Advanced Journal of Interdisciplinary Studies
Volume 10, Issue 1, 2026 | Pages 3079–3098
Journal homepage: www.ajids.dmu.edu.et

Research Article

A Mixed-Methods Longitudinal Investigation of Writing Conferences in Higher Education: Differential Effects on Higher- and Lower-Order Concerns in Ethiopian EFL Students' Writing Performance

Sindu Wondemu Desta[1], Haile Kassahun Bewuket[2], Sualih Mussa Akalu[1]

[1]*Department of English Language and Literature, Debre Markos University, Debre Markos, Amhara, Ethiopia*
[2]*Department of English Language and Literature, Injibara University, Injibara, Amhara Ethiopia*

ARTICLE INFO

Received: 30 March 2026
Revised: 16 June 2026
Accepted: 14 September 2026
Available online: 15 September 2026

KEYWORDS

Academic Writing Performance;
Writing Conferences;
Higher-Order Concerns;
Lower-Order Concerns;
Longitudinal Trajectories;
EFL Instruction;

ABSTRACT

This study evaluated the longitudinal trajectory of higher education EFL students' academic writing across higher-order concerns (Content/Organization) and lower-order conventions during a 16-week structured Writing Conference intervention. A quantitative-dominant embedded mixed-methods (QUAN → qual) single-group quasi-experimental design was deployed using a census population of 47 third-year English majors at an Ethiopian public university. Quantitative data from calibrated parallel essay forms scored via an analytical rubric were modelled using a Multi-Wave Linear Mixed Model (LMM) with Restricted Maximum Likelihood (REML) estimation, adjusting for baseline writing performance (AWPPrM) as a continuous fixed covariate. Qualitative focus groups contextualized these trends. Writing scores increased significantly across the repeated measurement period, with adjusted marginal means rising from 73.84 to 86.32. The non-significant Time × Baseline interaction ($p = .388$) indicated no detectable difference in growth slopes based on initial proficiency tiers. Descriptive sub-construct gains in higher-order domains outpaced conventions. Qualitative themes supported these trends, showing reported shifts toward rhetorical awareness and independent revision strategies. Writing metrics improved progressively across waves. However, because the study utilized a single-cohort framework without a parallel control group, score gains cannot be attributed exclusively to the intervention. The findings provide localized empirical evidence of process-scaffolding viability in resource-constrained higher education settings, warranting future controlled comparative validation.

1. Introduction

Academic writing proficiency is fundamental to success in higher education because it enables students to demonstrate knowledge, develop arguments, and participate in scholarly and professional communities (Hyland, 2019; Tardy, 2016). In contemporary higher education, English academic writing is particularly important as a major medium of international scholarly communication and knowledge exchange (Canagarajah, 2019). This need is especially relevant in the Global South, where expanding higher education systems must prepare graduates for academic and professional contexts requiring advanced literacy skills (Kioko & Muthwii, 2018).

Despite its importance, writing instruction can remain oriented toward the final written product rather than the recursive processes through which texts are planned, developed, reviewed, and revised (Richards & Rodgers, 2014). Such an orientation may restrict opportunities for individualized, process-level support during composition. Dialogic feedback and scaffolding offer an alternative by positioning learners as active participants who discuss writing

Corresponding author: senduwondemu@gmail.com *(Desta S.W.)*

DOI: http://doi.org/10.20372/ajids.2026.3079

problems, clarify intentions, consider alternatives, and make revision decisions. In the present study, dialogic scaffolding refers to adaptive, interaction-based assistance through which instructors and students jointly examine writing problems and progressively work toward their resolution. However, empirical research examining how sustained dialogic support relates to specific dimensions of academic writing across multiple measurement points remains limited.

These concerns are particularly relevant in Ethiopian public higher education. Research has documented difficulties in students' academic writing, including low self-efficacy and formulaic writing practices (Geberew, 2019). Teshome (2020) identified a **"**reproduction mind-set," characterized by reliance on rigid templates and reproduction of memorized material rather than independent argumentation. Research at Mizan Tepi University has also identified vocabulary and sentence-level difficulties while pointing to broader challenges in sustaining writing support (Mesfin & Wondimu, 2024). These findings suggest that students' difficulties involve not only linguistic knowledge but also the instructional conditions under which they develop and revise academic texts. Debre Markos University (DMU) was therefore selected as the research site to examine this issue within an authentic public-university setting.

The provision of individualized writing support is further constrained by institutional conditions. Negash (2022) reports large class sizes exceeding 50 students per section and instructor workloads of approximately 15–20 hours per week. Tedeneke (2022) reported that approximately 90% of instructional feedback in the examined context focused on surface-level error correction rather than substantive engagement with students' ideas. Such conditions may leave students with information about linguistic errors but fewer opportunities to discuss content, organization, argumentation, and rhetorical choices during drafting and revision. The resulting mismatch between the demands of academic writing and available instructional support provides a contextual rationale for examining structured writing conferences.

The intervention draws on three complementary perspectives: Sociocultural Theory (SCT), the Cognitive Process Model of Writing, and the Higher-Order Thinking Skills (HOTS) framework. They are not treated as independently tested causal mechanisms; rather, they explain different dimensions of the writing-conference process.

Sociocultural Theory (SCT) provides the principal account of social mediation. Drawing on Vygotsky (1978) and its application to L2 pedagogy (Lantolf & Poehner, 2014), learning can occur through mediated interaction in which assistance enables learners to perform beyond what they can accomplish independently. In writing conferences, teacher and learner jointly examine writing problems within the learner's Zone of Proximal Development (ZPD). Questioning, prompting, modelling, explanation, and negotiated revision can provide contingent support that is progressively adjusted to learner needs.

The Cognitive Process Model of Writing (Flower & Hayes, 1981) complements this perspective by conceptualizing writing as a recursive process involving planning, translating, reviewing, and revising (Graham, 2018). Conferencing provides opportunities for students to discuss these processes while working on their texts, allowing instructors to support difficulties in planning, developing, organizing, and revising ideas.

Table 1. *Theoretical Mapping of Writing Conference Intervention*

Theoretical perspective	Intervention Component	Expected Writing-Related Outcome
Sociocultural Theory (Vygotsky, 1978; Lantolf & Poehner, 2014)	Dialogic teacher–student conferencing within the ZPD through questioning, prompting, modelling, and negotiated revision	Supported participation in writing problem-solving and gradual development of independent revision strategies
Cognitive Process Model (Flower & Hayes, 1981; Graham, 2018)	Discussion of planning, translating, reviewing, and revising during drafting and revision	Greater engagement with recursive composing processes and overall writing development
HOTS framework (Aslam et al., 2025; Chavda et al., 2024)	Analytical questioning, evaluation of alternatives, and reflective discussion	Greater attention to higher-order writing processes, particularly Content and Organization

The HOTS framework provides a further perspective on analytical and evaluative thinking in writing. Academic writing requires students to analyze information, evaluate alternatives, establish relationships among ideas, and construct arguments. Dialogic questioning and reflection can encourage these processes, particularly in relation to Content and Organization, while Language Use and Conventions remain part of the overall writing assessment (Aslam et al., 2025; Chavda et al., 2024).

Taken together, these perspectives provide an integrated rationale for the intervention. SCT explains how learning may be socially mediated through dialogue; the Cognitive Process Model explains the composing processes on which that dialogue operates; and the HOTS framework explains the analytical and evaluative thinking encouraged through interaction. The study uses these perspectives to guide intervention design and interpretation rather than directly testing ZPD, cognitive load, or HOTS as mediating variables.

International research suggests that interactive feedback and writing conferences can support L2 writing development, but findings are not uniform. Differences concern the writing dimensions affected, learner characteristics, intervention structure, and duration. Three issues are particularly relevant: construct-specific development, development across initial performance levels, and change over time.

Research differs regarding whether conferences produce stronger effects on higher-order than lower-order writing concerns. Leung (2008) and Çepni (2016) emphasized the value of oral interaction for clarification, self-correction, and revision. Afdila et al. (2023), however, reported improvements involving both higher- and lower-order concerns, whereas Kazemi et al. (2025) found particularly strong effects of dialogic interaction on content clarity, content development, cohesion, and coherence. These differences suggest that construct development may depend on how conferencing is structured and which writing problems are emphasized. Thus, whether Content, Organization, Language Use, and Conventions develop in the same way remains an empirical question.

A second issue concerns whether students entering with different levels of writing performance develop similarly. Çepni (2016) suggests that scaffolding can support learners across proficiency levels, whereas Saito (2008) raises concerns that internationally demanding feedback may favour learners with stronger linguistic resources. Average improvement may therefore conceal differences associated with initial performance. Longitudinal analysis that accounts for baseline performance can provide information about this issue beyond a simple pre-test/post-test comparison.

A third issue concerns writing development over time. Chen (2024) reported sustained gains following an 18-week feedback-integrated intervention, including improvement at a delayed post-test. In contrast, Skar et al. (2024), in a large longitudinal study, found no developmental advantage from simply increasing writing opportunities. Together, these findings suggest that writing quantity alone may not explain development and that the nature of instructional support warrants attention. Nevertheless, differences among studies in intervention structure, participants, measures, and observation periods leave uncertainty about how structured dialogic support relates to multi-wave writing development.

The international evidence must be considered alongside Ethiopian conditions. Local research identifies formulaic writing practices, limited confidence, and linguistic and textual difficulties (Geberew, 2019; Teshome, 2020). The reproduction-oriented tendency is particularly problematic because university writing requires students to develop, organize, evaluate, and communicate ideas independently.

Institutional conditions may further restrict individualized support. Negash (2022) reports large classes and substantial teaching workloads, while Tedeneke (2022) reported that approximately 90% of feedback in the examined context focused on surface-level error correction. These conditions create a tension between the complex demands of academic writing and the opportunities available for sustained interaction around developing texts. Legesse and Tegegne (2025) further indicate that writing difficulties may affect professional activities requiring proposal and report writing.

The convergence of international and Ethiopian evidence indicates a connected instructional and empirical problem. Students are expected to develop writing involving content, organization, language use, and conventional accuracy, yet Ethiopian evidence suggests that feedback may remain predominantly corrective and surface-oriented while

institutional conditions constrain sustained individualized interaction. International research indicates that dialogic conferencing may offer a useful alternative, but important questions remain.

First, Ethiopian EFL writing support remains strongly oriented toward mechanical features of language, potentially limiting sustained engagement with substantive writing decisions (Geberew, 2019; Tedeneke, 2022). This is related to the reproduction-oriented tendency identified by Teshome (2020), yet relatively little intervention research has examined whether sustained dialogic interaction can provide an alternative environment for students to reconsider and develop their writing decisions.

Second, more longitudinal evidence is needed concerning writing development during sustained feedback interventions. Much research examines immediate or short-term outcomes, leaving uncertainty about how writing performance changes across multiple measurement points during an extended intervention (Chen, 2024; Skar et al., 2024).

Third, overall writing scores may conceal differences among writing dimensions. There is therefore a need for evidence showing how Content, Organization, Language Use, and Conventions develop when examined through aggregated pre- and post-intervention construct measures.

Fourth, differences in students' initial writing performance may confound interpretation of longitudinal change. Students do not begin an intervention with identical writing ability. In the present study, aggregated pre-intervention writing performance (AWPPrM**)** is therefore entered as a continuous covariate in the LMM to statistically control for initial differences and reduce their potential confounding influence when estimating the overall longitudinal trajectory. The associated Time × AWPPrM interaction additionally examines whether the observed longitudinal pattern varies systematically with baseline performance. This analysis does not treat baseline performance as an outcome or establish equivalent developmental effects across proficiency levels.

These gaps are interconnected rather than independent. A predominantly corrective feedback environment may limit sustained discussion of substantive writing decisions; without repeated measurement, developmental change cannot be examined across time; without construct-level analysis, overall change may conceal differences among writing dimensions; and without statistical control for baseline variation, differences in initial writing performance may complicate interpretation of the longitudinal trajectory. The present study brings these dimensions together in a longitudinal investigation of structured Writing Conferences.

Why the Proposed Intervention Is an Appropriate Response?

Table 2 Alignments between Identified problems and intervention Response.

Identified challenge	How the intervention responds
Predominantly mechanical feedback	Dialogic scaffolding creates opportunities to discuss substantive as well as surface-level writing problems during drafting and revision.
Reproduction-oriented writing	Guided questioning and evaluation of alternatives encourage students to articulate, reconsider, and justify writing decisions.
Episodic opportunities for writing support	Repeated conferencing across the 16-week intervention provides sustained engagement through recurring cycles of dialogue, feedback, revision, and reflection.
Limited evidence on construct-specific development	Construct-level pre/post analysis compares aggregated pre-intervention and post-intervention means for Content, Organization, Language Use, and Conventions.
Potential confounding influence of initial writing performance	AWPPrM is entered as a continuous covariate in the LMM to control statistically for initial performance differences when estimating the overall longitudinal trajectory; the Time × AWPPrM interaction examines whether the pattern varies with baseline performance.

The intervention therefore does not assume that writing conferences automatically resolve these problems. Rather, it provides a structured context in which they can be examined. Repeated conferencing addresses the need for sustained interaction, the dialogic structure moves feedback beyond one-way correction, construct-level

measurement permits examination of writing dimensions, and the LMM accounts statistically for baseline variation in estimating longitudinal change.

The present study investigated the longitudinal development of academic writing performance among higher education EFL students participating in a structured Writing Conference intervention at an Ethiopian public university. It examined overall writing performance across repeated measurement points and construct-level development in Content, Organization, Language Use, and Conventions.

Accordingly, the study pursued the following objectives:

1. To examine the continuous, multi-wave developmental trajectory of students' overall writing performance over consecutive measurement points during the Writing Conference intervention.

2. To measure the overall magnitude of change in student writing performance by comparing the aggregated pre-intervention mean with the aggregated post-intervention mean.

3. To compare the differential impact of the intervention on higher-order concerns (Content and Organization) versus lower-order concerns (Conventions**).**

Grounded in these objectives, the study was guided by the following research questions:

RQ1: What is the continuous, multi-wave developmental trajectory of students' overall writing performance over consecutive measurement points during the Writing Conference intervention?

RQ2: To what extent does student writing performance change between the aggregated pre-intervention mean and the aggregated post-intervention mean following the Writing Conference intervention?

RQ3: Is there a significant difference in the magnitude of improvement between higher-order concerns (Content and Organization) and lower-order concerns (Conventions)?

2. METHODOLOGY

2.1. Research Design

This study employed an embedded mixed-methods design with quantitative dominance (QUAN → qual) within a single-group quasi-experimental longitudinal framework. The design was selected because the intervention was implemented within an intact university course without a parallel comparison group. Quantitative data provided the primary basis for examining writing development across repeated measurement points, while qualitative Focus Group Discussions (FGDs) provided contextual information about students' experiences and reported revision practices.

The intervention lasted 16 weeks. Four pre-intervention writing assessments were aggregated to establish **AWPPrM**, a stable continuous baseline measure. This aggregation reduced the influence of occasion-specific fluctuations and was subsequently used as a covariate in the LMM.

2.2. Participants and Setting

The study involved a complete cohort of 47 third-year English majors at Debre Markos University. The intervention was embedded in regular university instruction rather than delivered as a separate experimental course.

2.3. Writing Conference Intervention

The intervention followed a structured six-phase protocol implemented through repeated conferencing and writing activities. The phases involved preparation, individual or peer examination of texts, peer comparison, strategic discussion, rubric-oriented peer feedback, and revision planning/closure.

The intervention was delivered by the course instructor following a standardized session manual. Before implementation, the instructor completed 4.5 hours of preparation over three days covering the intervention protocol, dialogic questioning, and procedures for maintaining consistency. Pilot activities were used to refine phase boundaries.

Dialogic feedback was defined as reciprocal discussion in which students interpreted and responded to feedback rather than simply receiving corrections. Dialogic scaffolding involved contingent questioning, prompting, modelling, explanation, and graduated assistance. The intervention prioritized substantive writing concerns before surface-level conventions, while still addressing Language Use and Conventions.

Peer feedback was structured using analytical criteria rather than unrestricted discussion. Students identified strengths, offered specific suggestions, and articulated revision responses.

2.3.1 Intervention Fidelity

Fidelity was monitored through instructor self-checklists and independent observation. The instructor completed a 12-item checklist after each session, covering phase completion, timing, dialogic questioning, attention to higher-order concerns, and revision planning. An independent EFL colleague audited 22% of contact hours. Adherence scores were 79.2% in the early session and 78.6% in the later session, with the later variation reflecting increased student-led participation.

2.4. Data Collection Instruments

2.4.1 Writing Performance Tests

Writing performance was assessed using researcher-developed genre-specific parallel writing tests aligned with the Intermediate Writing Skills course. Texts were scored on a 100-point analytical rubric:

- Content: 30 points
- Organization: 30 points
- Language Use: 25 points
- Conventions: 15 points

Six external experts reviewed the assessment prompts against a table of specifications. Parallel forms were calibrated for comparable complexity and administered under standardized 100-minute conditions. Four pre-intervention writing scores were averaged to create AWPPrM.

Pilot exploratory factor analysis used Principal Axis Factoring with Promax rotation. KMO values ranged from .518 to .725 and Bartlett's tests were significant ($p < .001$).

2.4.2 Scoring and Inter-Rater Reliability

The 376 writing scripts were scored independently by two EFL instructors following rubric calibration. Scripts were anonymized and randomized, and raters were blinded to measurement wave and initial-performance grouping. Discrepancies exceeding five points were resolved through adjudication.

Table 3 *Component Specific Interclass Correlations*

Scoring dimension	ICC
Overall analytical rubric	.94
Content	.89
Organization	.87
Language Use	.84
Conventions	.78

Overall inter-rater reliability was high, with Single Measures ICC values ranging from .923 to 1.000 and Average Measures ICC values from .960 to 1.000 ($p < .001$). Construct-specific ICCs were .89 for Content, .87 for Organization, .84 for Language Use, and .78 for Conventions.

2.4.3 Focus Group Discussions

Three semi-structured FGDs were conducted after the intervention with 21 purposively selected participants representing different baseline performance levels. Discussions addressed experiences of conferencing, perceived writing changes, teacher and peer feedback, and revision practices. Sessions followed a common protocol, were externally moderated, recorded, and transcribed verbatim.

2.5. Data Analysis

2.5.1 Quantitative Analysis

Descriptive statistics were calculated for all measurement occasions. Subgroup trajectories were illustrated using the baseline median of 43.13 as a descriptive threshold. These groups were not treated as experimental conditions and were not used to make causal claims.

The primary inferential analysis was a Linear Mixed Model (LMM) estimated using Restricted Maximum Likelihood (REML). Time was specified as a fixed factor and participant ID as a random intercept. AWPPrM, the mean of the four pre-intervention assessments, was entered as a continuous fixed covariate to control for initial differences in writing performance when estimating longitudinal change. The Time × AWPPrM interaction examined whether the longitudinal pattern varied systematically according to baseline performance.

Estimated marginal means and Bonferroni-adjusted pairwise comparisons were used to examine differences among measurement points. Six pairwise comparisons were evaluated using an adjusted alpha of .008.

A paired-samples *t*- test was used as a supplementary analysis to compare the aggregated pre-intervention writing mean with the aggregated post-intervention mean and to estimate the standardized magnitude of the overall pre-post difference. Construct-level analyses compared corresponding aggregated pre- and post-intervention means for Content, Organization, Language Use, and Conventions.

As an a priori G*Power analysis indicated a minimum sample of 34 participants under the specified assumptions, the final sample of 47 participants was considered adequate. All participants completed the study, so no missing-data imputation was required.

2.5.2 Qualitative Analysis

FGD data were analysed using Braun and Clarke's (2006) six-phase thematic analysis, involving familiarization with the transcripts, initial coding of meaningful segments, development and review of themes, definition of themes, and reporting. The transcripts were repeatedly reviewed to ensure that emerging themes reflected participants' accounts rather than isolated statements, while divergent accounts were considered where present. **Member checking with five participants and peer debriefing** further supported the trustworthiness of the interpretation, and comparison with the quantitative findings facilitated integration across the two data sources. Qualitative findings were interpreted as **contextual evidence of students' reported experiences and perceptions,** rather than as causal confirmation of the quantitative effects.

2.6. Ethical Considerations

The study received approval from the Debre Markos University Institutional Review Board (DMU IRB 078/2024; 10 September 2024). Participants provided written informed consent and were informed of voluntary participation and withdrawal rights. Data were anonymized using participant codes and securely stored.

3. RESULTS

3.1. Longitudinal Changes in Overall Academic Writing Performance

The primary objective of this analysis was to examine changes in students' academic writing performance across the four successive measurement points following the Writing Conference intervention. In accordance with Research Objective 1, the analysis focused on whether performance changed systematically across time and whether the pattern of change varied according to students' baseline performance. Because the study employed a single-group longitudinal design without a parallel comparison group, the results are interpreted as evidence of observed longitudinal change rather than definitive evidence of an intervention-specific causal effect.

3.1.1. Descriptive Development Across Measurement Points

Table 4 presents the descriptive statistics for academic writing performance across the baseline and four successive measurement points. The observed mean increased from 43.54 at baseline to 73.84 at Time 1, 78.16 at Time 2, 82.20 at Time 3, and 86.32 at Time 4. The corresponding medians increased from 43.13 at baseline to 73.00, 78.00, 82.00, and 86.00, respectively. The standard deviations at the four post-intervention measurement points remained relatively narrow, ranging from 2.96 to 3.14, indicating limited dispersion around the respective means.

Table 4. Descriptive Statistics of Academic Writing Performance Across Measurement Points.

Measurement point	M	SD	Median	Range
Pre-test (baseline)	43.54	3.50	43.13	12.63
Post-test 1	73.84	3.14	73.00	11.00
Post-test 2	78.16	3.00	78.00	10.00
Post-test 3	82.20	3.10	82.00	12.00
Post-test 4	86.32	2.96	86.00	10.00

The descriptive pattern indicates a sustained increase across the four measurement waves. The successive mean gains were 4.32 points from Time 1 to Time 2, 4.04 points from Time 2 to Time 3, and 4.12 points from Time 3 to Time 4. Thus, the descriptive evidence shows a relatively regular upward trajectory rather than a single isolated increase. The close correspondence between the means and medians at each measurement point further indicates that the observed progression was not driven solely by a small number of unusually high scores.

These descriptive findings establish the observed pattern but do not, by themselves, establish that the Writing Conference intervention caused the change. The longitudinal inferential analysis below therefore provides the principal test of whether performance differed significantly across measurement points.

3.1.2. Descriptive Performance Trajectories by Initial Performance Level

In relation to Research Objective 1, which concerned the performance trajectories of students with different initial levels of writing performance, students were descriptively divided using the baseline median of 43.13. Students scoring below 43.13 were classified as the Low-Initial group, whereas those scoring 43.13 or above were classified as the High-Initial group.

The subgroup classification was used only to visualize and describe performance trajectories. It was not treated as an independent experimental condition, and no causal or developmental conclusion was drawn from the subgroup comparison.

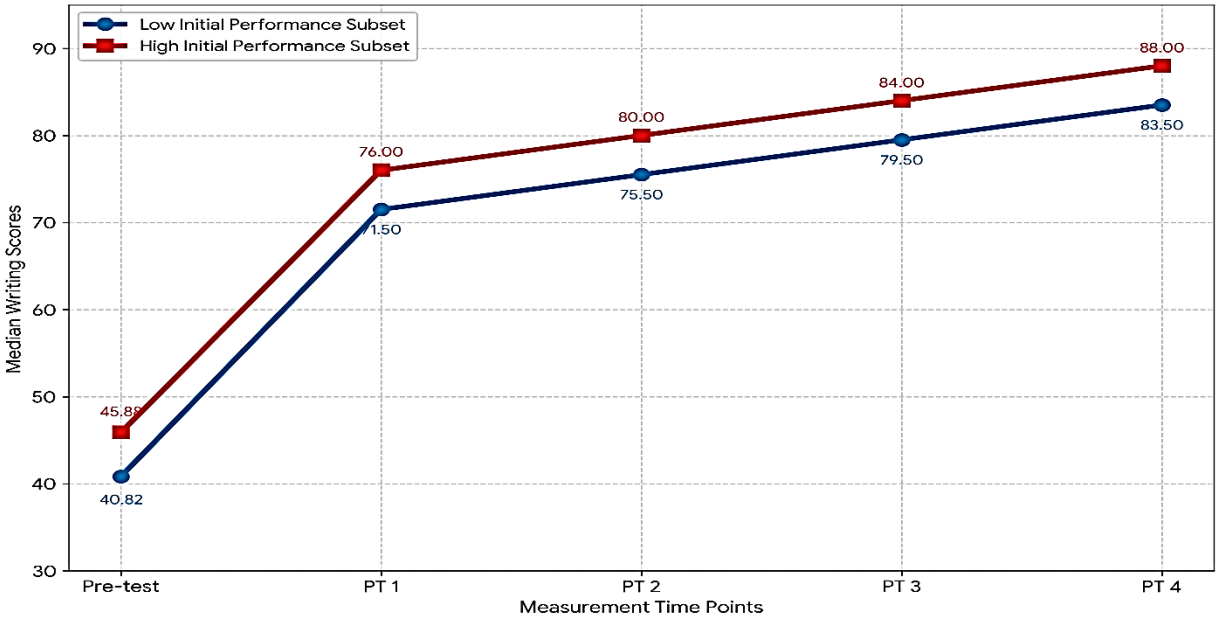


Figure 1. Median Writing Performance Trajectories by Initial Performance Group.

The descriptive trajectories show continued improvement across the four measurement points for students entering the intervention at different levels of baseline performance. The trajectories therefore provide a visual indication that improvement was observed across the range of initial performance represented in the data.

However, descriptive subgroup trajectories cannot establish whether the rate of improvement differed statistically between students with lower and higher initial performance. This question was therefore examined through the Time × baseline-performance interaction in the LMM.

3.1.3. Linear Mixed Model Analysis

To examine longitudinal change while accounting for the repeated observations within students, a Linear Mixed Model (LMM) was estimated using Restricted Maximum Likelihood (REML). Time was specified as a fixed effect, the aggregated baseline pre-test mean (AWPPrM) was entered as a continuous covariate, and a Time × AWPPrM interaction was included to determine whether the longitudinal pattern varied according to baseline performance. Participant ID was specified as a random intercept.

Fixed effects

The LMM showed a statistically significant main effect of Time, $F(3, 28.222) = 25.034$, $p < .001$. Baseline writing performance was also statistically significant, $F(1, 21.577) = 620.466$, $p < .001$. In contrast, the Time × AWPPrM interaction was not statistically significant, $F(3, 28.222) = 1.045$, $p = .388$.

Thus, the model indicates statistically significant differences in academic writing performance across the measurement points after accounting for baseline performance. At the same time, the non-significant interaction provides no statistical evidence that the rate of longitudinal change differed systematically according to students' baseline performance.

The parameter estimates provide further information about the direction of the time effect. With Time 4 as the reference category, the coefficients for Time 1 ($b = −14.932$, $p < .001$), Time 2 ($b = −8.249$, $p < .001$), and Time 3 (b

= −6.204, p = .001) were all negative. This indicates that predicted performance at each earlier measurement point was significantly lower than at Time 4.

Table 5 Test of Fixed Effect for Academic Writing Performance

Source	Df	F	P
Intercept	1, 21.577	846.706	< .001
Time	3, 28.222	25.034	< .001
AWPPrM	1, 21.577	620.466	< .001
Time × AWPPrM	3, 28.222	1.045	.388

The baseline-performance coefficient was positive and statistically significant (b = 0.821, p < .001), indicating a strong association between initial performance and subsequent performance levels. Importantly, the interaction terms were not statistically significant. Accordingly, higher baseline performance was associated with higher subsequent performance, but the model did not detect a statistically significant difference in the rate of change across baseline-performance levels.

Table 6 Estimate of Fixed Effect for academic writing performance

Parameter	Estimate	*SE*	*Df*	*t*	*P*
Intercept	50.580	1.744	33.009	29.007	< .001
Time 1	−14.932	1.742	29.915	−8.572	< .001
Time 2	−8.249	1.847	24.442	−4.465	< .001
Time 3	−6.204	1.654	24.330	−3.752	.001
Time 4	0.000	—	—	—	—
AWPPrM	0.821	0.040	33.009	20.560	< .001
Time 1 × AWPPrM	0.056	0.040	29.915	1.412	.168
Time 2 × AWPPrM	0.002	0.042	24.442	0.048	.962
Time 3 × AWPPrM	0.048	0.038	24.330	1.264	.218

Note. Time 4 was the reference category.

3.1.4. Estimated Marginal Means and Pairwise Progressions

Estimated marginal means were evaluated to examine the absolute magnitude of the longitudinal shifts while fixing the continuous baseline covariate at its total sample mean value of 43.54. The adjusted grand mean calculated across all post-intervention observation windows was 80.13 (*SE* = 0.117, 95% CI [79.888, 80.372]). To deliver complete empirical transparency and integrate standard sequential tracking models, Table 7 unifies the baseline marginal profiles alongside the Bonferroni-adjusted post-hoc pairwise matrix (6 structural contrasts, α~adjusted~ = .008).

Table 7. Integrated Estimated Marginal Mean and Bonferroni Adjusted Marks.

Wave Window	Marginal Mean (*M*)	Std. Error (*SE*)	Denominator *df*	95% Confidence Interval [CI]	Comparative Contrast Pair	Mean Difference (Δ)	Pairwise Contrast Sig. (p)
Time 1	73.840	0.148	28.079	[73.536, 74.144]	Time 1 − 2	−4.320	< .001
Time 2	78.160	0.156	34.874	[77.843, 78.477]	Time 2 − 3	−4.040	< .001
Time 3	82.200	0.142	23.089	[81.906, 82.494]	Time 3 − 4	−4.120	< .001
Time 4	86.320	0.137	33.009	[86.042, 86.598]	Time 1 − 4	−12.480	< .001

Note. Core macro baseline contrasts across intermediate waves indicate matching significance profiles: Time 1 − Time 3 (Δ = −8.360, *df* = 33.537, p < .001) and Time 2 − Time 4 (Δ = −8.160, *df* = 24.442, p < .001).

The integrated data parameters reveal highly significant adjustments between every single successive measurement point, with all pairwise comparisons yielding p < .001. Performance increased by 4.32 points from Time 1 to Time 2, by 4.04 points from Time 2 to Time 3, and by 4.12 points from Time 3 to Time 4. The total cumulative adjusted displacement achieved across the 16-week timeline between Time 1 and Time 4 was 12.48 points. These sequential

steps confirm that score development was continuous and incremental across waves rather than an isolated increase between front-end and terminal checkpoints. Figure 2 provides a visual representation of these adjusted marginal paths and step-wise adjustments.

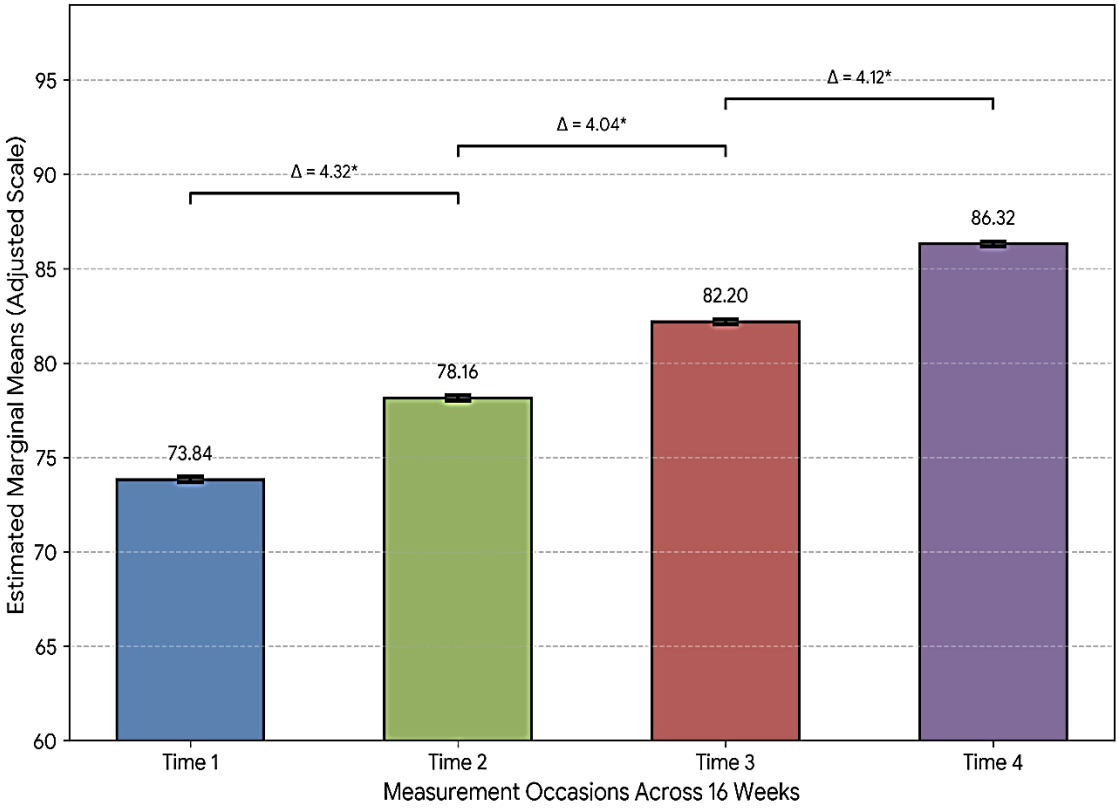


Figure 2 Marginal Mean Wave Progressions and Bonferroni Brackets

3.1.5. Supplementary Pre-Post Analysis

As a supplementary assessment of overall pre-post change, a paired-samples t-test was conducted comparing the aggregated baseline pre-test (M = 43.54, SD = 3.50) with the aggregated post-test composite (M = 68.20, SD = 4.27). The post-test composite was calculated by first averaging each student's four post-intervention writing assessments (PT1, PT2, PT3, PT4) and then averaging these student-level composites across the cohort. This aggregation procedure reduces measurement error and provides a stable estimate of overall post-intervention performance.

The mean difference was 24.66 points, 95% CI [22.42, 26.90], t (46) = -22.84, $p < .001$. Cohen's d was -4.57, 95% CI [-5.25, -3.89], with the negative sign resulting from the calculation order (pre-test minus post-test). In magnitude, this corresponds to d = 4.57. Hedges' corrected estimate was 4.42. Given the single-group design, this very large standardized pre-post effect should be interpreted as evidence of substantial observed change rather than as definitive evidence of a causal intervention effect.

Because this t-test represents a single pre-post comparison (aggregated baseline vs aggregated post-test composite), the standard $\alpha = .05$ was retained without adjustment. This test was not used to draw conclusions about longitudinal trajectories; rather, it served exclusively as a global magnitude check, complementing the LMM analysis which provided the primary inferential test of change across measurement waves.

3.1.6. Consistency of Change Across Initial Performance Levels

The findings related to Research Objective 3 were considered by integrating the descriptive subgroup trajectories with the inferential Time × AWPPrM interaction.

The descriptive median trajectories indicated continued improvement among students classified as Low-Initial and High-Initial according to the baseline median of 43.13. However, the LMM provided the appropriate inferential test of whether the rate of change differed according to baseline performance. The Time × AWPPrM interaction was not statistically significant, $F(3, 28.222) = 1.045$, $p = .388$.

Accordingly, the present analysis found no statistically significant evidence of differential longitudinal change associated with baseline performance. This finding should not be interpreted as demonstrating identical growth among all students, nor as evidence of a compensatory "catch-up" mechanism. It indicates only that the model did not detect a statistically significant difference in the rate of change as a function of baseline performance.

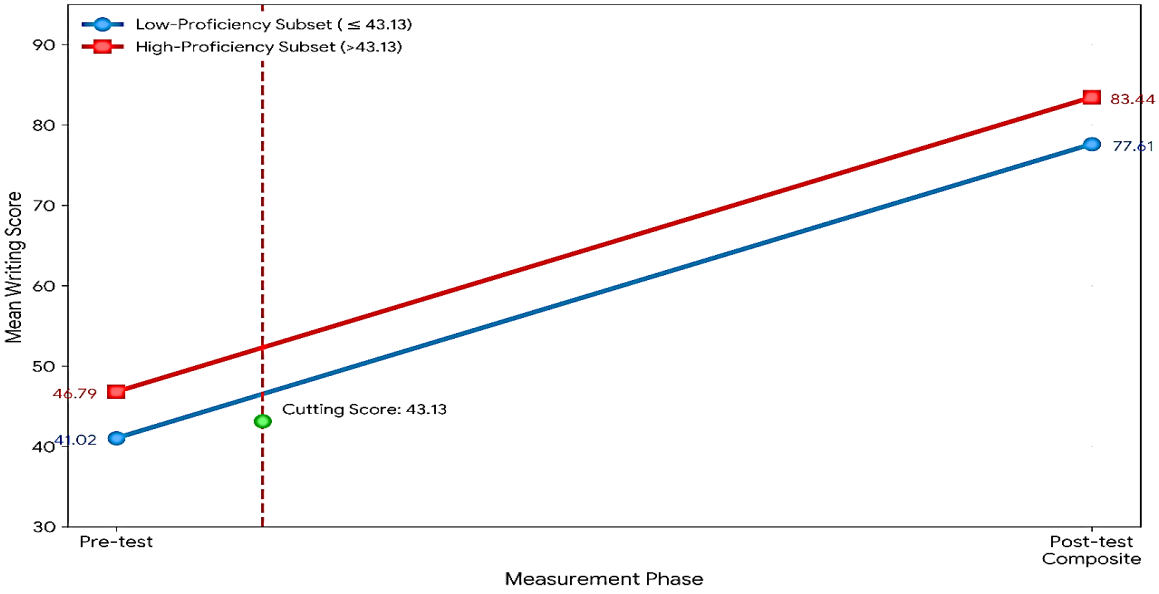


Figure 3 Descriptive Performance Trajectories by Initial Performance Level

The descriptive trajectories and the non-significant interaction therefore provide complementary evidence. The former illustrates the observed pattern across the two descriptively defined starting-level groups, whereas the latter provides the statistical test of differential change. This distinction is important because the subgroup classification was exploratory and descriptive rather than an experimental grouping variable.

3.2. Construct-Specific Performance Trends

The analysis of the writing constructs indicated that the magnitude of observed change was not uniform across the assessed dimensions. Content and Organization, representing higher-order aspects of writing, showed larger observed gains than the lower-order Conventions dimension, while Language Use occupied an intermediate position.

The original construct-level analysis reported the following changes:

- **Content:** +8.53, 68.0%, $d = 6.16$
- **Organization:** +6.98, 68.7%, $d = 4.95$
- **Language Use:** +6.69, 44.9%, $d = 4.38$
- **Conventions:** +2.47, 42.1%, $d = 1.94$

These results suggest a descriptive hierarchy in which the largest observed changes occurred in higher-order dimensions, particularly Content and Organization. However, the standardized effect sizes are extremely large, and therefore should not be interpreted as independent causal effects of the Writing Conference intervention. They represent standardized within-group differences and are especially sensitive to the variability of the paired scores.

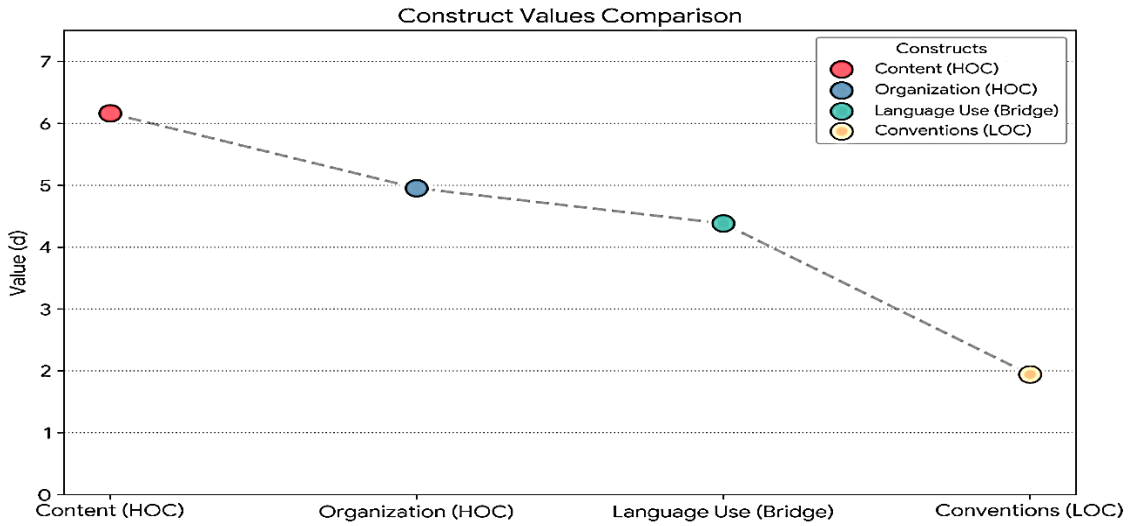


Figure 4 Observed Construct Level Change In Academic Writing Performance

The construct-level pattern is theoretically compatible with the intervention's emphasis on dialogic attention to meaning, argumentation, organization, and revision before surface-level correction. However, the present design cannot establish that the intervention itself caused the differential magnitude of change across constructs. Alternative explanations, including differential construct difficulty, repeated exposure to writing tasks, maturation, and instructional practice, cannot be completely excluded.

Thus, the construct-level findings are best interpreted as evidence of differential observed development across writing dimensions, rather than proof of a hierarchical causal mechanism.

3.3. Qualitative Findings

The thematic analysis of the Focus Group Discussions (FGDs) provided contextual insight into how students experienced the Writing Conference intervention and how they perceived changes in their writing practices. Two prominent themes emerged from the discussions: (1) enhanced rhetorical awareness and (2) developing revision strategies. The qualitative accounts generally supported the direction of the quantitative findings, particularly the stronger development observed in higher-order writing dimensions. However, these accounts are interpreted as participants' reported experiences rather than as independent evidence of intervention effects or developmental mechanisms.

Theme 1: Enhanced Rhetorical Awareness

Participants described a shift from a predominantly sentence-level orientation toward greater attention to meaning, argumentation, and overall text organization. One participant explained, *"I think about my ideas first... good writing is not just correct, it is clear communication"* (FGD2, Female, 21). Another reported, *"Before, I only checked grammar. Now I ask: Does my argument make sense? Is my evidence strong?"* (FGD1, Male, 22).

These accounts are consistent with the quantitative pattern in which Content and Organization showed comparatively pronounced gains across the measurement period. The reported shift toward considering ideas, evidence, and argument structure also corresponds with the intervention's emphasis on Higher-Order Concerns during dialogic conferencing. Nevertheless, the qualitative evidence indicates students' perceptions of their

changing writing practices; it does not establish that the intervention alone produced these changes. Other influences associated with continued course participation and repeated writing practice cannot be excluded within the single-group design.

Theme 2: Developing Revision Strategies

A second theme concerned changes in how participants approached revision. Rather than describing revision exclusively as correction of teacher-marked errors, several participants reported becoming more attentive to the relationship between claims, evidence, paragraph development, and overall clarity. One student stated, *"I ask myself the questions my teacher asked me... I check if evidence supports my claim"* (FGD3, Female, 23). Another explained, *"I used to just fix the sentences the teacher marked. Now I rewrite whole paragraphs if the idea is not clear"* (FGD2, Male, 20).

These reports provide qualitative context for the observed longitudinal improvement and, in particular, for the comparatively greater gains in higher-order dimensions. They are also compatible with the intervention's dialogic scaffolding sequence, in which questioning and guided reflection were intended to shift some responsibility for evaluating and revising texts toward the learner. However, the use of terms such as **"**internalization" or "self-regulation**"** should remain cautious because the FGD data document students' accounts of their practices rather than directly measuring the internal cognitive processes underlying those practices.

Variability and Constraints in Students' Experiences

The qualitative findings should not be interpreted as indicating that the conference process was uniformly experienced as beneficial or equally accessible to all students. Dialogic writing activities may involve practical and pedagogical challenges, including differences in students' willingness to participate, difficulties in interpreting peer feedback, dependence on teacher prompting, and the additional time required for extended discussion and revision. Where such concerns were expressed in the FGDs, they should be treated as part of the findings rather than excluded because they do not support the intervention.

This qualification is important for interpreting the generally positive accounts reported above. The presence of positive perceptions does not demonstrate that all students experienced the intervention in the same way, nor does the absence of statistically significant differential change by baseline performance establish equivalent individual benefit. Rather, the quantitative interaction result, F (3, 28.222) = 1.045, p = .388, indicates that the LMM did not detect a statistically significant difference in the rate of change associated with baseline writing performance. The qualitative evidence can therefore be used to illuminate students' reported experiences of the intervention, but not to establish uniformity of experience or a compensatory developmental mechanism.

3.4. Integration of Qualitative and Quantitative Findings

Taken together, the quantitative and qualitative findings show a broadly convergent pattern within the study cohort. Quantitatively, overall writing performance increased across the four measurement waves, with significant differences between measurement points. At the construct level, the descriptive pre-post results indicated greater gains in higher-order dimensions than in lower-order conventions. Qualitatively, participants reported increased attention to rhetorical meaning, argumentation, evidence, and substantive revision, providing contextual support for this pattern.

The qualitative findings therefore provide explanatory context rather than causal confirmation. Participants' accounts suggest that dialogic questioning, peer interaction, and guided revision may have redirected attention from isolated surface corrections toward broader rhetorical considerations. This interpretation is consistent with the theoretical rationale of the intervention, but it remains provisional because the study did not experimentally isolate these processes and the qualitative evidence was based on participant reports.

The convergence of the two datasets should nevertheless be interpreted cautiously. The LMM demonstrates change across measurement waves but does not establish that the Writing Conference intervention alone caused that change. Similarly, the large supplementary pre-post standardized difference indicates substantial observed change within the

cohort rather than a definitive intervention effect. Without a comparison group, maturation, repeated exposure to writing tasks, assessment familiarity, and other concurrent instructional influences cannot be ruled out.

Overall, the integrated findings suggest that participation in the structured conference environment was associated with sustained improvement in overall writing performance and with participants' reported shifts toward more rhetorical and substantive approaches to revision. The qualitative evidence strengthens the interpretation by illustrating how students perceived changes in their writing practices, while the quantitative evidence documents the observed pattern of performance change. Future controlled longitudinal studies are needed to determine the extent to which these changes can be attributed specifically to writing conferences.

4. DISCUSSION

This study examined academic writing development during a 16-week structured Writing Conference intervention. The findings are interpreted as observed longitudinal change within a single cohort, not as definitive evidence of intervention-specific causation.

4.1. Overall Longitudinal Development

Writing performance increased progressively across the four measurement waves. The LMM showed a significant main effect of Time, $F(3, 28.222) = 25.034$, $p < .001$. The supplementary paired-samples analysis also showed a significant difference between the aggregated pre-intervention and post-intervention means, $t(46) = -22.84$, $p < .001$, with $d = 4.57$.

These analyses address different aspects of change. The LMM evaluates the longitudinal pattern across repeated measurement points while accounting for within-student dependence and baseline performance. The paired-samples *t* test summarizes the overall aggregated pre-to-post difference. The large standardized effect should therefore be interpreted as a within-group observed magnitude, not as an isolated estimate of causal intervention effectiveness.

The upward trajectory is broadly consistent with research emphasizing interactive and process-oriented feedback (Leung, 2008; Çepni, 2016; Chen, 2024; Kazemi et al., 2025). However, cross-study comparisons should be made cautiously because participant characteristics, intervention structures, assessment procedures, and research designs differ substantially.

Most importantly, the absence of a comparison group means that maturation, testing familiarity, regression toward the mean, concurrent language learning, and other time-related influences cannot be separated statistically from intervention-related change. The results therefore demonstrate substantial observed development during the study period rather than definitive causal effects of the Writing Conference intervention.

4.2. Construct-Level Development

The construct-level findings indicate comparatively greater improvement in Content and Organization than in Conventions. This pattern is consistent with the intervention's emphasis on discussing meaning, argument development, organization, and rhetorical decisions before addressing surface-level errors.

The pattern corresponds with research suggesting that dialogic feedback can support higher-order writing concerns (Kazemi et al., 2025), while also accommodating findings showing improvement in both higher- and lower-order dimensions (Afdila et al., 2023). The present findings therefore do not establish that conferences inherently produce a fixed hierarchy of effects. Rather, the observed construct pattern may reflect the priorities and structure of the intervention.

The results should also be distinguished from the theoretical claim that higher-order gains were caused by reduced cognitive load or internalized HOTS processes. Those mechanisms were not directly measured. The construct-level findings instead provide empirical evidence of different observed changes across the assessed writing dimensions.

4.3. Interpretation of Baseline Performance and the Time × AWPPrM Interaction

Baseline writing performance was statistically significant in the LMM, reflecting substantial variation in students' initial writing levels. Including AWPPrM as a covariate allowed the longitudinal Time effect to be estimated while statistically controlling for this baseline variation.

The Time × AWPPrM interaction was not significant, $F(3, 28.222) = 1.045$, $p = .388$. Thus, within this cohort, the model did not detect statistically significant variation in the longitudinal pattern associated with baseline performance.

This finding should **not** be interpreted as proof that students of different proficiency levels experienced identical or equivalent development. Nor does it demonstrate that the intervention equalized performance differences. Given the sample size and single-group design, the result is better understood as an absence of detectable moderation by baseline performance within the observed cohort.

The finding is nevertheless relevant to the proposed psychological leveller concept. It provides a preliminary empirical basis for investigating whether structured dialogic support can be associated with similar longitudinal patterns across different starting levels, but the leveller proposition remains a hypothesis requiring stronger testing with larger samples, formal comparison groups, and more direct moderation analyses.

4.4. Theoretical Interpretation of the Findings

The findings are broadly consistent with the three complementary perspectives underpinning the intervention. From a Sociocultural Theory perspective, the observed improvement accords with evidence that dialogic feedback can support writing development through interaction, clarification, and negotiated revision (Hyland & Hyland, 2006; Alexander, 2020). The findings also align with the Cognitive Process Model, as conferencing created opportunities for learners to plan, review, and revise their texts through recursive engagement with writing (Flower & Hayes, 1981; Graham, 2018). Finally, the stronger descriptive gains in Content and Organization are compatible with the HOTS perspective, given evidence that dialogic interaction can promote attention to meaning, evaluation, and rhetorical decisions (Alexander, 2020; Aslam et al., 2025). Thus, the quantitative and qualitative patterns provide empirical support consistent with the theoretical rationale, but do not establish these processes as causal mechanisms because they were not independently measured or experimentally tested.

4.5. Dialogic Feedback and Scaffolding

The intervention also addresses the distinction between providing feedback and enabling students to use feedback. Traditional corrective approaches can identify errors, whereas dialogic conferencing allows learners to question, explain, evaluate, and respond to feedback. The present protocol operationalized this orientation through structured teacher-student and peer interaction.

The findings are consistent with previous research emphasizing dialogue, clarification, and active revision (Çepni, 2016; Leung, 2008). They also reinforce the point that interaction alone does not guarantee effective feedback. Its value depends on the quality of questioning, participants' evaluative knowledge, instructional time, and the structure provided for peer interaction.

This issue is particularly important in resource-constrained settings. The present study demonstrates that structured conferencing can be implemented within the participating course, but it does not establish large-scale institutional scalability. Workload, staffing, class size, training, and cost require separate investigation.

4.6. Mixed-Methods Interpretation

The quantitative and qualitative findings show broad descriptive convergence. Quantitative analyses demonstrated progressive writing development, while FGD accounts indicated reported changes in rhetorical awareness and revision practices.

This convergence strengthens the contextual interpretation of the findings but does not constitute independent causal confirmation. The quantitative data provide evidence of observed writing change, whereas the qualitative data illuminate students' reported experiences of the instructional process.

5. CONCLUSION AND IMPLICATIONS

5.1. Summary of Findings

The study examined the development of academic writing performance during a 16-week structured Writing Conference intervention among Ethiopian undergraduate EFL students. Overall writing performance showed a significant longitudinal increase across the four measurement waves. The aggregated pre-intervention and post-intervention comparisons also indicated substantial overall improvement. At the construct level, the observed gains were not uniform, with higher-order dimensions, particularly Content and Organization, showing stronger improvement than Conventions. The qualitative findings further indicated that participants perceived greater attention to argumentation, rhetorical meaning, evidence, and substantive revision during the conferencing process.

The longitudinal analysis additionally accounted for differences in students' initial writing performance by incorporating aggregated baseline performance (AWPPrM) as a covariate in the LMM. The findings therefore provide evidence of substantial observed development in writing performance over the study period while controlling statistically for baseline differences, but they do not establish that the Writing Conference intervention alone caused the observed changes. The single-group design requires this interpretation to remain cautious.

5.2. Conclusion

The study documented substantial improvement in academic writing performance across a 16-week period of structured Writing Conference implementation. Overall performance increased progressively across four measurement waves, and the LMM showed a significant effect of Time. The supplementary aggregated pre-post analysis also indicated a substantial within-group change, while construct-level results showed comparatively greater development in Content and Organization than in Conventions. FGD findings provided complementary accounts of increased attention to rhetorical and revision processes.

These findings are consistent with research emphasizing interactive feedback, dialogic mediation, and process-oriented writing instruction. They also extend the literature by providing longitudinal evidence from an Ethiopian public-university context.

However, the single-group design prevents the observed gains from being attributed exclusively to the Writing Conference intervention. Similarly, the non-significant Time × AWPPrM interaction indicates only that the model did not detect statistically significant variation in longitudinal change associated with baseline performance within this cohort; it does not establish equivalent development or a levelling mechanism.

The principal contribution is therefore localized evidence of substantial writing development during a structured dialogic intervention, together with an empirically specified approach for examining longitudinal and construct-level writing change. Stronger causal conclusions require controlled, multi-site, and longer-term research.

5.3. Theoretical and Empirical Contributions

The study contributes to L2 writing research by examining structured writing conferencing through a longitudinal, mixed-methods design in an Ethiopian higher education context. Its main empirical contribution is the documentation of writing development across four repeated measurement points, supplemented by aggregated pre-post and construct-level analyses.

The intervention translates SCT, the Cognitive Process Model, and HOTS into a structured pedagogical procedure involving dialogue, recursive writing processes, analytical questioning, peer feedback, and revision. The findings are compatible with these theoretical perspectives, but they should not be interpreted as direct tests of their underlying mechanisms.

The study also contributes context-specific evidence from an Ethiopian public university, where large classes and workload constraints can make sustained individualized writing support difficult. The findings demonstrate the feasibility of implementing a structured conference model in the participating course while leaving broader questions of scalability open.

5.4. Implications for Practice

Curriculum and instruction: Writing courses may benefit from structured process-oriented cycles that combine drafting, dialogue, feedback, and revision, with substantive concerns addressed before extensive surface correction.

Teacher development: Training should emphasize dialogic questioning, contingent scaffolding, and techniques that require students to explain and justify revision decisions before receiving direct corrections.

Assessment: Analytical rubrics separating Content, Organization, Language Use, and Conventions can support diagnostic feedback and make construct-level development more visible. Parallel prompts and rater calibration are important in longitudinal assessment.

Peer feedback: Peer review should be guided by explicit criteria and structured response procedures rather than left as unrestricted discussion.

Institutional implementation: Writing conferences require time and trained personnel. Their broader adoption should therefore be accompanied by consideration of class size, workload, staffing, and institutional resources.

5.5. Limitations

Several limitations constrain interpretation. First, the absence of a control or comparison group means that maturation, history, testing familiarity, regression toward the mean, and concurrent language-learning experiences cannot be separated from intervention-related change. Second, the sample came from one cohort at one university, limiting generalization to other institutions and student populations. Third, the sample size of 47 limits precision for detecting subtle interaction effects. Consequently, the non-significant Time × AWPPrM interaction should not be interpreted as proof of equivalent development across baseline levels. Fourth, repeated writing assessments may introduce practice and familiarity effects, although parallel forms were used. Fifth, the large pre-post effect size (d = 4.57) comes from an uncontrolled within-group comparison and therefore should not be interpreted as a causal estimate of intervention effectiveness. Sixth, FGD data represent participants' reported experiences and may be affected by recall or social desirability. Finally, no delayed post-test was administered, so the durability and transfer of observed writing gains cannot be established.

5.6. Directions for Future Research

Future research should employ controlled or randomized designs to strengthen causal inference, replicate the intervention across multiple Ethiopian universities, and include delayed post-tests to examine retention and transfer.

Larger samples would permit more precise investigation of Time × Baseline and Construct × Time interactions. Direct analysis of conference transcripts could also identify which questioning, scaffolding, and peer-feedback practices are associated with subsequent writing changes. Finally, implementation studies should examine workload, training requirements, class size, cost, and institutional sustainability before broader scaling is recommended.

REFERENCES

Afdila, I., Muniroh, S., & Anugerahwati, M. (2023). The implementation of writing conferences in an EFL writing classroom: A case study. Journal of English Language Teaching and Linguistics, 8(2), 145–160.

Alexander, R. (2020). A dialogic teaching companion. Routledge.

Aslam, S., Khan, M. S., & Ahmed, R. (2025). Beyond grammar correction: Cultivating higher-order thinking skills in L2 writing pedagogy. International Journal of Applied Linguistics & English Literature, 14(1), 44–58.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101.

Canagarajah, A. S. (2019). Translingual practice: Global Englishes and cosmopolitan relations. Routledge.

Carroll, L. A. (2002). Rehearsing new roles: How college students develop as writers. Southern Illinois University Press.

Çepni, G. (2016). The effect of writing conferences on Turkish EFL students' self-correction of language errors. Journal of Language and Linguistic Studies, 12(1), 168–181.

Chavda, J., Miller, S., & Thompson, K. (2024). Aligning feedback with Higher-Order Thinking Skills (HOTS): Impact on L2 academic literacy. Journal of Second Language Writing, 63, Article 101082.

Chen, W. (2024). Peer feedback and Chinese medical students' English academic writing development: A longitudinal intervention study (Doctoral dissertation, University of California). eScholarship.

Creswell, J. W., & Plano Clark, V. L. (2017). Designing and conducting mixed methods research (3rd ed.). SAGE Publications.

Elbow, P. (1981). Writing with power: Techniques for mastering the writing process. Oxford University Press.

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. College Composition and Communication, 32(4), 365–387.

Geberew, T. (2019). Practices and challenges of teaching writing skills: The case of public universities in Ethiopia. Journal of Languages and Culture, 10(1), 1–12.

Graham, S. (2018). The Writer(s)-within-Community model of writing. Educational Psychologist, 53(4), 258–279.

Graham, S., Hebert, M., & Harris, K. R. (2025). Effectiveness of writing interventions in higher education: A best-evidence meta-analysis. Review of Educational Research, 95(2), 310–345.

Heaton, J. B. (1990). Writing English language tests. Longman.

Hyland, K. (2000). Disciplinary discourses: Social interactions in academic writing. Longman.

Hyland, K. (2019). Second language writing (2nd ed.). Cambridge University Press.

Hyland, K., & Hyland, F. (2006). Feedback in second language writing: Contexts and issues. Cambridge University Press.

Inoue, A. B. (2012). Writing assessment as social justice: Theorizing a model for social justice in writing assessment. University Press of Colorado.

Kazemi, M., Thompson, R., & Farrokhi, F. (2025). Negotiating meaning in the digital ZPD: A comparative analysis of monologic and dialogic technology-mediated feedback. Language Learning & Technology, 29(1), 112–134.

Kioko, A., & Muthwii, M. (2018). English language teaching in East Africa: Challenges and opportunities. African Journal of Language Education, 5(1), 23–41.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of Chiropractic Medicine, 15(2), 155–163.

Lantolf, J. P., & Poehner, M. E. (2014). Sociocultural theory and the pedagogical imperative in L2 education: Vygotskian praxis and the research/practice space. Routledge.

Legesse, A., & Tegegne, W. (2025). Professional writing challenges and career advancement: A longitudinal tracking of Ethiopian graduates. Ethiopian Journal of Higher Education, 12(1), 45–63.

Leung, C. H. (2008). Oral feedback in writing conferences: A study of Hong Kong secondary students' perspectives and performance. The Asia-Pacific Education Researcher, 17(2), 211–226.

Li, D., & Qian, J. (2024). Engaging in dialogic peer feedback in L2 writing development: A microgenetic approach. System, 126, Article 103513.

Lunsford, A. A., & Ede, L. (2007). Writing together: Collaboration in theory and practice. Bedford/St. Martin's.

Mesfin, A., & Wondimu, T. (2024). Identifying obstacles to academic writing success at Mizan Tepi University: A large-scale analysis. Advanced Journal of Interdisciplinary Studies, 13(2), 78–95.

Miller, D. M., & McVee, M. B. (2022). Self-regulation and writing: A review of the literature. Educational Psychology Review, 34(2), 567–592.

Murray, D. M. (1980). Writing as process: How writing finds its own meaning. In T. R. Donovan & B. W. McClelland (Eds.), Eight approaches to teaching composition (pp. 3–20). NCTE.

Negash, N. (2022). Structural barriers to individualized feedback in Ethiopian higher education: Large classes and teaching loads. Choke Journal of Education and Behavioral Studies, 5(2), 112–129.

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. Language Learning, 64(4), 878–912.

Richards, J. C., & Rodgers, T. S. (2014). Approaches and methods in language teaching (3rd ed.). Cambridge University Press.

Saito, H. (2008). Agency and authority in the writing conference: A cultural perspective. Journal of Second Language Writing, 17(2), 112–130.

Skar, G. B., Aasen, A. J., & Jølle, L. (2024). Testing the "writing is caught" hypothesis: A large-scale longitudinal study on unscaffolded writing opportunities. Reading and Writing, 37, 412–435.

Sommers, N. (1982). Responding to student writing. College Composition and Communication, 33(2), 148–156.

Straub, R. (2000). The student, the teacher, and the text: Variations in teacher comments on student writing. In R. Straub (Ed.), The practice of response (pp. 1–34). Hampton Press.

Tardy, C. M. (2016). Beyond genre: Melding genre and practice in second language writing. University of Michigan Press.

Tedeneke, M. (2022). Teachers' and students' perceptions and practices of process-oriented writing feedback: The case of Ethiopian secondary schools (Unpublished doctoral dissertation). Addis Ababa University.

Teshome, D. (2020). The "reproduction mindset" in Ethiopian undergraduate writing: Challenges for critical argumentation. Journal of Education and Behavioral Sciences, 8(1), 22–39.

Topping, K. J. (2009). Peer assessment from primary to tertiary education. Assessment & Evaluation in Higher Education, 34(1), 20–27.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

Weigle, S. C. (2002). Assessing writing. Cambridge University Press.

Yu, S., & Poehner, M. E. (2025). Dynamic Assessment and L2 writing development: Negotiating stance and support in the ZPD. Modern Language Journal, 109(1), 88–110.

Zamel, V. (1982). Writing: The process of discovering meaning. TESOL Quarterly, 16(2), 195–209.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. Theory into Practice, 41(2), 64–70