). Routledge. https://doi.org/10.4324/9781003135725

Falloon, G. (2020). From digital literacy to digital competence: the teacher digital competency (TDC) framework. Educational technology research and development, 68(5), 2449-2472. https://doi.org/10.1007/s11423-020-09767-4

Fan, Y., Bannert, M., & Molenaar, I. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. British Journal of Educational Technology, 56(2), 445-462. https://doi.org/10.1111/bjet.13544

Fawns, T. (2022). An entangled pedagogy: Looking beyond the pedagogy–technology dichotomy. Postdigital Science and Education, 4, 711–728. https://doi.org/10.1007/s42438-022-00302-7

Field, A. (2018). Discovering statistics using IBM SPSS statistics (5th ed.). SAGE Publications.

Fleckenstein, J., Liebenow, L. W., & Meyer, J. (2023). Automated feedback and writing: A multi-level meta-analysis of effects on students' performance. Frontiers in Artificial Intelligence, 6, 1162454. https://doi.org/10.3389/frai.2023.1162454

Frye, K. E., Boss, D. L., Anthony, C. J., Du, H., & Xing, W. (2024). Content analysis of the CASEL framework using K–12 state SEL standards. School Psychology Review, 53(3), 208-222. https://doi.org/10.1080/2372966X.2022.2030193

Gharaibeh, M., Ayasrah, M. N., & Almulla, A. A. (2025). Supplemental role of ChatGPT in enhancing writing ability for children with dysgraphia in the Arabic language. Education and Information Technologies, 1-21.

Gharaibeh, M., Ayasrah, M. N., & Almulla, A. A. (2025). Supplemental role of ChatGPT in enhancing writing ability for children with dysgraphia in the Arabic language. Education and Information Technologies. https://doi.org/10.1007/s10639-025-13605-6

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. International journal of endocrinology and metabolism, 10(2), 486. https://doi.org/10.5812/ijem.3505.

Ginns, P., & Leppink, J. (2019). Special issue on cognitive load theory. Educational Psychology Review, 31(2), 255-259. https://doi.org/10.1007/s10648-019-09474-4

Gupta, N., Ali, K., Jiang, D., Fink, T., & Du, X. (2024). Beyond autonomy: Unpacking self-regulated and self-directed learning through the lens of learner agency, a scoping review. BMC Medical Education, 24, Article 1519.

Haidar, S. (2025). Impact of the integration of metacognitive prompts by Generative Artificial Intelligence (GenAI) in collaborative and individual learning in improving writing skills and metacognitive awareness. International Journal of Learning, Teaching and Educational Research, 24(1), 156-174. https://doi.org/10.26803/ijlter.24.1.9

Hastomo, T., Sari, A. S., Widiati, U., Ivone, F. M., Zen, E. L., & Andianto, A. (2025). Exploring EFL teachers' strategies in employing AI chatbots in writing instruction to enhance student engagement. World Journal of English Language, 15(7), 93. https://doi.org/10.5430/wjel.v15n7p93,

Hayes & Flower's (1980) model shows how planning, translating and reviewing draw on limited executive resources, laying groundwork for Kellogg's later work (1996).

Hayes, J. R., & Flower, L. S. (1980). Identifying the organisation of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), Cognitive processes in writing (pp. 3–30). Erlbaum.

Holec, H. (1981). Autonomy and foreign language learning. Pergamon Press.

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., Santos, O. C., Rodrigo, M. T., Cukurova, M., Bittencourt, I. I., & Koedinger, K. R. (2022). Ethics of AI in education: Towards a community-wide framework. International Journal of Artificial Intelligence in Education, 32(3), 504–526. https://doi.org/10.1007/s40593-021-00239-1

Jose, B., Cherian, J., Verghis, A. M., Varghise, S. M., S., M., & Joseph, S. (2025). The cognitive paradox of AI in education: Between enhancement and erosion. Frontiers in Psychology, 16. https://doi.org/10.3389/fpsyg.2025.1550621

Kellogg, R. T. (1996). The psychology of writing. Oxford University Press.

Kim, M., & Chon, Y. V. (2025b). The impact of self-revision, machine translation, and ChatGPT on L2 writing: Raters' assessments, linguistic complexity, and error correction [Duplicate entry]. Assessing Writing, 65. https://doi.org/10.1016/j.asw.2025.100950

Kim, M., Kim, J., Knotts, T. L., & Albers, N. D. (2025). AI for academic success: Investigating the role of usability, enjoyment, and responsiveness in ChatGPT adoption. Education and Information Technologies.

Koedinger, K. R. (2022). Ethics of AI in education: Towards a community-wide framework. International Journal of Artificial Intelligence in Education, 32(3), 504–526. https://doi.org/10.1007/s40593-022-00296-6

König, J., Krepf, M., Bremerich-Vos, A., & Buchholtz, C. (2021). Meeting cognitive demands of lesson planning: introducing the CODE-PLAN model to describe and analyze teachers' planning competence. The Teacher Educator, 56(4), 466–487. https://doi.org/10.1080/08878730.2021.1938324

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. https://doi.org/10.1016/j.jcm.2016.02.012

Koraishi, O. (2023). Teaching English in the age of AI: Embracing chatGPT to optimize EFL materials and assessment. Language Education and Technology, 3(1), Article 1. https://langedutech.com/letjournal/index.php/let/article/view/48

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X. H., Beresnitzky, A. V., ... & Maes, P. (2025). Your brain on chatgpt: Accumulation of cognitive debt when using an ai assistant for essay writing task. arXiv preprint arXiv:2506.08872. https://doi.org/10.48550/arXiv.2506.08872

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V.

Krashen, S. (1982). Principles and practice in second language acquisition.

Krashen, S. D. (1982). Principles and practice in second language acquisition. Pergamon Press.

Kuckartz, U., & Rädiker, S. (2021). Using MAXQDA for mixed methods research. In The Routledge Reviewer's Guide to Mixed Methods Analysis. Routledge.

Kumar, S. (2023). Developing human skills in the era of artificial intelligence: Challenges and opportunities for education and training. Scholedge International Journal of Multidisciplinary & Allied Studies ISSN 2394-336X, 10(2), 11–19. https://doi.org/10.19085/sijmas100201

Kurt, G., & Kurt, Y. (2024). Enhancing L2 writing skills: ChatGPT as an automated feedback tool. Journal of Information Technology Education: Research, 23, 024. https://doi.org/10.28945/5370

Li, D. J., Kao, Y. C., Tsai, S. J., Bai, Y. M., Yeh, T. C., Chu, C. S., ... & Su, K. P. (2024). Comparing the performance of ChatGPT GPT-4, Bard, and Llama-2 in the Taiwan psychiatric licensing examination and in differential diagnosis with multi-center psychiatrists. Psychiatry and Clinical Neurosciences, 78(6), 347-352. https://doi.org/10.1111/pcn.13656

Li, M., Wang, Y., & Yang, X. (2025). Can generative AI chatbots promote second language acquisition? A meta-analysis. Journal of Computer Assisted Learning, 41(4), e70060. https://doi.org/10.1111/jcal.70060

Lindebaum, D., & Fleming, P. (2024). ChatGPT undermines human reflexivity, scientific responsibility and responsible management research. British Journal of Management, 35(2), 566-575. https://doi.org/10.1111/1467-8551.12781

Little, D. (1995). Learning as dialogue: The dependence of learner autonomy on teacher autonomy. System, 23(2), 175-181.https://doi.org/10.1016/0346-251X(95)00006-6

Littlewood, W. (2019). Learner autonomy in the age of AI. ELT Journal, 73(2), 129–137. https://doi.org/10.1093/elt/ccy045

Liu, X., Wang, S., & Chen, Y. (2024). Impact of ChatGPT on ESL students' academic writing skills. Smart Learning Environments, 11(1), Article 26. https://doi.org/10.1186/s40561-024-00295-9

Luo, M., Hu, X. & Zhong, C. The collaboration of AI and teacher in feedback provision and its impact on EFL learner's argumentative writing. Educ Inf Technol (2025). https://doi.org/10.1007/s10639-025-13488-7

Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. Smart Learning Environments, 11, 9. https://doi.org/10.1186/s40561-024-00295-9

Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. Smart Learning Environments, 11(1), Article 9. https://doi.org/10.1186/s40561-024-00295-9

Marsman, M., & Wagenmakers, E.-J. (2017). Bayesian benefits with JASP. The European Journal of Developmental Psychology, 14(5), 545-555. https://doi.org/10.1080/17405629.2016.1259614

Marsman, M., & Wagenmakers, E.-J. (2017). Three insights from a Bayesian interpretation of the one-sided P value. Educational and Psychological Measurement, 77(3), 529-539. https://doi.org/10.1177/0013164416669201

Milberg, T. (2024, April). The future of learning: How AI is revolutionizing education 4.0. World Economic Forum. https://www.weforum.org/stories/2024/04/future-learning-ai-revolutionizing-education-4-0

Misirli, A. T., & Ergulec, F. (2021). Transformative agency: An investigation into how one teacher taught during the COVID-19 pandemic. Learning, Culture and Social Interaction, 30, 100584. https://doi.org/10.1016/j.lcsi.2021.100584

Namaziandost, E., & Rezai, A. (2024). Under the world of AI-generated feedback on writing: Mirroring motivation, foreign language peace of mind, trait emotional intelligence, and writing development. Language Testing in Asia, 14, 53. https://doi.org/10.1186/s40468-025-00343-2

Ng, D. T. K., Leung, J. K. L., Su, J., Ng, R. C. W., & Chu, S. K. W. (2023). Teachers' AI digital competencies and twenty-first century skills in the post-pandemic world. Educational Technology Research and Development, 71(1), 137–161. https://doi.org/10.1007/s11423-023-10203-6

Ng, D. T. K., Leung, J. K. L., Su, J., Ng, R. C. W., & Chu, S. K. W. (2021). Conceptualizing AI literacy: An exploratory review. Computers and Education: Artificial Intelligence, 2, 100041. https://doi.org/10.1016/j.caeai.2021.100041

Nga, P. T. (2014). Teacher perceptions and learner autonomy in Vietnamese EFL contexts. Journal of New Approaches in Educational Research, 3(2), 112–120. https://doi.org/10.7821/naer.2014.2.112

Nyaaba, M., Shi, L., Nabang, M., Zhai, X., Kyeremeh, P., Ayoberd, S. A., & Akanzire, B. N. (2024). Generative AI as a learning buddy and teaching assistant: Pre-service teachers' uses and attitudes (No. arXiv:2407.11983). arXiv. https://doi.org/10.48550/arXiv.2407.11983

Oberauer, K. (2009). Design for a working memory. Psychology of Learning and Motivation, 51, 45-100.

OECD (2019), Education at a Glance 2019: OECD Indicators, OECD Publishing, Paris, https://doi.org/10.1787/f8d7880d-en.

OECD. (2019). OECD Skills Outlook 2019: Thriving in a Digital World. OECD Publishing. https://doi.org/10.1787/df80bc12-en

Patton, M. Q. (2015). Qualitative research & evaluation methods: Integrating theory and practice (4th ed.). Sage Publications.

Peachey, N. (2024, October 24). Top 10 AI tools for planning EFL lessons • Oxford TEFL. https://oxfordtefl.com/blog/ai-tools-for-planning-efl-lessons/

Polakova, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. Cogent Education, 11(1), 2410101. https://doi.org/10.1080/2331186X.2024.2410101

Polyportis, A., & Pahos, N. (2024). Understanding students' adoption of the ChatGPT chatbot in higher education: the role of anthropomorphism, trust, design novelty and institutional policy. Behaviour & Information Technology, 44(2), 315–336. https://doi.org/10.1080/0144929X.2024.2317364

Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. Research and Practice in Technology Enhanced Learning, 12, Article 22.

Pu, X., Gao, M., & Wan, X. (2023). Summarization is (almost) dead. arXiv preprint arXiv:2309.09558. https://doi.org/10.48550/arXiv.2309.09558

Rahimi, A. R., & Teimouri, R. (2025). Advancing language education with ChatGPT: A path to cultivate 21st-century digital skills. Research Methods in Applied Linguistics, 4(2), 100218. https://doi.org/10.1016/j.rmal.2025.100218

Ranalli, J. (2021). L2 student engagement with automated feedback on writing: Potential for learning and issues of trust. Journal of Second Language Writing, 52, 100816.

Reed, J. (2025). The cognitive catastrophe of AI dependency: Why large language models are sabotaging our minds! Retrieved from https://blog.alor.org/the-cognitive-catastrophe-of-ai-dependency-why-large-language-models-are-sabotaging-our-minds-by-james-reed

Reinders, H. (2010). Towards a classroom pedagogy for learner autonomy: A framework of independent language learning skills. Australian Journal of Teacher Education, 35(5). https://doi.org/10.14221/ajte.2010v35n5.4

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? Journal of Applied Learning and Teaching, 6(1), 342-363. https://doi.org/10.37074/jalt.2023.6.1.9

Santika, R. (2025). AI integration in education: Balancing oversight and autonomy. Higher Education Review, 29(2), 49–71.

Selinker, L. (1972). Interlanguage. International Review of Applied Linguistics in Language Teaching, 10(1–4), 209–232. https://doi.org/10.1515/iral.1972.10.1-4.209

Selinker, L. (2015). Interlanguage. In Error analysis (pp. 31-54). Routledge.

Shi, H., & Aryadoust, V. (2024). A systematic review of AI-based automated written feedback research. ReCALL, 36(2), 187-209. https://doi.org/10.1017/S0958344023000265

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. Harvard Educational Review, 57(1), 1-23.

Socrat.ai – Responsible AI tools for teaching and learning. (n.d.). Retrieved July 10, 2025, from https://socrat.ai/

Song, C., & Song, Y. (2023a). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. Frontiers in Psychology, 14, Article 1260843. https://doi.org/10.3389/fpsyg.2023.1260843

Song, C., & Song, Y. (2023b). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students [Duplicate]. Frontiers in Psychology, 14. https://doi.org/10.3389/fpsyg.2023.1260843

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. Science, 333(6043), 776-778. https://doi.org/10.1126/science.1207745

SPL: A Socratic Playground for Learning Powered by Large Language Model. (n.d.). Retrieved July 10, 2025, from https://arxiv.org/html/2406.13919v1

Strauss, A., & Corbin, J. (1998). Basics of qualitative research: Techniques and procedures for developing grounded theory (2nd ed.). Sage Publications.

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). Newbury House.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. Educational Psychology Review, 31(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Thanasoulas, D. (2000). What is learner autonomy and how can it be fostered? The Internet TESL Journal, 6(11). Retrieved from https://iteslj.org/Articles/Thanasoulas-Autonomy.html

Turner, S. F., Cardinal, L. B., & Burton, R. M. (2017). Research design for mixed methods: A triangulation-based framework and roadmap. Organizational Research Methods, 20(2), 243–267. https://doi.org/10.1177/10944281156108

Voogt, J., & Pareja Roblin, N. (2010). 21st century skills: Discussion paper. Universiteit Twente.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

Wang, J. (2020). Fast identification of possible drug treatment of coronavirus disease-19 (COVID-19) through computational drug repurposing study. Journal of Chemical Information and Modeling, 60(6), 3277–3286. https://doi.org/10.1021/acs.jcim.0c00179

Watson, S., & Romic, J. (2025). ChatGPT and the entangled evolution of society, education, and technology: A systems theory perspective. European Educational Research Journal, 24(2), 205–224. https://doi.org/10.1177/14749041231221266

Wilson, T., & Robertson, K. (2024). Ethical implications of Turnitin's Feedback Studio in academic contexts. Journal of Academic Integrity, 22(3), 55–68.

Yan, D. (2024). Feedback seeking abilities of L2 writers using ChatGPT: A mixed method multiple case study. Kybernetes, 54(7), 3757–3781. https://doi.org/10.1108/K-09-2023-1933

Yan, Y., Sun, W., & Zhao, X. (2024). Metaphorical conceptualizations of generative artificial intelligence use by Chinese university EFL learners. Frontiers in Education, 9. https://doi.org/10.3389/feduc.2024.1430494

Yang, H., & Aryadoust, V. (2024). Creativity in AI tools: Bard's influence on learner confidence. Computers in Language Learning, 14(1), 78–99.

Yang, H., & Aryadoust, V. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. Smart Learning Environments, 11, 42. https://doi.org/10.1186/s40561-024-00342-5

Yang, H., & Aryadoust, V. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. Smart Learning Environments, 11, 42. https://doi.org/10.1186/s40561-024-00342-5

Yin, X. (2024). A critical thinking-oriented approach in English as a foreign language (EFL) writing: Using ChatGPT as an ideal peer. In Proceedings of the International CALL Research Conference (pp. 327–332). https://doi.org/10.29140/9780648184485-49

Yu, S.-M. (2006). Interrelationship between motivation, autonomy, and proficiency. Journal of Applied Language Learning and Research, 7(1), 1–12. Retrieved from https://www.jallr.com/index.php/JALLR/article/download/733/pdf733

Yu, S.-M. (2006). Interrelationship between motivation, autonomy, and proficiency. Journal of Applied Language Learning and Research, 7(1), 1–12. Retrieved from https://www.jallr.com/index.php/JALLR/article/download/733/pdf733

Yuan, Y., Li, H., & Sawaengdist, A. (2024). The impact of ChatGPT on learners in English academic writing: Opportunities and challenges in education. Language Learning in Higher Education, 14(1), 41-56. https://doi.org/10.1515/cercles-2023-0006

Yuan, Z., Shang, Y., Zhou, Y., Dong, Z., Zhou, Z., Xue, C., ... & Keutzer, K. (2024). Llm inference unveiled: Survey and roofline model insights. arXiv preprint arXiv:2402.16363.

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. Smart Learning Environments, 11(1), Article 28. https://doi.org/10.1186/s40561-024-00316-7

Zhai, X. (2022). ChatGPT user experience: Implications for education. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4312418

Zhang, K., & Lu, J. (2021). AI literacy in education: Definition, measurement, and implications. Computers & Education, 171, Article 104242. https://doi.org/10.1016/j.compedu.2021.104242

Zhang, L., Wei, M., & Lu, Y. (2025). Investigating students' uptake of teacher- and ChatGPT-generated feedback in EFL writing: A comparison study. Computer Assisted Language Learning, 38(2), 1-24. https://doi.org/10.1080/09588221.2024.2447279

Zhang, R., Liu, Y., & Chen, X. (2023). Evaluating the role of ChatGPT in enhancing EFL writing fluency: A mixed-methods study. Humanities and Social Sciences Communications, 12, Article 247. https://doi.org/10.1057/s41599-023-01972-1

Zhang, Z., Aubrey, S., Huang, X., & Chiu, T. K. F. (2025). The role of generative AI and hybrid feedback in improving L2 writing skills: A comparative study. Innovation in Language Learning and Teaching, 1–19. https://doi.org/10.1080/17501229.2025.2503890

Zhao, L., Wu, X., & Luo, H. (2022). Developing AI literacy for primary and middle school teachers in China: Based on a structural equation modeling analysis. Sustainability, 14(21), Article 21. https://doi.org/10.3390/su142114549

Zhao, L., Wu, X., & Luo, H. (2022). Developing AI literacy for primary and middle school teachers in China: Based on a structural equation modeling analysis. Sustainability, 14(21), Article 21. https://doi.org/10.3390/su142114549

Zhao, Y., & Watterston, J. (2021). The changes we need: Education post COVID-19. Journal of Educational Change. Advance online publication.

Zhou, X., Teng, D., & Al-Samarraie, H. (2024). The mediating role of generative AI self-regulation on students' critical thinking and problem-solving. Education Sciences. doi.org/10.3390/educsci14121302

Zimmerman, B. J., & Campillo, M. (2003). Motivating self-regulated problem solvers. The psychology of problem solving, 233262, 103.

Zimmerman, B. J., & Campillo, M. (2003). Motivating self-regulated problem solvers. The psychology of problem solving, 233262, 103.

Zimmerman, B. J., & Campillo, M. (2003). Motivating self-regulated problem solvers. The psychology of problem solving, 233262, 103.

Zou, S., Guo, K., Wang, J., & Liu, Y. (2024). Investigating students' uptake of teacher-and ChatGPT-generated feedback in EFL writing: A comparison study. Computer Assisted Language Learning, 1-30.

Zou, S., Guo, K., Wang, J., & Liu, Y. (n.d.). Investigating students' uptake of teacher- and ChatGPT-generated feedback in EFL writing: A comparison study. Computer Assisted Language Learning. https://doi.org/10.1080/09588221.2024.2447279

Zulianti, H., Hastuti, H., Nurchurifiani, E., Hastomo, T., Maximilian, A., & Ajeng, G. D. (2024). Enhancing novice EFL teachers' competency in AI-powered tools through a TPACK-based professional development program. World Journal of English Language, 15(3), Article 3. https://doi.org/10.5430/wjel.v15n3p117

Appendix 1. VELİ ONAM FORMU

Sayın Veli,

Çocuğunuzun katılacağı bu çalışma, "**Yapay Zekâ Destekli Yazma Eğitiminin Yabancı Dil Olarak İngilizce Öğrenen Öğrencilerin Yazma Becerileri ve Öğrenen Özerkliği Üzerindeki Etkisi**" başlıklı yüksek lisans tez araştırması kapsamında yürütülmektedir. Araştırma, 2023-2024 Eğitim-Öğretim yılı güz döneminde gerçekleştirilecektir.

Yukarıda belirtilen araştırmaya çocuğunuz davet edilmektedir. Araştırmaya katılım tamamen gönüllülük esasına dayanmaktadır. Katılmadan önce lütfen aşağıdaki bilgilendirmeyi dikkatlice okuyunuz, aklınıza takılan bir husus varsa sormaktan çekinmeyiniz.

Araştırmanın Amacı:

Bu çalışmanın amacı, İngilizceyi yabancı dil olarak öğrenen ortaokul öğrencilerinin yapay zekâ destekli yazma eğitimi yoluyla yazma performanslarında ve öğrenen özerkliklerinde bir gelişme olup olmadığını incelemektir.

Araştırma Süreci: Araştırma süreci 8 hafta sürecek olup, öğrenciler iki farklı gruba ayrılacaktır: Deney grubu: Yazma görevlerini yapay zekâ (örneğin ChatGPT) destekli biçimde tamamlayacaktır. Kontrol grubu: Yazma görevlerini geleneksel yöntemlerle (öğretmen desteğiyle) tamamlayacaktır. Her iki grup da haftalık yazma görevleri, yansıtıcı günlükler ve bireysel görüşmeler ile desteklenecektir.

Araştırma süresince öğrencilerin kimlikleri gizli tutulacak, sadece rumuzlar (örneğin E1, C2 gibi) kullanılacaktır. Elde edilen veriler sadece bilimsel amaçlarla kullanılacak, hiçbir şekilde öğrencinin ders notunu, akademik kaydını veya öğretmenleriyle ilişkisini etkilemeyecektir.

Katılım Koşulları ve Haklarınız:

Katılım gönüllülük esasına dayanır.

Çocuğunuz dilediği zaman çalışmadan ayrılabilir.

Ayrılmak ya da katılmamak herhangi bir hak kaybına yol açmaz.

Çalışma sırasında öğrencilerden kişisel veya özel bilgi toplanmayacaktır.

İletişim Bilgileri:

Araştırma hakkında herhangi bir sorunuz olursa aşağıdaki iletişim adresinden bana ulaşabilirsiniz:

Adı Soyadı: Fatma Şen

E-posta: xxxxx

Telefon: xxxxx

VELİ ONAM BEYANI

Velisi bulunduğum …… sınıfı …… numaralı öğrencisi ……'in yukarıda açıklanan araştırma kapsamında yapılacak çalışmalara gönüllü olarak katılmasına **izin veriyorum**.

Araştırmanın içeriği hakkında bilgilendirildim.

Katılımın gönüllü olduğunu ve istediğim zaman vazgeçebileceğimi biliyorum.

☐ Çocuğumun kimlik bilgilerinin gizli tutulacağını ve verilerin yalnızca bilimsel amaçla kullanılacağını anladım.

Veli Adı Soyadı: ……

İmza: ……

Tarih: …… / …… / 2024

Araştırmacı Adı Soyadı: Fatma Şen

İmza: ……

Tarih: …… / …… / 2024

(*Lütfen bu formu imzalayarak çocuğunuzla birlikte okula geri gönderiniz.*)

Appendix 2. Ethics Committee Approval

BURSA ULUDAĞ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
ENSTİTÜ YÖNETİM KURULU
TOPLANTISI

FR 3.8.2_01

ENSTİTÜ YÖNETİM KURULU KARARI

OTURUM TARİHİ	**OTURUM SAYISI**
22.10.2024	2024/43

KARAR NO: 1/g

Yabancı Diller Eğitimi Ana Bilim Dalı 18.10.2024 tarih ve 17939 sayılı 802293043 numaralı Yüksek Lisans Programı öğrencisi Fatma ŞEN'in tez konusu önerisi konulu yazısı görüşmeye açıldı.

Yapılan görüşmeler sonunda; Yabancı Diller Eğitimi Ana Bilim Dalı İngiliz Dili Eğitimi Bilim Dalı Öğretim Üyesi Prof. Dr. A. Amanda YEŞİLBURSA'nın danışmanlığını yürüttüğü 802293043 numaralı Yüksek Lisans Programı öğrencisi Fatma ŞEN'in "EFL Eğitiminde Üretken Yapay Zeka Entegrasyonunun Öğrenen Özerkliği ve Yazma Yeterliliği Üzerindeki Etkilerinin İncelenmesi" tez konusu önerisinin BUÜ Lisansüstü Eğitim-Öğretim Yönetmeliğinin 28/1 maddesi uyarınca uygun olduğuna oy birliği ile karar verildi.

Appendix 3. The Approval of the Institute of Educational Sciences

T.C.
BURSA ULUDAĞ ÜNİVERSİTESİ
Eğitim Bilimleri Enstitüsü Müdürlüğü

Sayı: E-20585590-044-18818 03.01.2025
Konu: Fatma ŞEN'in araştırma izni.

YABANCI DİLLER EĞİTİMİ ANABİLİM DALI BAŞKANLIĞINA

İlgi : 18.10.2024 tarihli ve 56643073-300-17940 sayılı yazınız.

Ana Bilim Dalınız yüksek lisans programı öğrencisi Fatma ŞEN'in tez çalışmasına ilişkin Ankara Üniversitesi Rektörlüğü Türkçe ve Yabancı Dil Uygulama ve Araştırma Merkezi Müdürlüğünden alınan yazı örneği ekte gönderilmiştir.

Bilgilerinizi ve öğrenci ile danışmanı Prof. Dr. Ayşegül Amanda YEŞİLBURSA'ya bildirilmesi konusunda gereğini rica ederim.

Prof. Dr. Gökhan ARI
Müdür

Ek:
Yazı Örneği

Bu belge, güvenli elektronik imza ile imzalanmıştır.
Belge Doğrulama Kodu:3tCjEPuXgkmkvoi-vreHpA Belge Doğrulama Adresi: https://udos.uludag.edu.tr/Teyit/

BUÜ Eğitim Bilimleri Enst. Üniversite 1 Cd. No: 221 Görükle Kampüsü 16059 Nilüfer/BURSA
Telefon No: 0224 29 42 884 Faks No: 0224 29 40 975 Bilgi için: Halil AYDOĞAN
e-Posta: egtbil@uludag.edu.tr İnternet Adresi: uludag.edu.tr/egitimbilimleri Bilgisayar İşletmeni
Kep Adresi: uludag.rektorluk@hs03.kep.tr Telefon No: 0224 29 40 978
Bu belge UDOS ile hazırlanmıştır.

1/1

Appendix 4. Informed Consent

Dear Participant,

I am Fatma ŞEN, an MA student at Bursa Uludağ University. I am conducting a research study for my MA thesis, titled "Investigating the Impact of Integrating Generative AI on Learner Autonomy and Writing Proficiency in EFL Instruction. This study aims to explore the role of Generative AI tools, such as ChatGPT, in fostering learner autonomy and improving writing proficiency among young adult and adult English language learners in educational settings.

Purpose of the Study:

The purpose of this study is to investigate how the integration of ChatGPT and the feedback from it improves the development of writing and learner autonomy in English as a Foreign Language (EFL) learning.

What Participation Involves:

If you choose to participate and be in the focus group, you will:

Attend a workshop on the use of the Generative AI tool ChatGPT.

Complete a pre-test essay before the study begins.

Engage in weekly AI-assisted writing tasks for eight weeks.

Maintain a reflective diary about your experience using AI for writing improvement.**

Receive feedback on your writing from AI and/or your instructor.

Participate in a post-test writing task at the end of the study.

✅ Take part in semi-structured interviews(optional) to share your experiences with AI in writing instruction. Your participation will help provide valuable insights into the effects of AI tools on writing skills and learner independence.

Confidentiality and Anonymity:

Please be assured that all responses will be kept strictly confidential. No personal identifiers will be included in the thesis or any related publication. All collected data will be anonymised and used solely for academic research purposes.

Voluntary Participation:

Your participation in this study is completely voluntary. You have the right to withdraw at any time without providing a reason. Withdrawing from the study will have no negative consequences on your academic status or relationship with the university.

Use of Data: The data collected from this study will be used for my MA thesis at Bursa Uludağ University. Potential academic publications (e.g., journal articles, conference presentations). Your anonymity will be maintained in all research publications.

Contact Information:

If you have any questions or concerns about the study or your participation, please feel free to contact me at:

Email:

Consent Statement: By proceeding with the workshop and related activities, you are indicating your consent to participate in this study.

Thank you for considering participation in this research. Your insights and experiences are vital to the success of this study.

Sincerely,

Fatma ŞEN

September 2024

Appendix 5. Demographic Questions

Do you agree to participate in this study?	● Yes, I agree to participate. ● No, I do not agree to participate.
Email Address:	
Full name:	
Gender:	● Male ● Female ● Other
How old are you?	
What is your nationality?	
What kind of institution you've completed your education in high school or university?	● State School ● Private School ● Vocational School
What is the highest degree you have achieved?	● PhD ● MA ● BA ● Still highschool
How long have you been learning English in years?	
Have you attended any training/workshops related to the use of AI in English language teaching? if yes, when and where?	

Appendix 6. Resume

ÖZGEÇMİŞ			
Name-Surname	FATMA ŞEN		
Education	**Start**	**Finish**	**Name of the Institute**
BA (English Language Teaching)	2017	2022	**University of Cukurova**
Work Experience	**Start**	**Finish**	**Name of the Institute**
IT Service	2014	2015	Delta Denizcilik Ticaret A.Ş. Incirlik/ AirBase
Book Fair Sales Assistant	2021	2021	Dokuz Publishing
English Teacher	2020	2021	Zirve Dershaneleri
English Teacher	2021	2022	Başarı+ Dershanesi
English Teacher	2021	2022	Verol Okulları
English Teacher	2021	2021	Son Adım Derhanesi
IELTS Invigilator	2020	2023	IDP IELTS
English Intructor & Head of the English and Foreign Languages Department	2023	2025	Ankara University TÖMER Bursa Branch
English Teacher	2025	Now	Bademli Kindergarten

Appendix 7. Reflective Diary

A. Reflective Diary

Name Surname: _____

Date: _____

Week No: _____

Topic of the Week: _____

Student Reflection Diaries

Instruction:

The topic of discussion for this week is:

"……?"

By using the prompts you learned and discovered, chat with ChatGPT, and get feedback for your essay based on the rubric given. If there are places you don't understand and feel confused, ask for details and explanations from ChatGPT to get answers.

You need to keep a digital reflection diary and share it with the teacher every Friday. Remember to share the document of your digital reflection diary notes with your teacher on Google Drive.

Section A

How was your general experience while getting feedback from AI chatbot ChatGPT in this week's topic? Please state at least three strengths, three challenges, and three suggestions for improvement.

Strengths:

1.
2.
3.

Challenges:

1.
2.
3.

Prompts You Used and Suggestions for Improvement:

1.
2.
3.

Section B

What are some particular prompts you discovered to be working very efficiently ChatGPT this week?

For example, If you don't understand a topic clearly, what method do you use to clarify the situation for GPT to explain to you?

Section C

How would you rate the effectiveness of your experience in getting feedback from the AI chatbot this week?

Please circle one number from 0 (Not effective) to 10 (Extremely effective).

Circle your choice:

0 1 2 3 4 5 6 7 8 9 10

(Adapted from Koç, 2024)

Appendix 8. Ankara Üniversitesi Araştırma İzni

T.C.
ANKARA ÜNİVERSİTESİ REKTÖRLÜĞÜ
Türkçe ve Yabancı Dil Uygulama ve Araştırma Merkezi Müdürlüğü
İngilizce Koordinatörlüğü

Sayı : E-38268124-622.01-1646753
Konu : Fatma ŞEN'in Araştırma İzni

ÖĞRENCİ İŞLERİ DAİRE BAŞKANLIĞINA

İlgi : 07.11.2024 tarihli ve E-94390400-302.08.01-187231 sayılı yazınız.

İlgide kayıtlı yazınıza istinaden İngiliz Dili Eğitimi Yüksek Lisanas Programı öğrencisi Fatma ŞEN'in "İngilizce Öğretiminde Üretken Yapay Zekanın Kullanılması:Öğrenci Özerkliğinin İncelenmesi, Yazma Gelişimi ve Üretken Kelime Hazinesinin Zenginleştirilmesi"konulu tez çalışması Merkezimizce uygun görülmüştür.

Gereği için bilgilerinize saygılarımla arz ederim.

Prof. Dr. Timur GÜLTEKİN
Müdür

Bu belge, güvenli elektronik imza ile imzalanmıştır.
Doğrulama Kodu: 107F55C8-A83A-442A-84D0-E266C73B4D66 Doğrulama Adresi: https://www.turkiye.gov.tr/ankara-universitesi-ebys

Ankara Üniversitesi Beşevler 10. Yıl Yerleşkesi Ord.Prof.Dr. Şevket Aziz Kansu Binası Emniyet Mahallesi İncitaşı Sokak No :10 Beşevler / ANKARA
Telefon No: 0(312) 214 13 50 Belge Geçer No: 0312 600 01 31 Kep Adresi:
ankunvrek@ankuni.hs01.kep.tr
www.tomer.ankara.edu.tr
KEP Adresi : ankunvrek@ankuni.hs01.kep.tr

Bilgi için:Seher KILIÇ
Sekreter

Appendix 9. 8 Weeks Essay Topics

Pre-Test / Post-Test Topic

“Is using Artificial Intelligence (AI) tools a more effective way to improve language learning experiences compared to traditional classroom methods? Discuss both views and give your opinion.”

Writing Exam Weekly Tasks B1 Level

Week 1 **Education & Learning**:

Should universities be free for everyone, or should students pay for their education?

Week 2 **Technology & Innovation**:

Do you think technology makes our lives easier, or does it make us too dependent on gadgets?

Week 3 **Health & Fitness**:

Is it better to exercise outdoors or in a gym?

Week 4 **Social media and government**:

Should social media usage be restricted by the government or not?

Week 5 **Work & Careers**:

Do you think people should choose a career based on passion or salary?

Week 6 **Travel & Experiences**:

Should people travel abroad for a better language-learning experience?

Week 7 **Culture & Society**:

Should cultural traditions be preserved, even if they no longer fit into modern society?

Week 8 **Social Media & Communication**:

Do the advantages of social media outweigh the disadvantages?