Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Examining Factor-Level Changes in Teacher Feedback, Feedback Orientation, and Writing Across AI, Human, and Hybrid Feedback Conditions

**SIYI ZHU**

s1147049@s.eduhk.hk

EdUHK: The Education University of Hong Kong https://orcid.org/0009-0000-8193-3192

**JianJun Zhu**

Minnan University of Science and Technology https://orcid.org/0009-0002-5114-2626

**Research Article**

**Keywords:** AI-assisted feedback, Chinese writing, feedback orientation, human–AI collaboration, teacher feedback

**Posted Date:** November 16th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11720044/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** The authors declare no competing interests.

**Examining Factor-Level Changes in Teacher Feedback, Feedback Orientation, and Writing Across AI, Human, and Hybrid Feedback Conditions**

Augustine Zhu SiYi[a], Zhu JinXin (Corresponding Author)[a *] and Zhu JianJun[b]

*[a] Department of Curriculum and Instruction, Education University of Hong Kong, Hong Kong, s1147049@s.eduhk.hk; [a] Department of Curriculum and Instruction, Education University of Hong Kong, Hong Kong, jxzhu@eduhk.hk ; [b] Faculty of Education, Minnan University of Science and Technology, FuJian, China, chinesezhu@126.com.*

# Examining Factor-Level Changes in Teacher Feedback, Feedback Orientation, and Writing Across AI, Human, and Hybrid Feedback Conditions


## Abstract

(Research problem) This study addressed limited empirical examination of the multidimensional structures of teacher feedback and feedback orientation and of how their dimensions change under different AI exposure conditions in Chinese L1 writing classrooms. (Objectives) It tested the proposed five-factor teacher feedback and four-factor feedback orientation structures and compared pre–post changes across AI-only, AI–human, and human-only conditions. (Method) Parallel analysis, EFA, and CFA supported both structures; repeated-measures ANOVA tested condition differences. (Key findings) AI–human produced the largest gains on three feedback dimensions: verification, directive, scaffolding, all feedback orientation dimensions, and writing; AI-only produced moderate gains in the same dimensions; and human-only improved only praise and criticism. (Implications) Findings validate the dimensional structures, show that AI–human feedback constitutes higher-quality AI exposure, and inform differentiated integration of AI and human feedback to support specific feedback perceptions, feedback orientation, and writing improvement.


## Introduction

Teacher feedback is multidimensional (Guo, 2017, 2020), and feedback orientation comprises utility, accountability, social awareness, and self-efficacy (Linderbaum & Levy, 2010). (Topic sentence) However, how these constituent dimensions change under different AI exposure conditions remains unclear. (Overall research issue) This matters because dimensions may respond distinctively, and because the three conditions represent different qualities of AI exposure: human involvement embeds AI in subject-specific tasks, builds AI literacy, and supplies verification strategies, making AI–human feedback higher-quality exposure than AI-only feedback, whereas human-only feedback provides relational support without AI mechanisms (Lim et al., 2026; Zhang et al., 2024; Borges et al., 2026). (Justification) Yet empirical examination of these dimensional distinctions remains limited, particularly in Chinese L1 writing classrooms, and constituent factors have rarely been treated as distinct analytical units across AI exposure conditions. (Previous limitations) This study therefore tests the proposed five-factor teacher feedback and four-factor feedback orientation structures, compares pre- to post-intervention changes across AI-only, AI–human, and human-only conditions. (Research purpose) Theoretically, it validates dimensional structures and clarifies which feedback dimensions are sensitive to AI exposure quality; practically, it informs how AI feedback can be integrated to support specific feedback perceptions, feedback orientation, and writing improvement. (Theoretical and practical benefits)


**Keywords:** AI-assisted feedback; Chinese writing; feedback orientation; human–AI collaboration; teacher feedback


## Literature Review

### The Multidimensional Structures of Teacher Feedback and Feedback Orientation

Examining a construct's underlying dimensions reveals their distinct and interrelated contributions to achievement, offering insights beyond treating the construct as a whole (Jin & Zhang, 2021). Accordingly, the present study examines teacher feedback and feedback orientation as theoretically established five- and four-factor constructs, respectively, investigating how different AI exposure interventions influence their constituent dimensions. The proposed factor structures are further examined through parallel analysis, EFA, and CFA to establish their empirical adequacy in the present sample.

Teacher feedback is multidimensional, with its effects depending on the information provided and learners' engagement with it (Hattie & Timperley, 2007; Shute, 2008). Guo (2017) identified five dimensions: verification, directive, scaffolding, teacher praise, and teacher criticism, with differentiated patterns across grade levels and associations with self-regulated learning (Guo, 2020). Verification indicates correctness but alone offers limited improvement guidance, whereas directive feedback provides answers or explicit instructions that facilitate correction but may constrain independent problem-solving when overused (Guo, 2017, 2020; Shute, 2008). Scaffolding offers cues, hints, or partial solutions to support learners' own answer construction and guided engagement in automated feedback and learning (Guo, 2017, 2020; Wang et al., 2022). Praise positively evaluates performance, effort, or learning qualities, whereas criticism identifies deficiencies; learners do not necessarily process criticism as praise's opposite (Worrall et al., 1983; Guo, 2017, 2020).

Feedback orientation concerns how individuals perceive, interpret, and respond to feedback. The Feedback Orientation Scale (FOS) identifies four related but distinct dimensions: utility, accountability, social awareness, and self-efficacy (Linderbaum & Levy, 2010), subsequently applied to educational feedback and learner engagement (Wang et al., 2015). Utility concerns feedback's perceived value for performance improvement and action (Wang et al., 2015); accountability reflects responsibility for attending to and acting on feedback (Linderbaum & Levy, 2010); social awareness involves interpersonal implications and sensitivity to social contexts (Miller et al., 2017; Wegner & Giuliano, 1982); and self-efficacy concerns confidence in responding to feedback and performing relevant tasks (Liu & Gumah, 2020).

**Instructional Intervention**

The study employed a three-condition instructional intervention over one academic semester: AI-only, Human-only, and Hybrid (ChatGPT + Human). All groups received two writing lessons per week (20 minutes per lesson), following the same curriculum, instructional objectives, schedule, and assessment criteria. Each weekly cycle comprised writing instruction, writing production and in Lesson 1, followed by feedback provision, feedback processing and reflection in Lesson 2.

The primary distinction among the three conditions was students' exposure to AI-mediated feedback. Feedback sources, interaction arrangements, and teacher involvement were structured to establish differentiated AI exposure conditions while maintaining consistent instructional arrangements.

| Intervention dimension | AI-only | Hybrid | Human-only |
|---|---|---|---|
| Feedback source | ChatGPT | ChatGPT + Teacher | Teacher |
| Overall AI exposure | Direct and sustained | Integrated with teacher guidance | None |
| AI interaction | Students submit drafts, examine suggestions, and conduct follow-up interactions | Students examine AI suggestions before receiving additional teacher guidance | No AI interaction |
| Teacher involvement | Procedural and technical assistance only | Reviews AI feedback and provides contextual guidance | Provides content-related feedback |
| AI interaction time | Throughout designated feedback activities | Within designated components of feedback activities | None |
| Interaction frequency | Repeated AI-feedback interactions | AI interactions alongside teacher feedback | None |
| Actual-use proportion | All designated feedback activities | Designated components of feedback activities | None |

AI exposure was operationalized through three observable indicators: interaction time, interaction frequency, and actual-use proportion (the proportion of designated feedback activities involving actual AI use).

### AI exposure, Feedback Perceptions, Feedback Orientation, and Writing Performance

# AI Exposure and Five Dimensions of Feedback Perceptions

### 2.1. Verification Feedback

AI exposure may strengthen students' perceptions of verification feedback through several complementary mechanisms. In AI-mediated learning environments, students may develop stronger perceptions of verification feedback by repeatedly requesting explanations and confirmation of AI-generated responses and cross-checking suggestions against assessment criteria (Li et al., 2026). Furthermore, AI exposure may strengthen these perceptions by providing opportunities to examine the accuracy and reliability of feedback against explicit task requirements (Dai et al., 2024). AI may also strengthen students' this kind of perceptions by exposing them to detailed, accurate assessments of their writing and enabling them to identify and evaluate specific textual issues during revision (Glüsing et al., 2026). In addition, interaction with AI-mediated feedback may support verification by helping students examine writing-related problems and confirm their understanding through proficiency-sensitive guidance (Qi et al., 2026). Finally, exposure to both AI- and human generated feedback may provide opportunities to compare feedback content and evaluate its objectivity and usefulness, thereby enhance perception of verification feedback (Nazaretsky et al., 2026). Collectively, AI exposure may strengthen verification feedback, while human involvement may further enrich how students engage with and evaluate AI-generated feedback.

## 2.2. Directive Feedback

AI exposure may strengthen students' perceptions of directive feedback by creating repeated opportunities to access specific, task-relevant guidance. AI-mediated feedback can adjust the explicitness of support to learners' evolving writing needs, moving between implicit prompts and explicit guidance as appropriate (Qi et al., 2026). This kind of feedback can also organize suggestions around essay subthemes, directing learners' attention to relevant content and possible directions for idea development (Engeness, 2025). Through iterative prompting, learners may further seek clarification and obtain more targeted writing support from educational chatbots (Sørhaug, 2026). Evidence from feedback-quality research confirms that ChatGPT can provide formative feedback with explicit directions for improvement (Steiss et al., 2024). Moreover, pedagogically enriched AI feedback can address higher-order concerns, such as coherence and organization, and offer targeted revision suggestions, extending beyond the primarily corrective focus of conventional automated writing evaluation (Zhang et al., 2026). Lastly, AI and human integrated exposure may strengthen directive feedback by combining AI-generated suggestions with teachers' contextualized explanations and actionable revision guidance (Asadi et al., 2025; Zhu et al., 2025)

Collectively, AI exposure may strengthen perceived directive feedback by increasing access to actionable guidance, while human involvement may further enrich this process.

## 2.3. Scaffolding Feedback

AI exposure may enhance scaffolding feedback by making support progressively more explicit, individualized, and responsive to emerging difficulties. Repeated engagement with and application of AI-generated revision guidance may strengthen perceived scaffolding quality by making support more actionable (Chong, 2025). Student-centered prompting may calibrate scaffolding to learners' proficiency and evolving needs by progressing from implicit cues to explicit explanations when initial guidance proves insufficient (Qi et al., 2026). In AI-supported peer review, feedback on learners' own comments may refine the scaffolding they provide by developing more descriptive, diagnostic, and actionable responses (Guo et al., 2024). Lastly, AI-human exposure may strengthen scaffolding feedback by combining AI-generated suggestions with teachers' contextualized guidance on content, structure, and learner needs (Wei et al., 2026). Collectively, AI exposure may strengthen scaffolding feedback, with human involvement further enriching its potential through contextualized and responsive guidance.

## 2.4. Teacher Praise

AI exposure may have a neutral or even negative association with students' perceptions of praise feedback because of many mechanisms. AI involvement may compromise praise quality by prioritising emotional reassurance over substantive feedback engagement and revision (Alsaiari et al., 2025), by reducing the motivational value of praise when interpersonal recognition is absent (Pi et al., 2026), and by limiting students' engagement with feedback despite frequent positive evaluative language (Wang et al., 2026). Furthermore, person-focused praise may undermine constructive motivational support by tying perceived competence to personal attributes rather than task-related progress, increasing vulnerability to subsequent failure (Skipper & Douglas, 2012). This interpretation is consistent with evidence that students tend to evaluate human-generated feedback more favorably than AI-generated feedback (Nazaretsky et al., 2024).

## 2.5. Teacher Criticism

AI feedback exposure may compromise perceived criticism quality by providing less relationally grounded corrective input than teacher feedback (Soyoof et al., 2026), by generating repetitive or inaccurate criticism that discourages affective and behavioural engagement (Wang et al., 2026), and by producing suggestions that conflict with learners' expectations or external evidence or impose excessive revision demands (Chen et al., 2026). These limitations may be compounded by inconsistent explanations and proficiency-inappropriate feedback, which erode trust and elicit negative emotions (Takahashi & Kadowaki, 2026). Moreover, criticism may lose its instructional value by emphasising personal evaluation rather than clarifying task requirements and actionable revision steps (Hattie & Timperley, 2007). On the contrary, teacher criticism may be more readily accepted when providing personalized human emotional support (Soyoof et al., 2026).

# 3. AI Exposure and Four Dimensions of Feedback Orientation

## 3.1. Feedback Utility

AI exposure may strengthen students' perceptions of feedback utility through several complementary mechanisms. Task-specific interaction with GenAI can enhance the practical value of feedback by enabling learners to obtain more specific, comprehensible, and actionable corrective comments tailored to their writing requirements (Zhang et al., 2024). Repeated AI interaction may further strengthen perceived utility by allowing learners to refine prompts, clarify initial responses, and obtain suggestions more closely aligned with their actual writing needs (Alaoui Mhamdi, 2026). Moreover, active engagement with AI-generated feedback may enhance its perceived usefulness by encouraging learners to evaluate, verify, modify, or reject suggestions according to task requirements rather than accept them uncritically (Alaoui Mhamdi, 2026; Liu et al., 2026). Through feedback-seeking interactions, AI exposure may also reinforce perceived utility by helping learners recognize the relevance of feedback to their learning goals and subsequent writing improvement (Ng & Cheung, 2026). Finally, sustained engagement with feedback and revision may strengthen learners' appreciation of its practical value by providing opportunities to apply suggestions and observe their contribution to improving written work (Lu et al., 2026). AI exposure may yield greater feedback utility when human teachers contextualize AI suggestions, address higher-order writing needs, and translate feedback into actionable revision guidance (Zhang et al., 2025). Thus, AI exposure may enhance feedback utility through iterative engagement, with human guidance further contextualizing feedback and translating it into actionable improvements.

## 3.2. Accountability

Regular engagement with ChatGPT may strengthen accountability by encouraging learners to consult AI feedback when encountering writing difficulties and use its suggestions to plan subsequent revisions (Stojanov et al., 2024). Repeated AI consultation may further strengthen accountability by prompting learners to consider how feedback can inform their responses to specific writing problems rather than treating suggestions as isolated information. Moreover, reflective and cautious forms of AI reliance may strengthen accountability by encouraging learners to examine the relevance of AI-generated suggestions before incorporating them into their problem-solving and revision processes (Hou et al., 2025). Individual and contextual factors shaping GenAI reliance may also influence accountability by determining how learners integrate AI-generated feedback into their decision-making and assume responsibility for subsequent writing actions (Tao et al., 2026). Moreover, AI exposure may strengthen perceived feedback accountability when human peer

dialogue requires learners to evaluate, justify, and act on AI-generated suggestions (Ates, 2026). Together, these findings suggest that repeated AI exposure may establish feedback as a sustained resource for writing and revision, with human involvement further strengthening learners' engagement and accountability.

### 3.3. Social Awareness

Repeated exposure to context-aware GenAI feedback may further strengthen social awareness by encouraging learners to evaluate how textual choices serve rhetorical purposes and address readers' expectations, as demonstrated by the greater number of rhetorical-level revisions prompted by feedback grounded in rhetorical situation theory (Tseng, 2026). Such contextualized AI interaction may also enhance learners' awareness of disciplinary and communicative conventions by directing their attention to how writing is interpreted within particular academic and social communities. Furthermore, AI literacy development may strengthen this process by encouraging learners to consider interpersonal communication, collaboration, and ethical judgment when engaging with AI-mediated feedback (Reicho et al., 2026). Moreover, AI exposure may create opportunities for feedback-related social awareness when human peer dialogue reveals how readers interpret and respond to learners' writing (Tsao & Liu, 2026). Collectively, AI exposure may foster feedback-related social awareness, with human interaction further enriching this process through authentic reader perspectives.

### 3.4. Feedback Self-Efficacy

AI exposure may strengthen feedback self-efficacy by enabling learners to receive personalized guidance that builds confidence in meeting learning demands (Lyu & Salam, 2025). Repeated engagement with ChatGPT feedback may further enhance this confidence by helping learners translate AI-generated comments into substantive revisions and recognize their capacity to improve their own work (Yang et al., 2025). Acceptance of generative AI may strengthen feedback self-efficacy by encouraging learners to engage with AI-mediated feedback, while the development of feedback literacy helps them understand and use such feedback effectively (Zhou & Wang, 2026). Similarly, intensive LLM co-editing may reinforce self-efficacy by facilitating feedback literacy and revision behaviours through which learners progressively develop confidence in their writing capabilities (Nguyen & Vo, 2026). AI exposure may strengthen feedback self-efficacy when human peer feedback helps learners generate, discuss, and reformulate feedback through sustained collaborative practice (Yan et al., 2026). Together, these findings suggest that AI exposure may strengthen feedback self-efficacy by developing learners' feedback skills, with human interaction further reinforcing their confidence through collaborative practice.

## AI Exposure in Chinese Writing Classrooms

By eliciting and developing AI-generated ideas, learners can broaden their initial thinking and build more developed arguments or compositions (Li et al., 2025; Levine et al., 2025). By selectively modifying, reorganising, and re-expressing AI-generated text, learners can transform initial suggestions into more coherent compositions while retaining control over authorial decisions (Li et al., 2025). Repeated engagement with GenAI during revision may further improve writing by helping learners refine their texts and sustain task engagement (Wang et al., 2026). AI-assisted brainstorming may enhance creativity by generating diverse ideas (Yao, 2026). Moreover, AI exposure may further strengthen writing by enabling learners to iteratively generate, evaluate, adapt, and integrate AI contributions while retaining human authorial control, suggesting the potential of human-AI collaboration (Nguyen et al., 2024).

## Effective AI Exposure Through Human–AI Collaboration

Human involvement may strengthen AI exposure by embedding AI use in subject-specific tasks, enabling students to engage with AI across curricular contexts (Lim et al., 2026). By integrating teacher-led AI literacy instruction into regular lessons, teachers can develop students' understanding of AI and its practical relevance, supporting more informed engagement (Zhang et al., 2024). By providing foundational knowledge and verification strategies, teachers can help students move beyond passive AI consumption towards iterative prompting, critical evaluation, and purposeful application (Borges et al., 2026). Accordingly, this study examines AI–human feedback as a higher-quality form of AI exposure, comparing it with AI-only exposure and human-only feedback to investigate how different level of AI exposure influence students' Chinese writing.

### Research Gaps, Aims and Questions

Although teacher feedback and feedback orientation have established theoretical factor structures, empirical examination of these dimensional distinctions remains limited, particularly in Chinese L1 writing classrooms. Furthermore, despite the established status of the Feedback Orientation Scale (FOS) and growing research on feedback and feedback orientation across different AI exposure conditions, relatively little attention has been paid to their constituent factors as distinct analytical units. Consequently, how individual dimensions vary across different AI exposure conditions remains underexplored.

Teacher feedback and feedback orientation are multidimensional constructs whose constituent factors may exhibit distinct patterns. Establishing their factorial structures is therefore essential before examining intervention effects. Building on this measurement foundation, the study compares pre- to post-intervention changes across three instructional conditions to investigate how different forms of AI exposure relate to changes in these dimensions and writing performance. The present study addresses one main research question:

**RQ.** How do students' teacher feedback, feedback orientation, and writing performance dimensions change from pre-test to post-test across the AI-only, AI–human, and human-only conditions? (Mixed-design repeated-measures ANOVA)

**Data Preparation**

**Table 1**

*Dimensionality of the FB Pre-test: Parallel-Analysis Analogue and Factor Retention*

**Panel A. Model fit of 1- to 6-factor EFA solutions**

| Factors | $\chi^2$ | df | p | RMSEA | CFI | TLI | SRMR |
|---|---|---|---|---|---|---|---|
| 1 | 4289.63 | 252 | < .001 | .211 | .594 | .555 | .204 |
| 2 | 3244.69 | 229 | < .001 | .191 | .697 | .634 | .168 |
| 3 | 2324.39 | 207 | < .001 | .169 | .787 | .716 | .133 |
| 4 | 1642.52 | 186 | < .001 | .147 | .853 | .783 | .088 |
| 5 | 167.42 | 166 | .455 | .005 | 1.000 | 1.000 | .016 |
| 6 | 127.83 | 147 | .871 | .000 | 1.000 | 1.004 | .014 |

**Panel B. First six sample eigenvalues**

| Component | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Eigenvalue | 8.14 | 2.91 | 2.72 | 2.46 | 2.18 | 0.51 |

*Note.* N = 360. Estimator = WLSMV; rotation = GEOMIN. Factors were retained following the parallel-analysis analogue in Mplus (comparison of 1- to 6-factor solutions). The first five eigenvalues exceeded 1.00 (8.14–2.18), whereas the sixth dropped to 0.51. Model fit improved substantially from four to five factors (RMSEA = .005, CFI = 1.000, SRMR = .016) and did not improve further at six factors (ΔCFI = .000, TLI = 1.004), indicating over-extraction. The 5-factor solution is therefore retained.

**Table 2**

*Five-Factor EFA of the FB Pre-test: GEOMIN-Rotated Loadings and Factor Correlations*

**Panel A. Rotated factor loadings**

| Item | F1 | F2 | F3 | F4 | F5 |
|---|---|---|---|---|---|
| 1 | **.79** | .01 | −.02 | −.01 | .07 |
| 2 | **.83** | −.02 | .03 | −.02 | .06 |
| 3 | **.82** | .01 | −.03 | .11 | −.07 |
| 4 | **.80** | .01 | .03 | .02 | −.02 |
| 5 | **.76** | .01 | .03 | −.03 | .01 |
| 6 | .06 | **.83** | −.03 | −.04 | −.04 |
| 7 | −.03 | **.82** | −.02 | .02 | .02 |
| 8 | −.01 | **.77** | .00 | .02 | .05 |
| 9 | .03 | **.83** | .08 | .01 | .03 |
| 10 | −.01 | **.82** | .03 | .03 | .00 |
| 11 | −.01 | .08 | **.83** | −.03 | −.08 |
| 12 | .07 | −.05 | **.75** | .04 | .06 |
| 13 | −.01 | .05 | **.91** | .01 | .01 |
| 14 | −.02 | −.05 | **.85** | .01 | .00 |
| 15 | .06 | −.04 | .01 | **.85** | .04 |
| 16 | −.06 | .04 | .03 | **.84** | .02 |
| 17 | .01 | −.02 | .02 | **.89** | .00 |
| 18 | .02 | .00 | .02 | **.82** | .00 |

| Item | F1 | F2 | F3 | F4 | F5 |
|---|---|---|---|---|---|
| 19 | −.01 | .12 | −.06 | **.76** | −.03 |
| 20 | .01 | .01 | .04 | −.02 | **.88** |
| 21 | −.01 | .02 | .01 | .00 | **.90** |
| 22 | −.01 | .00 | −.02 | .08 | **.81** |
| 23 | .02 | .07 | .01 | −.02 | **.84** |
| 24 | .02 | −.03 | −.02 | .02 | **.87** |

*Note.* N = 360. Estimator = WLSMV; rotation = GEOMIN. Items 1–24 correspond to FB_pre_item1–FB_pre_item24. Primary loadings are shown in bold. All primary loadings ≥ .76; no cross-loading exceeded .30. The pattern of loadings is consistent with the hypothesized five-factor structure (Items 1–5 → F1; 6–10 → F2; 11–14 → F3; 15–19 → F4; 20–24 → F5).

**Panel B. Factor correlations**

| Factor | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1. F1 | — | | | | |
| 2. F2 | .27 | — | | | |
| 3. F3 | .23 | .23 | — | | |
| 4. F4 | .29 | .34 | .34 | — | |
| 5. F5 | .29 | .36 | .19 | .35 | — |

*Note.* All inter-factor correlations were statistically significant ($p < .05$) and below .40, supporting discriminant validity among the five factors (maximum r = .36).

A parallel-analysis analogue (1- to 6-factor EFA solutions under WLSMV with GEOMIN rotation) and the eigenvalue rule converged on a **five-factor** structure: the first five eigenvalues exceeded 1.00, model fit was excellent at five factors

($\chi^2(166) = 167.42$, $p = .455$; RMSEA = .005; CFI = 1.000; SRMR = .016), and extracting a sixth factor yielded no substantive improvement (TLI = 1.004). Rotated loadings were clean (all primary loadings ≥ .76, no cross-loading > .30) and factor correlations ranged from .19 to .36, all below the .70 discriminant threshold. The five-factor structure is thus supported.

**Table 3**

*Dimensionality of the FB Post-test: Parallel-Analysis Analogue and Factor Retention*

**Panel A. Model fit of 1- to 6-factor EFA solutions**

| Factors | $\chi^2$ | df | p | RMSEA | CFI | TLI | SRMR |
|---|---|---|---|---|---|---|---|
| 1 | 6972.78 | 252 | < .001 | .272 | .619 | .583 | .322 |
| 2 | 4105.01 | 229 | < .001 | .217 | .780 | .735 | .213 |
| 3 | 2880.27 | 207 | < .001 | .189 | .848 | .798 | .155 |
| 4 | 1898.45 | 186 | < .001 | .160 | .903 | .856 | .092 |
| 5 | 159.10 | 166 | .636 | .000 | 1.000 | 1.001 | .012 |
| 6 | 122.06 | 147 | .934 | .000 | 1.000 | 1.003 | .010 |

**Panel B. First six sample eigenvalues**

| Component | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Eigenvalue | 7.31 | 4.64 | 3.09 | 2.71 | 2.13 | 0.38 |

*Note.* N = 360. Estimator = WLSMV; rotation = GEOMIN. Factors were retained following the parallel-analysis analogue in Mplus (comparison of 1- to 6-factor solutions). The first five eigenvalues exceeded 1.00 (7.31–2.13), whereas the sixth

dropped sharply to 0.38. Model fit improved substantially from four to five factors (RMSEA = .000, CFI = 1.000, SRMR = .012) and did not improve meaningfully at six factors (ΔCFI = .000, TLI = 1.003), indicating over-extraction. The 5-factor solution is therefore retained.

# Table 4

**Table 4**

*Five-Factor EFA of the FB Post-test: GEOMIN-Rotated Loadings and Factor Correlations*

**Panel A. Rotated factor loadings**

| Item | F1 | F2 | F3 | F4 | F5 |
|---|---|---|---|---|---|
| 1 | **.90** | −.07 | .02 | .02 | −.01 |
| 2 | **.86** | .03 | .04 | −.02 | .04 |
| 3 | **.86** | .07 | −.02 | .03 | −.02 |
| 4 | **.87** | .03 | .01 | .01 | .04 |
| 5 | **.87** | −.02 | .01 | −.03 | −.03 |
| 6 | −.02 | **.86** | .03 | −.01 | .02 |
| 7 | −.02 | **.86** | .01 | −.01 | −.03 |
| 8 | .02 | **.83** | .01 | .05 | −.04 |
| 9 | .02 | **.88** | .00 | .00 | .05 |
| 10 | .04 | **.89** | −.02 | −.01 | .00 |
| 11 | .04 | −.02 | **.88** | −.04 | −.03 |
| 12 | −.02 | .02 | **.81** | .07 | −.08 |
| 13 | −.02 | .02 | **.94** | −.01 | .04 |

| Item | F1 | F2 | F3 | F4 | F5 |
|---|---|---|---|---|---|
| 14 | .05 | .00 | **.86** | .00 | .04 |
| 15 | .02 | −.02 | −.02 | **.93** | .00 |
| 16 | .03 | −.02 | .01 | **.92** | −.02 |
| 17 | .03 | −.04 | −.01 | **.90** | .00 |
| 18 | −.07 | .05 | .07 | **.82** | .04 |
| 19 | −.01 | .06 | −.02 | **.84** | .01 |
| 20 | .05 | .00 | −.03 | .01 | **.94** |
| 21 | −.02 | .01 | .05 | .02 | **.92** |
| 22 | .01 | −.01 | −.01 | −.02 | **.87** |
| 23 | −.08 | .00 | .05 | .05 | **.88** |
| 24 | .03 | .01 | −.05 | −.02 | **.92** |

*Note.* N = 360. Estimator = WLSMV; rotation = GEOMIN. Items 1–24 correspond to FB_post_item1–FB_post_item24. Primary loadings are shown in bold. All primary loadings ≥ .81; no cross-loading exceeded .10 in absolute value. The pattern of loadings is consistent with the hypothesized five-factor structure (Items 1–5 → F1; 6–10 → F2; 11–14 → F3; 15–19 → F4; 20–24 → F5).

**Panel B. Factor correlations**

| Factor | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1. F1 | — | | | | |
| 2. F2 | .33 | — | | | |
| 3. F3 | .35 | .35 | — | | |

| Factor | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 4. F4 | .11 | .15 | .18 | — | |
| 5. F5 | .13 | .15 | .03 | .30 | — |

*Note.* All inter-factor correlations were statistically significant ($p < .05$) except F3–F5 ($r = .03$, ns), and all were below .40, supporting discriminant validity among the five factors (maximum $r = .35$).

A parallel-analysis analogue (1- to 6-factor EFA solutions under WLSMV with GEOMIN rotation) and the eigenvalue rule converged on a **five-factor** structure: the first five eigenvalues exceeded 1.00 (7.31–2.13), model fit was excellent at five factors ($\chi^2(166) = 159.10$, $p = .636$; RMSEA = .000; CFI = 1.000; SRMR = .012), and extracting a sixth factor yielded no substantive improvement (TLI = 1.003). Rotated loadings were clean (all primary loadings ≥ .81, no cross-loading > .10) and factor correlations ranged from .03 to .35, all below the .70 discriminant threshold. The five-factor structure is thus supported.

**Table 5**

*Confirmatory Factor Analysis of the FB Pre-test: Five-Factor Model*

**Panel A. Model fit**

| Model | $\chi^2$ | df | p | $\chi^2$/df | CFI | TLI | RMSEA [90% CI] | SRMR |
|---|---|---|---|---|---|---|---|---|
| 5-factor CFA | 249.13 | 242 | .363 | 1.03 | .999 | .999 | .009 [.000, .024] | .028 |

*Note.* N = 360. Estimator = WLSMV; parameterization = DELTA; link = probit. The model specified Items 1–5 → F1, 6–10 → F2, 11–14 → F3, 15–19 → F4, and 20–24 → F5. All fit indices met conventional criteria (CFI ≥ .95, TLI ≥ .95, RMSEA ≤ .06, SRMR ≤ .08), and the model was not rejected by the $\chi^2$ test ($p = .363$). The five-factor structure is therefore supported.

**Panel B. Standardized factor loadings (STDYX)**

| Item | F1 | F2 | F3 | F4 | F5 |
|---|---|---|---|---|---|
| 1 | **.81** | | | | |
| 2 | **.84** | | | | |
| 3 | **.82** | | | | |
| 4 | **.81** | | | | |
| 5 | **.76** | | | | |
| 6 | | **.78** | | | |
| 7 | | **.81** | | | |
| 8 | | **.80** | | | |
| 9 | | **.90** | | | |
| 10 | | **.84** | | | |
| 11 | | | **.79** | | |
| 12 | | | **.81** | | |
| 13 | | | **.95** | | |
| 14 | | | **.80** | | |
| 15 | | | | **.88** | |
| 16 | | | | **.85** | |
| 17 | | | | **.89** | |
| 18 | | | | **.83** | |
| 19 | | | | **.77** | |

| Item | F1 | F2 | F3 | F4 | F5 |
|---|---|---|---|---|---|
| 20 | | | | | **.88** |
| 21 | | | | | **.90** |
| 22 | | | | | **.84** |
| 23 | | | | | **.88** |
| 24 | | | | | **.86** |

*Note.* All loadings are standardized (STDYX) and significant at p < .001. Loadings range from .76 to .95. The marker item in each factor was fixed to 1.00 in the unstandardized solution.

**Panel C. Latent factor correlations (STDYX)**

| Factor | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1. F1 | — | | | | |
| 2. F2 | .30 | — | | | |
| 3. F3 | .26 | .27 | — | | |
| 4. F4 | .33 | .38 | .37 | — | |
| 5. F5 | .32 | .40 | .21 | .38 | — |

*Note.* All correlations are standardized (STDYX) and significant at p < .001. Correlations range from .21 to .40; all are below the .70 discriminant-validity threshold, supporting the distinctiveness of the five factors.

**Panel D. $R^2$ for observed items**

| Item | $R^2$ | Item | $R^2$ | Item | $R^2$ |
|---|---|---|---|---|---|
| 1 | .66 | 9 | .80 | 17 | .80 |

| Item | $R^2$ | Item | $R^2$ | Item | $R^2$ |
|---|---|---|---|---|---|
| 2 | .70 | 10 | .71 | 18 | .69 |
| 3 | .68 | 11 | .63 | 19 | .59 |
| 4 | .65 | 12 | .65 | 20 | .78 |
| 5 | .58 | 13 | .90 | 21 | .82 |
| 6 | .61 | 14 | .64 | 22 | .70 |
| 7 | .65 | 15 | .78 | 23 | .77 |
| 8 | .64 | 16 | .72 | 24 | .73 |

*Note.* $R^2$ = proportion of variance in each item explained by its latent factor. Values range from .58 to .90.

A five-factor CFA was estimated with WLSMV using the hypothesized item-to-factor assignment (Items 1–5, 6–10, 11–14, 15–19, and 20–24). The model fit the data well, $\chi^2(242) = 249.13$, $p = .363$; CFI = .999; TLI = .999; RMSEA = .009, 90% CI [.000, .024]; SRMR = .028. All standardized loadings were substantial (.76–.95, all $p < .001$), and latent factor correlations ranged from .21 to .40, all below the .70 discriminant threshold. The five-factor measurement model was therefore retained.

**Table 6**
*Confirmatory Factor Analysis of the FB Post-test: Five-Factor Model*

**Panel A. Model fit**

| Model | $\chi^2$ | df | p | $\chi^2$/df | CFI | TLI | RMSEA [90% CI] | SRMR |
|---|---|---|---|---|---|---|---|---|
| 5-factor CFA | 251.54 | 242 | .324 | 1.04 | .999 | .999 | .010 [.000, .024] | .026 |

*Note.* N = 360. Estimator = WLSMV; parameterization = DELTA; link = probit. The model specified Items 1–5 → F1, 6–10 → F2, 11–14 → F3, 15–19 → F4, and 20–24

→ F5. All fit indices met conventional criteria (CFI ≥ .95, TLI ≥ .95, RMSEA ≤ .06, SRMR ≤ .08), and the model was not rejected by the $\chi^2$ test ($p = .324$). The five-factor structure is therefore supported.

**Panel B. Standardized factor loadings (STDYX)**

| Item | F1 | F2 | F3 | F4 | F5 |
|---|---|---|---|---|---|
| 1 | **.88** | | | | |
| 2 | **.89** | | | | |
| 3 | **.88** | | | | |
| 4 | **.89** | | | | |
| 5 | **.85** | | | | |
| 6 | | **.87** | | | |
| 7 | | **.84** | | | |
| 8 | | **.84** | | | |
| 9 | | **.89** | | | |
| 10 | | **.89** | | | |
| 11 | | | **.87** | | |
| 12 | | | **.82** | | |
| 13 | | | **.95** | | |
| 14 | | | **.89** | | |
| 15 | | | | **.92** | |
| 16 | | | | **.91** | |
| 17 | | | | **.90** | |

| Item | F1 | F2 | F3 | F4 | F5 |
|---|---|---|---|---|---|
| 18 | | | | **.85** | |
| 19 | | | | **.85** | |
| 20 | | | | | **.95** |
| 21 | | | | | **.92** |
| 22 | | | | | **.86** |
| 23 | | | | | **.88** |
| 24 | | | | | **.92** |

*Note.* All loadings are standardized (STDYX) and significant at $p < .001$. Loadings range from .82 to .95. The marker item in each factor was fixed to 1.00 in the unstandardized solution.

**Panel C. Latent factor correlations (STDYX)**

| Factor | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1. F1 | — | | | | |
| 2. F2 | .35 | — | | | |
| 3. F3 | .38 | .37 | — | | |
| 4. F4 | .12 | .17 | .19 | — | |
| 5. F5 | .14 | .15 | .03 | .32 | — |

*Note.* All correlations are standardized (STDYX). Correlations range from .03 to .38; all are below the .70 discriminant-validity threshold. All correlations are significant at $p < .05$ except F3–F5 ($r = .03$, $p = .579$, ns). The distinctiveness of the five factors is therefore supported.

**Panel D. $R^2$ for observed items**

| Item | $R^2$ | Item | $R^2$ | Item | $R^2$ |
|---|---|---|---|---|---|
| 1 | .77 | 9 | .80 | 17 | .80 |
| 2 | .80 | 10 | .80 | 18 | .72 |
| 3 | .78 | 11 | .76 | 19 | .72 |
| 4 | .80 | 12 | .66 | 20 | .89 |
| 5 | .73 | 13 | .89 | 21 | .85 |
| 6 | .76 | 14 | .79 | 22 | .75 |
| 7 | .71 | 15 | .85 | 23 | .78 |
| 8 | .70 | 16 | .82 | 24 | .84 |

*Note.* $R^2$ = proportion of variance in each item explained by its latent factor. Values range from .66 to .89.

A five-factor CFA was estimated with WLSMV using the hypothesized item-to-factor assignment (Items 1–5, 6–10, 11–14, 15–19, and 20–24). The model fit the data well, $\chi^2(242) = 251.54$, $p = .324$; CFI = .999; TLI = .999; RMSEA = .010, 90% CI [.000, .024]; SRMR = .026. All standardized loadings were substantial (.82–.95, all $p < .001$), and latent factor correlations ranged from .03 to .38, all below the .70 discriminant threshold (only F3–F5 was non-significant). The five-factor measurement model was therefore retained.

**Table 7**

*Dimensionality of the FO Pre-test: Parallel-Analysis Analogue and Factor Retention*

**Panel A. Model fit of 1- to 4-factor EFA solutions**

| Factors | $\chi^2$ | df | p | $\chi^2$/df | CFI | TLI | RMSEA [90% CI] | SRMR |
|---|---|---|---|---|---|---|---|---|
| 1 | 1744.49 | 54 | < .001 | 32.31 | .672 | .600 | .295 [.283, .307] | .255 |

| Factors | $\chi^2$ | df | p | $\chi^2$/df | CFI | TLI | RMSEA [90% CI] | SRMR |
|---|---|---|---|---|---|---|---|---|
| 2 | 1073.19 | 43 | < .001 | 24.96 | .800 | .694 | .258 [.245, .271] | .195 |
| 3 | 681.03 | 33 | < .001 | 20.64 | .874 | .749 | .234 [.218, .249] | .125 |
| 4 | 33.17 | 24 | .101 | 1.38 | .998 | .995 | .033 [.000, .057] | .014 |

**Panel B. First six sample eigenvalues**

| Component | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Eigenvalue | 4.79 | 2.00 | 1.83 | 1.55 | 0.34 | 0.31 |

*Note.* N = 360. Estimator = WLSMV; rotation = GEOMIN. Factors were retained following the parallel-analysis analogue in Mplus (comparison of 1- to 4-factor solutions). The first four eigenvalues exceeded 1.00 (4.79, 2.00, 1.83, 1.55), whereas the fifth dropped sharply to 0.34. Model fit improved substantially from three to four factors (RMSEA = .033, CFI = .998, SRMR = .014), and the four-factor solution was not rejected by the $\chi^2$ test (p = .101). The 4-factor solution is therefore retained.

**Table 8**

*Four-Factor EFA of the FO Pre-test: GEOMIN-Rotated Loadings and Factor Correlations*

**Panel A. Rotated factor loadings**

| Item | F1 | F2 | F3 | F4 |
|---|---|---|---|---|
| 1 | **.90** | .02 | .02 | −.01 |

| Item | F1 | F2 | F3 | F4 |
|---|---|---|---|---|
| 2 | **.84** | −.03 | −.01 | .02 |
| 3 | **.87** | .02 | .01 | .00 |
| 4 | .10* | **.79** | −.05 | .03 |
| 5 | .00 | **.89** | .01 | −.01 |
| 6 | −.04 | **.86** | .07 | .00 |
| 7 | −.03 | .02 | **.89** | −.01 |
| 8 | .01 | .03 | **.91** | .00 |
| 9 | .05 | −.03 | **.88** | .02 |
| 10 | .02 | .02 | −.02 | **.86** |
| 11 | .04 | −.02 | −.01 | **.84** |
| 12 | −.05 | .01 | .03 | **.93** |

*Note.* N = 360. Estimator = WLSMV; rotation = GEOMIN. Items 1–12 correspond to FO_pre_item1–FO_pre_item12. Primary loadings are shown in bold. All primary loadings ≥ .79; no cross-loading exceeded .10 in absolute value (the only significant cross-loading was Item 4 on F1, $\lambda = .10$, $p < .05$). The pattern of loadings is consistent with the hypothesized four-factor structure (Items 1–3 → F1; 4–6 → F2; 7–9 → F3; 10–12 → F4).

**Panel B. Factor correlations**

| Factor | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1. F1 | — | | | |
| 2. F2 | .39* | — | | |
| 3. F3 | .32* | .30* | — | |

| Factor | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 4. F4 | .28* | .30* | .23* | — |

*Note.* All inter-factor correlations were statistically significant ($p < .05$) and ranged from .23 to .39, all below the .70 discriminant threshold, supporting discriminant validity among the four factors.

A parallel-analysis analogue (1- to 4-factor EFA solutions under WLSMV with GEOMIN rotation) and the eigenvalue rule converged on a **four-factor** structure: the first four eigenvalues exceeded 1.00 (4.79, 2.00, 1.83, 1.55), whereas the fifth dropped sharply to 0.34. Model fit was excellent at four factors ($\chi^2(24) = 33.17$, $p = .101$; RMSEA = .033; CFI = .998; SRMR = .014), whereas the three-factor solution still showed poor fit (RMSEA = .234, CFI = .874). Rotated loadings were clean (all primary loadings ≥ .79, no cross-loading > .10) and factor correlations ranged from .23 to .39, all below the .70 discriminant threshold. The four-factor structure is thus supported.

**Table 9**

*Dimensionality of the FO Post-test: Parallel-Analysis Analogue and Factor Retention*

**Panel A. Model fit of 1- to 4-factor EFA solutions**

| Factors | $\chi^2$ | df | p | $\chi^2$/df | CFI | TLI | RMSEA [90% CI] | SRMR |
|---|---|---|---|---|---|---|---|---|
| 1 | 2098.79 | 54 | < .001 | 38.87 | .757 | .703 | .324 [.313, .336] | .303 |
| 2 | 1239.03 | 43 | < .001 | 28.81 | .858 | .782 | .278 [.265, .291] | .226 |
| 3 | 761.97 | 33 | < .001 | 23.09 | .913 | .827 | .248 [.233, .263] | .143 |
| 4 | 14.85 | 24 | .925 | 0.62 | 1.000 | 1.003 | .000 [.000, .014] | .007 |

**Panel B. First six sample eigenvalues**

| Component | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Eigenvalue | 5.15 | 1.97 | 1.82 | 1.64 | 0.27 | 0.23 |

*Note.* N = 360. Estimator = WLSMV; rotation = GEOMIN. Factors were retained following the parallel-analysis analogue in Mplus (comparison of 1- to 4-factor solutions). The first four eigenvalues exceeded 1.00 (5.15, 1.97, 1.82, 1.64), whereas the fifth dropped sharply to 0.27. Model fit improved substantially from three to four factors (3-factor vs. 4-factor: $\Delta\chi^2(9) = 393.88$, $p < .001$; RMSEA = .000; CFI = 1.000; SRMR = .007), and the four-factor solution was not rejected by the $\chi^2$ test ($p = .925$). The 4-factor solution is therefore retained.

**Table 10**

*Four-Factor EFA of the FO Post-test: GEOMIN-Rotated Loadings and Factor Correlations*

**Panel A. Rotated factor loadings**

| Item | F1 | F2 | F3 | F4 |
|---|---|---|---|---|
| 1 | **.94** | −.02 | −.01 | .03 |
| 2 | **.88** | .02 | .00 | −.02 |
| 3 | **.91** | .02 | .02 | .00 |
| 4 | .01 | **.89** | .02 | .02 |
| 5 | .02 | **.98** | −.04 | −.05 |
| 6 | −.02 | **.80** | .05 | .06 |
| 7 | −.03 | −.01 | **.91** | −.03 |
| 8 | .03 | .03 | **.91** | −.01 |
| 9 | .01 | .00 | **.92** | .05 |
| 10 | .02 | .01 | .01 | **.86** |
| 11 | −.02 | .02 | .02 | **.87** |

| Item | F1 | F2 | F3 | F4 |
|---|---|---|---|---|
| 12 | .01 | −.01 | −.02 | **.95** |

*Note.* N = 360. Estimator = WLSMV; rotation = GEOMIN. Items 1–12 correspond to FO_post_item1–FO_post_item12. Primary loadings are shown in bold. All primary loadings ≥ .80; no cross-loading exceeded .06 in absolute value, and the only significant cross-loading was Item 5 on F4 ($\lambda = -.05$, $p < .05$). The pattern of loadings is consistent with the hypothesized four-factor structure (Items 1–3 → F1; 4–6 → F2; 7–9 → F3; 10–12 → F4).

**Panel B. Factor correlations**

| Factor | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1. F1 | — | | | |
| 2. F2 | .35* | — | | |
| 3. F3 | .32* | .38* | — | |
| 4. F4 | .27* | .29* | .32* | — |

*Note.* All inter-factor correlations were statistically significant (*p* < .05) and ranged from .27 to .38, all below the .70 discriminant threshold, supporting discriminant validity among the four factors.

A parallel-analysis analogue (1- to 4-factor EFA solutions under WLSMV with GEOMIN rotation) and the eigenvalue rule converged on a **four-factor** structure: the first four eigenvalues exceeded 1.00 (5.15, 1.97, 1.82, 1.64), whereas the fifth dropped sharply to 0.27. Model fit was excellent at four factors ($\chi^2(24) = 14.85$, $p = .925$; RMSEA = .000; CFI = 1.000; SRMR = .007), whereas the three-factor solution still showed poor fit (RMSEA = .248, CFI = .913). Rotated loadings were clean (all primary loadings ≥ .80; only a trivial cross-loading of −.05 emerged for Item 5 on F4) and factor correlations ranged from .27 to .38, all below the .70 discriminant threshold. The four-factor structure is thus supported.

**Table 11**
*Confirmatory Factor Analysis of the FO Pre-test: Four-Factor Model*

**Panel A. Model fit**

| Model | $\chi^2$ | df | p | $\chi^2$/df | CFI | TLI | RMSEA [90% CI] | SRMR |
|---|---|---|---|---|---|---|---|---|
| 4-factor CFA | 40.93 | 48 | .755 | 0.85 | 1.000 | 1.002 | .000 [.000, .025] | .019 |

*Note.* N = 360. Estimator = WLSMV; parameterization = DELTA; link = probit. The model specified Items 1–3 → F1, 4–6 → F2, 7–9 → F3, and 10–12 → F4. All fit indices met conventional criteria (CFI ≥ .95, TLI ≥ .95, RMSEA ≤ .06, SRMR ≤ .08), and the model was not rejected by the $\chi^2$ test (p = .755). No modification indices exceeded the threshold of 10. The four-factor structure is therefore supported.

**Panel B. Standardized factor loadings (STDYX)**

| Item | F1 | F2 | F3 | F4 |
|---|---|---|---|---|
| 1 | **.91** | | | |
| 2 | **.83** | | | |
| 3 | **.89** | | | |
| 4 | | **.84** | | |
| 5 | | **.87** | | |
| 6 | | **.86** | | |
| 7 | | | **.87** | |
| 8 | | | **.93** | |
| 9 | | | **.89** | |
| 10 | | | | **.87** |
| 11 | | | | **.84** |
| 12 | | | | **.92** |

*Note.* All loadings are standardized (STDYX) and significant at $p < .001$. Loadings range from .83 to .93. The marker item in each factor was fixed to 1.00 in the unstandardized solution.

**Panel C. Latent factor correlations (STDYX)**

| Factor | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1. F1 | — | | | |
| 2. F2 | .43 | — | | |
| 3. F3 | .34 | .32 | — | |
| 4. F4 | .30 | .32 | .25 | — |

*Note.* All correlations are standardized (STDYX) and significant at $p < .001$. Correlations range from .25 to .43; all are below the .70 discriminant-validity threshold, supporting the distinctiveness of the four factors.

**Panel D. $R^2$ for observed items**

| Item | $R^2$ | Item | $R^2$ | Item | $R^2$ |
|---|---|---|---|---|---|
| 1 | .83 | 5 | .76 | 9 | .80 |
| 2 | .69 | 6 | .74 | 10 | .76 |
| 3 | .79 | 7 | .76 | 11 | .71 |
| 4 | .70 | 8 | .86 | 12 | .84 |

*Note.* $R^2$ = proportion of variance in each item explained by its latent factor. Values range from .69 to .86.

A four-factor CFA was estimated with WLSMV using the hypothesized item-to-factor assignment (Items 1–3, 4–6, 7–9, and 10–12). The model fit the data well, $\chi^2(48) = 40.93$, $p = .755$; CFI = 1.000; TLI = 1.002; RMSEA = .000, 90% CI [.000, .025]; SRMR = .019. All standardized loadings were substantial (.83–.93, all $p < .001$), and latent factor correlations ranged from .25 to .43, all below the .70 discriminant threshold. No modification indices exceeded 10. The four-factor measurement model was therefore retained.

**Table 12**

*Confirmatory Factor Analysis of the FO Post-test: Four-Factor Model*

**Panel A. Model fit**

| Model | $\chi^2$ | df | p | $\chi^2$/df | CFI | TLI | RMSEA [90% CI] | SRMR |
|---|---|---|---|---|---|---|---|---|
| 4-factor CFA | 49.41 | 48 | .417 | 1.03 | 1.000 | 1.000 | .009 [.000, .036] | .019 |

*Note.* N = 360. Estimator = WLSMV; parameterization = DELTA; link = probit. The model specified Items 1–3 → F1, 4–6 → F2, 7–9 → F3, and 10–12 → F4. All fit indices met conventional criteria (CFI ≥ .95, TLI ≥ .95, RMSEA ≤ .06, SRMR ≤ .08), and the model was not rejected by the $\chi^2$ test (p = .417). Only one trivial modification index emerged (M.I. = 10.21 for a cross-loading of Item 5 on F4; StdYX E.P.C. = −.11), which does not alter the substantive conclusions. The four-factor structure is therefore supported.

**Panel B. Standardized factor loadings (STDYX)**

| Item | F1 | F2 | F3 | F4 |
|---|---|---|---|---|
| 1 | **.94** | | | |
| 2 | **.88** | | | |
| 3 | **.92** | | | |
| 4 | | **.92** | | |
| 5 | | **.93** | | |
| 6 | | **.84** | | |
| 7 | | | **.87** | |
| 8 | | | **.93** | |
| 9 | | | **.94** | |

| Item | F1 | F2 | F3 | F4 |
|---|---|---|---|---|
| 10 | | | | **.89** |
| 11 | | | | **.88** |
| 12 | | | | **.94** |

*Note.* All loadings are standardized (STDYX) and significant at $p < .001$. Loadings range from .84 to .94. The marker item in each factor was fixed to 1.00 in the unstandardized solution.

**Panel C. Latent factor correlations (STDYX)**

| Factor | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1. F1 | — | | | |
| 2. F2 | .37 | — | | |
| 3. F3 | .33 | .40 | — | |
| 4. F4 | .28 | .31 | .34 | — |

*Note.* All correlations are standardized (STDYX) and significant at $p < .001$. Correlations range from .28 to .40; all are below the .70 discriminant-validity threshold, supporting the distinctiveness of the four factors.

**Panel D. $R^2$ for observed items**

| Item | $R^2$ | Item | $R^2$ | Item | $R^2$ |
|---|---|---|---|---|---|
| 1 | .88 | 5 | .86 | 9 | .89 |
| 2 | .78 | 6 | .71 | 10 | .79 |
| 3 | .85 | 7 | .76 | 11 | .78 |
| 4 | .84 | 8 | .87 | 12 | .87 |

*Note.* $R^2$ = proportion of variance in each item explained by its latent factor. Values range from .71 to .89.

A four-factor CFA was estimated with WLSMV using the hypothesized item-to-factor assignment (Items 1–3, 4–6, 7–9, and 10–12). The model fit the data well, $\chi^2(48) = 49.41$, $p = .417$; CFI = 1.000; TLI = 1.000; RMSEA = .009, 90% CI [.000, .036]; SRMR = .019. All standardized loadings were substantial (.84–.94, all $p < .001$), and latent factor correlations ranged from .28 to .40, all below the .70 discriminant threshold. Only a trivial modification index (M.I. = 10.21, StdYX E.P.C. = −.11) emerged, which does not change the substantive conclusions. The four-factor measurement model was therefore retained.

**Table 13**

*Internal Consistency Reliability of the FB Scale at Pre-test and Post-test*

| Occasion | Dimension | Items | *N* | Cronbach's α | McDonald's ω |
|---|---|---|---|---|---|
| **Pre-test** | | | | | |
| | Overall FB (24 items) | 24 | 360 | .903 | .904 |
| | F1 (Items 1–5) | 5 | 360 | .879 | .880 |
| | F2 (Items 6–10) | 5 | 360 | .894 | .894 |
| | F3 (Items 11–14) | 4 | 360 | .879 | .880 |
| | F4 (Items 15–19) | 5 | 360 | .903 | .904 |
| | F5 (Items 20–24) | 5 | 360 | .919 | .920 |
| **Post-test** | | | | | |
| | Overall FB (24 items) | 24 | 360 | .891 | .892 |

| Occasion | Dimension | Items | *N* | Cronbach's α | McDonald's ω |
|---|---|---|---|---|---|
| | F1 (Items 1–5) | 5 | 360 | .930 | .930 |
| | F2 (Items 6–10) | 5 | 360 | .922 | .923 |
| | F3 (Items 11–14) | 4 | 360 | .914 | .915 |
| | F4 (Items 15–19) | 5 | 360 | .932 | .932 |
| | F5 (Items 20–24) | 5 | 360 | .943 | .943 |

*Note.* N = 360. Cronbach's α and McDonald's ω were computed for the overall 24-item FB scale and for each of the five factors at both time points.

**Table 14**. *Internal Consistency Reliability of the FO Scale at Pre-test and Post-test*

| Occasion | Dimension | Items | *N* | Cronbach's α | McDonald's ω |
|---|---|---|---|---|---|
| **Pre-test** | | | | | |
| | Overall FO (12 items) | 12 | 360 | .840 | .841 |
| | F1 (Items 1–3) | 3 | 360 | .871 | .872 |
| | F2 (Items 4–6) | 3 | 360 | .849 | .850 |
| | F3 (Items 7–9) | 3 | 360 | .885 | .885 |
| | F4 (Items 10–12) | 3 | 360 | .871 | .872 |
| **Post-test** | | | | | |

| Occasion | Dimension | Items | *N* | Cronbach's α | McDonald's ω |
|---|---|---|---|---|---|
| | Overall FO (12 items) | 12 | 360 | .863 | .864 |
| | F1 (Items 1–3) | 3 | 360 | .909 | .910 |
| | F2 (Items 4–6) | 3 | 360 | .892 | .894 |
| | F3 (Items 7–9) | 3 | 360 | .909 | .910 |
| | F4 (Items 10–12) | 3 | 360 | .901 | .902 |

*Note.* N = 360. Cronbach's α and McDonald's ω were computed for the overall 12-item FO scale and for each of the four factors at both time points.

**Table 15**

*Convergent Validity of the FB Scale: Composite Reliability (CR) and Average Variance Extracted (AVE)*

| Occasion | Factor | Items | CR | AVE |
|---|---|---|---|---|
| **Pre-test** | | | | |
| | F1 (Items 1–5) | 5 | .904 | .653 |
| | F2 (Items 6–10) | 5 | .914 | .682 |
| | F3 (Items 11–14) | 4 | .905 | .706 |
| | F4 (Items 15–19) | 5 | .926 | .715 |
| | F5 (Items 20–24) | 5 | .940 | .759 |
| **Post-test** | | | | |
| | F1 (Items 1–5) | 5 | .945 | .773 |

| Occasion | Factor | Items | CR | AVE |
|---|---|---|---|---|
| | F2 (Items 6–10) | 5 | .938 | .753 |
| | F3 (Items 11–14) | 4 | .932 | .775 |
| | F4 (Items 15–19) | 5 | .947 | .782 |
| | F5 (Items 20–24) | 5 | .958 | .821 |

*Note.* N = 360. CR = composite reliability; AVE = average variance extracted.

**Table 16**

*Discriminant Validity of the FB Scale: Fornell–Larcker Criterion and HTMT*

**Panel A. Fornell–Larcker criterion (pre-test)**

| Factor | F1 | F2 | F3 | F4 | F5 |
|---|---|---|---|---|---|
| F1 | **.808** | | | | |
| F2 | .302 | **.826** | | | |
| F3 | .260 | .273 | **.840** | | |
| F4 | .328 | .379 | .371 | **.845** | |
| F5 | .322 | .403 | .211 | .384 | **.871** |

**Panel B. Fornell–Larcker criterion (post-test)**

| Factor | F1 | F2 | F3 | F4 | F5 |
|---|---|---|---|---|---|
| F1 | **.879** | | | | |
| F2 | .354 | **.868** | | | |
| F3 | .381 | .368 | **.880** | | |
| F4 | .121 | .166 | .194 | **.884** | |

| Factor | F1 | F2 | F3 | F4 | F5 |
|---|---|---|---|---|---|
| F5 | .139 | .154 | .030 | .315 | **.906** |

**Panel C. HTMT (pre-test)**

| Factor | F1 | F2 | F3 | F4 | F5 |
|---|---|---|---|---|---|
| F1 | — | | | | |
| F2 | .296 | — | | | |
| F3 | .253 | .253 | — | | |
| F4 | .321 | .376 | .361 | — | |
| F5 | .316 | .394 | .196 | .382 | — |

**Panel D. HTMT (post-test)**

| Factor | F1 | F2 | F3 | F4 | F5 |
|---|---|---|---|---|---|
| F1 | — | | | | |
| F2 | .346 | — | | | |
| F3 | .376 | .367 | — | | |
| F4 | .111 | .164 | .187 | — | |
| F5 | .121 | .146 | .047 | .315 | — |

*Note.* N = 360.

For the five-factor FB scale (N = 360), the Fornell–Larcker criterion was satisfied at both occasions: each factor's square root of AVE exceeded all inter-factor correlations (pre-test diagonal .808–.871 vs. $|r| \leq .403$; post-test .868–.906 vs. $|r| \leq .381$). HTMT values were uniformly low, with maximum values of .394 at pre-test

and .376 at post-test, well below conventional thresholds, confirming discriminant validity.

**Table 17. Discriminant Validity of the FB Scale: Fornell–Larcker Criterion and HTMT**

**Panel A. Fornell–Larcker criterion (pre-test)**

| Factor | F1 | F2 | F3 | F4 |
|---|---|---|---|---|
| F1 | 0.877 | | | |
| F2 | 0.427 | 0.857 | | |
| F3 | 0.343 | 0.321 | 0.896 | |
| F4 | 0.298 | 0.316 | 0.245 | 0.878 |

Panel B. Fornell–Larcker criterion (post-test)

| Factor | F1 | F2 | F3 | F4 |
|---|---|---|---|---|
| F1 | 0.914 | | | |
| F2 | 0.367 | 0.896 | | |
| F3 | 0.333 | 0.398 | 0.917 | |
| F4 | 0.283 | 0.312 | 0.337 | 0.902 |

Panel C. HTMT (pre-test)

| Factor | F1 | F2 | F3 | F4 |
|---|---|---|---|---|
| F1 | 1.000 | | | |
| F2 | 0.421 | 1.000 | | |
| F3 | 0.339 | 0.318 | 1.000 | |
| F4 | 0.298 | 0.312 | 0.243 | 1.000 |

Panel D. HTMT (post-test)

| Factor | F1 | F2 | F3 | F4 |
|---|---|---|---|---|
| F1 | 1.000 | | | |
| F2 | 0.366 | 1.000 | | |
| F3 | 0.328 | 0.396 | 1.000 | |
| F4 | 0.281 | 0.312 | 0.332 | 1.000 |

Across the pre-test and post-test, the square root of AVE for every factor exceeded its correlations with the other factors, supporting Fornell–Larcker discriminant validity at both occasions. All HTMT values were below conventional cutoffs at pre-test and post-test;

the maximum HTMT was 0.421 at pre-test and 0.398 at post-test, respectively, indicating satisfactory discriminant validity.

**Table 19. Inter-rater Reliability of Writing Scores Across Groups and Measurement Occasions**

| Group | Measurement | N | ICC (2,1) | 95% CI | F |
|---|---|---|---|---|---|
| 1 | Pre-test | 120 | .900 | [.859, .929] | 19.056 |
| 1 | Post-test | 120 | .891 | [.846, .923] | 17.948 |
| 2 | Pre-test | 120 | .871 | [.819, .908] | 14.591 |
| 2 | Post-test | 120 | .885 | [.839, .919] | 16.366 |
| 3 | Pre-test | 120 | .843 | [.782, .888] | 11.634 |
| 3 | Post-test | 120 | .912 | [.876, .938] | 21.620 |

Note. ICC = intraclass correlation coefficient; CI = confidence interval. ICCs were estimated using a two-way random-effects model with absolute agreement and single measures. All ICC estimates were statistically significant ($p < .001$).

Inter-rater reliability was consistently high across all three groups and both measurement occasions (ICC = .843–.912), indicating strong agreement between the two raters and supporting the reliability of the writing scores.

**Research Findings**

**Table A. Descriptive Statistics by Group and Measurement Occasion**

*Table A reports descriptive statistics for all outcome variables separately by group and measurement occasion. It includes sample size, mean with standard deviation, observed minimum and maximum, skewness, and kurtosis. These indices characterize central tendency, dispersion, range, and distributional shape before inferential analyses.*

| Outcome | Group | Time | N | M (SD) | Min | Max | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|---|
| FB_F1 | G1 | Pre | 120 | 2.79 (0.79) | 1.00 | 4.40 | -0.152 | -0.986 |
| FB_F1 | G1 | Post | 120 | 3.86 (0.86) | 1.60 | 6.00 | -0.109 | 0.005 |

| Outcome | Group | Time | N | M (SD) | Min | Max | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|---|
| FB_F1 | G2 | Pre | 120 | 2.86 (0.85) | 1.20 | 5.00 | 0.091 | -0.638 |
| FB_F1 | G2 | Post | 120 | 3.27 (1.03) | 1.00 | 5.80 | -0.018 | -0.352 |
| FB_F1 | G3 | Pre | 120 | 2.80 (0.84) | 1.00 | 5.40 | 0.099 | 0.146 |
| FB_F1 | G3 | Post | 120 | 2.97 (1.05) | 1.00 | 5.60 | 0.085 | -0.511 |
| FB_F2 | G1 | Pre | 120 | 2.87 (0.92) | 1.00 | 5.20 | 0.193 | -0.315 |
| FB_F2 | G1 | Post | 120 | 3.82 (1.04) | 1.40 | 6.00 | -0.067 | -0.500 |
| FB_F2 | G2 | Pre | 120 | 2.92 (0.85) | 1.20 | 5.40 | 0.224 | -0.255 |
| FB_F2 | G2 | Post | 120 | 3.26 (0.95) | 1.20 | 6.00 | 0.164 | -0.361 |
| FB_F2 | G3 | Pre | 120 | 2.90 (0.86) | 1.00 | 5.20 | 0.113 | -0.274 |
| FB_F2 | G3 | Post | 120 | 3.03 (0.97) | 1.00 | 5.40 | 0.252 | -0.453 |
| FB_F3 | G1 | Pre | 120 | 2.86 (0.85) | 1.00 | 4.75 | -0.063 | -0.639 |
| FB_F3 | G1 | Post | 120 | 3.85 (0.99) | 1.50 | 6.00 | -0.148 | -0.579 |

| Outcome | Group | Time | N | M (SD) | Min | Max | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|---|
| FB_F3 | G2 | Pre | 120 | 2.83 (0.91) | 1.00 | 5.00 | 0.107 | -0.731 |
| FB_F3 | G2 | Post | 120 | 3.17 (0.96) | 1.00 | 5.75 | 0.104 | -0.135 |
| FB_F3 | G3 | Pre | 120 | 2.82 (0.92) | 1.00 | 4.75 | -0.098 | -0.880 |
| FB_F3 | G3 | Post | 120 | 2.97 (1.10) | 1.00 | 5.50 | 0.071 | -0.695 |
| FB_F4 | G1 | Pre | 120 | 2.86 (0.80) | 1.00 | 4.60 | -0.063 | -0.602 |
| FB_F4 | G1 | Post | 120 | 2.92 (0.94) | 1.00 | 5.00 | 0.068 | -0.732 |
| FB_F4 | G2 | Pre | 120 | 2.88 (0.91) | 1.20 | 5.00 | 0.206 | -0.715 |
| FB_F4 | G2 | Post | 120 | 2.96 (1.07) | 1.00 | 5.60 | 0.151 | -0.588 |
| FB_F4 | G3 | Pre | 120 | 2.91 (0.89) | 1.00 | 5.20 | 0.028 | -0.461 |
| FB_F4 | G3 | Post | 120 | 3.74 (1.06) | 1.00 | 6.00 | -0.166 | -0.643 |
| FB_F5 | G1 | Pre | 120 | 2.79 (0.85) | 1.00 | 5.00 | 0.027 | -0.455 |
| FB_F5 | G1 | Post | 120 | 2.89 (0.97) | 1.00 | 5.60 | -0.012 | -0.520 |

| Outcome | Group | Time | N | M (SD) | Min | Max | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|---|
| FB_F5 | G2 | Pre | 120 | 2.81 (0.91) | 1.00 | 5.00 | 0.152 | -0.552 |
| FB_F5 | G2 | Post | 120 | 2.90 (1.03) | 1.00 | 5.60 | 0.064 | -0.338 |
| FB_F5 | G3 | Pre | 120 | 2.81 (0.91) | 1.00 | 5.80 | 0.703 | 0.621 |
| FB_F5 | G3 | Post | 120 | 3.64 (1.04) | 1.40 | 6.00 | 0.077 | -0.718 |
| FO_F1 | G1 | Pre | 120 | 2.39 (0.89) | 1.00 | 4.67 | 0.122 | -0.671 |
| FO_F1 | G1 | Post | 120 | 3.19 (0.94) | 1.00 | 5.00 | -0.347 | -0.553 |
| FO_F1 | G2 | Pre | 120 | 2.34 (0.74) | 1.00 | 4.00 | 0.283 | -0.618 |
| FO_F1 | G2 | Post | 120 | 2.64 (0.87) | 1.00 | 4.67 | -0.046 | -0.580 |
| FO_F1 | G3 | Pre | 120 | 2.31 (0.89) | 1.00 | 5.00 | 0.396 | -0.358 |
| FO_F1 | G3 | Post | 120 | 2.29 (0.96) | 1.00 | 5.00 | 0.624 | -0.280 |
| FO_F2 | G1 | Pre | 120 | 2.35 (0.80) | 1.00 | 4.00 | -0.045 | -0.884 |
| FO_F2 | G1 | Post | 120 | 3.27 (0.87) | 1.33 | 5.00 | -0.188 | -0.803 |

| Outcome | Group | Time | N | M (SD) | Min | Max | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|---|
| FO_F2 | G2 | Pre | 120 | 2.39 (0.76) | 1.00 | 4.33 | 0.124 | -0.570 |
| FO_F2 | G2 | Post | 120 | 2.68 (0.97) | 1.00 | 5.00 | 0.250 | -0.669 |
| FO_F2 | G3 | Pre | 120 | 2.37 (0.77) | 1.00 | 4.00 | 0.222 | -0.713 |
| FO_F2 | G3 | Post | 120 | 2.39 (0.82) | 1.00 | 4.67 | 0.279 | -0.378 |
| FO_F3 | G1 | Pre | 120 | 2.32 (0.72) | 1.00 | 4.33 | 0.228 | -0.086 |
| FO_F3 | G1 | Post | 120 | 3.22 (0.83) | 1.00 | 5.00 | -0.254 | -0.085 |
| FO_F3 | G2 | Pre | 120 | 2.36 (0.88) | 1.00 | 4.33 | 0.116 | -0.779 |
| FO_F3 | G2 | Post | 120 | 2.66 (0.89) | 1.00 | 5.00 | 0.165 | -0.409 |
| FO_F3 | G3 | Pre | 120 | 2.41 (0.84) | 1.00 | 4.33 | 0.230 | -0.707 |
| FO_F3 | G3 | Post | 120 | 2.37 (0.89) | 1.00 | 4.33 | 0.175 | -0.873 |
| FO_F4 | G1 | Pre | 120 | 2.36 (0.82) | 1.00 | 4.67 | 0.300 | -0.538 |
| FO_F4 | G1 | Post | 120 | 3.23 (1.04) | 1.00 | 5.00 | -0.161 | -0.911 |

| Outcome | Group | Time | N | M (SD) | Min | Max | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|---|
| FO_F4 | G2 | Pre | 120 | 2.35 (0.90) | 1.00 | 4.67 | 0.344 | -0.564 |
| FO_F4 | G2 | Post | 120 | 2.64 (0.99) | 1.00 | 4.67 | -0.036 | -0.922 |
| FO_F4 | G3 | Pre | 120 | 2.35 (0.81) | 1.00 | 4.00 | 0.086 | -0.927 |
| FO_F4 | G3 | Post | 120 | 2.46 (0.87) | 1.00 | 5.00 | 0.206 | -0.422 |
| Writing | G1 | Pre | 120 | 12.72 (4.85) | 6.00 | 27.00 | 0.518 | -0.161 |
| Writing | G1 | Post | 120 | 17.35 (5.65) | 6.00 | 30.00 | -0.034 | -0.492 |
| Writing | G2 | Pre | 120 | 12.55 (4.61) | 6.00 | 23.50 | 0.361 | -0.857 |
| Writing | G2 | Post | 120 | 14.39 (5.63) | 6.00 | 30.00 | 0.502 | -0.364 |
| Writing | G3 | Pre | 120 | 12.44 (4.40) | 6.00 | 26.00 | 0.286 | -0.738 |
| Writing | G3 | Post | 120 | 13.00 (5.34) | 6.00 | 29.50 | 0.511 | -0.349 |

*Note.* N = 120 per group and occasion. M (SD) = mean and standard deviation. Skewness and kurtosis are sample-moment estimates.

**Table B. Baseline Group Differences**

*Table B summarizes baseline one-way analyses of variance comparing the three groups on each pre-test outcome. F ratios, degrees of freedom, p values, and eta-squared are reported. These tests evaluate pre-intervention equivalence; small eta-squared values indicate negligible between-group variance at baseline.*

| Outcome | F (df) | p | $\eta^2$ |
|---|---|---|---|
| FB_F1 | 0.280 (2, 357) | 0.756 | 0.002 |
| FB_F2 | 0.091 (2, 357) | 0.913 | 0.001 |
| FB_F3 | 0.063 (2, 357) | 0.939 | 0.000 |
| FB_F4 | 0.114 (2, 357) | 0.892 | 0.001 |
| FB_F5 | 0.009 (2, 357) | 0.991 | 0.000 |
| FO_F1 | 0.288 (2, 357) | 0.750 | 0.002 |
| FO_F2 | 0.099 (2, 357) | 0.906 | 0.001 |
| FO_F3 | 0.382 (2, 357) | 0.683 | 0.002 |
| FO_F4 | 0.003 (2, 357) | 0.997 | 0.000 |
| Writing | 0.114 (2, 357) | 0.892 | 0.001 |

**Table C. Main Effect of Group**

*Table C presents the main effect of Group from mixed-design analyses of variance for each outcome. F ratios, p values, partial eta-squared, Cohen's f, and partial omega-squared are shown. These effect-size indices quantify the magnitude of overall between-group differences across measurement occasions.*

| Outcome | F (df) | p | $\eta_p^2$ | Cohen's f | $\omega_p^2$ |
|---|---|---|---|---|---|
| FB_F1 | 7.967 (2, 357) | < .001 | 0.043 | 0.211 | 0.037 |
| FB_F2 | 5.951 (2, 357) | 0.003 | 0.032 | 0.183 | 0.027 |

| Outcome | F (df) | p | η_p² | Cohen's f | ω_p² |
|---|---|---|---|---|---|
| FB_F3 | 8.489 (2, 357) | < .001 | 0.045 | 0.218 | 0.040 |
| FB_F4 | 9.014 (2, 357) | < .001 | 0.048 | 0.225 | 0.043 |
| FB_F5 | 6.787 (2, 357) | 0.001 | 0.037 | 0.195 | 0.031 |
| FO_F1 | 10.526 (2, 357) | < .001 | 0.056 | 0.243 | 0.050 |
| FO_F2 | 9.268 (2, 357) | < .001 | 0.049 | 0.228 | 0.044 |
| FO_F3 | 7.352 (2, 357) | < .001 | 0.040 | 0.203 | 0.034 |
| FO_F4 | 6.895 (2, 357) | 0.001 | 0.037 | 0.197 | 0.032 |
| Writing | 7.175 (2, 357) | < .001 | 0.039 | 0.200 | 0.033 |

*Note.* $\eta_p^2$ = partial eta-squared; $\omega_p^2$ = partial omega-squared.

The main effect of Group was statistically significant for every outcome, indicating that the three groups differed overall across measurement occasions. Effect sizes, however, were consistently small: partial eta-squared and partial omega-squared both fell in the small range, and Cohen's f mirrored this pattern, showing that group membership accounted for only a modest share of residual variance despite reaching significance. Because the ten outcomes were tested without adjustment for multiple comparisons, some significant p values may reflect inflated Type I error. The group effect is therefore reliable in direction but modest in magnitude, and should be interpreted alongside effect sizes rather than significance alone.

**Table D. Main Effect of Time**

*Table D reports the main effect of Time from the same mixed-design models. It includes F ratios, p values, partial eta-squared, Cohen's f, and partial omega-squared. These indices describe the overall pre-to-post change across groups, irrespective of group membership.*

| Outcome | F (df) | p | η_p² | Cohen's f | ω_p² |
|---|---|---|---|---|---|
| FB_F1 | 276.166 (1, 357) | < .001 | 0.436 | 0.880 | 0.434 |

| Outcome | F (df) | p | η_p² | Cohen's f | ω_p² |
|---|---|---|---|---|---|
| FB_F2 | 191.017 (1, 357) | < .001 | 0.349 | 0.731 | 0.346 |
| FB_F3 | 213.265 (1, 357) | < .001 | 0.374 | 0.773 | 0.372 |
| FB_F4 | 99.730 (1, 357) | < .001 | 0.218 | 0.529 | 0.216 |
| FB_F5 | 115.787 (1, 357) | < .001 | 0.245 | 0.570 | 0.242 |
| FO_F1 | 138.266 (1, 357) | < .001 | 0.279 | 0.622 | 0.277 |
| FO_F2 | 172.458 (1, 357) | < .001 | 0.326 | 0.695 | 0.323 |
| FO_F3 | 148.740 (1, 357) | < .001 | 0.294 | 0.645 | 0.292 |
| FO_F4 | 165.468 (1, 357) | < .001 | 0.317 | 0.681 | 0.314 |
| Writing | 189.055 (1, 357) | < .001 | 0.346 | 0.728 | 0.344 |

*Note.* η_p² = partial eta-squared; ω_p² = partial omega-squared.

The main effect of Time was statistically significant for every outcome, indicating a consistent overall pre-to-post change across all groups combined. Unlike the group effect, these effects were moderate to large in magnitude: partial eta-squared and partial omega-squared both fell well above conventional thresholds for a large effect, and Cohen's f mirrored this pattern, showing that time accounted for a substantial share of residual variance. Because the ten outcomes were tested without adjustment for multiple comparisons, the unadjusted p values may slightly inflate Type I error; however, the uniformly large effect sizes make the time effect robust to this concern. The pre-to-post improvement is therefore both reliable and practically meaningful.

**Table E. Group × Time Interaction**

*Table E displays the Group × Time interaction effect for each outcome.*

| Outcome | Group | F (df) | p | η_p² | Cohen's f | ω_p² |
|---|---|---|---|---|---|---|
| **FB_F1** | G1 | 444.211 (1, 119) | < .001 | .789 | 1.932 | .786 |

| Outcome | Group | F (df) | p | η_p² | Cohen's f | ω_p² |
|---|---|---|---|---|---|---|
| | G2 | 39.810 (1, 119) | < .001 | .251 | 0.578 | .243 |
| | G3 | 9.372 (1, 119) | .003 | .073 | 0.281 | .065 |
| **FB_F2** | G1 | 267.654 (1, 119) | < .001 | .692 | 1.500 | .688 |
| | G2 | 38.422 (1, 119) | < .001 | .244 | 0.568 | .236 |
| | G3 | 4.190 (1, 119) | .043 | .034 | 0.188 | .026 |
| **FB_F3** | G1 | 319.274 (1, 119) | < .001 | .728 | 1.638 | .725 |
| | G2 | 36.543 (1, 119) | < .001 | .235 | 0.554 | .227 |
| | G3 | 5.703 (1, 119) | .019 | .046 | 0.219 | .037 |
| **FB_F4** | G1 | 1.274 (1, 119) | .261 | .011 | 0.103 | .002 |
| | G2 | 1.728 (1, 119) | .191 | .014 | 0.121 | .006 |
| | G3 | 223.975 (1, 119) | < .001 | .653 | 1.372 | .648 |
| **FB_F5** | G1 | 3.384 (1, 119) | .068 | .028 | 0.169 | .019 |

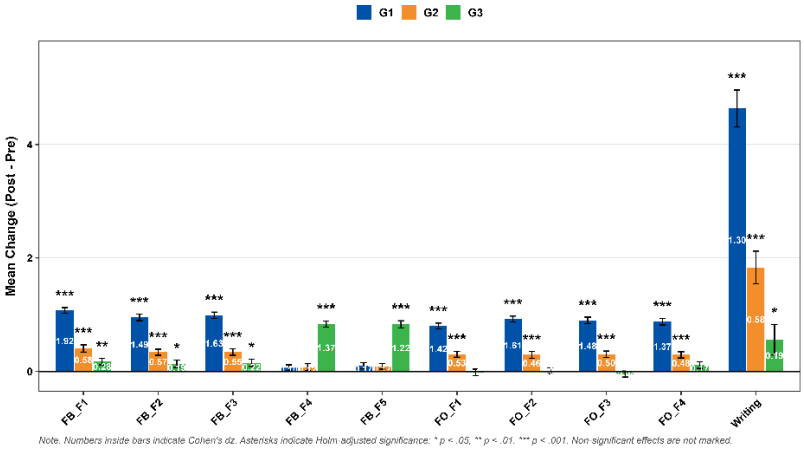
| Outcome | Group | F (df) | p | η_p² | Cohen's f | ω_p² |
|---|---|---|---|---|---|---|
| | G2 | 3.546 (1, 119) | .062 | .029 | 0.173 | .021 |
| | G3 | 178.048 (1, 119) | < .001 | .599 | 1.223 | .594 |
| **FO_F1** | G1 | 242.777 (1, 119) | < .001 | .671 | 1.428 | .666 |
| | G2 | 33.283 (1, 119) | < .001 | .219 | 0.529 | .211 |
| | G3 | 0.088 (1, 119) | .767 | .001 | 0.027 | .000 |
| **FO_F2** | G1 | 310.998 (1, 119) | < .001 | .723 | 1.617 | .719 |
| | G2 | 25.478 (1, 119) | < .001 | .176 | 0.463 | .168 |
| | G3 | 0.072 (1, 119) | .788 | .001 | 0.025 | .000 |
| **FO_F3** | G1 | 263.414 (1, 119) | < .001 | .689 | 1.488 | .684 |
| | G2 | 29.753 (1, 119) | < .001 | .200 | 0.500 | .192 |
| | G3 | 0.790 (1, 119) | .376 | .007 | 0.081 | .000 |
| **FO_F4** | G1 | 226.151 (1, 119) | < .001 | .655 | 1.379 | .650 |

| Outcome | Group | F (df) | p | η_p² | Cohen's f | ω_p² |
|---|---|---|---|---|---|---|
| | G2 | 27.393 (1, 119) | < .001 | .187 | 0.480 | .179 |
| | G3 | 3.634 (1, 119) | .059 | .030 | 0.175 | .021 |
| **Writing** | G1 | 202.434 (1, 119) | < .001 | .630 | 1.304 | .625 |
| | G2 | 40.999 (1, 119) | < .001 | .256 | 0.587 | .248 |
| | G3 | 4.270 (1, 119) | .041 | .035 | 0.189 | .026 |

*Note.* F tests represent simple effects of Time within each Group.

The Group × Time simple effects were decisive for the mechanism claim. Groups 1 and 2 showed large, significant improvements on FB_F1–F3 and all FO factors ($\eta_p^2$ = .176–.789, all p < .001) but no reliable change on FB_F4–F5 (p = .062–.261), whereas Group 3 showed the mirror image: no reliable change on FB_F1–F3 or the FO factors (p = .059–.788) but large improvements on FB_F4–F5 ($\eta_p^2$ = .599 and .653, both p < .001). This selective, disordinal pattern—not the modest overall Group main effect ($\eta_p^2$ = .032–.056; Cohen's f = .183–.243, approaching the medium benchmark)—is what justifies treating the three conditions as qualitatively distinct AI exposure regimes. The uniformly large Time main effects ($\eta_p^2$ = .218–.436) confirm that the pre–post design captured substantive change.

**Table F. Within-Group Pre–Post Changes**




Group 1 exhibited the largest mean changes, strong effect sizes (dz > 1), and highly significant p values across FB_F1 through FB_F3, all FO factors (FO_F1–FO_F4), and Writing. Group 2 showed moderate changes, medium effect sizes, and significant p values on these identical dimensions. Conversely, Group 3 demonstrated its primary significant effects with large effect sizes exclusively on FB_F4 and FB_F5, while showing negligible, non-significant changes on all FO factors. For FB_F4 and FB_F5, Groups 1 and 2 showed minimal, non-significant changes.

**Table G. Post-Test Pairwise Group Comparisons**

*Table G presents post-test pairwise group comparisons for each outcome, with Holm-adjusted p values. Mean differences are estimated as the first group minus the second; positive values indicate a higher post-test mean in the first group. Significant comparisons are those with adjusted p < .05; the table permits evaluation of which groups differed after intervention.*

| Outcome | Contrast | Estimate | p |
|---|---|---|---|
| FB_F1 | G1 − G2 | 0.593 | < .0001 |
| FB_F1 | G1 − G3 | 0.888 | < .0001 |
| FB_F1 | G2 − G3 | 0.295 | 0.0295 |
| FB_F2 | G1 − G2 | 0.565 | < .0001 |
| FB_F2 | G1 − G3 | 0.792 | < .0001 |
| FB_F2 | G2 − G3 | 0.227 | 0.0691 |

| Outcome | Contrast | Estimate | p |
|---|---|---|---|
| FB_F3 | G1 − G2 | 0.677 | < .0001 |
| FB_F3 | G1 − G3 | 0.877 | < .0001 |
| FB_F3 | G2 − G3 | 0.200 | 0.1352 |
| FB_F4 | G1 − G2 | -0.038 | 0.7678 |
| FB_F4 | G1 − G3 | -0.827 | < .0001 |
| FB_F4 | G2 − G3 | -0.788 | < .0001 |
| FB_F5 | G1 − G2 | -0.005 | 0.9690 |
| FB_F5 | G1 − G3 | -0.743 | < .0001 |
| FB_F5 | G2 − G3 | -0.738 | < .0001 |
| FO_F1 | G1 − G2 | 0.553 | < .0001 |
| FO_F1 | G1 − G3 | 0.897 | < .0001 |
| FO_F1 | G2 − G3 | 0.344 | 0.0038 |
| FO_F2 | G1 − G2 | 0.586 | < .0001 |
| FO_F2 | G1 − G3 | 0.883 | < .0001 |
| FO_F2 | G2 − G3 | 0.297 | 0.0111 |
| FO_F3 | G1 − G2 | 0.558 | < .0001 |
| FO_F3 | G1 − G3 | 0.856 | < .0001 |
| FO_F3 | G2 − G3 | 0.297 | 0.0104 |
| FO_F4 | G1 − G2 | 0.592 | < .0001 |

| Outcome | Contrast | Estimate | p |
|---|---|---|---|
| FO_F4 | G1 − G3 | 0.775 | < .0001 |
| FO_F4 | G2 − G3 | 0.183 | 0.1293 |
| Writing | G1 − G2 | 2.967 | 0.0001 |
| Writing | G1 − G3 | 4.358 | < .0001 |
| Writing | G2 − G3 | 1.392 | 0.0507 |

*Note.* p values are Holm-adjusted. Positive estimates indicate a higher post-test mean in the first-listed group. Comparisons with adjusted $p < .05$ are statistically significant.

# Discussion

G1 improved on FB_F1–F3, FO_F1–F4, and Writing, and exceeded both groups on FB_F1–F3, FO_F1–F3, and Writing. FB_F1 gains reflect opportunities to compare AI- and human-generated feedback and judge its objectivity (Nazaretsky et al., 2026); FB_F2 gains reflect AI suggestions combined with teachers' contextualized explanations and actionable guidance (Asadi et al., 2025; Zhu et al., 2025); FB_F3 gains reflect AI suggestions integrated with teachers' contextualized guidance on content, structure, and learner needs (Wei et al., 2026). FO_F1 rose because teachers contextualized AI suggestions into actionable revision guidance (Zhang et al., 2025); FO_F2 because peer dialogue required learners to evaluate, justify, and act on AI suggestions (Ates, 2026); FO_F3 because peer dialogue revealed how readers interpret writing (Tsao & Liu, 2026); FO_F4 because peer feedback supported generating, discussing, and reformulating feedback collaboratively (Yan et al., 2026). Writing improved because human involvement embedded AI in subject-specific tasks (Lim et al., 2026), built AI literacy (Zhang et al., 2024), and supplied verification strategies that moved learners beyond passive consumption (Borges et al., 2026). Yet G1 did not improve on FB_F4–F5 and scored below G3: individualized encouragement through AI-supported human feedback did not offset AI's relational deficit (Alsaiari et al., 2025), students evaluated human feedback more favorably (Nazaretsky et al., 2024), interpersonal recognition remained absent (Pi et al., 2026), positive language did not ensure engagement (Wang et al., 2026), and person-focused praise remained fragile (Skipper & Douglas, 2012); criticism stayed less relational (Soyoof et al., 2026), inconsistent (Takahashi & Kadowaki, 2026), and personally evaluative rather than task-guiding (Hattie & Timperley, 2007).

G2 improved moderately on the same dimensions but remained below G1. FB_F1 gains reflect cross-checking AI responses against criteria (Li et al., 2026), accuracy appraisal (Dai et al., 2024), detailed textual diagnosis (Glüsing et al., 2026), and proficiency-sensitive confirmation (Qi et al., 2026); FB_F2 gains reflect calibrated explicitness (Qi et al., 2026), subtheme organization (Engeness, 2025), iterative clarification (Sørhaug, 2026), explicit formative directions (Steiss et al., 2024), and higher-order revision suggestions (Zhang et al., 2026); FB_F3 gains reflect actionable application (Chong, 2025), calibrated prompting (Qi et al., 2026), and peer-comment refinement (Guo et al., 2024). FO_F1 rose through task-specific comments (Zhang et al., 2024), prompt refinement (Alaoui Mhamdi, 2026), active evaluation (Liu et al., 2026), feedback-seeking (Ng & Cheung, 2026), and observed application (Lu et al., 2026); FO_F2 through consultation (Stojanov et al., 2024), reflective reliance (Hou et al., 2025), and contextual integration (Tao et al., 2026); FO_F3 through rhetorical prompts (Tseng, 2026) and AI literacy (Reicho et al., 2026); FO_F4 through personalized guidance (Lyu & Salam, 2025), comment-to-revision translation (Yang et al., 2025), acceptance and literacy (Zhou & Wang, 2026), and LLM co-editing (Nguyen & Vo, 2026). Writing improved through idea elicitation (Li et al., 2025; Levine et al., 2025), selective rewriting (Li et al., 2025), repeated revision (Wang et al., 2026), brainstorming (Yao, 2026), and iterative integration (Nguyen et al., 2024). FB_F4–F5 showed no gains for the same relational reasons as in G1.

G3 improved only on FB_F4–F5 and was highest there: individualized encouragement operated without AI's relational deficit (Alsaiari et al., 2025), human feedback was evaluated more favorably (Nazaretsky et al., 2024), and criticism was accompanied by personalized emotional support and task guidance (Soyoof et al., 2026; Hattie & Timperley, 2007). Its non-gains on FB_F1–F3, FO, and Writing are consistent with the absence of the AI-mediated mechanisms detailed above.

The Group × Time simple effects showed that the two AI-exposed groups improved on the core FB and FO factors but not on FB_F4–F5, whereas the relational-feedback group showed the reverse. This disordinal, largely non-overlapping pattern, rather than differences in degree, is what justified treating the three conditions as qualitatively distinct AI exposure regimes. These contrasts align with the argument that human involvement raises AI exposure quality by embedding AI in subject-specific tasks (Lim et al., 2026), building teacher-led AI literacy (Zhang et al., 2024), and supplying verification strategies that move learners beyond passive consumption (Borges et al., 2026). G1 therefore represents higher-quality AI exposure, G2 AI mechanisms without quality enhancement, and G3 relational human feedback without AI exposure. The statistically robust but modest overall Group effect, together with the large Time effect, further indicates substantial change that was condition-dependent on the quality and existence of AI exposure.

**Conclusion**

This study tested the dimensional structures of teacher feedback and feedback orientation and examined how those dimensions changed under AI-only, AI–human, and human-only conditions. Both structures were empirically supported. AI–human feedback yielded the broadest gains, AI-only moderate gains, and human-only gains confined to praise and criticism. These findings suggest that teacher mediation is integral to effective AI integration, while relational human feedback remains essential for evaluative dimensions. The study contributes by validating the dimensional structures and reframing AI–human feedback as higher-quality AI exposure. Its single-context Chinese L1 writing sample and short intervention limit generalizability. Future research should adopt longitudinal and cross-context designs and process-level measures of feedback engagement.

## Ethics Statement

Ethical approval for this study was granted by the Ethics Committee of Minnan University of Science and Technology (No. KYLL2025-08-26-3). Written consent was obtained from all participating students after they had been informed of the study's purpose and procedures. Class teachers provided consent, as documented by their signatures. Parents or legal guardians were informed of the study by the class teachers and did not raise any objections. The study was conducted in accordance with the applicable institutional ethical requirements and relevant ethical principles for research involving human participants.

## Conflict of Interest

The authors declare that they have no conflict of interest.


## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Social Science Fund of China (General Project, Grant No. 20BMZ078).


## Data Availability Statement

The datasets (https://doi.org/10.5281/zenodo.23984810) generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

## Notes on contributor


*Augustine Zhu SiYi* (First Author): is a doctoral student in Curriculum Studies at The Education University of Hong Kong. He was responsible for the conceptualization, research design, literature review, methodology, data analysis, interpretation of findings, and original manuscript preparation. Contact: s1147049@s.eduhk.hk.
*Zhu JinXin* (Corresponding Author): is an assistant professor at The Education

University of Hong Kong with a PhD within the same university. He was responsible for supervision, and revision. Contact: jxzhu@eduhk.hk.
*Zhu JianJun* (Second Author): is a professor in Minnan University of Science and Technology with a PhD in Eastern China Normal University was responsible for data collection and ethical approval application. Contact: chinesezhu@126.com.
**ORCID**
Augustine Zhu SiYi https://orcid.org/0009-0000-8193-3192
Zhu JinXin https://orcid.org/0000-0001-5632-4440
Zhu JianJun https://orcid.org/0009-0002-5114-2626

## References

Alsaiari, O., Baghaei, N., Lahza, H., Lodge, J. M., Boden, M., & Khosravi, H. (2025). Emotionally enriched AI-generated feedback: Supporting student well-being without compromising learning. Computers & Education, 239, Article 105363. https://doi.org/10.1016/j.compedu.2025.105363

Alaoui Mhamdi, N. (2026). AI-mediated feedback in English learning: Moroccan graduate students' accounts of motivation, autonomy, and trust conditions. Computers and Education Open, 10, Article 100356. https://doi.org/10.1016/j.caeo.2026.100356

Ates, H. (2026). Human-centered GenAI feedback design in higher education: A multisite experiment on direct, reflective, and hybrid approaches to scientific argumentation. International Journal of Educational Technology in Higher Education, 23, Article 38. https://doi.org/10.1186/s41239-026-00614-9

Asadi, M., Ebadi, S., & Mohammadi, L. (2025). The impact of integrating ChatGPT with teachers' feedback on EFL writing skills. Thinking Skills and Creativity, 56, Article 101766. https://doi.org/10.1016/j.tsc.2025.101766

Borges, I., Sebastião, C., Caridade, C., & Pereira, V. (2026). From ChatGPT to classroom learning: Exploring the role of teacher mediation in AI-supported education. Australasian Journal of Educational Technology, 42(2), 116-137.

Chong, H. (2025). Exploring the role of AI-assisted revision with instructional scaffolding: A comparative study among Korean university students. Language Research, 41(1), 41–58. https://doi.org/10.18627/jslg.41.1.202505.41

Chen, Z., Wei, W., Zhu, X., & Yao, Y. (2025). Unpacking the rejection of L2 students toward ChatGPT-generated feedback: An explanatory research. ECNU Review of Education, 9(1). https://doi.org/10.1177/20965311241305140

Dai, W., Tsai, Y. S., Lin, J., Aldino, A., Jin, H., Li, T., Gašević, D., & Chen, G. (2024). Assessing the proficiency of large language models in automatic feedback generation: An evaluation

study. Computers and Education: Artificial Intelligence, 7, Article 100299. https://doi.org/10.1016/j.caeai.2024.100299

Engeness, I. (2025). Exploring AI-generated feedback in peer-discussion contexts: A mixed-methods study of essay writing in secondary classrooms. Computers and Education: Artificial Intelligence, 9, Article 100504. https://doi.org/10.1016/j.caeai.2025.100504

Glüsing, R., Fleckenstein, J., Schmidt, F. T. C., & Möller, J. (2026). LLM feedback for academic writing: Effects on students' performance and engagement. Contemporary Educational Psychology, 85, Article 102463. https://doi.org/10.1016/j.cedpsych.2026.102463

Guo, K., Pan, M., Li, Y., & Lai, C. (2024). Effects of an AI-supported approach to peer feedback on university EFL students' feedback quality and writing ability. The Internet and Higher Education, 63, Article 100962. https://doi.org/10.1016/j.iheduc.2024.100962

Guo, W. J. (2017). The relationships between Chinese secondary teachers' feedback and students' self-regulated learning [Doctoral dissertation, The Chinese University of Hong Kong].

Guo, W. (2020). Grade-level differences in teacher feedback and students' self-regulated learning. Frontiers in Psychology, 11, 783. https://doi.org/10.3389/fpsyg.2020.00783

Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1), 81–112. https://doi.org/10.3102/003465430298487

Hyland, F., & Hyland, K. (2001). Sugaring the pill: Praise and criticism in written feedback. Journal of Second Language Writing, 10(3), 185–212. https://doi.org/10.1016/S1060-3743(01)00038-8

Hou, C., Zhu, G., Sudarshan, V., Lim, F. S., & Ong, Y. S. (2025). Measuring undergraduate students' reliance on generative AI during problem-solving: Scale development and validation. Computers & Education, 234, Article 105329. https://doi.org/10.1016/j.compedu.2025.105329

Jin, Y., & Zhang, L. J. (2021). The dimensions of foreign language classroom enjoyment and their effect on foreign language achievement. International Journal of Bilingual Education and Bilingualism, 24(7), 948-962. https://doi.org/10.1080/13670050.2018.1526253

Lim, C., Hong, S., Eom, D., & Park, S. (2026). Bridging Teacher Competency and Student Outcomes: The Effects of ai-Integrated Instruction on ai-tpack and ai Literacy. Innovation and Education, 7(2), 224-249.

Li, Y., Jin, H., Dong, C., & Lu, H. (2025). 基于生成式人工智能的中学生人机协同写作行为及影响因素研究 [Human–AI collaborative writing behaviors and influencing factors among secondary school students using generative AI]. 现代远距离教育, (1), 22–33. https://doi.org/10.13927/j.cnki.yuan.20250401.001

Levine, S., Beck, S. W., Mah, C., Phalen, L., & Pittman, J. (2025). How do students use ChatGPT as a writing support? Journal of Adolescent & Adult Literacy, 68(5), 445–457. https://doi.org/10.1002/jaal.1373

Lu, Q., Yao, Y., Xiao, L., Zhu, X., & Yin, H. (2026). Exploring the impact of a GenAI-supported feedback practice on student feedback literacy in L2 writing. Computer Assisted Language Learning. Advance online publication. https://doi.org/10.1080/09588221.2025.2605541

Liu, J., Shi, Z., & Li, W. (2026). Developing students' feedback literacy in disciplinary academic writing through generative artificial intelligence. Assessing Writing, 68, Article 101030. https://doi.org/10.1016/j.asw.2026.101030

Lyu, W., & Salam, Z. A. (2025). AI-powered personalized learning: Enhancing self-efficacy, motivation, and digital literacy in adult education through expectancy-value theory. Learning and Motivation, 90, 102129.

Li, R., Zhang, C., & Ye, H. (2026). When does GenAI feedback support learning? Trust calibration and verification in L2 writing assessment. Assessing Writing, 69, Article 101087. https://doi.org/10.1016/j.asw.2026.101087

Linderbaum, B. A., & Levy, P. E. (2010). The development and validation of the Feedback Orientation Scale (FOS). Journal of Management, 36(6), 1372–1405. https://doi.org/10.1177/0149206310373145

Miller, M. K., Mandryk, R. L., Birk, M. V., Depping, A. E., & Patel, T. (2017). Through the looking glass: The effects of feedback on self-awareness and conversational behaviour during video chat. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (pp. 5271–5283). Association for Computing Machinery. https://doi.org/10.1145/3025453.3025548

Nguyen, A., Hong, Y., Dang, B., & Huang, X. (2024). Human-AI collaboration patterns in AI-assisted academic writing. Studies in Higher Education, 49(5), 847–864. https://doi.org/10.1080/03075079.2024.2323593

Ng, D. T. K., & Cheung, K. C. (2026). Interweaving generative AI technological and pedagogical design: Enhancing students' motivation, intentions, and AI literacy for feedback seeking. Educational Psychology. Advance online publication. https://doi.org/10.1080/01443410.2026.2668683

Nguyen, K. Q., & Vo, Q. N. (2026). LLM-supported EFL writing: AI feedback literacy, revision behaviour, and writing self-efficacy among Vietnamese undergraduates. Journal of English Language Teaching and Learning, 8(1). https://doi.org/10.18860/jetle.v8i1.44861

Nazaretsky, T., Gabbay, H., & Käser, T. (2026). Can students judge like experts? A large-scale study on the pedagogical quality of AI and human personalized formative feedback. Computers and

Education: Artificial Intelligence, 10, Article 100533. https://doi.org/10.1016/j.caeai.2025.100533

Nazaretsky, T., Mejia-Domenzain, P., Swamy, V., Frej, J., & Käser, T. (2024, September). AI or human? Evaluating student feedback perceptions in higher education. In European conference on technology enhanced learning (pp. 284-298). Cham: Springer Nature Switzerland.

Pi, Z., Xie, M., Guo, Q., Wang, L., et al. (2026). Human- or AI-delivered praise? The moderating role of praise publicity in synchronous video-based instruction. Educational Technology Research and Development. https://doi.org/10.1007/s11423-026-10671-6

Pahi, K., Hawlader, S., Hicks, E., Zaman, A., & Phan, V. (2024). Enhancing active learning through collaboration between human teachers and generative AI. Computers and Education Open, 6, Article 100183. https://doi.org/10.1016/j.caeo.2024.100183

Qi, Z., Liu, Y., Liu, W., & Xu, W. (2026). How does AI match students' proficiency levels to provide support? Student-driven AI-mediated scaffolding in EFL writing instruction. System, 142, Article 104119. https://doi.org/10.1016/j.system.2026.104119

Reicho, M., Otrel-Cass, K., Ebner, M., Brünner, B., Freinhofer, D., et al. (2026). Generative AI literacy across education and business: Competencies, obstacles, and benefits—a systematic literature review. International Journal of Educational Technology in Higher Education, 23, Article 23. https://doi.org/10.1186/s41239-026-00596-8

Tsao, J.-J., & Liu, C.-Y. (2026). Aligning AI and human feedback in EFL writing: A mixed-methods comparison of ChatGPT, peer, and teacher feedback. SAGE Open. Advance online publication. https://doi.org/10.1177/21582440261497008

Stojanov, A., Liu, Q., & Koh, J. H. L. (2024). University students' self-reported reliance on ChatGPT for learning: A latent profile analysis. Computers and Education: Artificial Intelligence, 6, Article 100243. https://doi.org/10.1016/j.caeai.2024.100243

Shute, V. J. (2008). Focus on formative feedback. Review of Educational Research, 78(1), 153–189. https://doi.org/10.3102/0034654307313795

Steiss, J., Tate, T. P., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. Learning and Instruction, 91, Article 101894. https://doi.org/10.1016/j.learninstruc.2024.101894

Sørhaug, J. O. (2026). Prompting for scaffolding: A thematic analysis of K-12 students' use of educational chatbots for writing support. Journal of Writing Research, 17(3), 453–481. https://doi.org/10.17239/jowr-2026.17.03.04

Soyoof, A., Reynolds, B. L., Rassaei, E., Kao, C.-W., & Ha, X. V. (2026). From teachers to chatbots: Scaffolded corrective feedback and student trust in online L2 English classrooms. Computers and Education: Artificial Intelligence, 10, Article 100530. https://doi.org/10.1016/j.caeai.2025.100530

Tseng, S.-S. (2026). Context-aware GenAI feedback for fostering student feedback literacy and informing revision decisions in social science writing. Assessment & Evaluation in Higher Education. Advance online publication. https://doi.org/10.1080/02602938.2026.2658636

Tao, S., Zhang, H., & Ding, L. (2026). Explainable machine learning for sustainable education: Predicting college students' reliance on generative artificial intelligence. Acta Psychologica, 265, Article 106622. https://doi.org/10.1016/j.actpsy.2026.106622

Takahashi, J., & Kadowaki, Y. (2026). Exploring student engagement with ChatGPT feedback: A classroom study of JFL learners. Foreign Language Annals. Advance online publication. https://doi.org/10.1111/flan.70072

Wang, Y., Hu, W., & Guo, Q. (2026). GPT versus teacher feedback on English writing: An analysis of feedback characteristics and student engagement. System, 138, Article 103967. https://doi.org/10.1016/j.system.2025.103967

Wang, M., Burlacu, G., Truxillo, D., James, K., & Yao, X. (2015). Age differences in feedback reactions: The roles of employee feedback orientation on social awareness and utility. Journal of Applied Psychology, 100(4), 1296–1308. https://doi.org/10.1037/a0038334

Wang, Q., Rose, C. P., Ma, N., Jiang, S., Bao, H., & Li, Y. (2022). Design and application of automatic feedback scaffolding in forums to promote learning. IEEE Transactions on Learning Technologies, 15(2), 150–166. https://eric.ed.gov/?id=EJ1339282

Wegner, D.M., Giuliano, T. (1982). The Forms of Social Awareness. In: Ickes, W., Knowles, E.S. (eds) Personality, Roles, and Social Behavior. Springer Series in Social Psychology. Springer, New York, NY. https://doi.org/10.1007/978-1-4613-9469-3_6

Wang, S. B., Wang, S. I.-C., & Liu, E. Z.-F. (2026). Role of generative AI literary assistants in enhancing ninth-grade students' writing motivation, flow and achievement. Computers and Education Open, 10, Article 100339. https://doi.org/10.1016/j.caeo.2026.100339

Wei, W., Zhang, X., & Zhu, X. (2026). Embedding, copilot, or agent? L2 teachers' modes of collaboration with generative AI in providing feedback on writing assignments: Rubric influences, workload dynamics, and barriers to integration. European Journal of Education, 61(3), Article e70795. https://doi.org/10.1111/ejed.70795

Yan, D. A., Tian, F., Li, H., Gao, Y., & Li, J. (2026). A GenAI-supported dialogic model to enhance the functions of peer feedback in EFL writing: A mixed-method interventional study. System, 136, Article 103895. https://doi.org/10.1016/j.system.2025.103895

Yang, Y., Hussin, S., & Hashim, H. (2025). Revising with intelligence: ChatGPT feedback and its impact on EFL students' revision and self-efficacy. Forum for Linguistic Studies, 7(7), 352–367. https://doi.org/10.30564/fls.v7i7.9845

Yao, G. (2026). Symbiosis or stagnation? Investigating the impact of AI on creativity in L1 Chinese writing. Thinking Skills and Creativity, 60, Article 102096. https://doi.org/10.1016/j.tsc.2025.102096

Zhang, H., Lee, I., & Moore, K. (2024, March). An effectiveness study of teacher-led AI literacy curriculum in K-12 classrooms. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 38, No. 21, pp. 23318-23325).

Zhou, C., & Wang, Y. (2026). University students' writing feedback literacy in the AI era: The interplay of generative AI acceptance, writing anxiety and writing self-efficacy in Chinese EMI educational contexts. European Journal of Education, 61(1), Article e70478. https://doi.org/10.1111/ejed.70478

Zhu, S., Li, J., Yao, Y., Guan, Y., & Zhu, X. (2025). What they provide and how: An intervention study on pre-service teachers' GenAI-assisted writing feedback. The Internet and Higher Education, 67, Article 101040. https://doi.org/10.1016/j.iheduc.2025.101040

Zhang, D. W., Hong, X., & Qi, Y. (2026). When and how does pedagogically enriched LLM feedback outperform corrective automated writing evaluation? A learning trajectory analysis of writing improvement. Computers in Human Behavior Reports, 22, Article 101121. https://doi.org/10.1016/j.chbr.2026.101121

Zhang, Z., Aubrey, S., Huang, X., & Chiu, T. K. F. (2025). The role of generative AI and hybrid feedback in improving L2 writing skills: A comparative study. Innovation in Language Learning and Teaching. Advance online publication. https://doi.org/10.1080/17501229.2025.2503890