

A Genre-based Approach with GenAI feedback to Teaching Written Exposition in a Tertiary EFL Context in Indonesia

May 2026 – Volume 30, Number 1

<https://doi.org/10.55593/ej.30117a8>

Zainurrahman

Universitas Pendidikan Indonesia, Indonesia
zainurrahman@upi.edu

Emi Emilia

Universitas Pendidikan Indonesia, Indonesia
emi.emilia@upi.edu

Rojab Siti Rodliyah

Universitas Pendidikan Indonesia, Indonesia
rojab@upi.edu

Abstract

Despite the inclusion of the Genre-based Approach (GBA) in Indonesian educational curricula, GBA studies remain limited, particularly when a Generative Artificial Intelligence (GenAI) tool is integrated as a feedback agent. This study investigated the implementation of GBA with GenAI feedback in teaching exposition writing to undergraduate EFL students in Indonesia through an embedded single-case study. Fifteen students' diagnostic tests, three drafts with GenAI feedback, and a final test were scored. Writing scores were analyzed using repeated-measures procedures across overall quality, schematic structure, and linguistic features based on Systemic Functional Linguistics (SFL) metafunctions. Students' experiences with GenAI feedback were analyzed through thematic analysis of students' notes and interviews. The results indicate that students' writing development appears to be associated with explicit instruction and teacher scaffolding alone, as development was observed only from the diagnostic test to the first draft. Despite students' appreciation of GenAI feedback for its immediacy and comprehensive coverage, no development was observed during revision and editing with the feedback. Students also reported challenges related to prompting, feedback language, and feedback volume. These findings suggest that integrating GenAI feedback into GBA requires critical consideration. Pedagogical implications and study limitations are briefly discussed

Keywords: EFL, exposition genre, GenAI feedback, genre-based approach, Indonesia

Writing and publishing academic work in reputable journals has been among the requirements for students at the tertiary level to complete their studies in Asia, including Indonesia and Vietnam (Alfianika et al., 2019; Castulo et al., 2025). In Indonesia, this requirement is strictly stated in the Indonesian Director General of Higher Education circular letter (No. 152/E/T/2012) regarding scientific publication. To fulfill this requirement, students need to possess good control of academic genres. In addition, given that most reputable journals accept only English manuscripts (Tulasi et al., 2025), writing academic genres in English becomes increasingly demanding.

Among academic genres is exposition, a family of argumentative genres to propose a point of view on an issue or topic and mount arguments to support it (Butt et al., 2012; Derewianka & Jones, 2016). In the journal publication world, this genre is known as an opinion piece (Meyer-Junco et al., 2023). Moreover, its argumentative nature often makes this genre part of other academic genres (Martin & Rose, 2009). However, argumentative writing is also one of the most challenging academic genres (Aldabbus & Almansouri, 2022; Ghanbari & Salari, 2022; Peloghitis, 2017), and this is also the case for tertiary EFL students in Indonesia (Nurlatifah & Yusuf, 2022), although this genre has been taught since the senior high school (Rodliyah & Liani, 2022) using Systemic Functional Linguistics Genre-based Approach (throughout this paper referred to as GBA).

GBA entered Indonesia in 2004 and is still recommended by the Indonesian government in the recent Indonesian educational curriculum for English teaching (Emilia, 2005; Hidayat et al., 2023; Zein et al., 2020). GBA research has also proliferated, particularly on its effects on students' writing improvement (Arifin et al., 2024; Hutabarat & Gunawan, 2021; Suryadi & Yulandari, 2022) and teachers' implementation of the approach (Hidayat et al., 2023; Suharyadi & Basthomi, 2020; Zebua & Rozimela, 2020). In general, this line of research reported that GBA is effective when implemented properly. Since GBA adopts Vygotsky's Zone of Proximal Development (ZPD) (Derewianka & Jones, 2016; Gibbons, 2015), proper GBA implementation entails supportive interaction, scaffolding, between teacher and students or between a student and more capable peers (Moll, 1990; Vygotsky, 1978). Feedback provision is one way in which this interaction manifests in the writing classroom, "where the support of a more knowledgeable person can enable a student writer to develop both his/her text and writing abilities" (Hyland & Hyland, 2019, p. 6), which is the core of GBA. Nevertheless, it is reported that teachers implementing GBA often skipped feedback provision (Hidayat et al., 2023; Nurlaelawati & Novianti, 2017), perhaps due to an unbearable increase in workload (Paris, 2022; Selvaraj et al., 2021) and trust issues in peer feedback (Banister, 2023; Durán, 2021). Arguably, since feedback is inherent in GBA, skipping it potentially mitigates its value.

Recently, studies claimed that Generative Artificial Intelligence (GenAI) can serve as a feedback agent, providing immediate, personalized feedback on students' writing, comparable to teacher feedback (Escalante et al., 2023; Law, 2024; Polakova & Ivenz, 2024). This line of studies may suggest the possibility of integrating GenAI as a feedback agent in GBA. However, despite the growing number of GBA studies, little attention has been given to how GBA with GenAI feedback can facilitate students' writing development and how students experience the feedback. To bridge this gap, this study integrates GenAI feedback into GBA writing instruction to investigate the extent to which the instruction facilitates students' writing development, as well as how students experience the feedback. This study addresses the following questions:

RQ1: *To what extent does GBA with GenAI feedback facilitate students' exposition writing development?*

RQ2: *How do students describe their experiences with GenAI feedback during the instruction?*

Because previous research has focused on GBA and GenAI feedback separately, this study is one of the first studies to integrate and examine the value of GenAI feedback in GBA writing instruction.

Literature Review

Genre-based Approach to Teaching Writing

In this study, GBA refers to a language pedagogy grounded in Systemic Functional Linguistics (SFL) and the Zone of Proximal Development (ZPD) (Derewianka & Jones, 2016; Gibbons, 2015). SFL views language as a resource for making meaning in context and simultaneously realizing three metafunctions (ideational, interpersonal, and textual), which are respectively reflected in the register variables of field (what the language is used to talk about), tenor (how language is used to enact interaction), and mode (how language is used to organize messages) (Butt et al., 2012). This concept provides a basis for describing the linguistic features of genres according to their social purposes and the context in which they are created (Derewianka & Jones, 2016; Nagao, 2020). The term genre itself is understood as a “staged, goal-oriented social process” (Christie & Martin, 2000, p. 13), meaning that texts unfold in predictable stages to achieve particular purposes (Hyland, 2004). These stages form the schematic structure of a genre, which differs across genres according to their social purposes (e.g., exposition aims to present arguments, whereas narrative aims to tell stories) (Derewianka, 2003). GBA, therefore, focuses on teaching both the schematic structure and linguistic features of texts based on social purposes.

However, GBA scholars argue that explicit genre instruction and teacher scaffolding are required to ensure equal learning success. They argue that students from diverse backgrounds enter the classroom with varying levels of academic language due to limited exposure outside the school (Schleppegrell, 2004). Moreover, students' knowledge of the genres they rarely encounter is vague (Hyland, 2004) and will not develop naturally, but rather through explicit genre instruction and scaffolding, where ZPD's influence is strong (Derewianka & Jones, 2016; Gibbons, 2015).

The concepts of genre and register (SFL) and explicit instruction (scaffolding, ZPD) lead to the formulation of the GBA *Teaching-Learning Cycle* (TLC). While there are several TLC models, Rothery's four-stage TLC model is implemented in Indonesia (Emilia, 2005; Suharyadi & Basthomi, 2020). The description of each stage is summarized below.

Stage 1 (Negotiating or Building Knowledge of the Field): The teacher builds students' content knowledge required to write the text in focus, including the field-related vocabulary (Derewianka & Jones, 2016; Gibbons, 2015). Activities include the teacher's presentation, classroom discussion, and research when possible (Emilia, 2011; Hyland, 2004).

Stage 2 (Deconstruction or Modeling): The teacher familiarizes students with the schematic structure and linguistic features of the genre by presenting the model texts (Emilia, 2005), where SFL provides *metalanguage* (language to talk about language), helping the teacher and students to describe the typical register features leveraged in the target genre (Derewianka & Jones, 2016; Schleppegrell, 2013).

Stage 3 (Joint Construction): the teacher invites the students to collaboratively write a text in the target genre, where ZPD provides the basis for scaffolding (Derewianka & Jones, 2016;

Gibbons, 2015). Students can either work in groups or with the teacher, where they contribute ideas/sentences, while the teacher (or more capable peers) scaffolds by providing alternatives and suggestions to refine the unfolding text.

Stage 4 (Independent Construction): Students write their own texts in the target genre, and the teacher's support is gradually reduced (Derewianka & Jones, 2016), except for feedback provision (Emilia, 2005; Hyland, 2004).

Exposition Genre

It was mentioned earlier that an exposition is a family of argumentative genres that proposes a point of view on an issue or topic and mounts arguments to support it (Butt et al., 2012; Derewianka & Jones, 2016; Gibbons, 2015). Exposition is divided into two subtypes: analytical and hortatory expositions, where the former aims to argue on whether something is or is not the case, while the latter aims to argue on whether something should or should not be done (Coffin, 2004; Emilia, 2011). In addition, although they differ in aim, both share a similar schematic structure and linguistic features.

Schematic Structure and Linguistic Features

In GBA, an exposition has three consecutive stages: a thesis, a series of arguments, and a reiteration of the thesis (Coffin, 2004; Gibbons, 2015). The thesis stage provides a background on the issue (which may include different views), states a thesis, and possibly previews the main arguments. In the second stage, a series of arguments, and possibly a counterargument, is provided to support the thesis statement. If the main arguments are previewed in the thesis stage, the number and order of the arguments should match the previewed ones to maintain coherence. Eventually, the reiteration stage summarizes the main arguments and reiterates the thesis more confidently (Coffin, 2004; Emilia, 2011).

In GBA, the linguistic features of a genre are described in terms of the SFL register.

Field-related features include the use of technical terms, generalized participants (non-personal entities involved as subjects), sensing (e.g., think, believe, feel, see) and relating (e.g., is, are, have, has) verbs, and nominalization (e.g., the government's decision, feedback provision, his retirement) (Derewianka & Jones, 2016; Schleppegrell, 2001).

Tenor-related features include the use of various speech functions (offer, statement, command, rhetorical question), modality (modal verbs: e.g., can, may, must, should, will; and modal adjuncts: e.g., effectively, extremely), evaluative language (appreciation and judgment), and attribution of external sources through citation (e.g., according to, scholars argue), endorsement (e.g., this study proves that...), and concession (e.g., however, this study has flaws...) (Butt et al., 2012; Derewianka & Jones, 2016; Martin & White, 2005).

Mode-related features include the use of cohesive devices (e.g., and, but, also, as well as, therefore, however, in conclusion) and a variety of thematic progression patterns (Butt et al., 2012; Derewianka & Jones, 2016; Emilia, 2011). The thematic progression can be identified through the sequence of the arguments against the preview.

Pessoa et al. (2018) and Nagao (2020) have compiled and used these features in their exposition rubrics. We adapted these rubrics by combining all categories, adding a category and criterion for assessing the schematic structure, while adhering to the rating scale proposed by Pessoa et al. (2018). The adapted rubric is shown in **Appendix A1**.

GenAI Feedback

Feedback refers to information from various sources (e.g., teachers, peers, texts) about a learner's performance or understanding (Hattie & Timperley, 2007). In writing classrooms, it highlights the gap between students' current understanding and expected standards by providing information about "(a) the learner's current knowledge set and (b) the desired knowledge set" (Lipnevich & Smith, 2009, p. 319). Feedback can reduce this gap, functioning not only as immediate problem-solving support through direct and explicit feedback, but also as a learning source through indirect and implicit feedback (Bitchener & Ferris, 2012).

However, feedback provision is often challenging in practice. Teachers frequently experience a substantial increase in workload (Paris, 2022; Selvaraj et al., 2021), while peer feedback is often constrained by trust issues, as students may lack confidence in their own feedback or doubt the value of feedback from peers (Banister, 2023; Durán, 2021). Consequently, feedback effectiveness may be reduced, or feedback may not be provided at all (Hidayat et al., 2023; Nurlaelawati & Novianti, 2017).

Recently, GenAI tools have been increasingly used for writing-related purposes, including feedback traditionally provided by humans (Urmeneta & Romero, 2024). Scholars claim that GenAI tools can offer immediate and personalized feedback (Law, 2024), reduce teachers' workload (Alsaedi, 2024; Lee & Moore, 2024), and, in some cases, perform comparably to teacher feedback (Alnemrat et al., 2025; Escalante et al., 2023). These tools can do so because, beyond content generation (Kalota, 2024), they can ground responses in contextual input such as essays and rubrics through *Retrieval Augmented Generation* (RAG), which aims "to generate a more informed, context-rich response using a generative model" (Gheorghiu, 2024, p. 11). When sufficient context is provided, GenAI feedback can therefore be relevant and personalized.

However, effective use of GenAI for writing feedback requires specific user knowledge (Anders, 2024; Tseng & Warschauer, 2023). Users need to understand the tool's capabilities, provide sufficient content-as-context (e.g., student texts and rubrics), configure key parameters (e.g., system instruction, temperature, Top-P) to reduce hallucination (Google, 2025), and formulate effective prompts, as the quality of responses depends heavily on them. Gemini allows free access to these parameters and is therefore used in this study (see **Appendix A3**).

Despite these affordances, several limitations have been identified. GenAI tools may lack contextual understanding (Kulkarni & Shivananda, 2019; West et al., 2024). Without adequate context, feedback may be distorted by training data rather than grounded in the intended task (Jain et al., 2024), resulting in irrelevant or overly complex feedback that does not align with students' proficiency levels (Bok & Cho, 2023; Campos, 2025). Additionally, the use of GenAI tools raises concerns about data privacy, as personal information may be exposed to tool vendors (Dignum, 2019; Hagerty & Rubinov, 2019). Finally, due to *model drift*, where performance degrades because of changes in the operating environment, GenAI feedback may be inconsistent when evaluating the same text at different times (Bui & Barrot, 2025).

Previous GBA Studies

Although GBA has been included in Indonesian curricula and GBA studies have proliferated, research on its implementation with GenAI feedback remains limited. While the TLC emphasizes feedback during the construction stages, how GenAI feedback functions as scaffolding within explicit instruction (and how students experience it) has received little attention.

Most studies have examined whether GBA (without GenAI feedback) facilitates measurable improvement in students' writing performance. For example, Arifin et al. (2024) reported post-

treatment gains in procedure writing, and Herman et al. (2020) documented improvements in descriptive writing ability and motivation, although grammatical inaccuracies persisted. Similarly, Aunurrahman (2017) demonstrated metafunctional development during joint construction but reported unresolved grammatical errors, attributed to ineffective self-feedback.

Another strand of research has examined GBA's effects on students' control of specific metafunctions and genre features. Studies conducted in Japan reported development in ideational resources (Nagao, 2022) and limited metafunctional awareness among lower-achieving students, particularly in textual meaning in exposition writing (Nagao, 2020). Other studies reported improvement in discourse markers and schematic structure in exposition and discussion genres, with gains often stronger at the phase level than at the overall genre stage level (Ariyanfar & Mitchell, 2020; Morgan et al., 2022). In China, Feng Chen (2021) similarly found improved performance in exposition writing, associated with positive student perceptions of GBA, while Nagao (2019) reported interpersonal metafunctions development, particularly in discussion genres. However, feedback processes underlying these developments were rarely examined explicitly.

Relatedly, a small number of studies have examined the role of SFL register knowledge in students' text revision. Zhang (2021) reported that increased SFL knowledge enabled students to revise their expositions beyond form-level concerns (this is the first study of feedback in SFL to our knowledge). Studies in Thailand similarly reported improved control of schematic structure and linguistic features across procedure and recount genres (Mingsakoon & Srinon, 2018; Srirakarn, 2020), yet feedback was not theorized as a central pedagogical mechanism.

While student outcomes have been widely reported, fewer studies have examined teachers' beliefs and classroom implementation of GBA, particularly with respect to feedback provision across the TLC stages. In Indonesia, Nurlaelawati and Novianti (2017) found that teachers often skipped certain stages (e.g., BKOF) and rarely provided feedback during construction stages, a pattern also observed by Suharyadi and Basthomi (2020) and Hidayat et al. (2023), who reported that feedback was frequently omitted during independent construction due to limited training and implementation constraints.

Despite the central role of feedback in the TLC, systematic research on feedback provision and its effects on students' writing development within GBA remains limited. Zhang (2021, p. 469) noted that "almost no research has been conducted on how teachers, in line with the SFL-based perspective on writing instruction, offer written feedback to their students," observing that limited metalanguage knowledge constrained feedback uptake, although teacher feedback supported meaning- and structure-level revision. Existing studies largely focus on teacher feedback during joint construction (Kuiper et al., 2017; Suharyadi & Basthomi, 2020; Syarifah & Gunawan, 2016). Studies on peer and self-feedback in GBA reported increased awareness of schematic and interpersonal resources, but limited ideational and textual development, alongside trust issues and vague revisions among lower-achieving students (Aunurrahman et al., 2017; Cahyono, 2018; Durán, 2021; Nagao, 2018, 2019).

In conclusion, these studies highlight persistent challenges in feedback provision within GBA. Although research on GenAI feedback suggests that such challenges may be alleviated through emerging technologies, the pedagogical use of GenAI feedback within GBA has not been well documented. At the time of writing, searches using the keywords "GenAI SFL GBA" and "ChatGPT SFL GBA" yielded no studies directly addressing this intersection, indicating an opportunity to explore the value of GenAI as a feedback agent within this pedagogy.

Methodology

Research Design

This study employed an embedded single case study design (Yin, 2007), with the *case* defined as a GBA with GenAI (Gemini) feedback, and two units of analysis: *students' exposition writing development* and *students' experiences*. This definition was guided by the explanation that a case in case studies is not only restricted to individuals or a group of people, but it can also be a program or intervention (Creswell, 2009; Dörnyei, 2007; Merriam, 1991). Given the similarity between a single case study and an experiment (Nunan, 1992; Yin, 2007), this study approached the first unit of analysis quantitatively to trace students' writing developmental trajectory. The second unit of analysis was approached qualitatively through a thematic analysis. In addition, this study also resembled a teaching program evaluation. A GBA with GenAI feedback was the teaching program to be evaluated to decide whether the program needs modifications or alterations in particular aspects so that the objectives of the program can be achieved more effectively in the future (Emilia, 2005; Nunan, 1992). This design was appropriate for evaluation (Stake, 1995), where quantitative and qualitative methods can be applied (Mills et al., 2010) to employ multiple data sources to gain more rounded and complete accounts to test the values of the teaching program (Emilia, 2005).

Context and Participants

The teaching program was implemented in an undergraduate Writing for Academic Purposes course, a regular course held in the English Education Program at a state University in Indonesia, focusing on teaching exposition. In addition, GBA has been implemented in this context due to its inclusion in the Indonesian educational curriculum (Emilia, 2011; Faradina & Gandana, 2024; Triastuti et al., 2022).

There were 41 students enrolled in this course. Due to the high volume of concurrent coursework and short deadlines, only 17 students agreed to participate by signing and returning the consent form; all were females, 19-24 years of age ($M=20, SD=0.92$). Of the 17 students, only 15 completed all writing tasks, and 11 were involved in the interviews. A pre-instructional open-ended questionnaire revealed that half of them were still unfamiliar with the GBA TLC and the exposition genre, but all of them were familiar with various GenAI tools, as they had used them for various writing purposes. Regarding their initial EFL proficiency, however, the data were insufficient because most of them had not taken any formal tests. Additionally, since this was a regular writing classroom, students were not grouped into experimental/control groups.

Procedures (Classroom Interventions)

This 13-week study was divided into a preliminary, implementation, and evaluation phase, summarized below (see **Appendix A2**):

In the Preliminary Phase (week 1), background information regarding students' familiarity with GBA and their initial exposition writing competence was collected through an open-ended online questionnaire and a diagnostic writing test. The test was designed to assess baseline writing quality (Nunan, 1992), using the target genre to ensure construct validity (Weigle, 2002). Students wrote on their own topics to minimize content constraints and enable the task to elicit authentic writing quality. The test took 60 minutes, following the duration allocated for the course. Questionnaire data informed students' needs, consistent with the need-based nature of GBA (Hyland, 2004), while diagnostic test data served as a baseline measure, contributing to the study's internal validity (Hatch & Lazaraton, 1994; Nunan, 1992).

The Implementation Phase followed Rothery's four-stage TLC model outlined earlier, overviewed below:

- Weeks 2-3 (BKOF): The teacher presented and invited students to discuss the topic to build students' content knowledge, using three of the six opinion pieces on the use of

AI in academic writing. These opinion pieces were instances of an exposition genre as they shared the social purpose, schematic structure, and linguistic features.

- Week 4 (Modeling): The teacher presented the social purpose, schematic structure, and linguistic features of the model texts by inviting students to those texts.
- Weeks 5-6 (Joint Construction): The teacher and students jointly constructed an exposition entitled “Is it ethical to use AI to write the entire paper?” Demonstrations were provided on planning, drafting, formulating prompts for feedback, and revising using the web app (students had created their accounts in a web app developed for this study).
- Weeks 7-12 (Independent Construction): Students wrote their own exposition on a self-selected topic and recorded their experiences with GenAI feedback in the web app. This phase was conducted fully online due to the Ramadan break. The datasets from this phase included students’ notes, successive drafts (first, revised, and edited), and their feedback logs.

In the Evaluation Phase (week 13), a 60-minute in-situ final writing test was administered, where students wrote another exposition on a different self-selected topic. After the test, 11 students joined a two-day semi-structured interview (in Bahasa Indonesia). The interview questions covered the perceived contribution of the teaching program, adapted from Emilia (2005), perceived development in writing, and challenges in integrating GenAI feedback, adapted from Kim et al. (2024) and Wang (2025). Because this study focused on experiences with GenAI feedback, the interview data and students’ notes were expected to reveal their overall experiences with GenAI feedback during the teaching program.

Data Analysis

To address the research questions, this study collected and analyzed (a) students’ texts produced before (diagnostic test), during (three successive drafts for each student), and after (final test) the program (n=75); (b) students’ notes (n=175); and (c) students’ interviews (n=11).

Analysis of Students’ Exposition Writing Development (RQ1)

To address RQ1, 75 exposition texts produced by 15 students across five writing stages were scored using an analytic rating scale designed specifically to assess an exposition text (**Appendix A1**). The scoring process was aided by a Gemini-based automated essay scoring (AES) tool integrated into the project’s web application. In line with a case-study epistemology, AES was used as a supportive analytic tool rather than as an objective scorer of writing quality. Previous research suggests that AI-based AES can provide efficient and systematic support for longitudinal writing assessment when interpreted cautiously (Mizumoto & Eguchi, 2023; Shermis & Wilson, 2024).

To ensure scoring stability, consistent GenAI parameter settings were maintained throughout the scoring process, with variation limited to task-specific system instructions and scoring prompts. An initial AES score of 20 texts was collaboratively reviewed by the research team to establish a shared interpretation of the rating criteria and a stable application of the scoring model. Following this calibration step, the same scoring configuration was applied to all texts. Scores for each component (overall performance, schematic structure, ideational, interpersonal, and textual meaning) were aggregated and converted into percentages for analysis.

Following Yin’s (2007) recommendation for time-series analysis in case-study research, repeated-measures analyses were conducted using SPSS (version 22) to trace developmental trajectories across writing stages, rather than to support statistical generalization. In this regard,

statistical analyses “can be used in case study research” to verify the validity of results, even if sampling requirements are not met (Mills et al., 2010, p. 893).

Before analysis, data normality was examined using the Shapiro–Wilk test to guide the selection of parametric or non-parametric procedures (Hatch & Lazaraton, 1994; Scheff, 2016). Repeated-measures ANOVA or the Friedman test was applied accordingly, with Greenhouse–Geisser corrections used when sphericity assumptions were violated. Bonferroni-adjusted post-hoc analyses were conducted where relevant. Regarding the effect size, the main within-subjects Friedman test and ANOVA used Kendall’s W^2 and partial eta squared (η^2), respectively (Tomczak & Tomczak, 2014). Meanwhile, for pairwise comparisons, correlation (r) and mean difference (d_z) were used, respectively (Lakens, 2013; Plonsky & Oswald, 2014). Effect size indices were reported to support the interpretation of within-case change rather than population-level inference. The benchmark of the effect size estimates is shown below.

Table 1. Effect size estimate benchmarks

	Small	Medium	Large	Source
Kendall’s W^2	< .3	≤ .5	> .5	Adapted from Tomczak & Tomczak (2014)
η^2	< 0.06	< 1.13	≥ 1.14	Adapted from Wei et al. (2019)
R	.25	.40	.60	Based on Plonsky (2015) and Plonsky and Oswald (2014)
d_z	.60	1.00	1.40	

Note. W^2 and η^2 were used in the main analyses, while r and d_z were reported in the post-hoc analysis/pairwise comparisons.

Analysis of Students’ Experiences with GenAI Feedback (RQ2)

To address RQ2, a typical seven-step data-driven thematic analysis was conducted on students’ journals and interviews (Braun & Clarke, 2013). The thematic analysis was chosen for its flexibility in terms of research questions, sample size, data collection method, and approaches to meaning generation (Clarke & Braun, 2017), as well as its adaptability to any theoretical frameworks and research paradigms (Clarke & Braun, 2017; Maguire & Delahunt, 2017). Prior to the analysis, all datasets were translated into English, and students’ identities were removed. The data were then familiarized through repeated reading, in which initial codes were generated inductively to capture meaningful features of students’ learning and writing challenges (e.g., perceived GenAI helpfulness, prompting difficulties). The codes were organized into themes, which were reviewed and named to represent data patterns. Representative extracts were chosen to illustrate each theme in the results.

Trustworthiness

In this study, one of the researchers responsible for implementing the TLC was originally an outsider. However, due to the ongoing interaction with participants during the implementation, the researcher’s role could gradually shift toward that of an insider (Creswell, 2009). This shift could introduce bias in both data collection and analysis, which may affect the trustworthiness of the study. To mitigate the risk, a number of steps were taken. First, the teaching program was jointly planned by all investigators, following the established TLC model to maintain consistency throughout the implementation. Second, all student drafts, GenAI feedback, and student notes were automatically collected through the web app to avoid selective data collection. Third, the open-ended questionnaire and interview protocol were validated by non-investigator peers familiar with SFL GBA and GenAI, even though these instruments were adapted from existing studies. Regarding the scoring of students’ drafts, an SFL-informed rubric for exposition was used to maintain the construct validity (Weigle, 2002). While an AES system is often said to be inconsistent (Bui & Barrot, 2025), it could support objectivity and

reduce human bias (Shermis & Wilson, 2024), and the suitability of the initial scoring results was checked by all investigators. Finally, investigator triangulation (Stahl & King, 2020) was conducted in the coding of the interview data and students' notes, where two investigators independently coded the dataset, and later the themes were reviewed by all investigators.

Results

Students' Exposition Writing Developmental Trajectories

The following results address RQ1 by describing students' writing developmental trajectories across five stages. Table 2 presents descriptive statistics for overall writing scores and sub-scores (schematic structure and linguistic features), while Figure 1 illustrates mean score trajectories across stages.

Table 2. Descriptive statistics

	Diagnostic Writing Test	First Draft	Revised Draft	Edited Draft	Final Writing Test
<i>Components</i>	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Overall	50.26 (9.21)	75.88 (6.87)	74.91 (5.18)	76.84 (4.92)	68.42 (6.01)
Schem. Struc	76.67 (6.45)	85.00 (12.68)	85.00 (12.68)	83.33 (12.20)	76.67 (6.45)
Ideational	49.33 (9.42)	70.33 (9.54)	70.33 (8.34)	72.00 (9.02)	66.33 (3.99)
Interpersonal	42.22 (12.59)	77.50 (9.16)	76.11 (7.63)	79.17 (8.33)	66.94 (10.02)
Textual	54.05 (10.00)	77.14 (7.85)	75.71 (6.64)	77.38 (5.34)	70.00 (7.85)

Note. SD = standard deviation; Schem. Struc = schematic structure

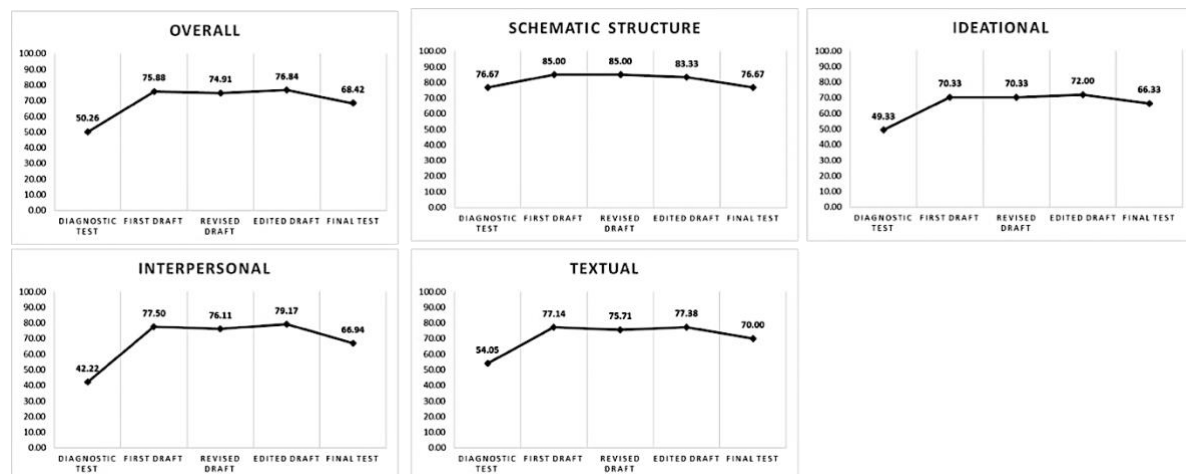


Figure 1. Means across writing stages

A one-way repeated measures ANOVA was performed to examine the differences in the overall writing scores, interpersonal meaning, and textual meaning, while a Friedman test was performed to examine the differences in the schematic structure and ideational meaning (Table 3). The tests revealed that significant differences were found in the overall scores, and ideational, interpersonal, and textual meaning scores, with effect sizes ranging from moderate to very large, while no difference was found in the schematic structure score.

Table 3. Differences in students' scores across writing stages

Components	Test	Test statistics	Df	P	Effect size
Overall	ANOVA	$F=53.42$	(1.46, 20.42)	<.001	$\eta p^2=.792$
Schematic structure	Friedman	$X^2=10.72$	4	.030	$W^2=.032$
Ideational meaning	Friedman	$X^2=31.34$	4	<.001	$W^2=.273$
Interpersonal meaning	ANOVA	$F=45.93$	(2.14, 29.91)	<.001	$\eta p^2=.77$
Textual meaning	ANOVA	$F=33.68$	(1.69, 23.70)	<.001	$\eta p^2=.706$

Note. Greenhouse-Geisser corrections were applied to repeated-measures ANOVA. Effect sizes are reported as partial eta squared (η^2) for ANOVA and Kendall's W for Friedman tests

Bonferroni-adjusted pairwise comparisons were conducted to find differences between stages, excluding schematic structure. Significant differences were noted only between students' diagnostic tests and first drafts, with no differences among the first, revised, and edited drafts. Additionally, overall writing scores, interpersonal meaning, and textual meaning declined from the edited draft to the final test. Results are summarized in Table 4 (see **Appendices A5-A9**).

Table 4. Summarized results of pairwise comparisons

Components	Comparison	Mean difference	p	Effect size
Overall	Diagnostic test-First draft	-25.612	<.001	$d_z=-2.094$
Overall	Edited draft-Final test	8.421	<.01	$d_z=1.180$
Ideational meaning	Diagnostic test-First draft	-3.246	.001	$r=.838$
Interpersonal meaning	Diagnostic test-First draft	-35.277	<.001	$d_z=-2.312$
Interpersonal meaning	Edited draft-Final test	12.224	<.05	$d_z=0.952$
Textual meaning	Diagnostic test-First draft	-23.096	<.001	$d_z=-1.695$
Textual meaning	Edited draft-Final test	7.380	<.05	$d_z=0.975$

Note. Effect sizes: correlation r for the Wilcoxon; mean difference d_z for parametric paired-samples comparisons

These results, in general, are consistent with the mean score trajectories (Fig. 1), indicating that the development of students' exposition writing was only observed from the diagnostic test (before the intervention) and the first draft (after the three early stages of the teaching-learning cycle). As discussed later, this development could be attributed to activities in the BKOF, Modeling, and Joint Construction stages, given that students wrote the first draft in the Independent Construction stage. However, it should be noted that this development may also partially reflect a task repetition effect, as students wrote the same genre in both stages (Kim & Li, 2025; Tabari et al., 2024), though the topics differed across the two stages. The insignificant changes in comparisons among the first, revised, and edited drafts indicated that GenAI feedback made no difference to students' writing. Meanwhile, the decline from the edited draft to the final test in the three components above indicated that students could not maintain the initial development.

Students' Experiences with GenAI Feedback

The following results address RQ2, categorized into positive and negative experiences, with subthemes for each. Figure 2 illustrates the hierarchy of these themes. A few selected excerpts are included, marked with sources ("Int." for Interview Transcript; "Note" for Student Note, with note number; "P" for Participant), while more supporting excerpts are provided in **Appendix A10**. In addition, form-level corrections were applied for clarity.

Positive Experiences

Despite the insignificant changes in students' writing development associated with GenAI feedback, students described GenAI feedback as supporting their *correction and revision*, particularly for identifying areas requiring correction that they found difficult to address independently (e.g., "AI really helped me to correct my writing when I asked for feedback, not to mention when I am not familiar with the task", Int. P03). They also said that GenAI feedback increased their *awareness of their writing errors*. For instance, "Sometimes, we were also unaware of the parts that needed revision, but the AI showed us the parts that need to be improved and enhanced" (Int. P06) and "I also found it interesting that Gemini could notice

misused or misspelled words, which I had not realized before” (Note #156, P36). *The immediacy of feedback* was also claimed to be useful when teacher or peer feedback was unavailable. For instance, “AI was accessible at any time, especially when there was no help from friends or lecturers who could give feedback directly” (Int. P27).

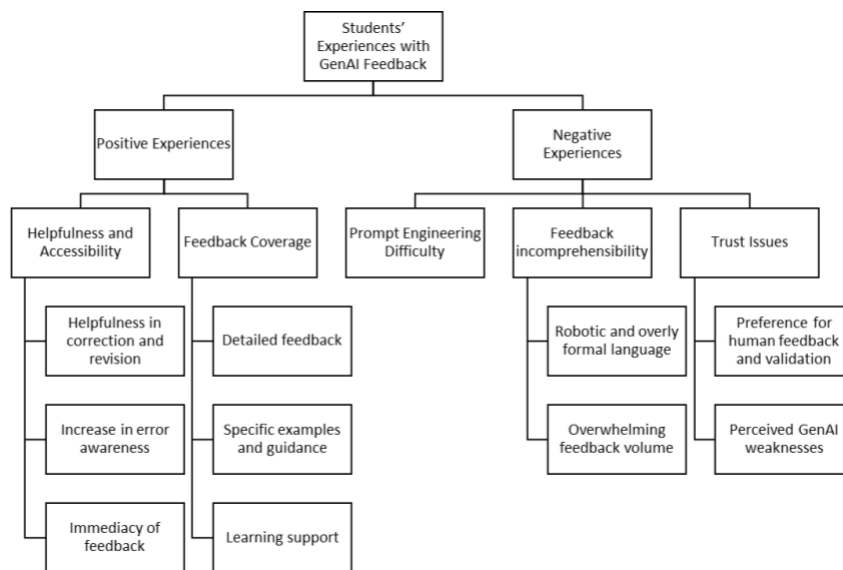


Figure 2. Themes and subthemes of students' experiences with GenAI feedback

In terms of feedback coverage, students often said that the feedback was *detailed* and addressed multiple aspects of their writing. For instance, a student said, “Feedback from AI happened to be detailed enough, which helped me much in revising” (Int. P05), while another student noted that “The feedback given was very detailed and clear if read carefully” (Note #124, P33). A few students mentioned that GenAI could provide *specific examples and guidance* for revision (e.g., “Whatever I asked for, AI gave the examples”, Int. P12), while others also mentioned that it provided *support for learning new knowledge* (e.g., “I think AI was very helpful for me to learn new knowledge that I was still confused about”, Int. P03).

Negative Experiences

Students also described their negative experiences. They faced *difficulties in prompt engineering* or formulating effective prompts (e.g., “Sometimes I had problems because I was sometimes confused about how to write the prompt so that AI could give the appropriate answer”, Int. P31). GenAI feedback was also *robotic and overly formal*, hindering comprehension (e.g., “The language of AI was quite formal, so it was difficult to understand. So, we needed to ask for another prompt so that the AI could explain in language that was easy to understand”, Int. P37) and the *voluminous* feedback often overwhelmed them (e.g., “Another challenge was managing the feedback process. At times, it felt overwhelming to revise based on multiple suggestions”, Note #171, P22).

While errors in GenAI feedback were not the focus of this study, some students expressed “trust issues.” Some students preferred *teacher/peer feedback and validation*, compared to GenAI feedback. For instance, “It was useful enough, but I think direct feedback from the lecturer or the peers is still better” (Int. P32), and “To be honest, I would prefer personalized feedback from an expert or lecturer” (Note #11, P22). Meanwhile, some students *perceived limitations of GenAI feedback* in higher-level argument development and evaluative judgment. For example, “AI is quite specific and very helpful in terms of grammar and writing, but in terms of arguments, I think it is lacking. In terms of providing feedback for arguments, it is

lacking” (Int. P32) and “I still have some trust issues with AI when it comes to assessment and correction” (Note #11, P22).

Discussions

This study reports on the results of implementing a genre-based approach with GenAI feedback to teaching exposition writing in a tertiary EFL context in Indonesia. To this end, the study addressed two research questions, aiming at capturing students’ exposition writing developmental trajectory and students’ experiences with GenAI feedback.

Students’ Exposition Writing Development

The statistical analyses indicated that students’ writing development could only be observed from the diagnostic test to the first draft. This development could be associated with the three stages of the TLC (BKOF, Modeling, and Joint Construction) that mediate the two writing stages. Students’ development in ideational meaning or field-related features could be due to their learning of the content knowledge, including field-related vocabulary, to build clear and grounded arguments during BKOF (Derewianka & Jones, 2016; Gibbons, 2015). The Modeling stage could have contributed to this development through familiarization with the genre structure and linguistic features (Emilia, 2005, 2011), helping them develop their control of the genre. Collaborative activities during the Joint Construction stage could also have equipped students with strategies to use their genre knowledge learned in the two previous stages, as well as writing process strategies (Emilia, 2005; Gibbons, 2015).

The development in this case could be identified in the linguistic features, especially the interpersonal meaning, followed by the textual meaning and ideational meaning. This finding corroborates the findings of the previous studies that the TLC could improve students’ writing across metafunctions (Arifin et al., 2024; Ariyanfar & Mitchell, 2020; Aunurrahman et al., 2017; Kuiper et al., 2017; Nagao, 2019). In terms of the interpersonal metafunction, the finding of this study corroborates Nagao’s (2019, 2020) finding in that it is the most developed category among all metafunctions. Her study included argumentative genres (discussion and exposition, respectively). Moreover, as with this study, Nagao (2020) also found that “the changes in ideational meaning, which is related to understanding of the background information of the essay topic, demonstrated only slight improvement” (p.155). The sharp improvement in interpersonal meaning might be due to the general purpose of the genre, which is to engage and persuade the reader, leading students to focus more on tenor-related linguistic features. This is in line with Derewianka and Jones (2016) that “persuasive language, because it involves negotiation between the speaker/writer and listener/reader, is concerned with interpersonal meanings” (p.241). Further investigation is necessary to examine whether students’ use of linguistic features across metafunctions is related to the genre in focus, given that these studies also found that ideational meaning experienced only slight improvement.

In this study, we observed that students’ interpersonal meaning development was mainly contributed by the use of modality, attribution, and evaluative language, especially positive appreciation. Students might have learned during the modeling and joint construction stages the importance of expressing modalities to express the degree of investment they put in their position through the use of various modal verbs (Butt et al., 2012; Derewianka & Jones, 2016). Moreover, compared to their diagnostic test, in which their arguments were mostly based on their opinions, their first drafts were more engaging and persuasive, given that they started to incorporate “other voices” by attributing external sources (Derewianka & Jones, 2016; Gibbons, 2015; Martin & White, 2005). Attribution (e.g., citing experts) is important in argumentative writing since it provides support (e.g., evidence) for the arguments, which in turn enhances their persuasiveness (Derewianka & Jones, 2016; Emilia, 2011). Nevertheless, students rarely showed how those attributed sources directly support their arguments (e.g., by employing

verbal processes, to use SFL terms, such as “argue” and “report”). This highlights the importance of explicitly teaching students the importance of *endorsement*, the way language is used to show how other texts relate to the texts they are writing.

In terms of ideational meaning, although students have developed their use of sensing and relating verbs, they rarely employed nominalization and technical terms, which are highly valued in academic and argumentative genres (Schleppegrell, 2001, 2004). Perhaps it was the reason why their ideational meaning was the least developed metafunction.

In terms of textual meaning, a sharp improvement could be seen from the diagnostic to the first draft, indicating that students have applied their knowledge of mode-related linguistic features learned, especially in the modeling stage. For instance, after the modeling stage, a student noted:

“...we examined how writers connect their ideas logically using transition words such as therefore, thus, and as a result. This aspect was especially useful because it demonstrated how a well-structured text can enhance clarity and coherence” (Part of Note #42, P36).

In the diagnostic test, although most of them used conjunctions correctly, they often combined multiple claims into a paragraph, making each claim undeveloped. Meanwhile, it is held that each claim should be developed in a separate paragraph to ensure that every claim can be sufficiently developed and supported (Derewianka & Jones, 2016; Emilia, 2005). In the first draft, they started to separate claims into paragraphs and supported each accordingly. Transitions were also problematic in the diagnostic test. For instance, they rarely used transition signals (e.g., “in conclusion”) to begin the reiteration stage. In the first draft, they started to use various transitions (e.g., “however,” “in conclusion”). In Figure 3 below, a student’s draft illustrates how students started using transition signals (and modality) in the first draft, compared to their diagnostic test. This finding is similar to one of the studies reported in Ariyanfar and Mitchell (2020), which reported that GBA “leads to across-the-board improvement in the use of metadiscourse markers” (p.252). It also coincides with the study by Kuiper et al (2017), reporting that “the comparison of students’ pre- and posttest further showed improvements in terms of field and mode” (p.50). This finding is inspiring, given that Nagao’s (2020) study found that there was no movement in students’ textual meaning, except for high-achieving students.

Storytelling as a learning method is effective in increasing students' interest in English, which in turn increases their vocabulary mastery. The many methods in storytelling, such as digital storytelling, puppet, picture make it easy for teachers to adapt it to the area where they teach. P02's last paragraph in the diagnostic test

In conclusion, this paper has shown some of the benefits of using AI and strategies for using it. By providing access to large amounts of information, AI increases curiosity and improves understanding. **Moreover**, AI **could** make writing more structured and coherent. Although its effectiveness depends on the prompt given, it **cannot be** denied that AI is a very promising software if used in accordance with applicable ethics and a sense of responsibility. P02's last paragraph in the first draft

Figure 3. P02's last paragraph in the diagnostic test and first draft (different topics)

Regarding the schematic structure, however, this study found no significant differences before and after the BKOF, Modeling, and Joint Construction stages. Previous studies reported that GBA could facilitate students’ control of the genre schematic structure (Mingsakoon & Srinon, 2018; Sritrakarn, 2020; Syarifah & Gunawan, 2016), particularly due to the analytical activities during the modeling stage. This finding partially corroborates Morgan et al.’s (2022) finding that students’ schematic structure remained the same across writing points; they found that, while the *stages* of the genre remained the same, the *phases* in each stage developed, but the present study did not focus on the *phase*. Looking at the results presented in Table 2 and Figure

1, it can be seen that there was a slight improvement from the diagnostic to the first draft. However, the score returned in the subsequent drafts. This fluctuation could be caused by two factors: (1) the rating scale used in this study involved counterargument as a criterion (**Appendix A1**, point 10), while (2) students learned that counterargument in an exposition is an optional stage (Coffin, 2004). Some students might have experimented with the structure by adding this optional stage, while others did not, due to their existing knowledge of the obligatory stages of an exposition. Indeed, students were taught in the BKOF and modeling stages that counterargument (and refutation) is an important, although optional, element in argumentative writing, given that older students need to demonstrate their critical stance (Derewianka & Jones, 2016). Consequently, several students added a counterargument in their texts, but fell short due to extremely limited refutation. In addition, as can be seen, the schematic structure has been scored highly among all categories. It can be assumed that when the counterargument is removed from the rating scale (adhering to the GBA original description of this genre), the results can capture students' real control of the genre schematic structure.

The development of students' writing could also be associated with task repetition. Although students wrote on different topics, the genre remained the same, as required by GBA (Derewianka & Jones, 2016; Gibbons, 2015). Coupled with the explicit genre instruction and teacher scaffolding, repeating the genre could relieve the burden on students' cognitive resources, compared to writing a different genre (Tabari et al., 2024). This is also in line with Kim and Li's study (2025) that repeating tasks with different content offers multiple learning opportunities for mastering linguistic structures, particularly with corrective feedback from a teacher (e.g., during the joint construction stage).

During the independent construction stage, students received feedback from GenAI. However, we failed to identify any significant changes among their first, revised, and edited drafts. Thus, this finding failed to resonate with the findings of other GenAI feedback studies that reported students' writing development (Escalante et al., 2023; Law, 2024; Lee & Moore, 2024; Lo et al., 2025; Polakova & Ivenz, 2024).

Finally, the decline in the final test indicated that students still needed support, although the means across writing stages (Fig. 1) showed that their final test was still relatively higher than the diagnostic test scores. This finding corroborates Nagao's (2022) finding that while students' skills improved immediately after the intervention, their scores often decreased in subsequent writing tests. This also resonates with the view that mastering a genre is challenging, especially for the genres students rarely encounter (Hyland, 2004). Therefore, it is highly recommended that the teaching-learning cycle be repeated throughout the semester or year to ensure students' readiness to write independently (Gibbons, 2015).

Students' Experiences with GenAI Feedback

The thematic analysis showed that students described both positive and negative experiences with the GenAI feedback. Understanding these experiences is vital in improving instruction in the future, since they contribute insights into the appropriateness of the activities for their learning needs (Kiely & Rea-Dickins, 2005). Importantly, students' positive experiences reported in this study do not imply GenAI feedback effectiveness, but rather illustrate how they perceived the feedback within a GBA pedagogy that was primarily shaped by explicit instruction and scaffolding.

Students' appreciation of the feedback immediacy that covered multiple aspects in their writing resonates with the findings of other studies, which reported students' satisfaction with using GenAI "for their paragraph revisions" (Bok & Cho, 2023, p. 15), covering "various aspects of writing, including surface-level issues such as grammar, vocabulary, or spelling and global-

level issues like content and organization” (Kurt & Kurt, 2024, p. 7). They also appreciated GenAI’s detailed feedback, resonating with the students in other studies who expressed the same appreciation (Polakova & Ivenz, 2024; Rafida et al., 2024). Their mentioning of the increase in error awareness also resonates with studies reporting that GenAI feedback helped students to notice unnoticed errors (Kurt & Kurt, 2024; Mun, 2024). However, these positive experiences did not translate well into students’ performance in the revision and editing processes. Due to the insignificant changes in students’ writing in the GenAI feedback stage, we could not discuss further. This gap between students’ self-reported experience and performance can be an opportunity for future research.

Students also described their negative experiences with GenAI feedback. Prompt engineering was mentioned by the students as one of the challenges, which was also found in Kurt and Kurt’s (2024) and Bok and Cho’s (2023) study, where students commented that their prompts were not properly understood, so GenAI gave them an answer completely different from what they intended. In this study, we provided students with a GenAI prompting guide and basic prompts (**Appendix A4**) for feedback, while encouraging them to create their own prompts. However, they rarely used their self-composed prompts, often relying instead on the provided prompts. This over-reliance on the provided prompt implies their limited GenAI literacy (Kim et al., 2025). Alternatively, it is also possible that students rarely used their self-composed prompts because they lacked trust in the feedback that GenAI could provide and therefore did not fully engage with it. While there have been no specific studies regarding the correlation between students’ trust in GenAI and their prompt engineering, scholars noted that students’ skeptical view of GenAI influenced their engagement in using the technology (Black & Tomlinson, 2025; Johnston et al., 2024).

Students in this study often mentioned that the GenAI feedback language was overly formal, complex, and confusing, a difficulty also reported in previous studies. For example, students in Bok and Cho’s (2023) study found that the vocabulary in ChatGPT feedback was sometimes “too advanced or difficult” to use (p. 21), while those in Campos’s (2025) study similarly noted that “the feedback could be unclear and complex” (p. 9). This difficulty may be related to students’ language proficiency, as highly proficient students did not find overly formal and complex GenAI feedback hindering their uptake of feedback (Toscu, 2024). One student in this study also described the feedback as too “robotic,” a term likewise used by participants in Kurt and Kurt’s (2024) study to express discomfort with the non-human tone of ChatGPT responses. As students may be more accustomed to teacher and peer feedback, which they perceive as more natural, scholars have noted that students may respond to this difficulty by prompting GenAI to simplify feedback or use simpler language (Kurt & Kurt, 2024; Mun, 2024) or by ignoring feedback they do not understand (Chen et al., 2024).

Students in this study also said that they were overwhelmed by the voluminous and lengthy GenAI feedback. GenAI feedback indeed has been reported to be overwhelming in quantity for learners to take up all at once due to its coverage (Chen et al., 2024; Toscu, 2024; Yu & Xie, 2025), and a high amount of feedback is known to negatively affect EFL students’ motivation and increase cognitive workload (Bitchener & Ferris, 2012). Consistent with this, some studies found that when faced with excessive GenAI feedback, ESL students chose to reject the feedback rather than use it in revision (Chen et al., 2024). This challenge may also be related to students’ GenAI prompting practices, as the nature of feedback depends partly on how prompts are formulated. In this study, students were allowed to use their native language (Indonesian) in prompts to facilitate more personalized feedback; however, their reliance on prepared generic prompts formulated in English may have shaped the GenAI’s responses. Moreover, although the GenAI was configured through system instructions to provide focused feedback using simple language, the generic prompts may have overridden these instructions,

resulting in feedback that was overly formal, complex, and voluminous. Together, these challenges may partly explain the plateau observed in students' writing performance across the revision and editing stages, despite their reported positive experiences with GenAI feedback.

Conclusions, Implications, and Limitations

Conclusions

The findings showed that GBA instruction could partially develop students' exposition writing in terms of linguistic features, particularly in the initial phase of independent construction. This development appears to be associated with explicit genre instruction and teacher scaffolding in the early stages of the teaching–learning cycle, given that students' writing development was observed only from the diagnostic test to the first draft. However, this development cannot be attributed to GenAI feedback, as no differences in students' writing during the later stages of independent construction were observed. The decline in the final test further indicates that students still require instructional support.

Regarding students' experiences, despite the insignificant change in the revised and edited draft, students expressed positive appreciation toward GenAI feedback, including its immediacy, coverage, and overall support. However, they also reported interrelated difficulties. Limited prompting skills led to overreliance on generic prompts, which in turn produced overly formal, robotic, voluminous, and lengthy feedback. These challenges made the feedback difficult to comprehend and manage, contributing to cognitive overload and reinforcing trust-related issues.

In conclusion, integrating GenAI feedback into GBA or the teaching–learning cycle requires careful consideration, including the need for relevant technical knowledge and further assistance in seeking, receiving, and understanding feedback. However, as no changes in students' writing were observed during revision and editing with GenAI feedback, this study emphasizes the continued importance of explicit instruction and teacher scaffolding.

Implications

This study was conducted in an Indonesian EFL context. However, the findings may be relevant to wider ESL/EFL settings, as GBA has been implemented in various educational contexts globally, and feedback provision remains a challenge in teaching EFL academic writing. The findings of this study suggest that explicit instruction and teacher scaffolding remain essential in GBA or the teaching-learning cycle (TLC). In this study, only one TLC cycle was implemented. While this single cycle allowed students to learn strategies to build content knowledge, familiarity with the exposition genre structure and linguistic features, and the writing process, GBA scholars typically recommend repeating the cycle to ensure the development of more stable genre knowledge across contexts. These forms of genre knowledge, developed through the TLC, are not dependent on continuous access to GenAI and may support students' future writing in similar academic contexts. However, this study does not claim improvement in general writing proficiency or performance, as the focus was limited to the exposition genre rather than writing across multiple genres.

Students reported positive experiences with GenAI feedback, but no development was observed during revision and editing, indicating that GenAI feedback alone may be insufficient to support writing development. When GenAI tools are integrated into GBA, explicit instruction in GenAI prompting and guidance in interpreting and using feedback are therefore necessary, although the results remain inconclusive. Thus, integrating GenAI feedback into GBA requires critical consideration.

Limitations

The conclusions above should be approached cautiously, due to the limitations of this study. First, the small sample size (n=15) and the classroom-based case study design restrict the generalizability of the findings beyond similar instructional contexts. Second, all participants were undergraduate EFL university students recruited through convenience sampling, which may limit the applicability of the results to a broader context. Third, the absence of a control group means that the observed improvements cannot be attributed solely to the teaching program, as other instructional factors may have contributed. Fifth, this study used a Gemini-based AES tool to score students' drafts, with limited human benchmarking. Although a GenAI-based AES could support evaluation, it could also suffer from scoring inconsistency due to model drift and scoring bias. Finally, the study did not analyze how students used GenAI feedback in their revisions, preventing conclusions about the extent to which the feedback was actually incorporated into the drafts. Considering these limitations, future research should therefore involve larger and more diverse cohorts, include human raters, control or comparison groups in experimental designs, and longitudinally track students' engagement with GenAI feedback in the revision process.

Acknowledgements

This study is conducted as part of the corresponding author's doctoral study, which is funded by the *Lembaga Pengelola Dana Pendidikan – LPDP* (Indonesia Endowment Fund for Education) under the Ministry of Finance of the Republic of Indonesia.

About the Authors

Zainurrahman is a lecturer in English Education at the *Institut Sains dan Ilmu Kependidikan (ISDIK) Kie Raha*, Indonesia, and a doctoral student at Universitas Pendidikan Indonesia in English Language Education. His research interests include academic writing, applied linguistics, and feedback provision in the second and foreign language writing instruction, and Generative AI in education. ORCID ID: 0000-0002-2576-8417

Emi Emilia is a professor of language and literacy education at Universitas Pendidikan Indonesia. Her expertise covers Systemic Functional Linguistics and its Genre-based Pedagogy, critical literacy, and critical pedagogy. ORCID ID: 0000-0002-4526-7740

Rojab Siti Rodliyah is an associate professor in English Language Education at Universitas Pendidikan Indonesia. Her research interests include literacy development, teacher training, digital literacy in EFL learning, and the integration of information and communication technology (ICT) into English instruction. ORCID ID: 0000-0001-7493-3354

To Cite this Article

Zainurrahman, Emilia, E. & Rodliyah, R. S. (2026). A genre-based approach with GenAI feedback to teaching written exposition in a tertiary EFL context in Indonesia. *Teaching English as a Second Language Electronic Journal (TESL-EJ)*, 30(1).
<https://doi.org/10.55593/ej.30117a8>

References

- Aldabbus, S., & Almansouri, E. (2022). Academic writing difficulties encountered by university EFL learners. *British Journal of English Language Linguistics*, 10(3), 1–11.
<https://doi.org/10.37745/bjel.2013/vol10n3111>
- Alfianika, N., Sunendar, D., Sastromiharjo, A., & Damaianti, V. S. (2019). Writing scientific work for Indonesia language and literature education study program students at university. *Proceedings of the International Conference on Education, Language and Society*, 349–358.
<https://doi.org/10.5220/0008998803490358>

- Alnemrat, A., Aldamen, H., Almashour, M., Al-Deaibes, M., & AlSharefeen, R. (2025). AI vs. teacher feedback on EFL argumentative writing: A quantitative study. *Frontiers in Education, 10*, Article 1614673. <https://doi.org/10.3389/educ.2025.1614673>
- Alsaedi, N. (2024). ChatGPT and EFL/ESL writing: A systematic review of advantages and challenges. *English Language Teaching, 17*(5), 41–50. <https://doi.org/10.5539/elt.v17n5p41>
- Anders, B. A. (2024). Student skills need to evolve to match our new AI society. *Patterns, 5*(10), Article 101062. <https://doi.org/10.1016/j.patter.2024.101062>
- Arifin, A. A. A., Arifin, W. L., & Rini, S. (2024). The investigation of students' writing skill improvement in GBA implementation using procedure text. *Proceedings of International Interdisciplinary Conference and Research Expo, 1*(1), 43–58. <https://doi.org/10.18326/iicare.v1i1.637>
- Ariyanfar, S., & Mitchell, R. (2020). Teaching writing skills through genre: Applying the genre-based approach in Iran. *International Research Journal of Management, IT and Social Sciences, 7*(1), 242–257. <https://doi.org/10.21744/irjmis.v7n1.843>
- Aunurrahman, Hamied, F. A., & Emilia, E. (2017). A joint construction practice in an academic writing course in an Indonesian university context. *Celt: A Journal of Culture, English Language Teaching and Literature, 17*(1), 27–44. <https://doi.org/10.24167/celt.v17i1.1137>
- Banister, C. (2023). Exploring peer feedback processes and peer feedback meta-dialogues with learners of academic and business English. *Language Teaching Research, 27*(3), 746–764. <https://doi.org/10.1177/1362168820952222>
- Bitchener, J., & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. Routledge.
- Black, R. W., & Tomlinson, B. (2025). University students describe how they adopt AI for writing and research in a general education course. *Scientific Reports, 15*(1), Article 8799. <https://doi.org/10.1038/s41598-025-92937-2>
- Bok, E., & Cho, Y. (2023). Examining Korean EFL college students' experiences and perceptions of using ChatGPT as a writing revision tool. *STEM Journal, 24*(4), 15–27. <https://doi.org/10.16875/stem.2023.24.4.15>
- Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners* (1st ed.). SAGE.
- Bui, N. M., & Barrot, J. S. (2025). ChatGPT as an automated essay scoring tool in the writing classrooms: How it compares with human scoring. *Education and Information Technologies, 30*(2), 2041–2058. <https://doi.org/10.1007/s10639-024-12891-w>
- Butt, D., Fahey, R., Feez, S., & Spinks, S. (2012). *Using functional grammar: An explorer's guide* (3rd ed.). Palgrave Macmillan.
- Cahyono, S. P. (2018). Teaching L2 writing through the use of Systemic Functional Linguistics (SFL). *Indonesian JELT: Indonesian Journal of English Language Teaching, 13*(1), 53–72. <https://doi.org/10.25170/ijelt.v13i1.1450>
- Campos, M. (2025). AI-assisted feedback in CLIL courses as a self-regulated language learning mechanism: Students' perceptions and experiences. *European Public & Social Innovation Review, 10*, 1–14. <https://doi.org/10.31637/epsir-2025-1568>

- Castulo, N. J., Marasigan, A. C., Buenaventura, Ma. L. D., De Vera, J. L., Bagaporo, E. C., Juan, M. P. C. S., & Dalida, N. S. (2025). Contextualizing the challenges of education graduate students in the Philippines: Translating needs analysis into strategic solutions. *Discover Education*, 4(1), 27. <https://doi.org/10.1007/s44217-025-00416-7>
- Chen, F. (2021). Exploring students' perceptions and attitudes towards genre-based pedagogy developed in persuasive writing teaching: The systemic functional linguistics perspective. *Arab World English Journal*, 12(4), 243–258. <https://doi.org/10.24093/awej/vol12no4.17>
- Chen, Z., Zhu, X., Lu, Q., & Wei, W. (2024). L2 students' barriers in engaging with form and content-focused AI-generated feedback in revising their compositions. *Computer Assisted Language Learning*, 1–21. <https://doi.org/10.1080/09588221.2024.2422478>
- Christie, F., & Martin, J. R. (Eds.). (2000). *Genre and institutions: Social processes in the workplace and school*. Continuum.
- Clarke, V., & Braun, V. (2017). Thematic analysis. *The Journal of Positive Psychology*, 12(3), 297–298. <https://doi.org/10.1080/17439760.2016.1262613>
- Coffin, C. (2004). Arguing about how the world is or how the world should be: The role of argument in IELTS tests. *Journal of English for Academic Purposes*, 3(3), 229–246. <https://doi.org/10.1016/j.jeap.2003.11.002>
- Creswell, J. W. (2009). *Qualitative inquiry and research design: Choosing among five approaches* (2nd ed.). SAGE.
- Derewianka, B. (2003). Trends and issues in genre-based approaches. *RELC Journal*, 34(2), 133–154. <https://doi.org/10.1177/003368820303400202>
- Derewianka, B., & Jones, P. (2016). *Teaching language in context* (2nd ed.). Oxford University Press.
- Dignum, V. (2019). *Responsible artificial intelligence: How to develop and use AI in a responsible way*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-30371-6>
- Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford University Press.
- Durán, D. C. (2021). Fostering written production of review texts among EFL university students through a genre-based approach. *Íkala*, 26(1), 117–138. <https://doi.org/10.17533/udea.ikala.v26n02a10>
- Emilia, E. (2005). *A critical genre-based approach to teaching academic writing in a tertiary EFL context in Indonesia* [Doctoral Thesis]. The University of Melbourne.
- Emilia, E. (2011). *Pendekatan berbasis teks dalam pengajaran Bahasa Inggris (Text-based approach in teaching English)*. PT. Kiblat Buku Utama.
- Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20(1), 1–20. <https://doi.org/10.1186/s41239-023-00425-2>
- Faradina, W. N., & Gandana, I. S. S. (2024). Inquiring into a teacher's understanding of genre-based pedagogy: A case study. In N. Haristiani, Y. Yulianeta, Y. Wirza, W. Gunawan, A. A. Danuwijaya, E. Kurniawan, S. Suharno, N. Nafisah, & E. D. A. Imperiani (Eds.), *Proceedings of the 7th International Conference on Language, Literature, Culture, and*

- Education (ICOLLITE 2023)* (Vol. 832, pp. 485–493). Atlantis Press International BV. https://doi.org/10.2991/978-94-6463-376-4_65
- Ghanbari, N., & Salari, M. (2022). Problematizing argumentative writing in an Iranian EFL undergraduate context. *Frontiers in Psychology, 13*, 862400. <https://doi.org/10.3389/fpsyg.2022.862400>
- Gheorghiu, A. (2024). *Building data-driven applications with llamaindex: A practical guide on retrieval-augmented generation (RAG) to enhance LLM applications* (1st ed.). Packt Publishing Ltd.
- Gibbons, P. (2015). *Scaffolding language, scaffolding learning: Teaching English language learners in the mainstream classroom* (2nd ed.). Heinemann.
- Google. (2025). *Experiment with parameter values*. Google Cloud. <https://cloud.google.com/vertex-ai/generative-ai/docs/learn/prompts/adjust-parameter-values>
- Hagerty, A., & Rubinov, I. (2019). *Global AI ethics: A review of the social impacts and ethical implications of artificial intelligence* (arXiv:1907.07892). arXiv. <https://arxiv.org/pdf/1907.07892>
- Hamied, F. A. (2023). *Research methods: A guide for first-time researchers* (3rd ed.). UPI Press.
- Hatch, E., & Lazaraton, A. (1994). *The research manual: Design and statistics for applied linguistics*. Heinle & Heinle.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Herman, Purba, R., Van Thao, N., & Purba, A. (2020). Using genre-based approach to overcome students' difficulties in writing. *Journal of Education and E-Learning Research, 7*(4), 464–470. <https://doi.org/10.20448/journal.509.2020.74.464.470>
- Hidayat, R. R. A., Fajriah, Y. N., Juanda, K. N. I., & Putri, I. (2023). Portraying systemic functional linguistics genre-based approach in Kurikulum Merdeka in EFL Senior High School. *Proceedings Virtual English Education Students Conference, 2*, 148–164. <https://ejournal.unsub.ac.id/index.php/vesco/article/download/1876/1419>
- Hutabarat, D. S. A., & Gunawan, W. (2021). *GBA in teaching writing to scaffold students in online learning*. Thirteenth Conference on Applied Linguistics (CONAPLIN 2020), Bandung, Indonesia. <https://doi.org/10.2991/assehr.k.210427.016>
- Hyland, K. (2004). *Genre and second language writing*. University of Michigan Press.
- Hyland, K., & Hyland, F. (2019). Contexts and issues in feedback on L2 writing. In K. Hyland & F. Hyland (Eds.), *Feedback in Second Language Writing* (2nd ed., pp. 1–22). Cambridge University Press. <https://doi.org/10.1017/9781108635547.003>
- Jain, S., Hitzig, Z., & Mishkin, P. (2024). *Contextual confidence and Generative AI* (arXiv:2311.01193). arXiv. <http://arxiv.org/abs/2311.01193>
- Johnston, H., Wells, R. F., Shanks, E. M., Boey, T., & Parsons, B. N. (2024). Student perspectives on the use of generative artificial intelligence technologies in higher education. *International Journal for Educational Integrity, 20*(1), Article 2. <https://doi.org/10.1007/s40979-024-00149-4>
- Kalota, F. (2024). A primer on generative artificial intelligence. *Education Sciences, 14*(2), Article 172. <https://doi.org/10.3390/educsci14020172>

- Kiely, R., & Rea-Dickins, P. (2005). *Program evaluation in language education*. Palgrave Macmillan UK. <https://doi.org/10.1057/9780230511224>
- Kim, J. (Claudia), & Li, S. (2025). The effects of task repetition and corrective feedback on L2 writing development. *The Language Learning Journal*, 53(6), 729–744. <https://doi.org/10.1080/09571736.2024.2390555>
- Kim, J., Yu, S., Detrick, R., & Li, N. (2024). Exploring students' perspectives on Generative AI-assisted academic writing. *Education and Information Technologies*, 30, 1265–1300. <https://doi.org/10.1007/s10639-024-12878-7>
- Kim, J., Yu, S., Lee, S.-S., & Detrick, R. (2025). Students' prompt patterns and its effects in AI-assisted academic writing: Focusing on students' level of AI literacy. *Journal of Research on Technology in Education*, 1–18. <https://doi.org/10.1080/15391523.2025.2456043>
- Kuiper, C., Smit, J., De Wachter, L., & Elen, J. (2017). Scaffolding tertiary students' writing in a genre-based writing intervention. *Journal of Writing Research*, 9(1), 27–59. <https://doi.org/10.17239/jowr-2017.09.01.02>
- Kulkarni, A., & Shivananda, A. (2019). *Natural language processing recipes: Unlocking text data with machine learning and deep learning using python*. Apress. <https://doi.org/10.1007/978-1-4842-4267-4>
- Kurt, G., & Kurt, Y. (2024). Enhancing L2 writing skills: ChatGPT as an automated feedback tool. *Journal of Information Technology Education: Research*, 23, Article 24. <https://doi.org/10.28945/5370>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, Article 863. <https://doi.org/10.3389/fpsyg.2013.00863>
- Law, L. (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. *Computers and Education Open*, 6, Article 100174. <https://doi.org/10.1016/j.caeo.2024.100174>
- Lee, S. S., & Moore, R. L. (2024). Harnessing Generative AI (GenAI) for automated feedback in higher education: A systematic review. *Online Learning*, 28(3), 82–106. <https://doi.org/10.24059/olj.v28i3.4593>
- Lipnevich, A. A., & Smith, J. K. (2009). Effects of differential feedback on students' examination performance. *Journal of Experimental Psychology: Applied*, 15(4), 319–333. <https://doi.org/10.1037/a0017841>
- Lo, N., Wong, A., & Chan, S. (2025). The impact of generative AI on essay revisions and student engagement. *Computers and Education Open*, 9, Article 100249. <https://doi.org/10.1016/j.caeo.2025.100249>
- Maguire, M., & Delahunt, B. (2017). Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars. *All Ireland Journal of Teaching and Learning in Higher Education*, 8(3), Article 335.
- Martin, J. R., & Rose, D. (2009). *Genre relations: Mapping culture*. Equinox.
- Martin, J. R., & White, P. R. (2005). *The language of evaluation: Appraisal in English*. Palgrave Macmillan.
- Merriam, S. B. (1991). *Case study research in education: A qualitative approach*. Jossey-Bass.

- Meyer-Junco, L., Waldfogel, J. M., & Duncan, N. (2023). Peer review questions & answers: How?: Part II: reviewing case reports, systematic reviews, narrative reviews, and opinion pieces. *Journal of Pain & Palliative Care Pharmacotherapy*, 37(3), 209–212. <https://doi.org/10.1080/15360288.2023.2245738>
- Mills, A. J., Durepos, G., & Wiebe, E. (Eds.). (2010). *Encyclopedia of case study research*. Sage.
- Mingsakoon, P., & Srinon, U. (2018). Development of secondary school students' generic structure execution in personal experience recount writing texts through SFL genre-based approach. *Advances in Language and Literary Studies*, 9(6), 112–119. <https://doi.org/10.7575/aiac.all.v.9n.6p.112>
- Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, 2(2), Article 100050. <https://doi.org/10.1016/j.rmal.2023.100050>
- Moll, L. C. (1990). Vygotsky's zone of proximal development: Rethinking its instructional implications. *Infancia y Aprendizaje*, 13(51–52), 157–168. <https://doi.org/10.1080/02103702.1990.10822276>
- Morgan, E., To, V., & Thomas, A. (2022). Using genre-based pedagogy to teach structural staging of short persuasive essays in a Japanese university context. *English as a Foreign Language International Journal*, 2(6), 49–79. <https://doi.org/10.56498/422262022>
- Mun, C. (2024). EFL learners' English writing feedback and their perception of using ChatGPT. *STEM Journal*, 25(2), 26–39. <https://doi.org/10.16875/stem.2024.25.2.26>
- Nagao, A. (2018). A genre-based approach to writing instruction in EFL classroom contexts. *English Language Teaching*, 11(5), 130–147. <https://doi.org/10.5539/elt.v11n5p130>
- Nagao, A. (2019). The SFL genre-based approach to writing in EFL contexts. *Asian-Pacific Journal of Second and Foreign Language Education*, 4, Article 6. <https://doi.org/10.1186/s40862-019-0069-3>
- Nagao, A. (2020). Adopting an SFL approach to teaching L2 writing through the teaching learning cycle. *English Language Teaching*, 13(6), 144–161. <https://doi.org/10.5539/elt.v13n6p64>
- Nagao, A. (2022). A genre-based approach to teaching descriptive report writing to Japanese EFL university students. *Teaching English as a Second or Foreign Language Journal--TESL-EJ*, 26(3), Article 13. <https://doi.org/10.55593/ej.26103a13>
- Nunan, D. (1992). *Research methods in language learning*. Cambridge University Press.
- Nurlaelawati, I., & Novianti, N. (2017). The practice of genre-based pedagogy in Indonesian schools: A case of preservice teachers in Bandung, West Java province. *Indonesian Journal of Applied Linguistics*, 7(1), 160–166. <https://doi.org/10.17509/ijal.v7i1.6869>
- Nurlatifah, L., & Yusuf, F. N. (2022). Students' problems in writing analytical exposition text in EFL classroom context. *English Review: Journal of English Education*, 10(3), 801–810. <https://doi.org/10.25134/erjee.v10i3.6633>
- Paris, B. (2022). Instructors' perspectives of challenges and barriers to providing effective feedback. *Teaching and Learning Inquiry*, 10, Article 3. <https://doi.org/10.20343/teachlearninqu.10.3>

- Peloghitis, J. (2017). Difficulties and strategies in argumentative writing: A qualitative analysis. *Transformation in Language Education*, 399–406. <https://jalt-publications.org/sites/default/files/pdf-article/jalt2016-pcp-052.pdf>
- Pessoa, S., Mitchell, T. D., & Miller, R. T. (2018). Scaffolding the argument genre in a multilingual university history classroom: Tracking the writing development of novice and experienced writers. *English for Specific Purposes*, 50, 81–96. <https://doi.org/10.1016/j.esp.2017.12.002>
- Plonsky, L. (Ed.). (2015). *Advancing quantitative methods in second language research* (1st ed.). Routledge. <https://doi.org/10.4324/9781315870908>
- Plonsky, L., & Oswald, F. L. (2014). How big is “Big”? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878–912. <https://doi.org/10.1111/lang.12079>
- Polakova, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students’ writing skills. *Cogent Education*, 11(1), Article 2410101. <https://doi.org/10.1080/2331186X.2024.2410101>
- Rafida, T., Suwandi, S., & Ananda, R. (2024). EFL students’ perception in Indonesia and Taiwan on using artificial intelligence to enhance writing skills. *Jurnal Ilmiah Peuradeun*, 12(3), 987–1016. <https://doi.org/10.26811/peuradeun.v12i3.1520>
- Rodliyah, R. S., & Liani, A. E. (2022). SFL analysis: An investigation of students’ use of cohesive devices in exposition text. *Indonesian Journal of Applied Linguistics*, 12(1), 235–246. <https://doi.org/10.17509/ijal.v12i1.46596>
- Scheff, S. W. (2016). Nonparametric statistics. In S. W. Scheff (Ed.), *Fundamental Statistical Principles for the Neurobiologist* (pp. 157–182). Elsevier. <https://doi.org/10.1016/B978-0-12-804753-8.00008-7>
- Schleppegrell, M. J. (2001). Linguistic features of the language of schooling. *Linguistics and Education*, 12(4), 431–459. [https://doi.org/10.1016/S0898-5898\(01\)00073-0](https://doi.org/10.1016/S0898-5898(01)00073-0)
- Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistics perspective*. Lawrence Erlbaum Associates.
- Schleppegrell, M. J. (2013). The role of metalanguage in supporting academic language development. *Language Learning*, 63(s1), 153–170. <https://doi.org/10.1111/j.1467-9922.2012.00742.x>
- Selvaraj, A. M., Azman, H., & Wahi, W. (2021). Teachers’ feedback practice and students’ academic achievement: A systematic literature review. *International Journal of Learning, Teaching and Educational Research*, 20(1), 308–322. <https://doi.org/10.26803/ijlter.20.1.17>
- Shermis, M. D., & Wilson, J. (Eds.). (2024). *The Routledge international handbook of automated essay evaluation* (1st ed.). Routledge. <https://doi.org/10.4324/9781003397618>
- Sritrakarn, N. (2020). Using the SFL genre-based approach to improve Thai learners’ writing of an explanation. *The New English Teacher*, 14(1), 56–77.
- Stahl, N. A., & King, J. R. (2020). Understanding and using trustworthiness in qualitative research. *Journal of Developmental Education*, 44(1), 26–28. <https://files.eric.ed.gov/fulltext/EJ1320570.pdf>
- Stake, R. E. (1995). *The art of case study research*. SAGE.

- Suharyadi, & Basthomi, Y. (2020). Patterns of the teaching and learning cycle of GBA by EFL teachers in Indonesia. *Journal of Education and E-Learning Research*, 7(1), 34–41. <https://doi.org/10.20448/journal.509.2020.71.34.41>
- Suryadi, H., & Yulandari, E. S. (2022). The influence of using genre based instruction (GBI) in writing skill exposition text in students SMPN 3 Pringgarata. *EDUKASIA: Jurnal Pendidikan Dan Pembelajaran*, 3(3), 617–628. <https://doi.org/10.62775/edukasia.v3i3.169>
- Syarifah, E. F., & Gunawan, W. (2016). Scaffolding in the teaching of writing discussion texts based on SFL genre-based approach. *English Review: Journal of English Education*, 4(1), 39–53. <https://doi.org/10.25134/erjee.v4i1.306>
- Tabari, M. A., Khezrlou, S., & Ghanbar, H. (2024). Task repetition versus task rehearsal: Understanding effects of task-readiness factors and elemental genres on L2 writing task performance. *Language Teaching Research*, Article 13621688241249689. <https://doi.org/10.1177/13621688241249689>
- Tomczak, M., & Tomczak, E. (2014). The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends in Sport Sciences*, 21(1), 19–25. <https://www.wbc.poznan.pl/publication/413565>
- Toscu, S. (2024). An investigation on the effectiveness of chatbots in evaluating writing assignments in EFL contexts. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi*, (72), 295–329. <https://doi.org/10.21764/maeuefd.1425384>
- Triastuti, A., Madya, S., & Chappell, P. (2022). Genre-based teaching cycle and instructional design for teaching texts and mandated curriculum contents. *Indonesian Journal of Applied Linguistics*, 12(1), 1–15. <https://doi.org/10.17509/ijal.v12i1.46563>
- Tseng, W., & Warschauer, M. (2023). AI-writing tools in education: If you can't beat them, join them. *Journal of China Computer-Assisted Language Learning*, 3(2), 258–262. <https://doi.org/10.1515/jccall-2023-0008>
- Tulasi, V. L., Rao, C. S., & Kumar, V. P. (2025). English as the language of research and worldwide academic journals. *Journal for Research Scholars and Professionals of English Language Teaching*, 9(47). <https://doi.org/10.54850/jrspelt.9.47.001>
- Urmeneta, A., & Romero, M. (Eds.). (2024). *Creative applications of artificial intelligence in education*. Springer Nature Switzerland. <https://doi.org/10.1007/978-3-031-55272-4>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press.
- Wang, C. (2025). Exploring students' generative AI-assisted writing processes: Perceptions and experiences from native and nonnative English speakers. *Technology, Knowledge and Learning*, 30, 1825–1846. <https://doi.org/10.1007/s10758-024-09744-3>
- Wei, R., Hu, Y., & Xiong, J. (2019). Effect size reporting practices in applied linguistics research: A study of one major journal. *Sage Open*, 9(2), Article 2158244019850035. <https://doi.org/10.1177/2158244019850035>
- Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press.
- West, P., Lu, X., Dziri, N., Brahman, F., Li, L., Hwang, J. D., Jiang, L., Fisher, J., Ravichander, A., Chandu, K. R., Newman, B., Koh, P. W., Ettinger, A., & Choi, Y. (2024). The generative AI paradox: “What it can create, it may not understand.” *The Twelfth International Conference on Learning Representations (ICLR)*. The International Conference on Learning Representations (ICLR), Austria. <https://openreview.net/pdf?id=CF8H8MS5P8>

- Yin, R. K. (2007). *Case study research: Design and methods* (3rd ed.). SAGE.
- Yu, H., & Xie, Q. (2025). Generative AI vs. teachers: Feedback quality, feedback uptake, and revision. *Language Teaching Research Quarterly*, 47, 113–137.
<https://doi.org/10.32038/ltrq.2025.47.07>
- Zebua, S., & Rozimela, Y. (2020). The implementation of genre-based approach in teaching writing analytical exposition text at SMAN 8 Padang. *Proceedings of the 7th International Conference on English Language and Teaching (ICOELT 2019)*, 411, 104–107.
<https://doi.org/10.2991/assehr.k.200306.018>
- Zein, S., Sukyadi, D., Hamied, F. A., & Lengkanawati, N. S. (2020). English language education in Indonesia: A review of research (2011–2019). *Language Teaching*, 53(4), 491–523. <https://doi.org/10.1017/S0261444820000208>
- Zhang, X. (2021). Exploring the interaction of EFL student writers with SFL-based teaching and teacher-written feedback. *Revista Signos*, 54(106), 465–486.
<https://doi.org/10.4067/S0718-09342021000200465>

Appendices

A1. An SFL-informed Exposition Rubric

Categories and Criteria	Scale
<i>Schematic Structure</i>	
1 The essay begins with an introduction to the topic/issue and the thesis statement, followed by a series of arguments to support the thesis (optionally a counterargument), and a conclusion reiterating the thesis.	0-4
<i>Ideational</i>	
2 The essay is grounded in accurate and relevant knowledge from the source text.	0-4
3 The thesis uses specialized/technical vocabulary to characterize an overarching claim.	0-4
4 The supporting claims are relevant and clearly create an analytical framework.	0-4
5 The thesis and Reinforcement stages are consistent.	0-4
6 Readers can clearly understand when, where, what, and who because relevant lexicogrammatical features are used in the essay (particularly in the first paragraph).	0-4
<i>Interpersonal</i>	
7 The essay has an argumentative thesis.	0-4
8 The essay uses expanding resources (attribution) to bring in the source text.	0-4
9 The essay uses contracting resources (endorsement) to show how cited material supports the claim.	0-4
10 The essay uses a concession-counter move (counterargument) to show awareness of a different perspective and bring the reader towards the writer's perspective.	0-4
11 The essay uses lexicogrammatical features related to the writer's opinion to show their support for or opposition to the topic (modalities, auxiliary verbs, and -ly adverbs).	0-4
12 The writer avoids using lexicogrammatical features related to personal pronouns, especially "I", and instead uses other words to replace subjective words.	0-4
<i>Textual</i>	
13 The introduction previews the content of the essay.	0-4
14 The conclusion sums up the content of the essay.	0-4
15 The language and order of sub-claims match the preview in the introduction.	0-4
16 Sub-claims are placed at the beginning of the paragraph.	0-4
17 The writer uses sensing verbs (e.g., believe, think, consider, etc.) and reporting verbs (e.g., claim that, argued, supported, explained, criticized that, etc.) along with citation, e.g., Author's name (year).	0-4
18 The essay demonstrates the use of nominalization.	0-4
19 Linking/signpost words are introduced in the body of the essay.	0-4

Note. Source: adapted from Pessoa et al. (2018) and Nagao (2020).

Scale description:

- 0: Poor execution or no use or almost no use of linguistic resources;
- 1: Limited use of linguistic resources;
- 2: Only fair or problematic use of linguistic resources;
- 3: Fairly good use with minor problems or inconsistencies using linguistic resources;
- 4: Excellent use of linguistic resources.

A2. Weekly Breakdown of the TLC

Phase	Week(s)	Stage / Activity	Description
Preliminary Phase	Week 1	Student profiling and diagnostic assessment	Students completed an online open-ended questionnaire and a 60-minute diagnostic exposition writing test. Supporting materials were provided (six opinion pieces, exposition handout, prompt-engineering guide, four argumentative-writing videos, web-app demonstration).
Implementation Phase	Weeks 2–3	BKOF	Topic: The use of AI in academic writing. Three opinion pieces were used in the teacher’s presentation to build students’ content knowledge.
	Week 4	Modeling of exposition	Schematic structure and linguistic features of exposition were introduced using the remaining three opinion pieces as model texts.
	Weeks 5–6	Joint construction	Students created web-app accounts and jointly wrote an exposition (“Is it ethical to use AI to write the entire paper?”). Demonstrations were provided on outlining, drafting, preparing prompts for feedback, using the web-app features, and revising.
	Weeks 7–12	Independent construction	Students wrote an exposition on a different self-selected topic and recorded their experiences with GenAI feedback in the web app. Drafts (first, revised, edited) and feedback logs were collected. All activities were conducted online due to the Ramadan break.
Evaluation Phase	Week 13	Final assessment and interviews	A 60-minute final writing test was administered on a different self-selected topic. Semi-structured interviews were conducted (11 students, Bahasa Indonesia), focusing on experiences with GenAI feedback.

Note. TLC = Teaching-Learning Cycle; BKOF = Building Knowledge of the Field

A3. GenAI Model Parameter Setting for Feedback

Model: gemini-2.0-flash

Temperature: 0

Top-P: 0

System Instruction:

You are a critical EFL writing teacher who is an expert in the Systemic Functional Linguistics Genre-based Approach. Your task is to meticulously read an exposition essay, evaluate the strengths and weaknesses, and provide corrective and suggestive feedback, based on the rubric/rating scale and my query. If the essay and rating scale do not provide sufficient context from which you can build your feedback, say so. Maintain a concise format, use words EFL students can easily comprehend, without missing important points. Your responses should be in bullets.

Here is the rubric/rating scale: [insert rubric/rating scale below]

Here is the exposition essay: [insert the essay below] (use this if the student seeks GenAI feedback by themselves; if the tool is used by the teacher to provide feedback on students’ essays, each essay should be provided with the first prompt).

A4. Sample Prompts for Feedback Provided to Students

- Have I introduced the topic or issue clearly? You should check this in the first paragraph of my draft
- Have I stated my thesis or position clearly? You should check this near the end of my last introductory paragraph
- Have I elaborated on arguments to support my thesis statement? My arguments should be in the body of my essay, before the conclusion at the end of it
- Have I cited external sources to support my arguments?
- Have I used transition signals and text connectives to create a smooth flow between ideas?
- Have I used modal verbs effectively and accurately throughout my essay to engage and convince my readers regarding my position?
- In the conclusion section, have I concluded my main arguments and reiterated my thesis statement more confidently?
- Overall, have I organized my essay based on an analytical exposition schematic structure?

A5. Pairwise comparisons: Overall writing scores

(I) Stage	(J) Stage	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b		d _z
					Lower Bound	Upper Bound	
Diagnostic Test	First Draft	-25.612*	3.158	.000	-36.114	-15.110	-2.094
	Revised Draft	-24.648*	3.128	.000	-35.050	-14.246	-2.035
	Edited Draft	-26.578*	2.892	.000	-36.196	-16.960	-2.373
	Final Test	-18.157*	1.724	.000	-23.889	-12.425	-2.719
First Draft	Diagnostic Test	25.612*	3.158	.000	15.110	36.114	2.094
	Revised Draft	.964	.965	1.000	-2.246	4.174	0.083
	Edited Draft	-.966	1.023	1.000	-4.369	2.437	-0.080
	Final Test	7.455*	2.193	.043	.162	14.748	0.878
Revised Draft	Diagnostic Test	24.648*	3.128	.000	14.246	35.050	2.035
	First Draft	-.964	.965	1.000	-4.174	2.246	-0.083
	Edited Draft	-1.930	.841	.376	-4.726	.866	-0.414
	Final Test	6.491	2.080	.075	-.428	13.410	0.806
Edited Draft	Diagnostic Test	26.578*	2.892	.000	16.960	36.196	2.373
	First Draft	.966	1.023	1.000	-2.437	4.369	0.080
	Revised Draft	1.930	.841	.376	-.866	4.726	0.414
	Final Test	8.421*	1.843	.004	2.293	14.549	1.180
Final Test	Diagnostic Test	18.157*	1.724	.000	12.425	23.889	2.719
	First Draft	-7.455*	2.193	.043	-14.748	-.162	-0.878
	Revised Draft	-6.491	2.080	.075	-13.410	.428	-0.806
	Edited Draft	-8.421*	1.843	.004	-14.549	-2.293	-1.180

Note. Based on estimated marginal means; *. The mean difference is significant at the .05 level; b. adjustment for multiple comparisons: Bonferroni; d_z = effect size

A6. Pairwise comparisons: Schematic structure

Pairwise Comparison	Z	Asymp. Sig. (2-tailed)	r
First Draft – Diagnostic Test	-1.890	.059	.49
Revised Draft – Diagnostic Test	-1.890	.059	.49
Edited Draft – Diagnostic Test	-1.633	.102	.42
Final Test – Diagnostic Test	.000	1.000	0
Revised Draft – First Draft	.000	1.000	0
Edited Draft – First Draft	-.577	.564	.15
Final Test – First Draft	-1.890	.059	.49
Edited Draft – Revised Draft	-.577	.564	.15
Final Test – Revised Draft	-1.890	.059	.49
Final Test – Edited Draft	-1.633	.102	.42

Note. $\alpha = .005$; r = effect size

A7. Pairwise comparisons: Ideational meaning

Comparison	Z	Asymp. Sig.(2-tailed)	R
First Draft – Diagnostic	-3.246*	.001	.838
Revised Draft – Diagnostic	-3.245*	.001	.838
Edited Draft – Diagnostic	-3.239*	.001	.837
Final Test – Diagnostic	-3.265*	.001	.843
Revised Draft – First Draft	-0.106	.916	.027
Edited Draft – First Draft	-0.682	.495	.176
Final Test – First Draft	-1.581	.114	.408
Edited Draft – Revised Draft	-1.289	.197	.333
Final Test – Revised Draft	-1.841	.066	.475
Final Test – Edited Draft	-2.070	.038	.534

Note: *p<.005 (Bonferroni); r = effect size

A8. Pairwise comparisons: Interpersonal meaning

(I) Stage	(J) Stage	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b		d _z
					Lower Bound	Upper Bound	
Diagnostic Test	First Draft	-35.277*	3.940	.000	-48.380	-22.174	-2.312
	Revised Draft	-33.889*	4.377	.000	-48.445	-19.333	-1.999
	Edited Draft	-36.944*	3.876	.000	-49.835	-24.053	-2.461
	Final Test	-24.720*	3.056	.000	-34.883	-14.557	-2.089
First Draft	Diagnostic Test	35.277*	3.940	.000	22.174	48.380	2.312
	Revised Draft	1.388	2.100	1.000	-5.595	8.371	0.171
	Edited Draft	-1.667	1.810	1.000	-7.685	4.351	-0.238
	Final Test	10.557	3.423	.081	-.829	21.942	0.796
Revised Draft	Diagnostic Test	33.889*	4.377	.000	19.333	48.445	1.999
	First Draft	-1.388	2.100	1.000	-8.371	5.595	-0.171
	Edited Draft	-3.055	1.546	.682	-8.197	2.087	-0.510
	Final Test	9.169	3.334	.156	-1.918	20.256	0.710
Edited Draft	Diagnostic Test	36.944*	3.876	.000	24.053	49.835	2.461
	First Draft	1.667	1.810	1.000	-4.351	7.685	0.238
	Revised Draft	3.055	1.546	.682	-2.087	8.197	0.510
	Final Test	12.224*	3.315	.024	1.198	23.250	0.952
Final Test	Diagnostic Test	24.720*	3.056	.000	14.557	34.883	2.089
	First Draft	-10.557	3.423	.081	-21.942	.829	-0.796
	Revised Draft	-9.169	3.334	.156	-20.256	1.918	-0.710
	Edited Draft	-12.224*	3.315	.024	-23.250	-1.198	-0.952

Note: Based on estimated marginal means; *. The mean difference is significant at the .05 level; b. adjustment for multiple comparisons: Bonferroni; d_z = effect size

A9. Pairwise comparisons: Textual meaning

(I) Stage	(J) Stage	Mean Difference (I- J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b		d _z
					Lower Bound	Upper Bound	
Diagnostic Test	First Draft	-23.096*	3.518	.000	-34.794	-11.398	-1.695
	Revised Draft	-21.667*	3.278	.000	-32.569	-10.766	-1.707
	Edited Draft	-23.333*	3.015	.000	-33.361	-13.304	-1.998
	Final Test	-15.953*	1.703	.000	-21.616	-10.289	-2.419
First Draft	Diagnostic Test	23.096*	3.518	.000	11.398	34.794	1.695
	Revised Draft	1.429	1.700	1.000	-4.225	7.082	0.217
	Edited Draft	-.237	1.498	1.000	-5.218	4.745	-0.041
	Final Test	7.143	2.654	.176	-1.684	15.971	0.695
Revised Draft	Diagnostic Test	21.667*	3.278	.000	10.766	32.569	1.707
	First Draft	-1.429	1.700	1.000	-7.082	4.225	-0.217
	Edited Draft	-1.665	1.149	1.000	-5.486	2.155	-0.374
	Final Test	5.715	2.306	.266	-1.956	13.385	0.640
Edited Draft	Diagnostic Test	23.333*	3.015	.000	13.304	33.361	1.998
	First Draft	.237	1.498	1.000	-4.745	5.218	0.041
	Revised Draft	1.665	1.149	1.000	-2.155	5.486	0.374
	Final Test	7.380*	1.955	.020	.879	13.881	0.975
Final Test	Diagnostic Test	15.953*	1.703	.000	10.289	21.616	2.419
	First Draft	-7.143	2.654	.176	-15.971	1.684	-0.695
	Revised Draft	-5.715	2.306	.266	-13.385	1.956	-0.640
	Edited Draft	-7.380*	1.955	.020	-13.881	-.879	-0.975

Note. Based on estimated marginal means; *. The mean difference is significant at the .05 level; b. adjustment for multiple comparisons: Bonferroni; d_z = effect size

A10. Excerpts: Students' Notes and Interviews

Themes / Sub-themes	Notes	Interviews
POSITIVE EXPERIENCES		
<i>GenAI Helpfulness and Accessibility</i>		
Helpfulness in correction and revision	<p>“The feedback helped me strengthen and refine my arguments” (Note #143, P36).</p> <p>“It still took a lot of time and focus, especially in checking grammar and word choice, but it was easier because Gemini AI helped me a lot” (Note #151, P05).</p> <p>“I think Gemini AI was very useful for writing skills, because through this AI, I was told what was lacking in my writing, what should be added to which parts” (Note #172, P20).</p>	<p>“AI really helped me to correct my writing when I asked for feedback, not to mention when I am not familiar with the task” (Int. P03).</p> <p>“I think feedback from AI can be very helpful, especially to fix small issues that we have missed” (Int. P37).</p>
Increase in error awareness	<p>“I also found it interesting that Gemini could notice misused or misspelled words, which I had not realized before” (Note #156, P36).</p> <p>“The feedback was really helpful and concise. I became more aware of my mistake” (Note #167, P41).</p> <p>“Based on the feedback I got in writing my revised draft, I became more aware that to write this text, we should use modal verbs or stronger words...” (Note #154, P17).</p>	<p>“Sometimes, we were also unaware of the parts that needed revision, but the AI showed us the parts that need to be improved and enhanced” (Int. P06).</p> <p>“AI was really helpful, [because it could identify] mistakes that I myself was unaware of, like grammatical mistakes, unclear sentences, or a thesis statement that needed to be strengthened” (Int. P41).</p> <p>“This AI really helped me to see mistakes that I was unaware of, such as sentence structure, grammar, and also word choices, which were inaccurate when writing” (Int. P31).</p>
Immediacy of feedback	<p>“I also immediately asked for feedback from Gemini” (Note #145, P13).</p> <p>“Last Wednesday, I made a rough draft and got direct feedback from Gemini AI. The feedback really helped me...” (Note #124, P33).</p> <p>“It was exciting to receive direct feedback through the platform [the web app], which made the learning experience feel more personal and interactive” (Note #166, P31).</p>	<p>“It made it [writing and revising] easier, because we could get the feedback anytime” (Int. P06)</p> <p>“In the classroom, I can still receive feedback, but because of the time limitation and the number of students, only some students receive feedback from the lecturer directly. However, if we use tools like AI, we can do it alone and get the result instantly” (Int. P12).</p> <p>“AI was accessible at any time, especially when there was no help from friends or lecturers who could give feedback directly” (Int. P27).</p> <p>“It was quite interesting because AI could give me timely feedback on what other people could not see in detail, but it was noticed by the AI” (Int. P37).</p>
<i>Feedback Coverage</i>		
Detailed feedback	<p>“Receiving detailed feedback from Gemini gave me a new perspective on how to improve” (Note #127, P26).</p> <p>“I am grateful for the detailed feedback that helped me see areas for improvement” (Note #129, P22).</p> <p>“The feedback given was very detailed and clear if read carefully” (Note #124, P33).</p>	<p>“I think AI provided instant and detailed suggestions when I provided it with questions regarding the weaknesses of my text” (Int. P31).</p> <p>“I also asked Gemini for feedback on the analytical exposition text, and I was quite satisfied with the result, because the AI gave me objective and more detailed feedback” (Int. P27).</p> <p>“Feedback from AI happened to be detailed enough, which helped me much in revising” (Int. P05).</p>
Specific examples and guidance	None	<p>“Whatever I asked for, AI gave the examples” (Int. P12).</p> <p>“AI also gave me suggestions regarding more appropriate words” (Int. P41).</p>
Learning support	<p>“After completing the first draft and using the feedback from Gemini as a reference, I learned about how to explain my position...I also learned to create supporting arguments to strengthen the previous argument” (Note #126, P06).</p> <p>“I also learned a lot of new words from the feedback given” (Note #154, P17).</p> <p>“Based on the feedback I received, I understand more about how to make analytical exposition text...I also learned</p>	<p>“I think AI was very helpful for me to learn new knowledge that I was still confused about” (Int. P03).</p>

	that every argument we build must have a detailed and clear explanation” (Note #152, P17).	
NEGATIVE EXPERIENCES		
<i>Difficulty with prompt engineering</i>	None	<p>“AI provided answers (responses) based on how detailed the prompt was that we use. Therefore, I sometimes felt confused when I had to type the prompt asking for feedback...” (Int. P03).</p> <p>“If the prompt was not detailed enough, the response from the AI won’t be detailed as well” (Int. P05).</p> <p>“Perhaps, one thing that becomes the main difficulty was the prompt. We need to be specific about what we want” (Int. P37).</p> <p>“Sometimes I had problems because I was sometimes confused about how to write the prompt so that AI could give the appropriate answer” (Int. P31).</p>
Feedback Incomprehensibility		
Robotic and overly formal language	None	<p>“AI language was too robotic for me, and it confused me with the feedback” (Int. P27).</p> <p>“The language was difficult to understand, in my opinion, so I had to reread the feedback from the AI itself several times to understand it better” (Int. P32).</p> <p>“The language of AI was quite formal, so it was difficult to understand. So we needed to ask for another prompt so that the AI could provide an explanation in language that was easy to understand” (Int. P37).</p> <p>“I had to go to another AI, ChatGPT, and copied the Gemini feedback to ask the meaning of the feedback” (Int. P41).</p>
Overwhelming volume of feedback	<p>“Gemini also provided suggestions on sentence structure, which was useful but sometimes a little bit confusing to apply” (Note #143, P36).</p> <p>“Another challenge was managing the feedback process. At times, it felt overwhelming to revise based on multiple suggestions” (Note #171, P22).</p>	<p>“Much of AI feedback was lengthy... So, we had to be selective on which feedback to include, which one was true and helpful...” (Int. P05).</p> <p>“I had to look at the feedback one by one, because AI feedback was often lengthy. I had to look at which parts needed to be revised” (Int. P29).</p> <p>“The suggestions from AI were sometimes too many or too complex, which made me have to choose which one was most relevant to my writing” (Int. P31).</p>
Trust Issues		
Preference for human feedback and validation	<p>“To be honest, I would prefer personalized feedback from an expert or lecturer” (Note #11, P22).</p>	<p>“I will consider feedback from AI, but I will also use feedback from friends and lecturers” (Int. P27).</p> <p>“I always get feedback from AI, but it also comes from lecturers, like those who understand better. It also needs validation from the lecturer. So, I will consider it, but it requires guidance from the lecturer or people who know better in that field” (Int. P29).</p> <p>“I also want to make sure to the lecturer if the suggestions from AI are on track or valid” (Int. P41).</p> <p>“It was useful enough, but I think direct feedback from the lecturer or the peers is still better” (Int. P32).</p>
Perceived GenAI weaknesses	<p>“I don’t know why my rough draft is better than my last final draft? Since actually I followed the AI suggestion” (Note #171, P22).</p> <p>“I still have some trust issues with AI when it comes to assessment and correction” (Note #11, P22).</p>	<p>“AI is quite specific and very helpful in terms of grammar and writing, but in terms of arguments, I think it is lacking. In terms of providing feedback for arguments, it is lacking” (Int. P32).</p>

Copyright of articles rests with the authors. Please cite TESL-EJ appropriately.