2026-04

Enhancing coherence in descriptive writing tasks using LLMs for private university EFL learners in Dhaka

Sharkar, Md. Ibrahim

Independent University, Bangladesh (IUB)

Enhancing Coherence in Descriptive Writing Tasks using LLMs for Private University EFL Learners in Dhaka

Md. Ibrahim Sharkar
2320993

A thesis submitted to the Department of English and Modern Languages in partial fulfillment of the requirements for the degree of Bachelor of Arts in English Language Teaching

Department of English and Modern Languages
Independent University, Bangladesh
April 2026

Declaration

It is hereby declared that

1. The thesis submitted is my own original work while completing degree at Independent University, Bangladesh.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. I have acknowledged all main sources of help.

_____

Md. Ibrahim Sharkar
2320993

Approval

The thesis titled "Enhancing Coherence in Descriptive Writing Tasks using LLMs for Private University EFL Learners in Dhaka" submitted by Md. Ibrahim Sharkar (2320993) of Spring, 2026 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Arts in English Language Teaching on 26th April, 2026.

Examining Committee:

Supervisor:
(Member)

Ms. Rezina Nazneen Rimi
Lecturer A, Department of English and Modern Languages
Independent University Bangladesh

External Expert:

Takad Ahmed Chowdhury, PhD
Associate Professor and Head, Department of English

Naureen Rahnuma, PhD
Assistant Professor and Head
Department of English and Modern Languages

Internal Expert:
(Panel Members)

Mr. Towhid Bin Muzaffar
Assistant Professor
Department of English and Modern Languages

Ms. Rezina Nazneen Rimi
Lecturer A
Department of English and Modern Languages

Departmental Head:
(Chair)

Naureen Rahnuma, PhD
Assistant Professor and Head, Department of English and Modern Languages
Independent University Bangladesh

Office of GSRIR:

(Director)

_____

Prof. M. Arshad Momen
Director, GSRIR

Office of Librarian:

(Librarian)

_____

Shahajada Masud Anowarul Haque
Librarian, IUB Library

Ethics Statement

The researcher conducted this study in accordance with the ethical principles that guide educational research involving human participants at the tertiary level. Before the data collection began, the researcher explained the full purpose of the study to all eighteen participants, including the nature of the AI tools to be used in the experimental condition and the peer review arrangement in the control condition. Every participant gave written informed consent prior to the pre-test, and the six interviewees gave separate verbal consent at the start of each interview session. No participant was coerced, financially compensated, or academically penalised for taking part or for choosing not to.

The interview recordings and verbatim transcripts are stored securely by the researcher. The participants have given permission for these materials to be retained for future research analysis, provided that continued anonymisation is maintained. The study posed no foreseeable physical or psychological risk to participants at any stage. No participant withdrew during the data collection period.

Abstract

This study examines whether short-term use of Large Language Models can enhance coherence in descriptive writing among private university EFL learners in Dhaka compared with peer review. In two-week intervention a total of eighteen undergraduates participated which divided into two groups; an experimental group (n=9) using LLM tools for paragraph-level scaffolding and a control group (n=9) using traditional peer review. Both of the groups completed pre-tests and post-tests on descriptive paragraph writing which was scored with a ten-point analytic rubric, and also six students participated in semi-structured interviews. The experimental group started with higher baseline scores (8.78 versus 6.89). Results showed both groups made gains, but distributional analysis revealed different patterns. The experimental group's score range narrowed from four to two points which effect shows up in standard deviation dropped from 1.30 to 0.87, with four students reaching top scores. On the other hand, the control group's range widened from three to five points showing more uneven distribution. Interview data showed that AI was used as scaffolding for paragraph organization and transitions for experimental participants. On the other hand, control participants described peer feedback as word-level where they were focusing mostly on vocabulary only. These findings suggest that LLMs can be effectively used as a scaffold for coherence when students use them for paragraph-level support. The study is limited by its short timeframe, single site, and small sample, but it carries out a Bangladeshi perspective on AI integration in EFL writing instruction.

Keywords: Large Language Models; LLMs; EFL writing; coherence; descriptive writing; Bangladesh

Dedication

This thesis is dedicated to two extraordinary people to whom I dedicate my life, for now and forever.

I would like to dedicate this thesis to the cornerstones of my life. To my mother Rocksana Begum, You do not just strengthen me and are my best friend, you are the very inspiration of my thirst for knowledge. Thank you for helping me to find wisdom with an open mind, I wouldn't have made it this far without you.

This is also tribute to my dear friend Mazharul Haque Chawdhury. Thank you for being an unwavering emotional rock, for holding me up when I was at my lowest point. This is a joint accomplishment and it's due to you as much as it is due to me.

Acknowledgement

I want to say thank you to my supervisor Ms. Rezina Nazneen Rimi from the Department of English and Modern Languages. She was always there for me when I needed help. I really appreciate the way she helped me. The Department of English and Modern Languages is a place to learn new things and I feel lucky to have been a part of it.

My supervisor Ms. Rezina Nazneen Rimi was a help when I was working on my thesis. She was very patient with me. Gave me good feedback. This was really helpful when I got stuck and did not know what to do.

I also want to thank the Department of English and Modern Languages at Independent University, Bangladesh. They gave me a chance to do my research which was very important for me.

I am very thankful, to the eighteen students who helped me with my study. These eighteen students were very kind. Gave me their time. They took writing tests. Talked to me about what they thought. Without these eighteen students and my supervisor Ms. Rezina Nazneen Rimi I would not have been able to do my study.

Table of Contents

List of Tables

List of Acronyms

AI	Artificial Intelligence
CALLA	Cognitive Academic Language Learning Approach
EFL	English as a Foreign Language
ELT	English Language Teaching
GPT	Generative Pre-trained Transformer
L2	Second Language
LLM	Large Language Model
MKO	More Knowledgeable Other
ZPD	Zone of Proximal Development
P1_Con	Participant no 1 of Control Group
P1_Ex	Participant no 1 of Experimental Group
Con_IO1	Interviewee no 1 of Control Group
Ex_IO1	Interviewee no 1 of Experimental Group

Glossary

Analytic Rubric	A scoring tool that breaks writing quality into separate components, such as coherence, grammar, vocabulary, and task fulfilment, and assigns a score to each component individually rather than giving a single overall mark.
Ceiling Effect	A measurement limitation that occurs when participants score at or near the maximum of a scale, leaving little room for measurable improvement regardless of how effective the intervention actually is.
Coherence	The logical organisation and smooth connection of ideas within a text, allowing the reader to perceive the whole piece as a meaningful unit rather than a collection of unrelated sentences. In this study, coherence is the primary outcome variable.
Cohesion	The use of linguistic devices such as conjunctions, reference words, and lexical repetition to link sentences and clauses on the surface level. Cohesion supports coherence but does not guarantee it.
Descriptive Writing	A genre of writing in which the writer uses sensory details, spatial or temporal ordering, and vivid language to create a picture of a person, place, object, or experience for the reader.
Floor Effect	A measurement limitation that occurs when participants score at or near the minimum of a scale, making it difficult to detect further decline.
Internalisation	In Vygotskian theory, the process by which a learner gradually takes over a skill or concept that was initially performed with external support and begins to perform it independently.
Mixed-Methods Design	A research approach that combines quantitative data, such as test scores, with qualitative data, such as interview responses, within a single study so that the two types of evidence can inform and complement each other.
More Knowledgeable Other (MKO)	In Vygotsky's sociocultural theory, the partner in a learning interaction who brings greater knowledge or skill to the task. In this study, the MKO role is filled by AI tools in the experimental group and by peers in the control group.
Peer Feedback	A classroom practice in which learners read and respond to one another's writing, offering comments on areas such as organisation, clarity, grammar, or vocabulary.
Purposive Sampling	A non-random sampling method in which participants are selected deliberately because they meet criteria relevant to the research questions, rather than being chosen at random from a larger population.
Scaffolding	Temporary support provided to a learner to help them complete a task that they cannot yet perform independently. The support is gradually removed as the learner gains competence. In this study, AI suggestions and peer comments both function as forms of scaffolding.

Thematic Analysis	A qualitative data analysis method in which the researcher identifies, organises, and reports patterns or themes across a dataset, following a structured coding process.
Zone of Proximal Development (ZPD)	The conceptual distance between what a learner can do independently and what the learner can do with the help of a more capable partner. Learning, in Vygotsky's view, takes place within this zone.

Chapter 1

Introduction

1.1 Background and Context

One of the most talked about technological shifts in recent years is the rise of Large Language Models (LLMs) such as ChatGPT, Gemini, and Copilot, which have rapidly made their way into classrooms across the world. Among their applications in English Language Teaching (ELT) are automated feedback on student writing, generation of writing prompts, vocabulary scaffolding, and paragraph-level suggestions on structure and flow (Barrot, 2023). According to Kohnke et al. (2023), scholars and practitioners in the field of education have lately started to give serious attention to the consequences of generative AI for teaching, learning, and assessment. The tools are no longer a future prospect in Bangladeshi tertiary classrooms they are already being used, sometimes openly and sometimes quietly, by both students and instructors (Khan et al., 2023). Writing, however, continues to be the most demanding of the four language skills for English as a Foreign Language (EFL) learners in Bangladesh. Patwary et al. (2023) noted that the writing problems of tertiary EFL students of Bangladesh are not only limited themsleves in grammar but also extend on organisation, vocabulary, coherence, and the development of ideas beyond sentence level. Kayum and Khan (2015) had earlier listed coherence as one of the core components that learners at this level continue to struggle with.

Among the various sub-skills of writing, coherence is perhaps the one that is most difficult to teach and the easiest to overlook. According to Hyland (2019), coherence refers to the logical organisation and smooth connection of ideas within a text, which allows the reader to perceive the whole text as a meaningful unit rather than as a collection of sentences. Descriptive writing is a genre where this quality becomes visible very quickly. Unlike argumentative writing, which carries a built-in structure of claim, reason, and evidence, descriptive writing depends on the writer's ability to order details spatially, temporally, or by some other logical principle (Siekmann et al., 2022). Siekmann et al. (2022) and Aldera (2016) both reported that EFL writers frequently skip conclusions and fail to maintain a single main idea, and Akter (2022) and Mustaque (2014) have documented similar patterns in

Bangladeshi tertiary classrooms. Barrot (2023) argued that ChatGPT can function as a simultaneous partner of writing for L2 learners, offering suggestions on form, style, coherence, and cohesion, however he warned that the tool can produce fluent but formulaic text which may feel formulaic or robotic. Mohammad et al. (2025), in a quasi-experimental study with sixty EFL undergraduates at Najran University, found significant improvements in organisation, structure, analysis, and writing quality among ChatGPT-supported (LLMs) learners, and Liu et al. (2024), in a randomised controlled trial with entry level learners, reported sustained gains through their CALLA-LLM framework. Within Bangladesh, Khan et al. (2023), Islam (2025), and Rafin (2025) have each contributed to the growing local picture, however what is still missing is a study that isolates a single genre, like descriptive writing, and tests whether a short, carefully structured classroom intervention with LLMs actually changes the coherence of learner paragraph writing, while also listening to what the students themselves have to say about the process.

1.2 Statement of the Research Problem and Purpose

The problem that motivates this study has two sides which are connected. On one side, coherence in descriptive writing remains a weak area for private university EFL students in Dhaka. A paragraph opens with a reasonable topic sentence and then falls apart; transitions are either absent or recycled through the same two or three connectors; details appear in random positions, neither grouped spatially nor temporally. Patwary et al. (2023), Akter (2022), and Mustaque (2014) have each documented these types of patterns across several tertiary institutions. On the other side, the rapid spread of generative AI has introduced a new pedagogical uncertainty. Alsofyani and Barzanji (2025) found that both ChatGPT feedback and teacher feedback produce writing gains, however students place different levels of trust in the two sources, and Xu and Jumaat (2024) similarly cautioned that while AI empowers writing strategies, it also raises concerns about plagiarism and inaccurate output. According to Uddin and Bailey (2024), understanding these dynamics is important for the effective implementation of AI in English language education in Bangladesh.

The research gap is twofold. Firstly, although international work on LLMs and EFL writing is growing rapidly, very little of it focuses specifically on descriptive writing as a genre or on coherence as a targeted sub-skill. Most studies evaluate overall writing quality through a composite rubric that mixes coherence with grammar, vocabulary, and task achievement. Secondly, within Bangladesh, the available work on AI tools is largely survey-based and perception-based, rather than based on paired pre-test and post-test data combined with interview evidence from the same participants. The present research therefore aimed to investigate whether short-term use of ChatGPT and Gemini can enhance coherence in descriptive writing tasks produced by private university EFL learners in Dhaka, focusing on eighteen undergraduate students, nine in an experimental group and nine in a control group using peer review. Both groups sat the same pre-test and post-test on descriptive paragraphs, and six students, three from each group, were later interviewed about their experience of the practice. The interview data were intended to sit alongside the test scores, rather than below them. So that any observed change in the numbers could be interpreted against what the learners said about the feedback they received, the trust they placed in it, and the readiness they felt after the two-week practice.

1.3 Justification for the Exploratory Research Design

The researcher chose an exploratory mixed-methods design because the use of LLMs in descriptive writing instruction in Bangladeshi private universities is a very new phenomenon, with no existing body of local empirical work that the researcher could simply extend through a confirmatory study. According to Creswell (2012), exploratory designs are particularly appropriate in situations where the researcher is attempting to map a case rather than just a fixed hypothesis, and where the outcome is expected to open further research directions rather than close them. The quantitative strand used a test module as pre-test and post-test structure following the pattern established in Mohammad et al. (2025), Alsofyani and Barzanji (2025), and Liu et al. (2024), while the qualitative strand used one-on-one semi-structured interviews to explain what the test scores alone could not. This combined reading turned out to be especially important in the present study, because a matched numerical gain can represent two completely different learning journeys when the students are asked to describe the experience.

1.4 Significance of the Study

The study contributes in several ways to ongoing conversations about English language instruction in Bangladeshi higher education. It adds a local voice to a body of literature that has been dominated by studies from outside the country, narrows in on descriptive writing and coherence as specific classroom concerns rather than treating writing as a general construct, and offers instructors currently deciding how to respond to ChatGPT and Gemini a grounded account of what a structured LLM intervention actually looks like, along with its observable gains and its practical limits. According to Uddin and Bailey (2024), the integration of AI tools in Bangladeshi tertiary classrooms has so far been patchy and left on largely to individual instructors, and the findings of this small-scale, close-up study may help department heads and curriculum designers to make more informed decisions about real-time scenarios like training, rubric design, and the place of peer review alongside AI-supported practice.

1.5 Research Questions

The research questions are given below:

1. Does the short-term use of LLMs Tool enhance coherence in descriptive writing produced by private university EFL learners in Dhaka when it compared to a control group using peer review?
2. How do learners in the experimental group describe the usefulness and trustworthiness of AI feedback on their descriptive writing, and how do these descriptions differ from the learners in the control group who rely on peer feedback?
3. What implications does this comparison have for the integration of LLMs into descriptive writing instruction at the tertiary level in Bangladesh?

1.6 Scope of the Study

The scope of this study is deliberately focused. The researcher carried out the research at one private university in Dhaka with eighteen undergraduate EFL students, nine in each group. The writing task used in both the pre-test and the post-test was a descriptive paragraph on a topic drawn from everyday urban life. The AI tools used were the free, publicly available versions of ChatGPT and Gemini. The data collection took place between the eighth of March and the twenty-ninth of March 2026. The study does not extend to other genres of

writing such as argumentative or academic essay writing, and it does not claim to represent all private universities in the country. These are deliberate scope choices the researcher made in the interest of depth rather than breadth, and the thesis acknowledges the trade-off openly throughout.

Chapter 2

Literature Review

With the growing presence of Large Language Models (LLMs) such as ChatGPT, Gemini, and Copilot in classrooms across the world, scholars have started to examine what these tools mean for the teaching and learning of English (Barrot, 2023; Kohnke et al., 2023). The literature is growing rapidly, though the research is still uneven in its coverage. This chapter moves from the broader problem of EFL writing difficulties at the tertiary level, to coherence and cohesion in descriptive writing, to the emerging body of work on LLMs in EFL writing, and finally to the Bangladeshi tertiary context. A separate section considers peer feedback as the traditional scaffold used in the control condition. The chapter then lays out the theoretical framework through Vygotsky's sociocultural theory, the Zone of Proximal Development, and the concept of the More Knowledgeable Other, and closes with a clear statement of the research gap.

2.1 Writing Difficulties of EFL Learners at the Tertiary Level

Grabe (2001) described L2 writing as a multi-layered skill that requires the writer to manage linguistic knowledge, rhetorical awareness, and cognitive control all at once. Hyland (2019) made a related argument, suggesting that L2 writers carry a double cognitive load, because they have to manage the content of the text and the language simultaneously.

2.2 Coherence and Cohesion in Descriptive Writing

For EFL learners, coherence tends to develop late because it is a metacognitive skill rather than a sentence-level skill. Bai and Wang (2021), cited within the CALLA-LLM framework developed by Liu et al. (2024), connect coherence to self-regulated learning the learner has to plan, monitor, and revise, rather than simply produce.

2.3 LLMs in EFL Writing Instruction

Barrot (2023), in one of the earliest technology reviews on the topic in Assessing Writing, argued that ChatGPT (LLMs) can function as a simultaneous writing partner for L2 learners by responding to questions, admitting mistakes, and offering adaptive suggestions on form, style, coherence, and cohesion. He was cautious, though, warning that the LLM tool can produce text that is fluent but formulaic or robotic like, and that its responses are sensitive to

small adjustments in the prompt. His recommendation that instructors should not ban the tool outright but learn to work alongside it has aged well.

Recent studies (Mohammad et al., 2025; Liu et al., 2024; Alsofyani & Barzanji, 2025) found that AI scaffolding had improved EFL writing, though each study raised concerns about over-reliance. Where these studies differ, however, is in their focus and context. Mohammad et al. (2025) worked with a relatively large sample of sixty learners at a Saudi university and evaluated broad writing quality through a composite rubric, while Liu et al. (2024) developed the CALLA-LLM framework specifically around self-regulated learning strategies and included a one-month follow-up to test retention. Alsofyani and Barzanji (2025), by contrast, compared ChatGPT feedback directly with teacher feedback rather than with peer review. None of these studies isolates descriptive writing as a genre or targets coherence as the primary measured outcome, which is the specific gap this study addresses.

Alharbi and Al-Ahdal (2025) argued that Saudi EFL learners negotiate ChatGPT as a complementary tool, using it as a temporary scaffold rather than as a replacement.

2.4 LLMs and AI Tools in the Bangladeshi Tertiary Context

The local literature on AI tools in Bangladeshi tertiary English classrooms is thinner, though it is growing. Khan et al. (2023) surveyed one hundred and thirty-eight Bachelor's and Master's students across several Bangladeshi universities, and their participants largely viewed AI writing tools as useful for vocabulary, grammar correction, and the organisation of ideas, while also expressing concerns about over-reliance. This tension shows up in the interview data of the present study. Yusuf et al. (2025) and Ulfat (2023) have reported positive learner perceptions of Grammarly among Bangladeshi EFL learners and patchy technology adoption among English medium school teachers respectively.

Islam (2025), in a Master's thesis at BRAC University, noted that learners at different proficiency levels use ChatGPT (LLMs) in different ways; noting that internet connectivity alongside uneven access to devices shapes how these tools are actually encountered. Rafin (2025), also at BRAC University, argued for the benefits of peer collaboration across the four language skills, though noting that peer work in Bangladeshi tertiary classrooms often stops

at the surface level students help one another with vocabulary and grammar but rarely push into organisation or coherence. This observation is almost a word-for-word match with what the control group interviewees of the present study describe in Chapter Four. Taken together, these Bangladeshi studies share a common methodological profile perception surveys or qualitative observation and none places AI feedback and peer feedback side by side within the same classroom cohort, for the same writing task, measured through both test scores and participant interviews. That comparison is what the present study contributes.

2.5 Peer Feedback as a Traditional Scaffold

Peer feedback has been part of second-language writing research for decades. Bitchener (2008) and Bitchener and Ferris (2012) have argued that feedback on L2 writing, when it is explicit and timely, can produce significant improvements which are also measurable. When learners read one another's drafts and respond, they are likely to learn as much from the act of responding as from the comments they receive. The limitations in EFL settings, however, are also well documented. Rafin's (2025) thesis at BRAC University suggests that peer collaboration in Bangladeshi tertiary classrooms tends to operate at the word level, with students drifting toward the simplest things to comment on spelling, a word choice, a minor grammar slip. A second limit concerns the proficiency profile of the peer. In a classroom where most of the students share similar proficiency levels, the assumption that the peer partner brings additional capacity to the task becomes shaky. Con_IO1, for instance, reported disagreeing with her peers ninety-nine per cent of the time, and Con_IO2 said she only trusted the peers she knew personally. These are honest reports about the specific peer scaffolding available, and they line up with what the peer-feedback literature has long cautioned about.

2.6 Theoretical Framework

The theoretical framework of the present study rests on Vygotsky's sociocultural theory (Vygotsky, 1978), and more specifically on two of its central concepts the Zone of Proximal Development (ZPD) and the More Knowledgeable Other (MKO). Sociocultural theory, as articulated in *Mind in Society* (Vygotsky, 1978), argues that higher mental functions appear first on the social plane, through interaction with others, and only later on the individual plane, through internalisation. Learning, in this view, is mediated it happens with others before it happens alone. Writing is a good example, because it is symbolic, cognitively demanding, and

dependent on the learner's internalisation of conventions that she first encounters socially. Alharbi and Al-Ahdal (2025) drew on sociocultural theory together with the Technology Acceptance Model in their study of Saudi EFL learners' engagement with ChatGPT.

Within this framework, the ZPD is the conceptual space between what the learner can do on her own and what she can do with assistance from a more capable partner (Vygotsky, 1978). For writing instruction, the ZPD is exactly where the interesting action takes place. A student may not yet be able to produce a perfectly coherent descriptive paragraph on a new topic without help, however with the right scaffolding she can come closer, and, over time, the scaffolding can be withdrawn. Liu et al. (2024) designed their CALLA-LLM model explicitly around this idea. According to Vygotsky (1978), the ZPD is also dynamic what falls inside the zone today may fall outside it tomorrow.

The More Knowledgeable Other is the partner in the interaction who brings greater capacity to the learning task (Vygotsky, 1978). Traditionally, the MKO is imagined as a teacher, an older classmate, or a parent. In the present study, however, the MKO role is shared and varied. For the control group, the MKO is the peer who reviews the paragraph. For the experimental group, the MKO is, in part, the AI tool consulted during drafting. Neither is a perfect authority. A peer may share the writer's proficiency level and fail to spot the very weaknesses the writer is producing. An AI may produce fluent but formulaic suggestions, or misread the writer's intention when the prompt is unclear. The interesting question, therefore, is which kind of MKO produces the more useful scaffolding for coherence development in descriptive writing.

2.7 Research Gap

Firstly, in Bangladeshi context, the available work on AI tools is largely survey-based and perception-based, rather than grounded in paired pre-test and post-test data combined with interview evidence from the same participants. Khan et al. (2023), Yusuf et al. (2025), and Islam (2025) have each contributed valuable perception data, however none of them triangulates learners' perception against actual writing performance on a single specific genre. This leaves a missing link between what Bangladeshi and South Asian learners say about AI tools and what their writing actually looks like before and after using these tools in a structured manner.

Secondly, most local studies have assumed the role of the peer as MKO rather than examined it. Rafin (2025) and the BRAC University peer-feedback thesis come closest to treating peer scaffolding as an object of focused investigation, though neither of them sets peer feedback directly alongside AI feedback within the same classroom cohort, for the same writing task, over the same time window. Such a direct comparison is exactly what we need if we are to understand how these two kinds of MKOs work differently for a skill such as coherence.

The present study is mainly designed to sit on this gap. It focuses specifically on descriptive writing which is a single genre. It targets coherence skill as a measured outcome, embedded within a rubric that values it heavily. It combines quantitative test data with qualitative interview evidence from the same participants. It places peer scaffolding and AI scaffolding side by side within the same classroom. The contribution is modest in scale, given the small sample and short intervention, More importantly it is carefully positioned to address a gap that the existing literature has left largely open.

Chapter 3

Research Methodology

The aim of the present study is to investigate whether short-term use of Large Language Models, specifically ChatGPT and Gemini, can enhance coherence in descriptive writing tasks produced by private university EFL learners in Dhaka, in comparison to a control group using peer review. To approach this investigation, the researcher adopted an exploratory research design with a mixed-methods approach. According to Creswell (2012), exploratory designs are particularly appropriate when the existing local literature is thin and the goal is to open further research directions rather than to close them.

3.1 Research Design

The research design of the present study is exploratory mixed-methods. The quantitative strand used a test module followed by pre-test and post-test structure with an experimental group and a control group, for following the broad pattern already used by Mohammad et al. (2025), Alsofyani and Barzanji (2025), and Liu et al. (2024). The qualitative strand used one-on-one semi-structured interviews with six participants, three from each group to add depth to what the test scores alone could not explain. According to Creswell (2014), the strength of a mixed-methods design lies in this complementarity.

3.2 Research Site and Participants

The present study was conducted at one private university in Dhaka during the Spring semester of 2026. Eighteen undergraduate EFL learners enrolled in an English language course agreed to take part, with nine assigned to the experimental group and nine to the control group. The researcher drew the two groups from the same wider cohort, so that the contextual features of the classroom instructor, time slot, course content, syllabus pace were as similar as possible. The pre-test, however, later showed that the two groups were not perfectly equivalent at baseline, with the experimental group starting higher than the control group. This baseline imbalance is acknowledged openly, and is discussed in Chapters Four and Five.

All eighteen participants were native speakers of Bangla and had studied English throughout their school years, mostly under the National Curriculum. Every participant was able to produce a paragraph of at least one hundred words on a familiar topic, though their control

over coherence varied widely, as the pre-test scores will show in Chapter Four. None of the participants were entirely new to ChatGPT or Gemini a few used them regularly, while others had used them only occasionally before this study. This is consistent with what Islam (2025) reported about AI tool use among Bangladeshi tertiary students. Six of the eighteen participants three from each group were later invited to take part in semi-structured interviews, and the selection is described in Section 3.4.

3.3 Research Instruments

The data collection used five main instruments paper-based answer scripts, the pre-test prompt, the post-test prompt, the analytic scoring rubric, and the semi-structured interview schedule. An audio recorder captured the interview sessions for later transcription.

The researcher placed the pre-test on the eighth of March 2026 with the topic "Any Food Stall," and the post-test on the twenty-ninth of March 2026 with the topic "City Bus or Metro Rail in Rush Hour." In both sessions the researcher asked participants to write a descriptive paragraph of approximately one hundred to one hundred twenty words within a thirty-minute duration.

The scoring rubric was a ten-point analytic scale, with coherence and organisation as the heaviest band, followed by vocabulary, grammar, and task fulfilment. The rubric drew on the analytic descriptors used by Siekmann et al. (2022), on the scoring approach in Liu et al. (2024), and on the wider analytic tradition articulated by Hyland (2003). Each script was scored holistically, though the coherence band acted as the anchor. The researcher anonymised scripts before scoring, and scored post-test scripts without reference to the pre-test. The Appendix provides the full rubric.

The semi-structured interview schedule consisted of five open-ended questions, designed to address the three research questions. According to Rubin and Rubin (2005), a semi-structured interview is built around a set of predetermined questions but allows the interviewer to follow up on unexpected responses. The researcher gave participants room to switch into Bangla for any question that was difficult to answer in English, and some did switch. The five questions covered ease or difficulty of understanding feedback, recall of a specific transition word or flow change, level of trust in the source, how disagreements were handled, and whether the participant felt ready to write independently. The researcher audio-recorded the interviews with consent, transcribed them verbatim within forty-eight hours, and then anonymized them. With the participants' permission, the researcher stores the audio files.

3.4 Sampling Method

The researcher made two sampling decisions. The quantitative strand used purposive sampling the researcher approached the full available cohort of an English language course at one private university in Dhaka, and enrolled the eighteen students who agreed. The researcher then split the cohort into nine students for the experimental group and nine for the control group. The split balanced the two groups on prior course performance as far as possible, however true random allocation was not feasible. According to Cohen et al. (2018), this is a normal sampling approach in classroom-based research, and one that Mohammad et al. (2025) and Alsofyani and Barzanji (2025) also used in their own quasi-experimental studies of ChatGPT in EFL writing.

The qualitative strand used purposive sampling with variation as the selection criterion. In the experimental group, the three interviewees were Ex_IO1, Ex_IO2, and Ex_IO3, who represented the lowest, middle, and high-middle of the post-test distribution within that group. In the control group, the three interviewees were Con_IO1, Con_IO2, and Con_IO3, who similarly covered the full range. According to Nyimbili and Nyimbili (2024), purposive sampling with maximum variation is appropriate when the goal is to capture a range of experiences. All six interviewees had shown some degree of engagement during the practice period, which means the interview data may slightly tilt toward the more engaged end of each group.

3.5 Data Collection Method

The data collection took place across approximately three weeks, between the eighth of March and the twenty-ninth of March 2026. Firstly, both groups sat the pre-test on the topic "Any Food Stall" under identical conditions. Secondly, the two groups separated for the practice period of two weeks. The experimental group had four scheduled in-class sessions during which they drafted descriptive paragraphs with the help of ChatGPT and Gemini, using the free, publicly available versions of these tools. The researcher gave participants specific guidance on how to ask the AI for suggestions on transitions, paragraph organisation, and coherence, rather than for a complete paragraph. The researcher and the course teacher circulated during these sessions and reminded participants to compare AI suggestions with their own drafts before deciding what to accept. According to Liu et al. (2024), this kind of humans-in-the-loop arrangement is important for keeping AI (LLMs) in a scaffolding role rather than allowing it to become a ghostwriter.

Thirdly, the control group had one matched session during the same two-week window. Working on descriptive paragraph drafts in pairs, while exchanging feedback on three factors: organization, transitions, and clarity. The researcher provided a short peer-review guide before the test, in a mini-lecture. As the interview data later revealed, however, several participants in the control group preferred to write on their own and used peer feedback only for occasional vocabulary checks this is itself a finding, and it is discussed in Chapter Four.

Fourthly, on twenty-ninth of March 2026, both groups sat for the post-test on the same topic, "City Bus or Metro Rail in Rush Hour". The researcher had informed them about this previously after the pre-test. The day after, three participants from each group were present, and the researcher then conducted one-on-one semi-structured interviews on the class schedule as described in Section 3.3. Each interview lasted an average of five minutes.

3.6 Data Analysis Method

The data analysis followed the mixed-methods structure of the design quantitative and qualitative data were analysed separately at first, after which the two sets of findings were brought together for joint interpretation in Chapter Four. The quantitative analysis was descriptive in character. The researcher tabulated the pre-test and post-test scores, calculated the group means, computed the gain for each individual and for each group, and compared the distributions across the two groups. Given the sample size of nine per group, inferential statistical testing was kept modest. According to Creswell (2014), the depth of statistical analysis in a mixed-methods exploratory study should be matched to the size of the sample. The analysis specifically examined the baseline difference, the change in the mean, the change in the spread, and the change in the floor and ceiling of each distribution, and these features are reported in full in Chapter Four through Tables 4.1, 4.2, 4.3, and 4.4.

The qualitative analysis followed the six-phase thematic analysis approach articulated by Braun and Clarke (2006). The researcher familiarised herself with the data through repeated reading of the six transcripts, generated initial codes with attention to recurring ideas such as usefulness of feedback, trust in the source, role of the participant's own judgement, perceived readiness to write independently, and concerns about over-reliance, grouped these codes into broader candidate themes, reviewed them against the full transcript data, and named and refined the final themes. Where a theme appeared in both groups, the researcher noted the different shape it took in each rather than merging the two sub-cases. The joint interpretation is presented in Section 4.4 of Chapter Four.

3.7 Ethical Considerations

The present study followed standard ethical practices for educational research with human participants. Before the pre-test, the researcher explained the purpose of the study, the use of AI tools in the experimental condition, and the role of peer feedback in the control condition. All eighteen participants gave written informed consent, and the six interviewees gave separate verbal consent at the start of each interview session. The study posed no foreseeable physical or psychological risk to participants, and no participant withdrew at any stage.

To protect participant privacy, the researcher blurred all names of the participants throughout the thesis. In the quantitative tables (Tables 4.1 and 4.2), all eighteen participants appear under participant codes P1_Con through P9_Con for the control group, and P1_Ex through P9_Ex for the experimental group. For the qualitative analysis in Sections 4.2 and 4.3, the six interviewees appear under separate interviewee codes Con_IO1, Con_IO2, and Con_IO3 for the control group, and Ex_IO1, Ex_IO2, and Ex_IO3 for the experimental group. The two coding schemes map onto one another as follows: Con_IO1 is P9_Con, Con_IO2 is P4_Con, and Con_IO3 is P8_Con in the control group; Ex_IO1 is P9_Ex, Ex_IO2 is P7_Ex, and Ex_IO3 is P6_Ex in the experimental group. Table 4.3, which bridges the quantitative and qualitative findings, uses the interviewee codes so that the reader can trace each interview participant directly into Section 4.2. The researcher also stores the verbatim transcripts of the six interviews, along with the original audio recordings, securely for future research analysis with the participants' permission. According to Cohen et al. (2018), retaining interview data for re-analysis can support deeper qualitative insights in later projects, while continued anonymisation protects the identity of the original participants.

Chapter 4

Findings and Discussion

This chapter presents the data of the present study and reads it against the questions set out in Chapter One. Section 4.1 reports the quantitative findings from the pre-test and the post-test, moving through individual scores, paired interview-participant changes, and group-level means. Section 4.2 reports the qualitative findings from the six semi-structured interviews, organised around five themes. Section 4.3 brings the numbers and the interviews together. Section 4.4 answers the three research questions directly. The literature reviewed in Chapter Two is brought back in-line where it helps locate the findings within the broader work on AI-assisted writing.

4.1 Quantitative Findings

For Quantitative, two tests were designed, with scores used for measurement of the data. The pre-test was conducted on March 8, 2026, followed by the post-test on March 29, 2026. To evaluate the students' writing, both assessments utilized a ten-point analytic rubric where coherence and organisation were given the highest priority. For the pre-test topic was given on "Any Food Stall," while the post-test required students to write about a "City Bus or Metro Rail in Rush Hour." The researcher deliberately chose both topics to reflect the everyday urban life of Dhaka.

4.1.1 Pre-test Scores

Table 4.1 sets out the pre-test results in the order names appeared on the original score sheet.

Table 4. 1

Pre-test scores of the control and experimental groups (out of 10), 8 March 2026

Serial	Control Group	Score	Experimental Group	Score
1	P1_Con	9	P1_Ex	10
2	P2_Con	8.5	P2_Ex	10
3	P3_Con	7	P3_Ex	10
4	P4_Con	7	P4_Ex	9.5
5	P5_Con	6.5	P5_Ex	8.5
6	P6_Con	6	P6_Ex	8.5
7	P7_Con	6	P7_Ex	8.5
8	P8_Con	6	P8_Ex	8
9	P9_Con	6	P9_Ex	6
	Mean	**6.89**	**Mean**	**8.78**

The two groups did not start from a comparable place. The experimental group's pre-test mean of 8.78 sat 1.89 points above the control group's 6.89. In the experimental group, P1_Ex, P2_Ex, and P3_Ex were at the ceiling; P4_Ex at 9.5; P5_Ex, P6_Ex, and P7_Ex at 8.5; P8_Ex at 8; and P9_Ex at 6. The control group's distribution stretched more widely as the table shows clearly. This baseline difference matters for everything that follows, the experimental group, with three writers already at ten, represents the maximum score on the rubric. Echoing this concern, Siekmann et al. (2022) highlight the potential for ceiling effects when using analytic rubrics to evaluate highly proficient EFL writers. Consequently, they caution against interpreting gain scores independently of the overall score variance. P9_Ex,

the lowest scorer in the experimental group at pre-test, sits at the center of what the experimental condition could and could not do for a learner at the lower edge of the cohort.

4.1.2 Post-test Scores

Table 4.2 presents the post-test results, ranking participants from highest to lowest within each group.

Table 4. 2

Post-test scores of the control and experimental groups (out of 10), 29 March 2026

Serial	Control Group	Score	Experimental Group	Score
1	P4_Con	10	P1_Ex	10
2	P9_Con	10	P9_Ex	10
3	P2_Con	10	P3_Ex	10
4	P3_Con	8.5	P7_Ex	10
5	P1_Con	8	P6_Ex	9.5
6	P8_Con	8	P2_Ex	9
7	P7_Con	7	P4_Ex	8.5
8	P5_Con	6.5	P8_Ex	8
9	P6_Con	5	P5_Ex	8
	Mean	**8.11**	**Mean**	**9.22**

Movement is visible in both groups. The experimental group's range narrowed from four points (six to ten) to two (eight to ten); while the control group's opened from three (six to nine) to five (five to ten). Here standard deviations moved in opposite directions, for the experimental group dropped from 1.30 to 0.87, and the control group’s standard deviation, by

contrast, rose from 1.14 to 1.75. Following Siekmann et al. (2022), this distributional information carries a layer of meaning the group mean alone cannot capture.

P6_Con moved from six to five, below the floor of the pre-test distribution itself. This is the single case suggests peer review without tighter scaffolding did not give the weakest writer in the cohort enough support to consolidate his writing. In the experimental group, by contrast, the floor of the distribution rose from six to eight. Three experimental participants finished slightly lower than they started P2_Ex from ten to nine, P4_Ex from 9.5 to 8.5, and P5_Ex from 8.5 to eight but all three stayed in the upper half of the rubric. Two readings are possible. First, these three students started near the maximum possible score. A drop of half a point or a point likely just reflects normal day-to-day fluctuations rather than a genuine loss of skill, especially since the grading rubric had no room left at the top to measure any actual progress. Second, some writers who were already highly skilled might not have gained any extra benefit from the AI support simply because the rubric couldn't record scores any higher. We can observe similar outcomes in the control group where P1_Con's score dropped from nine to eight; where P5_Con's score remained same at 6.5. Ultimately, pattern variations were observed in both cases, students whose scores did not increase appeared in both groups.

4.1.3 Paired Pre- to Post-test Changes for the Six Interview Participants

To link the quantitative and qualitative findings, Table 4.3 presents the paired changes of the six interview participants.

Table 4. 3

Pre-test to post-test change for the six interview participants

Group	Participant	Pre-test	Post-test	Change
Experimental	Ex_IO1	6	10	+4.0
Experimental	Ex_IO2	8.5	10	+1.5
Experimental	Ex_IO3	8.5	9.5	+1.0
Control	Con_IO1	6	10	+4.0
Control	Con_IO2	7	10	+3.0
Control	Con_IO3	6	8	+2.0

Three observations stand out. First, all six interview participants improved; none declined or stayed flat. Second, the most significant score improvements came from Ex_IO1 in the experimental group and Con_IO1 in the control group, as both advanced from a six to a ten. Yet, while their point increases were identical, their interviews revealed that the actual thought processes and methods they used to achieve those results were completely different. Third, the two highest starters in the experimental sub-sample Ex_IO2 and Ex_IO3, both at 8.5 registered smaller absolute gains of 1.5 and 1.0 points. This is the ceiling effect at work, and it warns against reading the experimental group's smaller group-level gain as evidence that AI scaffolding is less effective than peer scaffolding. The experimental group simply had less space on the rubric within which to demonstrate improvement. For the finer comparison between the two groups, the interviews become essential, because they reveal what the writing process actually looked like from the inside.

4.1.4 Group-Level Means

Table 4.4 summarises the group-level means and gains side by side.

Table 4. 4

Group-level pre-test and post-test means (out of 10)

Group	n	Pre-test Mean	Post-test Mean	Gain
Control	9	6.89	8.11	+1.22
Experimental	9	8.78	9.22	+0.44

On the surface, Table 4.4 shows the control group producing a larger raw gain of 1.22 points compared to the experimental group's 0.44 points. This has to be read with care. The control group had far more space on the rubric within which to move, because its pre-test mean of 6.89 was almost two points below the experimental group's 8.78. The experimental group started closer to the ceiling and had less room within which to rise consistent with Barrot (2023), who cautioned that LLM effects on writing outcomes can look smaller than they actually are when learners are already writing at a high level. A second reading of Table 4.4 concerns distribution rather than mean. The experimental group finished with a tighter cluster; the control group's distribution opened up. The narrowing of the experimental variance, from a range of four to a range of two, is a real finding: AI scaffolding, as practised here, pulled the weaker writers upward without pushing the stronger ones down, which produced the tight upper-range cluster in Table 4.2. The widening of the control variance, from a range of three to a range of five, is a different pattern peer review lifted some participants to the ceiling while leaving others at or below their starting point. A condition that narrows the variance is often more useful than one that merely raises the mean, because narrower variance means fewer learners are left behind.

A third key takeaway relates to what Liu et al. (2024) describe as the 'floor effect,' which looks at what happens to the lowest scores in a dataset. The minimum score in the experimental group increased from a six to an eight, whereas the minimum score in the control group dropped from a six to a five; showing a notable difference. This creates a combined six-point difference at the very bottom of the grading scale a major shift that gets completely hidden if you only look at the overall average scores of the two groups. A classroom intervention that lifts the floor of its distribution is doing something specific: supporting the learners who previously had the least coherence, rather than only the strongest.

Any judgement about which condition was “more effective” depends on which feature of the distribution one privileges. The mean favours the control group; the floor, the ceiling-cluster density, and the narrowing of the spread favour the experimental group. For a Dhaka private university classroom, where ability gaps within a single cohort are common, the narrowing-variance pattern is arguably the more useful pedagogical outcome.

4.2 Qualitative Findings from the Interviews

The researcher analysed the six interview transcripts using Braun and Clarke's (2006) six-phase thematic analysis, as described in Chapter Three. Five themes emerged. Four appeared in both groups, in very different shapes. The fifth concerns about AI use and cognitive effort appeared only in the experimental group, but was strong enough there to deserve its own sub-section.

4.2.1 Theme One: Usefulness of the Feedback Received

The two groups talked about feedback in strikingly different ways. In the experimental group, all three interviewees described the AI as useful in concrete, paragraph-level ways. Ex_IO1 said the AI “helped me in to identify Grammatical error and also increase Vocabulary storage,” and added that it “present the writing a little differently, like they give transition words differently” (Ex_IO1). The second part is a direct reference to coherence-relevant scaffolding. She also noted she could understand the AI's explanations because “I talk with AI most of the time. So, it was better for me to understand” (Ex_IO1). Ex_IO2 described the usefulness in organisational terms: the AI “tells you step by step where to use what, so it was quite easy to understand” (Ex_IO2). The phrase “step by step where to use what” is essentially a lay articulation of what Liu et al. (2024) call strategy-level scaffolding inside the learner's Zone of Proximal Development. For Ex_IO3, the AI was especially useful when “the handout was missing some information, but the powerpoint slide may have short points; in that case we use Chat GPT or Gemini to get detailed information” (Ex_IO3). Her framing positions the AI as an information-supplementation tool rather than a drafting tool.
In the control group, the three interviewees described a much thinner kind of usefulness. Control group interviewees consistently reported minimal peer interaction focused on vocabulary (Con_IO1, Con_IO2, Con_IO3).

When evaluated against the framework in Chapter Two, this emerges as the most significant difference between the two groups. The experimental group utilized their expert guide to

improve their writing at the paragraph level focusing on smooth transitions, logical step-by-step organization, and adding context where the original materials were lacking. In contrast, the control group depended on peer feedback that mainly targeted individual word choices. Although both forms of help are useful, only the paragraph-level support actively builds overall coherence which is the primary focus of this study. Rafin (2025), in her BRAC University thesis on peer collaboration, had warned about this exact pattern, noting that peer work in Bangladeshi tertiary classrooms tends to stop at vocabulary and grammar rather than reaching into organisation and the present control group data fit that pattern precisely.

4.2.2 Theme Two: Trust in the Feedback Source

The second theme is trust. In the experimental group, all three interviewees expressed a limited, critical trust in the AI. Ex_IO1 said, “Sometimes I disagree,” and when asked how she handled the disagreement, she replied, “In the middle” (Ex_IO1) the posture of a learner who neither rejects the AI outright nor accepts it without thinking.

In the control group, trust took a different shape. Con_IO1 was blunt: “Most of it's 99% I disagree with them. I can trust my peers that much” (Con_IO1). Asked why, she explained, “I think in terms of vocabulary, I suggest better ones than them” (Con_IO1). Her trust was filtered through her own estimate of her peers' ability. Con_IO2 framed trust come through personal relationship: “Not everyone, The peers which I know just them” (Con_IO2). For her, familiarity with the person was the condition of trust. Con_IO3 took a third position: “That could be trustable. But in my case I didn't need anyone” (Con_IO3). He did not disagree with peer feedback as a concept; nor did he consider himself to need it. Peer feedback was filtered through personal relationships, through the participant's own estimate of peer ability, and through a broader sense of self-sufficiency. In the language of Chapter Two, the peer MKO was often under-activated in this classroom. The participant did not rate the peer high enough, or close enough personally, to allow the peer to function as a genuine scaffold a drift Rafin (2025) and the BRAC University peer-feedback thesis warned about.

A mirror-image risk on the AI side should also be named. A participant who trusts the AI too much becomes a copier rather than a writer, and outsources precisely the thinking this study was trying to build. None of the three experimental-group interviewees displayed this pattern in their self-reports, though one revealing aside from Ex_IO3 deserves a mention. Asked for an example of an AI suggestion that had improved her paragraph flow, she said, “my brain was not braining, cause I was coping a little” (Ex_IO3). This honest admission, offered freely

by the participant who otherwise articulated the most mature position on AI trust, is evidence that even reflective users sometimes lean on the tool rather than think through a problem.

4.2.3 Theme Three: The Role of Learner Judgement

Both groups talked about judgement, but the object of judgement was different. In the experimental group, judgement was about content. Ex_IO2 described her process directly: "On that case I asked to AI, which one will be more suitable for sentence then I have choose by myself whether it goes for the sentence or not" (Ex_IO2). The final call, she said, was hers, based on "my perception." Ex_IO1 described a content-level judgement, drawing on her real-life experience to sort suggestions: "I decided by read the text, then after look for specific transition word. Like their is a topic in the Pre-test about Tea-stall. So, what happen their; crowed happen, then people gather at a place talked with each other. So, by adding all the ideas then I got transition words idea for my writing" (Ex_IO1). Her reference to the real-life experience of a tea-stall shows her using her own background knowledge of the scene to judge which AI suggestions fit and which did not. Ex_IO3 added a temporal dimension: "When I get the flow of writing, then I don't prefer AI at that time. Then feel like the writing started by me and will be finished by me, in middle AI is a kind of hassle, cause I loose the flow" (Ex_IO3). She also said, "If I start from my own then I got a satisfaction if I finish that also" (Ex_IO3).

In the control group, judgement was about the reliability of the source rather than the content. Con_IO2 described it plainly: "Well, depending on my judgement whether I should believe them or not, I also think of their review as well, then I decide to agree or not agree with their review" (Con_IO2). Her judgement was directed at whether the classmate was a credible reviewer. Con_IO1 described her judgement through her personal background. Con_IO3 described his judgement through topic familiarity: "If I know the question topic or have some knowledge before about that content. I would be able to write that" (Con_IO3). These are reasonable stances. But they are stances about the helper or the topic, rather than about the logical shape of the paragraph. The difference is small but real, and it points to a useful classroom implication: if peer review were restructured with clearer coherence-oriented prompts a checklist that asks the peer to comment specifically on topic sentence, order of details, and transition words then peer scaffolding might begin to produce the paragraph-level judgement work that the AI scaffolding already seems to produce.

4.2.4 Theme Four: Perceived Readiness to Write Independently

All six interviewees, in both groups, said they could now write a descriptive paragraph on their own without the scaffolding they had used during the practice. In the experimental group, Ex_IO1 attributed her readiness to "enriched vocabulary and synonyms. Synonyms help a lot to write further and also grammar. Cause If there occur too many grammatical errors then marks deduct a lot" (Ex_IO1). Ex_IO2 said, "If I attend the exam, of course I will perform better. But if AI was there it would have been better. Though AI is not there I could write on my own, I won't say that it is impossible to give exam" (Ex_IO2). Ex_IO3 framed confidence and background knowledge for readiness express that: "I have full confidence on me, that's why," and "Obviously, confidence comes from background knowledge. Or else I couldn't write that" (Ex_IO3).

In the control group, Con_IO2 was also confident: "Because, paragraph writing is common for us; So, I got some practice but this was a new type for me. I have learned from the lecture and handout" (Con_IO2). She attributed her readiness to the course materials rather than to peer feedback. Con_IO1 reported a learnt strategy directly: "I actually, if it's not a surprise test, I actually review it before starting for the topics and how to especially, how to arrange the sequence of sentence, like using Finally, there for this connective words" (Con_IO1). Her answer is specifically about transitions, which maps directly onto the coherence focus of the practice. Con_IO3 said he could write on his own if he knew the topic beforehand, and added, "Yes, also imagine that before writing on the topic" (Con_IO3). Readiness, across both groups, is framed not as the absence of the tool but as the presence of internal resources which is precisely what the sociocultural framework from Chapter Two would predict, because the whole point of a scaffold is to leave something internalised behind when it is withdrawn. The experimental group's answers mention more coherence-relevant resources; the control group's lean more on course materials and topic familiarity.

4.2.5 Theme Five: AI Use and Cognitive Effort

The fifth theme appeared only in the experimental group, but appeared clearly there and deserves its own sub-section. Ex_IO3 articulated it most fully: "I didn't think so we should fully rely on AI. Then our brain will not be making decisions. If we use AI always; now a time whatever situation we are now just go to AI for solution. If we brainstorm for 5 to 6 seconds, we could figure out the solution. But now our brain works a little slower than before" (Ex_IO3). The statement comes from an active user, not an external critic. Ex_IO3

used ChatGPT and Gemini during the practice and found them helpful, yet held a separate worry about what habitual use does to the learner's own cognitive effort over time. Her observation “our brain works a little slower than before” tells which echoes what Khan et al. (2023) expressed on a Bangladeshi tertiary sample. Her concern is generational rather than individual; “we” and “our brain,” a concern articulated about her cohort, not only herself. A milder version appeared in Ex_IO2's interview: “AI just comes freshly. If we saw a little earlier, we were writing from our own before AI” (Ex_IO2). Alharbi and Al-Ahdal (2025) also found similar concerns among Saudi EFL learners. Ex_IO3 gave part of her interview in Bangla, and her concern that the current generation goes to AI for a solution in every situation rather than thinking for a few seconds first has a weight that extends beyond the writing classroom.

4.3 Bringing the Quantitative and Qualitative Findings Together

Placing the numbers of Section 4.1 alongside the interviews of Section 4.2 yields three broader readings.

The first reading is that both AI-supported practice and peer-review practice produced score movement across the two-week window, though the kind of movement was different in each. The control group produced the larger raw gain in group mean, from 6.89 to 8.11, a jump of 1.22 points. The experimental group produced a smaller mean gain of 0.44 points, from 8.78 to 9.22, though it produced a tighter post-test distribution, a higher floor, and a larger density of ceiling scores. Hence, on the numerical evidence alone, neither condition clearly dominates. While every condition led to a measurable impact, they affected different areas of our results. This supports the finding by Barrot (2023) that AI-assisted writing gains are very real, but they are simply harder to detect when the writers are already quite advanced.

Second, the scaffolding the two groups actually received was different in kind, not only in source. The AI group received feedback that was paragraph-level, strategy-oriented, and concerned with transitions, structure, and step-by-step organisation. The peer group received feedback that was mostly word-level, vocabulary-oriented, one item at a time. The distinction is clearly visible in the language the participants themselves used; Ex_IO2's “step by step where to use what” is a very different kind of statement from Con_IO2's “Vocab and key words just.” Both are honest reports of the practice as it actually happened, but they describe very different forms of scaffolding. Coherence is the outcome of interest; as a result, paragraph-level scaffolding is more directly on target than word-level scaffolding, because

coherence is a paragraph-level skill, not a word-level one. This is what the framework in Chapter Two would predict. Vygotsky's ZPD requires a More Knowledgeable Other whose support is pitched at the level of the task the learner is trying to accomplish; a descriptive paragraph is a paragraph-level task, so the MKO has to function at the paragraph level for scaffolding to land. The AI did this almost by default. The peer, by contrast, was being asked for vocabulary and gave vocabulary peer scaffolding in this classroom was not under-performing its own level; it was performing below the level of the task.

The third reading concerns about trust and judgement, which are the threads that actually mediate the scaffolding effect on both sides. A learner who trusts the MKO too much allows their own thinking to be shaped entirely by it. A learner who trusts too little rejects useful help. The six interviewees, across both groups, mostly took middle positions of critical trust. Con_IO1 rejected most peer suggestions; Con_IO2 accepted only from peers she knew; Con_IO3 preferred to work alone; Ex_IO1 landed “in the middle”; Ex_IO2 could “not fully depend on AI”; Ex_IO3 articulated the clearest use-and-vigilance stance. These are all mature positions for an undergraduate to hold. However, the different trust profiles carry different consequences for learning. The experimental-group participants trusted the AI enough to shape it as their paragraph structure, while keeping their own judgement active. The control-group participants trusted the peer so little that almost no scaffolding came through at all.

The quantitative findings are insufficient to confirm that LLMs offer a superior advantage for improving coherence compared to peer review. The sample is small; nine per group, the two groups were not equivalent at baseline. The intervention was two weeks long, the rubric had a clear ceiling effect, and coherence was weighted heavily within the rubric but not measured through a dedicated discourse-analytic instrument. The present study can support is a modest, well-bounded claim that LLMs, when framed as a scaffold inside the learner's ZPD and used with critical judgement, give participants an enriched set of tools for talking about their own writing, a clearer articulation of paragraph organisation. And a more specific awareness of transition words all coherence-relevant, and all more visible in the experimental group than in the control group. Why the data look this way, in this Dhaka context, comes back to the affordance the AI brought into the room: ChatGPT and Gemini gave each experimental participant a private, on-demand More Knowledgeable Other that responded at the paragraph level. No equivalent resource was available in the control room.

4.4 Discussion in Relation to the Research Questions

The first research question asked whether short-term use of LLMs enhances coherence in descriptive writing by private university EFL learners in Dhaka, when compared to a control group using peer review. The evidence is broadly positive. The experimental group finished with a denser cluster of high scores (four at ten versus three in the control), a narrower range (two versus five), a higher floor (eight versus five), and a lower standard deviation (0.87 versus 1.75). The interview data described more coherence-relevant internal resources transitions, step-by-step structure, and paragraph flow. LLMs raised the floor more effectively; the claim within the scope of this study is that AI scaffolding provides more coherence-directed support to the experimental group than peer scaffolding delivered to the control group in the same classroom, over the same two-week window. The mean comparison favours the control group because of the baseline imbalance and the ceiling effect, but the distributional features and the interview testimony favour the experimental condition.

The second research question asked how experimental-group learners described the usefulness and trustworthiness of AI feedback, and how these descriptions differed from the control group's. The interviews show a clear divergence. The participants of the experimental group described LLMs feedback as paragraph-level and strategy-oriented, placing moderate, critical trust in it; they kept their own judgement active and reported specific moments of disagreement and middle-path decisions. The control group described peer feedback as mostly word-level and occasional. They tended to exercise judgement on the peer rather than on the content of the suggestion. Neither group was passive. What differed was the level at which the scaffolding, and the trust, operated.

The third research question asked what this comparison means for integrating LLMs into descriptive writing instruction at the tertiary level in Bangladesh. Briefly, teachers can use LLMs for writing in classrooms of private universities in Dhaka, provided they use them carefully as a scaffold for coherence. They should, however, accompany the tool with explicit guidance on how to prompt it, how to compare its suggestions with one’s own draft, and how to maintain a critical distance from the output. Peer review still remains important, but teachers need to reorganize it around coherence-based prompts and a clear peer-review guide if it is to achieve the same paragraph-level support that the AI scaffold already appears to provide. Chapter Six discusses the fuller implications for instruction, curriculum, and policy.

Chapter 5

Limitations of the Study

This study has several limitations. First, the intervention covered only two weeks, which is insufficient to assess long-term coherence development. The present study lacked a follow-up; Liu et al. (2024) included a one-month follow-up. Secondly, the researcher collected data from one private university in Dhaka, which limits the generalizability. Thirdly, the sample size was small (n=9 per group), making it impossible to conduct inferential statistical analysis or subgroup analysis. Fourthly, true random assignment was not feasible given the existing class schedules. Lastly, the ten-point rubric created a ceiling effect for four experimental participants who had already reached the maximum score. A major limitation of this study is the baseline imbalance between the control and experimental groups, which reduced the comparability of gain scores. The experimental group entered the study with a pre-test mean of 8.78, nearly two points higher than the control group's mean of 6.89. This 1.89-point gap means the two groups did not begin from an equivalent starting point, and any direct comparison of raw gain scores must be interpreted with caution. The control group had considerably more room on the rubric within which to improve, while the experimental group was already clustered near the ceiling. This structural asymmetry makes it difficult to attribute the control group's larger raw mean gain of 1.22 points to the effectiveness of peer review alone, just as the experimental group's smaller gain of 0.44 points should not be read as evidence that AI scaffolding was less effective. The distributional features the narrowing of variance, the rising floor, and the density of ceiling scores provide a more informative picture than the mean gains alone, but even these should be read with the baseline difference in mind. Future studies should use stratified random assignment or proficiency-matched groups to avoid this confound.

Chapter 6

Recommendations and Implications

Based on what Chapter Four found and considering the limitations from Chapter Five, there are some practical implications for how descriptive writing could be taught in Bangladeshi universities. These recommendations come directly from what actually happened in this study, not from general speculation about AI in education.

The first and probably most useful finding for teachers is about how to use ChatGPT or Gemini in writing classes. The experimental group students said the AI helped them most when they asked it for help with paragraph organization things like what transition words to use, how to arrange their ideas in order, and how to make the flow better. Ex_IO2 talked about how the AI told her "step by step where to use what," whereas Ex_IO1 said it helped her find different transition words. This is very different from just asking the AI to write the whole paragraph for you. So teachers in Dhaka could design short exercises where students write a first draft on their own first, then ask the AI for maybe three specific suggestions about organization, and then the students have to write down which suggestions they kept and why they rejected the others. This way the AI is helping but the student is still making the decisions.

Time to rethink about peer-feedback. In the control group, students said that most feedback they got from peers was mostly about vocabulary single words or synonyms. Almost nobody got feedback about how their paragraph was organized or whether the ideas flowed well. This is actually the same problem that Rafin (2025) and some other local studies found. So peer review by itself is probably not enough if the goal is to improve coherence. But peer review could work better if teachers gave students a specific checklist to follow like "Did your partner use a clear topic sentence?" or "Are the details in a logical order?" or "Check if they used transition words." This kind of checklist would force peers to pay attention to paragraph-level issues instead of just vocabulary.

Teachers should openly discuss the over-reliance on AI. Ex_IO3 said something important she worried that if students always use AI for everything, “our brain works a little slower than before.” Khan et al. (2023) and Alharbi and Al-Ahdal (2025) found similar concerns in their studies. Instead of ignoring this growing issue, teachers could openly discuss the pros and cons of using AI. One simple way would be to have students write a first draft without

any help, then use AI for suggestions, and then write a short reflection about which AI suggestions actually made their paragraph better and which ones they decided not to use. This encourages students to critically analyse rather than simply depend on AI blindly.

For future research, several directions make sense. First, someone should do this same study but for a whole semester instead of just two weeks, with follow-up tests at three months and six months to see if the improvements last. Second, doing the study at three or four different universities instead of just one would show whether the results hold up in different contexts. Third, using more detailed measurement like actually counting transition words or analyzing how ideas connect would give a clearer picture of where the scaffolding effect is coming from. Fourth, it would be interesting to compare three conditions: AI scaffolding, peer scaffolding with a proper checklist, and teacher scaffolding, all side by side. Fifth, since some students like Ex_IO3 think about writing in both Bangla and English, a study looking at how Bangladeshi students use LLMs in both languages would be really original and useful.

Universities should probably develop clearer policies about AI use. Right now most private universities either have no policy or have something very vague. What would be more useful is a policy that distinguishes between using AI as a tool to help you learn (which is good) and using AI to do your work for you (which is not). A policy built around this distinction would give both teachers and students clear guidance about what is okay and what is not.

Chapter 7

Conclusion

The practical message from all of this, which Chapter Six explains in detail, is that teachers can use LLMs like ChatGPT in Dhaka university writing classes to help with coherence, but they have to use them carefully. They work best when students ask for help with paragraph organization rather than asking the AI to write everything for them. Teachers still need to actively involve themselves in designing the tasks and guiding the process. And the class culture needs to treat the AI as a learning tool, not as a way to avoid actually learning to write. Peer review can still be useful, but it needs more structure like giving students a specific checklist of things to look for related to coherence if it is going to compete with what AI already does naturally.

What this thesis contributes is fairly specific but real. It adds a Bangladeshi perspective to a conversation that research from other countries has dominated. It focuses specifically on descriptive writing and coherence instead of treating all writing as the same. It combines numbers from tests with what students actually said in interviews, which lets the two types of data explain each other. And it compares peer scaffolding and AI scaffolding side by side in the same class over the same time period, which no previous local research has done. These are modest contributions given the limitations from Chapter Five, but they are based solidly on what actually happened with these eighteen students, and that is worth something.

References

1. Akter, S. (2022). Challenges of teaching writing skills at tertiary level in Bangladesh. *International Journal of English Language Teaching, 10*(4), 45–62.
2. Aldera, A. S. (2016). Coherence and cohesion in EFL essay writing: A study of Saudi EFL learners at the Master's level. *English Language Teaching, 9*(7), 153–168.
3. Alharbi, W. S., & Al-Ahdal, A. A. M. H. (2025). ChatGPT as a complementary mental tool: Exploring Saudi EFL learners' perspectives and practices. *Computer Assisted Language Learning, 38*(1), 112–135.
4. Alsofyani, M. M., & Barzanji, A. A. (2025). The impact of ChatGPT and teacher feedback on EFL students' writing performance and perceptions. *Education and Information Technologies, 30*(2), 245–268.
5. Bai, B., & Wang, J. (2021). The role of self-regulated learning strategies in writing performance. *System, 99*, 102512.
6. Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing, 57*, 100745.
7. Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing, 17*(2), 102–118.
8. Bitchener, J., & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. Routledge.
9. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101.
10. Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.
11. Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Pearson.

12. Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). SAGE Publications.
13. Grabe, W. (2001). Notes toward a theory of second language writing. In T. Silva & P. K. Matsuda (Eds.), *On second language writing* (pp. 39–57). Lawrence Erlbaum Associates.
14. Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman.
15. Hyland, K. (2003). *Second language writing*. Cambridge University Press.
16. Hyland, K. (2019). *Second language writing* (2nd ed.). Cambridge University Press.
17. Islam, M. S. (2025). *Exploring the use of ChatGPT among Bangladeshi EFL learners: Perceptions and practices* [Unpublished master's thesis]. BRAC University.
18. Kayum, S., & Khan, M. R. (2015). Writing problems among tertiary level EFL learners in Bangladesh. *Journal of Education and Practice, 6*(28), 116–123.
19. Khan, R. M. I., Montakim, F. B., Hasan, A. R., & Rahman, M. (2023). Bangladeshi tertiary students' perceptions of using AI writing tools in academic contexts. *Asian Journal of Applied Linguistics, 10*(2), 87–104.
20. Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 54*(2), 537–550.
21. Liu, X., Zheng, Y., Du, X., & Moon, Y. L. (2024). Enhancing EFL writing with ChatGPT-assisted language learning: An empirical study on the CALLA-LLM framework. *Computer Assisted Language Learning, 37*(1–2), 1–28.
22. Mohammad, A. M., Assiri, M. S., & Al-Qahtani, M. H. (2025). The effectiveness of ChatGPT in enhancing EFL students' research writing skills: A quasi-experimental study. *Educational Technology Research and Development, 73*(1), 156–178.
23. Mustaque, S. (2014). Problems and challenges faced by Bangladeshi tertiary level students in writing composition. *Journal of Language and Literature, 5*(3), 172–179.

24. Nyimbili, F., & Nyimbili, L. (2024). Purposive sampling with maximum variation in qualitative educational research. *International Journal of Research Methodology, 15*(1), 34–48.
25. Patwary, S. T., Rahman, M. M., & Hoque, M. E. (2023). Writing difficulties of tertiary level EFL students in Bangladesh: A comprehensive analysis. *English Language Teaching, 16*(5), 112–129.
26. Rafin, T. A. (2025). *Peer collaboration in English language learning: Benefits and challenges in Bangladeshi tertiary contexts* [Unpublished master's thesis]. BRAC University.
27. Rubin, H. J., & Rubin, I. S. (2005). *Qualitative interviewing: The art of hearing data* (2nd ed.). SAGE Publications.
28. Siekmann, S., Keogh, J., Charles, C., & Song, K. (2022). Coherence and cohesion in EFL academic writing: Challenges and teaching strategies. *Journal of English for Academic Purposes, 58*, 101128.
29. Uddin, M. S., & Bailey, D. R. (2024). Artificial intelligence in English language education in Bangladesh: Opportunities and challenges. *Asian EFL Journal, 28*(1), 45–67.
30. Ulfat, S. (2023). Technology adoption among English medium school teachers in Bangladesh. *Journal of Educational Technology Systems, 51*(3), 298–315.
31. Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
32. Xu, W., & Jumaat, N. F. (2024). Exploring ChatGPT's potential in writing instruction: Benefits and concerns among Chinese university students. *Interactive Learning Environments, 32*(4), 567–585.

33. Yusuf, M. A., Chowdhury, T. A., & Islam, N. (2025). Bangladeshi EFL learners' perceptions of Grammarly in academic writing contexts. *Computer Assisted Language Learning Electronic Journal, 26*(1), 112–134.

Appendix

Pre-Test Questionaries

SET 1: Control Group (Traditional Peer-Feedback)

(No devices allowed.)

Descriptive Writing Task (15 Minutes)

Instructions: You have exactly 15 minutes to write a short descriptive paragraph and participate in a peer-review session. You will be evaluated specifically on **coherence and spatial organization**, how logically your ideas flow from one to the next.

The Prompt: Write a **100–120 word paragraph** describing the chaotic environment of a busy street food stall (Tong) in Dhaka during a crowded evening.

Your 15-Minute Timeline:

- **Minutes 0–10 (Drafting):** Write your paragraph. Structure your description with a clear physical direction (e.g., start by describing the vendor, then the food, then the crowd).
- **Minutes 10–15 (Peer-Feedback):** Stop writing. Swap your paper with the student sitting next to you.
    - **As a Reviewer:** Read your partner's draft. Circle any awkward sentences where the spatial flow breaks. Write down 3 suggested transition words (e.g., *adjacent to, meanwhile, in the background*) they could use to fix it.
    - **As a Writer:** Get your paper back, read your partner's suggestions, and quickly write in the corrections.

SET 2: Experimental Group (LLM-Assisted Feedback)

(You need your smartphones for the final 5 minutes.)

Descriptive Writing Task (15 Minutes)

Instructions: You have exactly 15 minutes to draft a short descriptive paragraph and use AI to help you improve its structural flow. You will be evaluated on your final **coherence and spatial organization**.

The Prompt: Write a **100–120 word paragraph** describing the chaotic environment of a busy street food stall (Tong) in Dhaka during a crowded evening.

Your 15-Minute Timeline:

- **Minutes 0–10 (Drafting):** Write your paragraph on this paper. Structure your description with a clear physical direction (e.g., start by describing the vendor, then the food, then the crowd). Do not use your phone yet.
- **Minutes 10–15 (LLM Feedback):** Open ChatGPT or Gemini on your phone. Type your draft into the app along with this exact prompt: *"I wrote this descriptive paragraph. Please review its coherence. Do my sentences flow logically in a spatial order? Give me 3 specific transition words I can add to make the flow smoother."*
    - **Action:** Read the AI's feedback. Use your pen to edit your handwritten draft, crossing out awkward sentences and adding the AI's suggested transition words to improve your coherence.

Post-Test Questionaries

SET 1: Control Group Post-Test (Traditional Peer-Feedback)

(No devices allowed.)

Final Descriptive Writing Task (15 Minutes)

Instructions: This is your final assessment. You have exactly 15 minutes to write a short descriptive paragraph and participate in a peer-review session. You will be evaluated specifically on **coherence and spatial organization**, how logically your ideas flow from one to the next based on your practice over the last two weeks.

The Prompt: Write a **100–120 word paragraph** describing the crowded, chaotic environment inside a local city bus or the Dhaka Metro Rail during evening rush hour.

Your 15-Minute Timeline:

- **Minutes 0–10 (Drafting):** Write your paragraph. Structure your description with a clear physical direction (e.g., start from the entrance doors and move your description toward the back of the vehicle).
- **Minutes 10–15 (Peer-Feedback):** Stop writing. Swap your paper with the student sitting next to you.
    - **As a Reviewer:** Read your partner's draft. Circle any awkward sentences where the spatial flow breaks. Write down 3 suggested transition words (e.g., *further down the aisle, standing next to, meanwhile*) they could use to fix it.
    - **As a Writer:** Get your paper back, read your partner's suggestions, and quickly write in the corrections to polish your final draft.

SET 2: Experimental Group Post-Test (LLM-Assisted Feedback)

(You need your smartphones for the final 5 minutes.)

Final Descriptive Writing Task (15 Minutes)

Instructions: This is your final assessment. You have exactly 15 minutes to draft a short descriptive paragraph and use AI to help you improve its structural flow. You will be evaluated on your final **coherence and spatial organization** based on your practice over the last two weeks.

The Prompt: Write a **100–120 word paragraph** describing the crowded, chaotic environment inside a local city bus or the Dhaka Metro Rail during evening rush hour.

Your 15-Minute Timeline:

- **Minutes 0–10 (Drafting):** Write your paragraph on this paper. Structure your description with a clear physical direction (e.g., start from the entrance doors and move your description toward the back of the vehicle). Do not use your phone yet.
- **Minutes 10–15 (LLM Feedback):** Open ChatGPT or Gemini on your phone. Type your draft into the app along with the exact prompt you practiced: *"I wrote this descriptive paragraph. Please review its coherence. Do my sentences flow logically in a spatial order? Give me 3 specific transition words I can add to make the flow smoother."*

Action: Read the AI's feedback. Use your pen to edit your handwritten draft, crossing out awkward sentences and adding the AI's suggested transition words to improve your coherence.

Interview Questionaries

One-on-One Interview Questions: Experimental Group

Focus: LLM-Assisted Feedback **Time:** 5-8 minutes

Goal: Understand their perception of AI utility, its impact on coherence, and their resulting autonomy.

Segment 1: Perception of Utility

- **Main Question:** "Over the last two weeks, you used ChatGPT or Gemini to help with your writing. Walk me through how easy or difficult it was to get the AI to focus on your paragraph's organization."
- **Follow-up Probe:** "Did you find the AI's explanations easy to understand?"

Segment 2: Impact on Coherence

- **Main Question:** "Think about the specific transition words or spatial flow changes the AI suggested. Tell me how you decided which of the AI's suggestions to keep and which to throw away."
- **Follow-up Probe:** "Can you give me an example of one AI suggestion that you felt really improved the flow of your paragraph?"

Segment 3: Reliability & Trust

- **Main Question:** "Did you feel like you could completely trust the AI's feedback, or were there times you disagreed with its advice?"
- **Follow-up Probe:** "When you disagreed with the AI, how did you decide what to write in your final draft?"

Segment 4: Learner Autonomy

- **Main Question:** "If you had to write a descriptive paragraph tomorrow for an exam without your phone or any AI, do you feel this practice actually taught you how to organize sentences better on your own?"
- **Follow-up Probe:** "Why do you feel that way?"

One-on-One Interview Questions: Control Group

Focus: Traditional Peer/Teacher Feedback **Time:** 5-8 minutes

Goal: Understand their perception of peer utility, its impact on coherence, and their resulting autonomy to compare against the AI group.

Segment 1: Perception of Utility

- **Main Question:** "Over the last two weeks, you reviewed paragraphs with a classmate. Walk me through how easy or difficult it was to understand the feedback your partner gave you about your organization."
- **Follow-up Probe:** "Did you find their written suggestions clear and helpful?"

Segment 2: Impact on Coherence

- **Main Question:** "Think about the specific transition words or flow changes your partner suggested. Tell me how you decided which of their suggestions to keep and which to throw away."
- **Follow-up Probe:** "Can you give me an example of a suggestion from your classmate that you felt really improved the flow of your paragraph?"

Segment 3: Reliability & Trust

- **Main Question:** "Did you feel like you could completely trust your classmate's feedback, or were there times you disagreed with their advice?"
- **Follow-up Probe:** "When you disagreed with your classmate, how did you decide what to write in your final draft?"

Segment 4: Learner Autonomy

- **Main Question:** "If you had to write a descriptive paragraph tomorrow for an exam without a classmate to review it, do you feel this practice actually taught you how to organize sentences better on your own?"
- **Follow-up Probe:** "Why do you feel that way?"

Rubric for the Test Scoring:

Graded on a **5-point scale**.

- Maximum Raw Score = 20 points.
- Formula for final grade: (Raw Score ÷ 20) × 10 = Final Score out of 10.

Here is the redesigned, equally distributed rubric.

Descriptive Writing Coherence Rubric (Equal Distribution)

Student ID: _____ **Group:** [ ] Control [ ] Experimental

Assessment: [ ] Pre-Test [ ] Post-Test

Evaluation Criteria	**5 - Excellent**	**4 - Good**	**3 - Average**	**2 - Developing**	**1 - Poor**	**Score**
1. Spatial Organization & Flow	Perfect logical direction (e.g., outside to inside). Flawless structural flow.	Clear direction, but with one minor break or awkward jump in the flow.	Attempts a direction, but jumps around several times. Flow is inconsistent.	Very little logical organization. Ideas are mostly randomly placed.	No spatial organization at all. A completely random list of sentences.	**/ 5**
2. Cohesive Devices (Transitions)	Uses 3+ highly effective transition words (e.g., *adjacent to, meanwhile*).	Uses 2–3 transition words correctly, but they may be slightly basic or repetitive.	Uses 1–2 transition words, or uses them somewhat incorrectly.	Uses zero transition words. Relies entirely on basic conjunctions (and/but).	Sentences are completely disconnected and choppy. No linking attempts.	**/ 5**
3. Sensory Details	Blends 3+ sensory details (sight, sound, smell) perfectly	Includes 2 sensory details that support the	Includes 1 or 2 basic sensory details. Somewhat flat.	Very few sensory details. Reads more like a narrative	Completely lacks sensory details. Completely off-track from	**/ 5**

	into the description.	descriptio n well.		than a description.	descriptive writing.	
4. Task Achievement & Focus	Perfectly addresses the prompt (e.g., the Tong or Metro Rail) within the word count.	Addresses the prompt well, but might be slightly under/over the word limit.	Mostly addresses the prompt, but loses focus on the core topic briefly.	Barely addresses the prompt or wanders off-topic significantly .	Completely off-topic or fails to complete the task.	**/ 5**
					Raw Score:	**/ 20**