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.

Reading–Writing Integration Through ChatGPT-Supported Feedback: Growth, Associations, and Mediated Pathways of Student Engagement in Chinese Classrooms

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: ChatGPT-supported Feedback, Reading–Writing Integration, Student Engagement, Chinese Language Education, Cross-lagged Path Analysis

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11562000/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.

Abstract

Although reading–writing integration has been increasingly advocated in Chinese language education, limited empirical evidence explains how reading, writing, and student engagement become integrated, particularly under AI-supported feedback. To address this issue, this study examined whether ChatGPT-supported feedback improved reading outcomes, writing outcomes, and student engagement, while investigating their interrelationships, the mediating role of student engagement, and the corresponding pathways in reading–writing integration. A total of 400 middle school students from eight intact classes in Shenzhen participated, with 200 students receiving ChatGPT-supported feedback and 200 receiving human-generated feedback. Using a quasi-experimental pretest–posttest design, data were analyzed within a quantitative longitudinal framework integrating baseline equivalence tests, 2 × 2 mixed-design ANOVAs, factor-level correlation analyses, and multi-group cross-lagged path analysis to examine changes and temporal relationships among reading, student engagement, and writing across two groups (N = 400). Despite comparable pretest scores, findings showed that Group 1 demonstrated significantly greater pretest–posttest gains in reading, student engagement, and writing, with significant Time × Group interactions and large effects. Correlations further indicated stronger posttest associations among engagement, reading, and writing in Group 1, whereas these relationships remained weak in Group 2. Longitudinal analyses revealed stronger T2 within-time direct and mediating pathways in Group 1, alongside several significant cross-lagged effects, but cross-lagged mediating effects were generally nonsignificant in both groups. The findings theoretically highlight student engagement as a potentially time-sensitive mechanism through which AI-supported learning may strengthen the integration of reading and writing. Practically, they suggest the pedagogical value of AI-supported learning environments for fostering engagement and connecting literacy processes, while emphasizing sustained instructional exposure to consolidate these relationships over time.

Introduction

Reading and writing are increasingly conceptualized as reciprocal literacy processes that jointly support students' language development (Graham & Hebert, 2011; Graham et al., 2018; Fitzgerald & Shanahan, 2000). Research has demonstrated both reading-to-write and writing-to-read effects, showing that reading provides linguistic, discourse, and content resources for writing (Graham et al., 2018; Fitzgerald & Shanahan, 2000; Graham & Sandmel, 2011), while writing activities boost reading fluency, word reading, and comprehension outcomes (Graham & Hebert, 2011; Graham et al., 2012; Graham & Sandmel, 2011). At the same time, engagement is positively associated with reading achievement (e.g., Connor et al., 2009; Lee, 2014; Zhu et al., 2023) and writing performance (e.g., Ndruru et al., 2025; Rad et al., 2024; Philp & Duchesne, 2016) and functions as an important mechanism through which reading and writing mutually reinforce one another (e.g., Alexander, 2018; Ng & Graham, 2018).

However, these theoretical and empirical advances face challenges in mainland Chinese reading and writing. Previous studies have noted that traditional Chinese instruction has often emphasized teacher-centered activities such as explanation, and question-answering, while providing limited opportunities for

authentic student participation and frequently treating reading and writing as separate instructional domains (e.g., Zhu, 2008, 2009a, b; Pan, 1997, 2013, 2017). As a result, students may have fewer opportunities to engage actively in integrated literacy practices that support reciprocal reading–writing development. This situation highlights the need for innovative instructional approaches capable of simultaneously strengthening reading–writing connections and student engagement.

To address this gap, the present study applies ChatGPT-supported feedback to support reading–writing integration and student engagement. Recent research suggests that ChatGPT can enhance reading comprehension (e.g., BALANTEKİN, 2025; Etkin et al., 2025; Heung & Chiu, 2025), improve writing quality and feedback processes (Bai, 2024; Teng & Huang, 2026; Cipta et al., 2025), and increase learners' engagement and interaction in educational activities (e.g., Lo et al., 2024; Meşe et al., 2025; Wu & Ho, 2025), suggesting its potential to support the interconnected development of reading, writing, and student engagement. Building on these findings, the present study investigates whether ChatGPT-supported feedback can promote reading–writing integration by enhancing student engagement in Chinese language classrooms.

Literature review

Reading-to-Write and Writing-to-Read in First Language Education

Reading and writing are closely connected literacy practices that jointly support students' academic development in first language education, and recent research has increasingly conceptualized them as reciprocal rather than independent processes (Graham & Hebert, 2011; Graham et al., 2018; Fitzgerald & Shanahan, 2000). Extensive research has demonstrated that reading and writing rely on shared linguistic, cognitive, and metacognitive resources, highlighting the importance of integrated literacy instruction in contemporary literacy research (Berninger et al., 2016; Graham et al., 2018; Duke & Cartwright, 2021).

Research on reading-to-write has shown that reading provides students with vocabulary knowledge, genre awareness, discourse structures, and content knowledge that directly support writing development (Graham et al., 2018; Fitzgerald & Shanahan, 2000; Graham & Sandmel, 2011). A large-scale meta-analysis found that reading interventions significantly improved writing quality, writing output, and spelling performance among school-aged learners, demonstrating the transfer of literacy knowledge from reading to writing (Graham et al., 2018). Studies of literacy instruction have further reported that exposure to high-quality texts enhances students' organization, argumentation, and overall text quality by strengthening their understanding of written discourse conventions (Graham et al., 2018; Duke & Cartwright, 2021).

Research on writing-to-read has generated equally strong evidence that writing activities improve reading achievement and comprehension (Graham & Hebert, 2011; Graham & Hebert, 2010; Hebert et al., 2012). Meta-analytic findings indicate that writing about texts through summarizing, note-taking, responding,

and analytical writing produces significant gains in reading comprehension compared with reading-only activities (Graham & Hebert, 2011; Graham & Hebert, 2010). Explicit instruction in writing processes, sentence construction, and text structure has also been shown to strengthen reading fluency, word reading, and comprehension outcomes (Graham & Hebert, 2011; Graham et al., 2012; Graham & Sandmel, 2011). These effects are commonly explained by deeper processing, comprehension monitoring, and increased metacognitive engagement during writing tasks (Graham & Hebert, 2011; Duke & Cartwright, 2021; Berninger et al., 2016).

There is a potential reciprocal literacy model in which reading and writing develop through mutually reinforcing processes rather than unidirectional transfer (Graham et al., 2018; Duke & Cartwright, 2021). Although most influential studies have been conducted in Western first-language literacy contexts, emerging research has begun to examine whether reciprocal reading–writing relationships operate similarly in Chinese literacy education. In mainland Chinese language arts classrooms, Zhu (2008, 2009a, 2009b, 2010a, 2010b, 2019) argues that reading and writing should be understood as mutually constitutive rather than separate instructional domains. Drawing on both traditional Chinese literacy thought and contemporary classroom practices, he contends that extensive reading cultivates audience awareness, evaluative judgment, and discourse knowledge that support writing development (Zhu, 2008, 2010b). At the same time, writing activities provide opportunities for students to deepen textual understanding, participate in writing communities, and transform reading experiences into meaningful literacy production (Zhu, 2010a, 2019). Zhu's historical analysis further suggests that the long-standing debate between prioritizing reading and prioritizing writing should be replaced by an integrated perspective that views literacy learning as a dynamic process of reciprocal reading–writing development (Zhu, 2009a, 2009b).

Overall, evidence from both Western literacy research and mainland Chinese language education suggests that reading and writing develop through reciprocal and mutually reinforcing processes (Graham & Hebert, 2011; Graham et al., 2018; Zhu, 2019).

Relationships Between Student Engagement and Literacy Outcomes

Student engagement has long been regarded as both a central goal of literacy instruction and a key mechanism through which instructional practices shape reading-related outcomes. In particular, Alvermann (2002) argued that engagement mediates the relationship between instruction and student outcomes, including reading performance and beliefs about reading, a view that is also reflected in earlier engagement research (Guthrie et al., 1998). Consistent with this view, a recent review found that cognitive and behavioral engagement are reliable predictors of reading achievement, whereas emotional engagement primarily facilitates achievement through its influence on other forms of engagement (Zhu et al., 2023). Empirical studies have likewise reported positive associations between engagement and reading outcomes. For example, engagement scores were generally positively related to reading comprehension performance in elementary classrooms, although the relationships were complex and shaped by instructional factors (Connor et al., 2009). Similarly, behavioral engagement, reflected in effort

and perseverance, and emotional engagement, reflected in a sense of belonging, significantly predicted reading performance among students (Lee, 2014).

Student engagement is a multidimensional construct that shapes learners' participation in academic tasks and influences how they engage with instructional opportunities, including writing-related feedback and learning activities (Fredricks et al., 2004; Philp & Duchesne, 2016). Similarly, writing engagement strategies have been found to positively predict academic writing performance with a significant positive effect ($\beta = 0.294$) (Ndruru et al., 2025), while active participation in writing tasks is associated with improved linguistic outcomes, including greater proficiency, fluency, accuracy, and complexity of written expression (Rad et al., 2024). Beyond direct performance effects, engagement also appears to function as an important mechanism in writing development. The depth of students' engagement has been identified as a crucial mediating factor influencing how feedback is utilized and how feedback contributes to writing improvement over time (Zhang & Hyland, 2018). Additional evidence from engagement-oriented instructional approaches indicates that gamification can significantly enhance language learning outcomes with a medium effect size ($g = 0.517$) (Lee & Baek, 2023).

Taken together, prior research suggests that student engagement is closely associated with both reading and writing outcomes. Accordingly, the interrelationships among reading, writing, and student engagement merit further examination in the present study.

Student Engagement as a Mediator of Reading–Writing Development

Student engagement has been widely conceptualized as a central mechanism linking reading and writing development. Research suggests that students can transfer techniques encountered during reading into their own writing (Kletzien & Hushion, 1992), while growth in both reading and writing depends on learners' willingness to engage in these literacy activities (Roe, 1997). Empirical evidence increasingly supports the crucial role of engagement in supporting the linkage between reading and writing development (e.g., de Milliano, 2013).

On the one hand, meaningfully integrated reading and writing activities can promote one another, with interaction fostering sustained engagement and positive literacy outcomes (Park, 2016), while instruction supporting both reading and writing is associated with higher levels of student engagement and cognitive demand (Bock & Erickson, 2015). Recent scholarship has emphasized that engagement is critical to students' growth as readers and writers (Alexander, 2018; Ng & Graham, 2018), and enriches understanding of both reading and writing. (Olson, 1996).

On the other hand, insufficient student engagement may limit opportunities for learners to connect reading and writing experiences, as interest, motivation, and active participation have been identified as important contributors to the development of both literacy domains (Hidi & Anderson, 2014; Mason et al., 2023; Aktaş et al., 2023).

However, the close relationship between engagement and learning outcomes is not fixed, it should be considered in light of the developmental nature of student engagement (Li & Lerner, 2011), and also reciprocal reading-writing connection unfolds nonlinearly over an extended period, requiring long-term, longitudinal development to fully manifest across lexical and text levels (Kim et al., 2018). Accordingly, the present study examined engagement-reading-writing correlation and pathways across pretest, posttest and cross-lag rather than treating it as a static characteristic measured at a single time point.

Traditional Chinese Instruction and Its Constraints on Student Engagement in Reading and Writing

Research on Chinese language education has long raised concerns about the relationship between reading, writing, and instructional practice in mainland China. Existing scholarship suggests that traditional Chinese language instruction has often emphasized textual interpretation, teacher explanation, and examination-oriented learning, while providing comparatively limited opportunities for students' active participation in literacy practices (Pan, 1997, 2013, 2017). Pan (1997, 2013, 2017) examined the historical development of writing education and questioned the long-standing imbalance between reading and writing in Chinese language teaching.

Within the field of reading–writing integration, Zhu (2009b) analyzed the reading–writing relationship in traditional Chinese language education, while Zhu (2009a) revisited historical debates regarding the relative status of reading and writing. Zhu (2008) argued that cultivating audience awareness is essential for meaningful writing development. Extending this perspective, Zhu (2010a) proposed writing communities as a means of increasing interaction and participation in writing instruction, and Zhu (2010b) illustrated how evaluative reading can encourage more active engagement with texts. More recently, Zhu (2019) called for reading–writing integration research to move beyond experience-based practices toward stronger theoretical and empirical foundations. Related work by Zheng (2007, 2012, 2014) similarly emphasized instructional reflection, learner participation, and the integration of literacy activities. Taken together, these studies suggest that reading and writing are mutually supportive literacy processes. However, relatively limited research has explicitly examined student engagement as the mechanism linking reading–writing integration and literacy development in mainland Chinese classrooms.

Taken together, these studies suggest that traditional Chinese language instruction has often separated reading from writing and provided limited opportunities for authentic student participation. Although scholars have increasingly advocated reading–writing integration and more student-centered literacy practices, empirical classroom research in mainland China remains relatively limited, and many traditional instructional patterns continue to dominate Chinese language education.

ChatGPT-supported feedback in supporting Reading, Writing, and Engagement

Recent scholarship highlights the growing role of ChatGPT in supporting reading-related learning. Evidence suggests that ChatGPT-provided reading materials can enhance reading comprehension and learners' interaction with texts, particularly when reading activities are tailored to learners' needs (BALANTEKİN, 2025; Etkin et al., 2025; Ali et al., 2025). Research has also documented positive influences on reading motivation, attitudes, habits, and engagement, indicating that ChatGPT can foster more active participation in reading activities (BALANTEKİN, 2025; Heung & Chiu, 2025). Additionally, ChatGPT has been shown to facilitate the development of personalized and scalable reading resources, expanding opportunities for extensive reading and diversified reading experiences (BALANTEKİN, 2025).

Recent scholarship on ChatGPT and writing has converged around four major themes. First, effectiveness studies consistently report improvements in writing quality, linguistic accuracy, and overall performance when ChatGPT is used as a writing support tool (Bai, 2024; Cipta et al., 2025; Teng & Huang, 2026; Bauer et al., 2026). Second, pedagogical research highlights its potential to facilitate feedback, idea generation, and collaborative writing while emphasizing the need for responsible instructional integration (Su et al., 2023; Stockwell, 2024). Third, review and meta-analytic studies synthesize growing evidence supporting ChatGPT's educational value while identifying methodological and ethical concerns (Alvizo, 2025). Finally, emerging research examines how ChatGPT-assisted writing influences learners' metacognition, authorship, and writer identity, suggesting that ChatGPT can support reflective writing development without necessarily diminishing learner agency (Bauer et al., 2026; De Matas, 2023).

Engagement has been positioned as a key outcome of ChatGPT-supported learning. Empirical and review studies consistently indicate that ChatGPT can enhance students' participation, involvement, and interaction in learning activities (Lo et al., 2024; Heung & Chiu, 2025; Meşe et al., 2025; Wu & Ho, 2025). Attention has also shifted toward learners' engagement with feedback, with evidence suggesting that the uptake and use of ChatGPT feedback are closely linked to feedback literacy and learning processes (Teng & Huang, 2026; Zhan & Yan, 2025). Complementing these findings, studies have reported positive effects of ChatGPT on learner motivation, highlighting its potential to encourage greater participation and sustained involvement in learning activities (Ali et al., 2023; Yildiz, 2023; Wu & Ho, 2025).

Methodology

Research Design

Guided by an effect–relationship–mechanism perspective, this study aimed to develop an explanatory understanding of how AI-supported feedback facilitates Chinese reading and writing development and the role of student engagement in this process. Specifically, the study examined changes in reading, writing, and engagement across feedback conditions (Research Question 1), explored the interrelationships among these constructs (Research Question 2), evaluated whether student engagement served as a mechanism linking reading and writing outcomes, and the pathways among these variables (Research Question 3).

Figure 1 illustrates the conceptual framework of the present study, comprising two separately conducted research inquiries. The double-headed arrow represents the correlational relationship between Reading Comprehension and Writing Performance (Research Question 2). The solid single-headed arrows denote the hypothesized direct pathways (Path A, Path B, and Path C) and the indirect (mediating) effect (Path A × Path B) tested via path analysis (Research Question 3). These two sets of analyses—correlation and pathway analysis—were conducted independently rather than simultaneously.

Participants, Sampling, and Intervention

In the September 2025, the study was conducted in a middle school in Shenzhen, China, involving 400 students from eight intact classes (approximately 50 students per class). Four classes (n = 200) were assigned to the AI-supported feedback condition (Group 1), and four classes (n = 200) to the human-generated feedback condition (Group 2). Intact-class sampling was used to preserve the natural classroom structure and minimize disruption to regular instruction. To reduce potential teacher-related confounding, all participating teachers were novice teachers with fewer than five years of teaching experience, held a master's degree, and had not yet obtained professional teaching titles or received teaching-related awards.

The intervention was implemented over approximately four months (one semester) during regular Chinese-language classes. Both groups received approximately six to seven Chinese lessons per week and participated in the same two-part integrated reading–writing intervention, which targeted both reading-to-writing and writing-to-reading processes. The shared instructional structure and the modality-specific implementation are summarized in Table X.

Table 1
Intervention Design and Instructional Modalities Across Groups

Dimension	Group 1: AI-Supported Feedback	Group 2: Human-Generated Feedback
Participants	n = 200; four intact classes	n = 200; four intact classes
Classroom setting	School computer room	Regular Chinese-language classroom setting
Core intervention	Same two-part integrated reading–writing intervention	Same two-part integrated reading–writing intervention
Instructional objectives	Develop reading-to-writing and writing-to-reading processes	Develop reading-to-writing and writing-to-reading processes
Instructional modality	AI-supported instructional and feedback modality	Conventional teacher-led instructional and feedback modality
AI access	Computers equipped with VPN software enabling access to ChatGPT	No ChatGPT support
Primary instructional / feedback resource	ChatGPT functioned as the primary AI-supported instructional and feedback resource	Human teachers provided the instructional guidance and feedback
Instructional sequence	ChatGPT was used across the semester to design a formative sequence of integrated reading–writing activities aligned with students' cognitive developmental needs	The same instructional objectives and tasks were implemented through the school's conventional instructional approach
Student activities	Students completed ChatGPT-designed reading and writing tasks and engaged in post-task reflection	Students completed the corresponding reading and writing tasks through conventional instruction
Teacher role	Regular Chinese-language teacher remained present and provided necessary subject-matter and technical guidance	Teachers provided instructional guidance and feedback throughout the intervention
Feedback source	Primarily AI-supported feedback through ChatGPT, with teacher support as needed	Human-generated feedback from the Chinese-language teachers

Both groups completed the same two intervention components throughout the semester using a range of culturally and thematically relevant reading and writing topics. The first component adopted a **reading-to-writing** approach: students read culturally or thematically relevant texts and subsequently developed their own writing based on the content, themes, or perspectives encountered in the reading. For example, students might read texts concerning childhood memories of one's hometown and then write a narrative depicting a remembered hometown scene.

The second component adopted a **writing-to-reading** approach incorporating dynamic reading–writing interaction. Students first produced an inquiry-based text using a framework of problem identification, hypothesis generation, reasoning, and verification. They then applied this framework to the reading of related texts before completing a written inquiry composition. These examples illustrate the instructional processes rather than representing the sole content or topics covered during the semester.

Table 2
Research Questions, Hypotheses, and Statistical Methods

Research Question (RQ)	Hypotheses	Statistical Analysis
RQ1. To what extent do reading outcomes, writing outcomes, and student engagement change from pre-test to post-test in the AI-supported feedback and human-generated feedback groups? **RQ2. How are reading outcomes, writing outcomes, and student engagement associated across AI-supported feedback and human-generated feedback groups?**	**$H0_1$**: There is no significant changes in reading outcomes, writing outcomes, or student engagement within either group, nor significant differences in growth between groups. **$H1_1$**: Significant changes occur in reading outcomes, writing outcomes, and/or student engagement within groups, and the magnitude of growth differs significantly between groups. **$H0_2$**: Reading outcomes, writing outcomes, and student engagement do not demonstrate significant associations or meaningful reciprocal explanatory relationships, and these patterns remain comparable across groups and measurement occasions. **$H1_2$**: Reading outcomes, writing outcomes, and student engagement demonstrate significant associations and reciprocal explanatory relationships, with these patterns differing across groups and measurement occasions.	Descriptive analyses, within-group growth analyses, between-group comparisons of change scores, confidence interval estimation, and effect size evaluation (Cohen's d). Pearson correlations were first examined to characterize the association patterns among reading outcomes, writing outcomes, and student engagement across groups and measurement occasions. To quantify the degree of construct integration, separate simple regression models were subsequently conducted to estimate reciprocal variance explained (R^2) among the three domains. Correlation matrices, network visualizations, and R^2-based integration trajectories were compared descriptively across instructional conditions and time points to examine differences in relational structures and developmental patterns.
RQ3. To what extent do the within-time and cross-time mediated pathways among reading, writing and student engagement differ between AI-supported feedback and human-generated feedback groups?	**$H0_3$**: There are no significant differences between the AI-supported feedback group and the human-generated feedback group in the within-time or cross-time mediated pathways among reading, writing, and student engagement. **$H1_3$**: The AI-supported feedback group exhibits significantly stronger within-time or cross-time mediated pathways than the human-generated feedback group	Multi-group cross-lagged path analyses were conducted to examine within-time and cross-time mediated pathways. Maximum likelihood estimation was used for direct, autoregressive, cross-lagged, and indirect effects. Variables were standardized within group. Indirect effects were computed as the product of coefficients with delta-method standard errors. Group-specific estimates were obtained for the AI-supported and human-generated feedback groups. Standardized coefficients (β), SE, z, p, and 95% CIs were reported for all pathways.

Instruments

Content validity was established through expert review by three specialists, including two university professors in Chinese language education and an experienced practitioner holding PhD in Chinese education. All experts evaluated the reading, writing, and engagement instruments and reached full agreement (100%) regarding their relevance and appropriateness.

Student Engagement

Student engagement is widely recognized as a multifaceted construct encompassing several interrelated yet distinguishable dimensions. Consistent with previous research, engagement can be conceptualized through four domains: behavioral engagement, which reflects students' observable participation in learning activities; affective engagement, which concerns emotional reactions toward learning experiences; cognitive engagement, which represents the degree of psychological investment in learning; and agentic engagement, which captures students' active efforts to influence and enrich the learning process (Fredricks et al., 2004; Wang et al., 2019). Evidence from prior educational studies has supported the applicability and validity of this four-factor structure across diverse learning contexts. Accordingly, the present study adopts this framework to assess student engagement. Appendix 1 outlines the four dimensions and provides illustrative items used to represent each construct (5-point Likert scale), and scores could range from 8 to 40.

Table 3. Internal consistency reliability of the Engagement Scale

Group	Measurement occasion	*N*	No. of items	Cronbach's α	McDonald's ω
Group 1	Pre-test	200	8	.849	.915
	Post-test	200	8	.849	.913
Group 2	Pre-test	200	8	.848	.913
	Post-test	200	8	.866	.922

Note. Cronbach's α and McDonald's ω were estimated separately for each group and measurement occasion using the eight Engagement Scale items. Higher coefficients indicate greater internal consistency. Group 1 and Group 2 correspond to the two study conditions defined in the research design.

The Engagement Scale demonstrated strong internal consistency across both groups and measurement occasions. Both Cronbach's α and McDonald's ω indicated that the eight-item scale provided consistently reliable measurement at pre-test and post-test. The close correspondence between the two coefficients further suggests that the reliability estimates were robust to the choice of internal consistency estimator

Table 4
Construct Validity for Engagement Scale Parallel Analysis of the Engagement Scale (Pre-test and Post-test)

Factor	Pre-test			Post-test		
	Actual	Random	Random 95th	Actual	Random	Random 95th
1	3.313	0.398	0.401	4.710	0.392	0.398
2	0.562	0.151	0.152	0.427	0.154	0.154
3	0.481	0.095	0.097	0.304	0.097	0.096
4	0.383	0.047	0.047	0.240	0.047	0.046
5	−0.300	0.001	0.002	−0.218	0.002	0.001
6	−0.355	−0.044	−0.043	−0.235	−0.045	−0.045
7	−0.371	−0.095	−0.095	−0.245	−0.096	−0.095
8	−0.400	−0.156	−0.159	−0.274	−0.158	−0.157

Note. Extraction: principal axis; 1,000 random datasets. At both time points, only the first four actual eigenvalues exceeded their corresponding 95th-percentile random eigenvalues, supporting a four-factor solution.

Exploratory Factor Analysis of the Engagement Scale: Pre-test and Post-test

Panel A. Pre-test (n = 400)

Item	D1	D2	D3	D4	*h*²	*u*²
E_pre_item1	**.87**				.81	.19
E_pre_item2	**.89**				.76	.24
E_pre_item3		**.86**			.75	.25
E_pre_item4		**.86**			.73	.27
E_pre_item5			**.89**		.78	.22
E_pre_item6			**.88**		.78	.22
E_pre_item7				**.87**	.77	.23
E_pre_item8				**.88**	.78	.22
SS loadings	1.56	1.49	1.57	1.55		
% of variance	20	19	20	19		
Cumulative %	20	39	58	77		

Factor correlations (Pre-test)

D1	D1	D2	D3	D4
	1.00	**.46**	**.45**	**.40**
D2	.46	1.00	.46	.51
D3	.45	.46	1.00	.44
D4	.40	.51	.44	1.00

Fit: $\chi^2(2) = 4.65$, *p* = .098; TLI = .978; RMSEA = .057 [0, .128]; RMSR = .00; BIC = − 7.33.

Panel B. Post-test (n = 400)

Item	D1	D2	D3	D4	*h*²	*u*²
E_post_item1	**.88**				.83	.17
E_post_item2	**.89**				.81	.19
E_post_item3		**.90**			.82	.18
E_post_item4		**.89**			.85	.15
E_post_item5			**.88**		.82	.18
E_post_item6			**.89**		.82	.18
E_post_item7				**.90**	.85	.15
E_post_item8				**.90**	.86	.14
SS loadings	1.64	1.68	1.63	1.70		
% of variance	21	21	20	21		
Cumulative %	21	42	63	83		

Factor correlations (Post-test)

D1	D1	D2	D3	D4
	1.00	**.65**	**.64**	**.63**
D2	.65	1.00	.59	.62
D3	.64	.59	1.00	.67
D4	.63	.62	.67	1.00

Fit: $\chi^2(2) = 5.86$, *p* = .053; TLI = .978; RMSEA = .069 [0, .138]; RMSR = .00; BIC = − 6.12.

Parallel analysis and EFA were conducted separately for pre-test and post-test data, without splitting by group. At both time points, both methods consistently supported a clean four-factor structure, with each pair of items loading on its intended dimension and no cross-loadings, confirming the hypothesized four-dimension construct validity.

Confirmatory Factor Analysis of the Engagement Scale: Pre-test and Post-test

Panel A. Pre-test (n = 400)

Item	Standardized loading	*SE*	*z*	R^2
F1				
i1	.948	.030	31.210	.899
i2	.822	.031	26.842	.676
F2				
i3	.873	.028	30.928	.763
i4	.847	.029	29.638	.718
F3				
i5	.876	.028	31.317	.767
i6	.895	.028	32.192	.802
F4				
i7	.880	.029	30.735	.774
i8	.879	.029	30.716	.773

Factor correlations (Pre-test)

F1	F1	F2	F3	F4
	1.000			
F2	.477	1.000		
F3	.455	.468	1.000	
F4	.394	.516	.447	1.000

Reliability (Pre-test)

α	F1	F2	F3	F4
	.876	**.850**	**.879**	**.872**
ω	.882	.851	.879	.872
AVE	.791	.740	.783	.774

Model fit (Pre-test)

$\chi^2(14) = 8.842$, *p* = .841; CFI = 1.000; TLI = 1.006; RMSEA = .000 [.000, .028]; SRMR = .009; AIC = 6493.057; BIC = 6612.801.

Panel B. Post-test (n = 400)

Item	Standardized loading	*SE*	*z*	R^2
F1				
i1	.919	.017	55.227	.845
i2	.892	.018	50.794	.796
F2				
i3	.898	.017	51.696	.807
i4	.928	.016	56.367	.861
F3				
i5	.908	.017	52.819	.824
i6	.898	.018	51.255	.807
F4				
i7	.917	.015	60.344	.840
i8	.926	.015	62.363	.858

Factor correlations (Post-test)

F1	F1	F2	F3	F4
	1.000			
F2	.679	1.000		
F3	.662	.620	1.000	
F4	.653	.648	.703	1.000

Reliability (Post-test)

α	F1	F2	F3	F4
	.901	**.909**	**.898**	**.918**
ω	.902	.910	.899	.919
AVE	.821	.834	.816	.849

Model fit (Post-test)

$\chi^2(14) = 20.014$, *p* = .130; CFI = .998; TLI = .995; RMSEA = .033 [.000, .063]; SRMR = .011; AIC = 6400.862; BIC = 6520.606.

Note

All standardized loadings were significant at *p* < .001.

The four-factor model fit the data well at both time points, with strong loadings, adequate AVE, high reliability, and moderate-to-high factor correlations, supporting the hypothesized four-dimension structure of engagement.

Convergent Validity of the Engagement Scale: Pre-test and Post-test

Panel A. Pre-test (n = 400)

Factor	Item	Standardized loading	AVE	CR
F1	i1	.948	.787	.880
	i2	.822		
F2	i3	.873	.740	.851
	i4	.847		
F3	i5	.876	.784	.879
	i6	.895		
F4	i7	.880	.774	.872
	i8	.879		

Note. All standardized loadings were significant at *p* < .001. AVE = average variance extracted; CR = composite reliability. All AVE values exceeded .50 and all CR values exceeded .70, supporting strong convergent validity.

Panel B. Post-test (n = 400)

Factor	Item	Standardized loading	AVE	CR
F1	i1	.919	.821	.902
	i2	.892		
F2	i3	.898	.834	.909
	i4	.928		
F3	i5	.908	.815	.898
	i6	.898		
F4	i7	.917	.849	.918
	i8	.926		

Note. All standardized loadings were significant at *p* < .001. All AVE values exceeded .50 and all CR values exceeded .70, indicating excellent convergent validity at post-test.

Discriminant Validity of the Engagement Scale: Pre-test and Post-test (Fornell-Larcker Criterion)

Panel A. Pre-test (n = 400)

Factor	F1	F2	F3	F4	AVE	MSV
F1	**.887**				.787	.228
F2	.477	**.860**			.740	.266
F3	.455	.468	**.886**		.784	.219
F4	.394	.516	.447	**.880**	.774	.266

Note. Diagonal (bold) = √AVE; off-diagonal = factor correlations. MSV = maximum shared variance. Every √AVE exceeded its correlations with all other factors, and every AVE exceeded MSV, supporting discriminant validity.

Panel B. Post-test (n = 400)

Factor	F1	F2	F3	F4	AVE	MSV
F1	**.906**				.821	.461
F2	.679	**.913**			.834	.461
F3	.662	.620	**.903**		.815	.495
F4	.653	.648	.703	**.922**	.849	.495

Note. Diagonal (bold) = √AVE; off-diagonal = factor correlations. All √AVE values exceeded the corresponding factor correlations, and all AVE values exceeded MSV, confirming discriminant validity at post-test.

Both convergent and discriminant validity were fully supported at pre-test and post-test. All standardized loadings exceeded .70, all AVE values exceeded .50, and all CR values exceeded .70, indicating strong convergent validity. Fornell-Larcker criterion was satisfied for every factor at both time points (√AVE > all inter-factor correlations), and all AVE values exceeded MSV, confirming that the four engagement dimensions are empirically distinct. These results collectively demonstrate that the engagement scale has sound construct validity.

Chinese Reading Comprehension

To measure students' Chinese reading outcomes, this study employed the Six Types of Reading Comprehension Process Model developed by Zhu (2005) and Zhu et al. (2020). The model comprises six hierarchical cognitive processes—Retrieving, Explaining, Summarizing, Elaborating, Evaluating, and Creating—and describes the progression from learning to read to reading to learn. It has been widely adopted by educational authorities and teacher educators for almost twenty years. Based on Cheong et al. (2017), a 50-point assessment scheme was used, including one 5-point item for Retrieving and Explaining respectively, two 5-point items each for Summarizing, Elaborating, and Evaluating, and five 2-point items for Creating. The possible score distribution extended from 0 to 50 (Appendix 2). The pre-test and post-test were constructed using different reading passages and corresponding comprehension items. One passage was A Walk, a curriculum text that students had not previously studied in class. The pre-test and post-test were constructed using different reading passages and corresponding comprehension items. One passage was A Walk, a curriculum text that students had not previously studied in class. The other was My Late Father, an extracurricular reading text selected for its comparable length and difficulty to the curriculum passage.

Table 5
Inter-rater reliability of Six Types of Reading Comprehension across groups

Group	Measurement Occasion	*N*	Number of Raters	ICC(2,1)	ICC(2,3)
Group 1	Pre-test	200	3	.949	.982
	Post-test	200	3	.944	.981
Group 2	Pre-test	200	3	.951	.983
	Post-test	200	3	.960	.986

Note. ICC = intraclass correlation coefficient. ICC(2,1) represents a two-way random-effects model assessing absolute agreement for a single rater, whereas ICC(2,3) represents the corresponding reliability estimate based on the average scores of the three raters.

The Reading scores demonstrated very strong inter-rater reliability across both groups and measurement occasions. The high and closely clustered ICC estimates indicate that the scoring process produced consistent ratings across raters and remained stable across measurement occasions and groups. The reliability estimates based on the three-rater average were consistently stronger than those based on a single rater, as expected when multiple independent ratings are aggregated.

Table 6
Exploratory Factor Analysis and Parallel Analysis of Reading Across Measurement Occasions

Rater 1–Rater 2	Panel A. Pre-test: Correlations	Panel B. Pre-test: Factor Evidence	Panel C. Post-test: Correlations	Panel D. Post-test: Factor Evidence
	.961	**—**	**.978**	**—**
Rater 1–Rater 3	.955	—	.974	—
Rater 2–Rater 3	.954	—	.978	—
Overall KMO	—	.79	—	.79
Bartlett's χ^2	—	2070.526	—	2557.611
df	—	3	—	3
Bartlett's p	—	< .001	—	< .001
Parallel analysis	—	1 factor	—	1 factor
Rater 1 loading	—	.98	—	.99
Rater 2 loading	—	.98	—	.99
Rater 3 loading	—	.97	—	.99
Communality range (h^2)	—	.95–.96	—	.97–.98
Proportion of variance explained	—	.96	—	.98
RMSR	—	0.00	—	0.00

Note. EFA was conducted using minimum residual extraction. Parallel analysis used 1,000 iterations.

Both measurement occasions consistently supported a single common reading-rating dimension, with very high inter-rater correlations and uniformly strong factor loadings, providing convergent structural evidence for combining the three ratings into a composite Reading score.

Chinese Writing Comprehension

Writing performance was evaluated using the 6 + 1 Traits framework, which defines quality writing through seven dimensions and has been supported by evidence of reliability, validity, and instructional value (Culham, 2003, 2006; Kozlow & Bellamy, 2004). The framework has been successfully applied in Chinese-language writing instruction (Ge, 2016; Wang, 2016). Participants were asked to compose a 500-character essay on the topic of a family member. The writing task was intentionally aligned with the themes of the reading passages (A Walk and My Late Father), both of which focus on family relationships and interpersonal experiences. This design aimed to facilitate reading-to-writing transfer by providing students with opportunities to draw upon ideas, discourse features, and narrative techniques encountered in the reading texts when composing their own essays (Fitzgerald & Shanahan, 2000; Graham et al., 2018). Their writing was scored with a five-point analytic rubric adapted from Kozlow and Bellamy (2004), focusing on the six principal traits while excluding the presentation dimension. Participants could obtain total scores between 6 and 30. (Appendix 3)

Table 7
Inter-rater reliability of 6 + 1 Writing Assessment Scale

Group	Measurement Occasion	*N*	Number of Raters	ICC(2,1)	ICC(2,3)
Group 1	Pre-test	200	3	.899	.964
	Post-test	200	3	.881	.957
Group 2	Pre-test	200	3	.860	.949
	Post-test	200	3	.888	.960

Note. ICC = intraclass correlation coefficient. ICC(2,1) represents a two-way random-effects model assessing absolute agreement for a single rater, whereas ICC(2,3) represents the corresponding reliability estimate when the scores from all three raters are averaged.

The writing scores demonstrated strong inter-rater reliability across both groups and measurement occasions.

Table 8
Exploratory Factor Analysis and Parallel Analysis of Writing Scores Across Measurement Occasions

Rater 1–Rater 2	Panel A. Pre-test: Correlations	Panel B. Pre-test: Factor Evidence	Panel C. Post-test: Correlations	Panel D. Post-test: Factor Evidence
	.908	**—**	**.938**	**—**
Rater 1–Rater 3	.890	—	.937	—
Rater 2–Rater 3	.893	—	.938	—
Overall KMO	—	.78	—	.79
Bartlett's χ^2	—	1403.694	—	1784.493
df	—	3	—	3
Bartlett's p	—	< .001	—	< .001
Parallel analysis	—	1 factor	—	1 factor
Rater 1 loading	—	.95	—	.97
Rater 2 loading	—	.95	—	.97
Rater 3 loading	—	.94	—	.97
Communality range (h^2)	—	.88–.91	—	.94
Proportion of variance explained	—	.90	—	.94
RMSR	—	0.00	—	0.00

Note. EFA used minimum residual extraction. Parallel analysis used 1,000 iterations. Mean scores were excluded; the three raters were analyzed separately. Bartlett's *p* values are reported as < .001. Given three indicators, results provide evidence for a common rating dimension rather than full-scale structural validation.

Both occasions supported a single common writing-rating dimension, with strong inter-rater correlations and uniformly high factor loadings, supporting aggregation of the three ratings into a composite Writing score.

RQ1:

Table 9
Descriptive Statistics for Reading, Student Engagement, and Writing by Group and Measurement Occasion

Outcome	Group	Time	*N*	*M*	*SD*	Skewness	Kurtosis	Min	Max
Engagement	Group 1	Pretest	200	21.27	4.80	0.06	0.11	10.00	36.00
		Posttest	200	30.10	4.24	−0.45	0.20	16.00	40.00
	Group 2	Pretest	200	22.05	4.89	0.15	−0.41	11.00	35.00
		Posttest	200	21.91	5.14	0.04	−0.45	10.00	34.00
Reading	Group 1	Pretest	200	28.99	5.10	0.12	−0.19	17.00	44.00
		Posttest	200	37.72	4.51	0.18	−0.10	24.67	49.67
	Group 2	Pretest	200	29.16	5.11	0.01	−0.50	16.33	40.00
		Posttest	200	29.24	5.38	0.02	0.12	11.33	42.67
Writing	Group 1	Pretest	200	17.95	3.42	−0.16	0.12	7.33	26.33
		Posttest	200	22.85	2.98	−0.20	−0.12	15.00	30.00
	Group 2	Pretest	200	18.09	3.06	−0.17	−0.06	9.67	26.00
		Posttest	200	18.21	3.06	−0.11	−0.13	9.33	26.00

Across the two assessment occasions, the descriptive patterns suggest increases in all three outcomes for Group 1, whereas Group 2 shows relatively little change over time. The distributions appear approximately symmetric, with skewness and kurtosis remaining within commonly acceptable descriptive ranges. More specific between-group and longitudinal comparisons are presented in the subsequent analyses.

Table 10
Baseline Equivalence Tests Across Groups

Outcome	Mean Difference (G1 − G2)	95% CI	*t*	*df*	*p*
Engagement	−0.79	[− 1.74, 0.17]	−1.62	397.86	.106
Reading	−0.17	[− 1.17, 0.84]	−0.32	398.00	.747
Writing	−0.13	[− 0.77, 0.51]	−0.41	393.17	.685

Note. G1 = Group 1; G2 = Group 2. Mean difference = Group 1 − Group 2; negative values indicate slightly lower pretest means in Group 1. CI = confidence interval.

Independent-samples *t* tests showed no significant baseline differences on any outcome (all *p*s > .10), confirming that Group 1 and Group 2 were comparable at pretest. Mean differences are computed as Group 1 minus Group 2, so all three negative values indicate that Group 1 started marginally (non-significantly) lower.

Table 11
Pre–Post Changes by Group and Omnibus 2 × 2 Mixed-Design ANOVA Panel A. Descriptive statistics and paired effect sizes by group and overall

Outcome	Group	Pretest *M* (*SD*)	Posttest *M* (*SD*)	Mean Change [95% CI]	Cohen's *dz*
Engagement	Group 1	21.27 (4.80)	30.10 (4.24)	8.84 [8.13, 9.54]	1.76
	Group 2	22.05 (4.89)	21.91 (5.14)	−0.14 [− 0.94, 0.66]	−0.02
	Overall	21.66 (4.86)	26.01 (6.24)	4.35 [3.66, 5.04]	0.62
Reading	Group 1	28.99 (5.10)	37.72 (4.51)	8.73 [8.09, 9.37]	1.90
	Group 2	29.16 (5.11)	29.24 (5.38)	0.08 [− 0.59, 0.75]	0.02
	Overall	29.08 (5.10)	33.48 (6.53)	4.41 [3.78, 5.03]	0.69
Writing	Group 1	17.95 (3.42)	22.85 (2.98)	4.90 [4.43, 5.37]	1.46
	Group 2	18.09 (3.06)	18.21 (3.06)	0.13 [− 0.30, 0.56]	0.04
	Overall	18.02 (3.24)	20.53 (3.81)	2.51 [2.12, 2.91]	0.63

Panel B. Omnibus 2 × 2 mixed-design ANOVA

Outcome	Time *F*	Group *F*	Time × Group *F*	Interaction partial η²
Engagement	130.69***	87.82***	139.25***	.26
Reading	175.31***	87.27***	168.87***	.30
Writing	121.60***	70.26***	109.49***	.22

Note. All *F* tests had *df* = 1, 398. *** *p* < .001. Panel A reports each group's and the combined sample's pre- and posttest means, standard deviations, mean change (posttest − pretest) with 95% CIs, and paired Cohen's *dz*. Panel B reports the omnibus mixed-design ANOVA effects (single whole-model values; not group-specific). Partial η² refers to the Time × Group interaction.

Group 1 improved substantially on all three outcomes (Δ*M* = 4.90–8.84; *dz* = 1.46–1.90), whereas Group 2 remained essentially unchanged (Δ*M* ≈ 0; *dz* ≈ 0). Baseline levels were comparable across groups

(Table 10); the divergence emerged only at posttest. Omnibus ANOVA confirmed significant Time, Group, and Time × Group effects for every outcome, with large interaction effects (partial η² = .22–.30).

Table 12
Posttest Between-Group Comparisons

Outcome	**Group 1 *M* (*SD*)**	**Group 2 *M* (*SD*)**	**Mean Difference (G1 − G2) [95% CI]**	***t* (*df*)**	***p***	**Cohen's *d***
Engagement	30.10 (4.24)	21.91 (5.14)	8.19 [7.26, 9.12]	17.38 (384.30)	< .001	1.74
Reading	37.72 (4.51)	29.24 (5.38)	8.48 [7.51, 9.46]	17.09 (386.45)	< .001	1.71
Writing	22.85 (2.98)	18.21 (3.06)	4.64 [4.04, 5.23]	15.34 (397.74)	< .001	1.53

Note. Mean difference = Group 1 − Group 2; positive values indicate that Group 1 scored higher than Group 2 at posttest. CI = confidence interval.

At posttest, Group 1 significantly outperformed Group 2 on every outcome, with large effects (*d* = 1.53–1.74). Mean differences are calculated as Group 1 minus Group 2, so all positive values directly reflect Group 1's superior posttest performance across engagement, reading, and writing.

RQ2:

Table 13
Factor-Based Correlations Among Engagement, Reading, and Writing Across Groups and Measurement Occasions Panel A. Group 1 — Pre-test

E1	**E1**	**E2**	**E3**	**E4**	**Reading**	**Writing**
	—					
E2	.384	—				
E3	.421	.397	—			
E4	.385	.443	.332	—		
Reading	.101	.041	−.009	−.068	—	
Writing	.005	.066	.066	.065	.134	—

Panel B. Group 1 — Post-test

E1	E1	E2	E3	E4	Reading	Writing
	—					
E2	.476	—				
E3	.409	.335	—			
E4	.396	.323	.513	—		
Reading	.452	.465	.480	.421	—	
Writing	.539	.573	.428	.467	.782	—

Panel C. Group 2 — Pre-test

E1	E1	E2	E3	E4	Reading	Writing
	—					
E2	.427	—				
E3	.375	.410	—			
E4	.325	.444	.449	—		
Reading	−.007	.007	.101	.054	—	
Writing	.044	.021	.091	.037	.071	—

Panel D. Group 2 — Post-test

E1	E1	E2	E3	E4	Reading	Writing
	—					
E2	.399	—				
E3	.468	.411	—			
E4	.430	.446	.492	—		
Reading	.094	.009	.064	−.030	—	
Writing	.050	.115	.036	.035	.166	—

Note. *N* = 200 in each group at each measurement occasion. E1–E4 represent the four Engagement factors, each computed as the mean of its corresponding pair of Engagement items. Reading and Writing represent the respective one-factor scores operationalized using the mean scores across the three raters.

The correlation patterns revealed a clear distinction between the internal structure of Engagement and its associations with the academic performance measures. Across groups and measurement occasions, the four Engagement factors showed generally positive and moderate inter-factor associations, indicating meaningful relatedness while retaining differentiation among the factors.

In Group 1, the post-test pattern showed a marked strengthening of the associations between all four Engagement factors and both Reading and Writing, accompanied by a strong positive relationship between Reading and Writing. Group 2 displayed considerably weaker associations between Engagement and the two performance measures, with these relationships remaining relatively limited at both measurement occasions.

The heatmaps (Fig. 2) provide a visual representation of the factor-level correlation structure across the four group-by-time conditions. The visual pattern indicates that the four Engagement factors maintain broadly coherent relationships with one another across conditions, whereas their associations with Reading and Writing vary more substantially across groups and measurement occasions. The post-test heatmap for Group 1 shows a visibly stronger concentration of associations linking Engagement with the two performance constructs, together with a pronounced Reading–Writing relationship. In contrast, the Group 2 panels remain comparatively light in the Engagement–performance region, indicating weaker associations.

RQ3:

Table 14
Across Groups of direct, indirect cross-lagged, and autoregressive pathways

Path	Role	Group	Pathways per Group
Within-Time Direct and Indirect Effects (T1)			
RDpre → ENGpre	Within-Time Direct (path a)	EQ1, EQ2	4
ENGpre → WTpre	Within-Time Direct (path b)	EQ1, EQ2	4
RDpre → WTpre	Within-Time Direct (path c)	EQ1, EQ2	1
RDpre → ENGpre → WTpre	Within-Time Indirect (path a×b)	EQ1, EQ2	4
Autoregressive Effects			
RDpre → RDpost	Autoregressive	EQ1, EQ2	1
ENGpre → ENGpost	Autoregressive	EQ1, EQ2	4
WTpre → WTpost	Autoregressive	EQ1, EQ2	1
Cross-Lagged Effects and Cross-Time Direct and Indirect Effects			
RDpre → ENGpost	Cross-lag / Direct (path a)	EQ1, EQ2	4
ENGpre → WTpost	Cross-lag / Direct (path b)	EQ1, EQ2	4
ENGpre → RDpost	Cross-lag	EQ1, EQ2	4
RDpre → WTpost	Cross-lag / Direct (path c)	EQ1, EQ2	1
WTpre → RDpost	Cross-lag	EQ1, EQ2	1
WTpre → ENGpost	Cross-lag	EQ1, EQ2	4
RDpre → ENGpost→ WTpost	Cross-lag / Indirect (path a×b)	EQ1, EQ2	4
Within-Time Direct Effects (T2)			

Path	Role	Group	Pathways per Group
Within-Time Direct and Indirect Effects (T1)			
RDpost → ENGpost	Within-Time Direct (path a)	EQ1, EQ2	4
ENGpost → WTpost	Within-Time Direct (path b)	EQ1, EQ2	4
RDpost → WTpost	Within-Time Direct (path c)	EQ1, EQ2	1
RDpost → ENGpost → WTpost	Within-Time Indirect (path a×b)	EQ1, EQ2	4

Note. Indirect effects are counted as distinct mediational chains, not as the product of *a* and *b* path counts. Only $a_i \times b_i$ (same engagement factor) forms a valid chain (e.g., RD → ENG_1 → WT); cross-factor products are not model paths. Hence 4 chains, not 16.

Table 15

Direct, Indirect, Cross-Lagged, and Autoregressive Pathways Across Two Groups Panel A. EQ1 (Group 1, *N* = 200)

Section	Path	Role	β	SE	z	p	95% CI
Within-Time Direct and Indirect Effects (T1)	RDpre → ENG1pre	Within-Time Direct (path a)	0.101	0.070	1.431	.1525	[−0.037, 0.239]
	RDpre → ENG2pre	Within-Time Direct (path a)	0.041	0.071	0.585	.5589	[−0.097, 0.180]
	RDpre → ENG3pre	Within-Time Direct (path a)	−0.009	0.071	−0.127	.8988	[−0.148, 0.130]
	RDpre → ENG4pre	Within-Time Direct (path a)	−0.068	0.071	−0.969	.3323	[−0.207, 0.070]
	ENG1pre → WTpre	Within-Time Direct (path b)	−0.075	0.070	−1.075	.2823	[−0.212, 0.062]
	ENG2pre → WTpre	Within-Time Direct (path b)	0.033	0.070	0.479	.6319	[−0.103, 0.170]
	ENG3pre → WTpre	Within-Time Direct (path b)	0.064	0.070	0.912	.3619	[−0.073, 0.200]
	ENG4pre → WTpre	Within-Time Direct (path b)	0.068	0.070	0.979	.3275	[−0.068, 0.205]
	RDpre → WTpre	Within-Time Direct (path c)	0.145	0.070	2.064	.0390	[0.007, 0.283]
	RDpre → ENG1pre → WTpre	Within-Time Indirect (path a × b)	−0.008	0.009	−0.859	.3901	[−0.025, 0.010]
	RDpre → ENG2pre → WTpre	Within-Time Indirect (path a × b)	0.001	0.004	0.371	.7110	[−0.006, 0.009]
	RDpre → ENG3pre → WTpre	Within-Time Indirect (path a × b)	−0.001	0.005	−0.126	.8997	[−0.009, 0.008]
	RDpre → ENG4pre → WTpre	Within-Time Indirect (path a × b)	−0.005	0.007	−0.689	.4909	[−0.018, 0.009]

Section	Path	Role	β	SE	z	p	95% CI
Autoregressive Effects	RDpre → RDpost	Autoregressive	0.532	0.056	9.584	< .001	[0.423, 0.641]
	ENG1pre → ENG1post	Autoregressive	0.237	0.061	3.894	< .001	[0.118, 0.357]
	ENG2pre → ENG2post	Autoregressive	0.073	0.063	1.169	.2424	[−0.050, 0.196]
	ENG3pre → ENG3post	Autoregressive	−0.039	0.062	−0.626	.5311	[−0.161, 0.083]
	ENG4pre → ENG4post	Autoregressive	0.043	0.065	0.667	.5050	[−0.084, 0.171]
	WTpre → WTpost	Autoregressive	0.263	0.036	7.274	< .001	[0.192, 0.334]
Cross-Lagged Effects and Cross-Time Direct and Indirect Effects	RDpre → ENG1post	Cross-lag / Direct (path a)	−0.046	0.073	−0.627	.5305	[−0.189, 0.097]
	RDpre → ENG2post	Cross-lag / Direct (path a)	−0.042	0.075	−0.567	.5710	[−0.189, 0.104]
	RDpre → ENG3post	Cross-lag / Direct (path a)	−0.020	0.075	−0.269	.7877	[−0.167, 0.126]
	RDpre → ENG4post	Cross-lag / Direct (path a)	−0.103	0.078	−1.326	.1850	[−0.255, 0.049]
	ENG1pre → WTpost	Cross-lag / Direct (path b)	0.083	0.036	2.307	.0210	[0.013, 0.154]
	ENG2pre → WTpost	Cross-lag / Direct (path b)	0.030	0.035	0.868	.3856	[−0.038, 0.099]
	ENG3pre → WTpost	Cross-lag / Direct (path b)	−0.036	0.035	−1.046	.2954	[−0.104, 0.032]
	ENG4pre → WTpost	Cross-lag / Direct (path b)	0.020	0.035	0.562	.5744	[−0.049, 0.089]
	ENG1pre → RDpost	Cross-lag	0.048	0.055	0.872	.3833	[−0.060,

Section	Path	Role	β	SE	z	p	95% CI
							0.156]
	ENG2pre → RDpost	Cross-lag	0.110	0.055	2.023	.0431	[0.003, 0.217]
	ENG3pre → RDpost	Cross-lag	0.066	0.055	1.205	.2282	[−0.041, 0.173]
	ENG4pre → RDpost	Cross-lag	0.156	0.055	2.841	.0045	[0.048, 0.263]
	RDpre → WTpost	Cross-lag / Direct (path c)	−0.029	0.043	−0.673	.5009	[−0.112, 0.055]
	WTpre → RDpost	Cross-lag	0.131	0.055	2.364	.0181	[0.022, 0.239]
	WTpre → ENG1post	Cross-lag	0.120	0.062	1.935	.0530	[−0.002, 0.242]
	WTpre → ENG2post	Cross-lag	0.089	0.064	1.402	.1609	[−0.035, 0.214]
	WTpre → ENG3post	Cross-lag	−0.052	0.063	−0.827	.4082	[−0.177, 0.072]
	WTpre → ENG4post	Cross-lag	0.089	0.065	1.374	.1695	[−0.038, 0.217]
	RDpre → ENG1post → WTpost	Cross-lag / Indirect (path a × b)	−0.004	0.006	−0.605	.5450	[−0.016, 0.009]
	RDpre → ENG2post → WTpost	Cross-lag / Indirect (path a × b)	−0.001	0.003	−0.474	.6352	[−0.007, 0.004]
	RDpre → ENG3post → WTpost	Cross-lag / Indirect (path a × b)	0.001	0.003	0.261	.7942	[−0.005, 0.006]
	RDpre → ENG4post → WTpost	Cross-lag / Indirect (path a × b)	−0.002	0.004	−0.517	.6051	[−0.010, 0.006]
Within-Time Direct Effects (T2)	RDpost → ENG1post	Within-Time Direct (path a)	0.396	0.076	5.217	< .001	[0.247, 0.544]
	RDpost → ENG2post	Within-Time Direct (path a)	0.449	0.079	5.722	< .001	[0.295, 0.603]

Section	Path	Role	β	SE	z	p	95% CI
	RDpost → ENG3post	Within-Time Direct (path a)	0.510	0.078	6.553	< .001	[0.357, 0.662]
	RDpost → ENG4post	Within-Time Direct (path a)	0.448	0.081	5.534	< .001	[0.290, 0.607]
	ENG1post → WTpost	Within-Time Direct (path b)	0.096	0.040	2.377	.0174	[0.017, 0.175]
	ENG2post → WTpost	Within-Time Direct (path b)	0.177	0.039	4.520	< .001	[0.100, 0.254]
	ENG3post → WTpost	Within-Time Direct (path b)	−0.022	0.039	−0.567	.5709	[−0.099, 0.055]
	ENG4post → WTpost	Within-Time Direct (path b)	0.087	0.038	2.265	.0235	[0.012, 0.162]
	RDpost → WTpost	Within-Time Direct (path c)	0.563	0.057	9.864	< .001	[0.451, 0.674]
	RDpost → ENG1post → WTpost	Within-Time Indirect (path a × b)	0.038	0.018	2.163	.0305	[0.004, 0.072]
	RDpost → ENG2post → WTpost	Within-Time Indirect (path a × b)	0.080	0.022	3.547	< .001	[0.036, 0.124]
	RDpost → ENG3post → WTpost	Within-Time Indirect (path a × b)	−0.011	0.020	−0.565	.5724	[−0.051, 0.028]
	RDpost → ENG4post → WTpost	Within-Time Indirect (path a × b)	0.039	0.019	2.096	.0361	[0.003, 0.075]

Panel B. EQ2 (Group 2, N = 200)

Section	Path	Role	β	SE	z	p	95% CI
Within-Time Direct and Indirect Effects (T1)	RDpre → ENG1pre	Within-Time Direct (path a)	−0.007	0.071	−0.093	.9260	[−0.145, 0.132]
	RDpre → ENG2pre	Within-Time Direct (path a)	0.007	0.071	0.104	.9174	[−0.131, 0.146]
	RDpre → ENG3pre	Within-Time Direct (path a)	0.101	0.070	1.439	.1501	[−0.037, 0.239]
	RDpre → ENG4pre	Within-Time Direct (path a)	0.054	0.071	0.759	.4476	[−0.085, 0.192]
	ENG1pre → WTpre	Within-Time Direct (path b)	0.023	0.070	0.327	.7440	[−0.115, 0.161]
	ENG2pre → WTpre	Within-Time Direct (path b)	−0.024	0.070	−0.336	.7366	[−0.161, 0.114]
	ENG3pre → WTpre	Within-Time Direct (path b)	0.087	0.071	1.225	.2206	[−0.052, 0.225]
	ENG4pre → WTpre	Within-Time Direct (path b)	−0.002	0.070	−0.029	.9771	[−0.140, 0.136]
	RDpre → WTpre	Within-Time Direct (path c)	0.062	0.071	0.880	.3786	[−0.076, 0.201]
	RDpre → ENG1pre → WTpre	Within-Time Indirect (path a × b)	0.000	0.002	−0.089	.9288	[−0.003, 0.003]
	RDpre → ENG2pre → WTpre	Within-Time Indirect (path a × b)	0.000	0.002	−0.099	.9210	[−0.004, 0.003]
	RDpre → ENG3pre → WTpre	Within-Time Indirect (path a × b)	0.009	0.009	0.933	.3509	[−0.010, 0.027]
	RDpre → ENG4pre → WTpre	Within-Time Indirect (path a × b)	0.000	0.004	−0.029	.9771	[−0.008, 0.007]
Autoregressive Effects	RDpre → RDpost	Autoregressive	0.580	0.057	10.165	< .001	[0.468, 0.692]

Section	Path	Role	β	SE	z	p	95% CI
	ENG1pre → ENG1post	Autoregressive	0.236	0.069	3.426	< .001	[0.101, 0.371]
	ENG2pre → ENG2post	Autoregressive	0.263	0.068	3.882	< .001	[0.130, 0.397]
	ENG3pre → ENG3post	Autoregressive	0.214	0.070	3.075	.0021	[0.078, 0.351]
	ENG4pre → ENG4post	Autoregressive	0.189	0.069	2.725	.0064	[0.053, 0.325]
	WTpre → WTpost	Autoregressive	0.484	0.061	7.927	< .001	[0.364, 0.604]
Cross-Lagged Effects and Cross-Time Direct and Indirect Effects	RDpre → ENG1post	Cross-lag / Direct (path a)	0.188	0.084	2.248	.0246	[0.024, 0.352]
	RDpre → ENG2post	Cross-lag / Direct (path a)	0.108	0.083	1.292	.1962	[−0.056, 0.271]
	RDpre → ENG3post	Cross-lag / Direct (path a)	0.026	0.085	0.300	.7642	[−0.141, 0.192]
	RDpre → ENG4post	Cross-lag / Direct (path a)	−0.017	0.085	−0.195	.8450	[−0.183, 0.150]
	ENG1pre → WTpost	Cross-lag / Direct (path b)	−0.061	0.063	−0.975	.3298	[−0.185, 0.062]
	ENG2pre → WTpost	Cross-lag / Direct (path b)	0.085	0.062	1.366	.1718	[−0.037, 0.208]
	ENG3pre → WTpost	Cross-lag / Direct (path b)	0.000	0.062	0.006	.9951	[−0.122, 0.122]
	ENG4pre → WTpost	Cross-lag / Direct (path b)	−0.056	0.061	−0.906	.3647	[−0.176, 0.065]
	ENG1pre → RDpost	Cross-lag	0.143	0.057	2.519	.0118	[0.032, 0.254]

Section	Path	Role	β	SE	z	p	95% CI
	ENG2pre → RDpost	Cross-lag	−0.020	0.057	−0.349	.7272	[−0.131, 0.091]
	ENG3pre → RDpost	Cross-lag	−0.045	0.057	−0.783	.4336	[−0.157, 0.067]
	ENG4pre → RDpost	Cross-lag	0.029	0.057	0.512	.6086	[−0.082, 0.140]
	RDpre → WTpost	Cross-lag / Direct (path c)	0.025	0.076	0.329	.7425	[−0.124, 0.174]
	WTpre → RDpost	Cross-lag	0.060	0.057	1.061	.2887	[−0.051, 0.172]
	WTpre → ENG1post	Cross-lag	−0.056	0.068	−0.825	.4095	[−0.190, 0.077]
	WTpre → ENG2post	Cross-lag	0.056	0.068	0.816	.4147	[−0.078, 0.189]
	WTpre → ENG3post	Cross-lag	−0.029	0.070	−0.423	.6727	[−0.166, 0.107]
	WTpre → ENG4post	Cross-lag	0.049	0.070	0.709	.4781	[−0.087, 0.186]
	RDpre → ENG1post → WTpost	Cross-lag / Indirect (path a × b)	−0.012	0.013	−0.894	.3712	[−0.037, 0.014]
	RDpre → ENG2post → WTpost	Cross-lag / Indirect (path a × b)	0.009	0.010	0.939	.3477	[−0.010, 0.028]
	RDpre → ENG3post → WTpost	Cross-lag / Indirect (path a × b)	0.000	0.002	0.006	.9951	[−0.003, 0.003]
	RDpre → ENG4post → WTpost	Cross-lag / Indirect (path a × b)	0.001	0.005	0.191	.8485	[−0.009, 0.010]
Within-Time Direct Effects (T2)	RDpost → ENG1post	Within-Time Direct (path a)	−0.039	0.084	−0.461	.6450	[−0.204, 0.127]

Section	Path	Role	β	SE	z	p	95% CI
	RDpost → ENG2post	Within-Time Direct (path a)	−0.070	0.083	−0.842	.3999	[−0.233, 0.093]
	RDpost → ENG3post	Within-Time Direct (path a)	0.035	0.085	0.417	.6769	[−0.131, 0.201]
	RDpost → ENG4post	Within-Time Direct (path a)	−0.041	0.085	−0.477	.6336	[−0.207, 0.126]
	ENG1post → WTpost	Within-Time Direct (path b)	0.041	0.063	0.650	.5155	[−0.082, 0.164]
	ENG2post → WTpost	Within-Time Direct (path b)	0.085	0.063	1.356	.1750	[−0.038, 0.208]
	ENG3post → WTpost	Within-Time Direct (path b)	0.003	0.062	0.048	.9618	[−0.118, 0.124]
	ENG4post → WTpost	Within-Time Direct (path b)	−0.040	0.061	−0.650	.5154	[−0.160, 0.080]
	RDpost → WTpost	Within-Time Direct (path c)	0.104	0.075	1.377	.1686	[−0.044, 0.252]
	RDpost → ENG1post → WTpost	Within-Time Indirect (path a × b)	−0.002	0.004	−0.376	.7069	[−0.010, 0.007]
	RDpost → ENG2post → WTpost	Within-Time Indirect (path a × b)	−0.006	0.008	−0.715	.4745	[−0.022, 0.010]
	RDpost → ENG3post → WTpost	Within-Time Indirect (path a × b)	0.000	0.002	0.048	.9620	[−0.004, 0.004]
	RDpost → ENG4post → WTpost	Within-Time Indirect (path a × b)	0.002	0.004	0.384	.7006	[−0.007, 0.010]

Both groups show weak T1 within-time associations among reading, engagement, and writing, with negligible indirect effects. By T2, however, the patterns diverge. EQ1 shows substantially stronger within-time pathways from reading to all four engagement factors, accompanied by several meaningful reading→engagement→writing indirect effects. This pattern suggests a more evident within-time mediational mechanism in which engagement may link reading to writing under the intervention condition. EQ2, in contrast, shows generally weak T2 direct and indirect pathways, providing little

evidence for a comparable within-time mediational process. Autoregressive pathways indicate that the constructs retain varying degrees of temporal stability in both groups.

The cross-lagged results provide a more differentiated picture. In EQ1, several cross-lagged direct pathways are evident, particularly from prior engagement to subsequent outcomes, although other cross-lagged associations remain weak. By contrast, EQ2 shows generally limited cross-lagged direct effects. Importantly, cross-lagged indirect pathways remain weak in both groups, indicating that the reading→engagement→writing mechanism is more apparent within the same measurement occasion than across time. Thus, the longitudinal results do not yet provide clear evidence that this mediational process consistently carries forward from T1 to T2. A longer observation interval or greater time for learning processes to consolidate may be needed to determine whether these within-time mechanisms subsequently develop into stable longitudinal pathways.

Discussion

RQ1.

Baseline comparability, Group 2 stability, substantial Group 1 gains (dz = 1.46–1.90 vs. ≈0), significant Time × Group interactions (partial η^2 = .22–.30), and large posttest effects (*d* = 1.53–1.74) support claims that ChatGPT-supported feedback drives improvement. Reading gains align with evidence that ChatGPT enhances comprehension, text interaction, motivation, and participation when materials are tailored to learners' needs (BALANTEKİN, 2025; Etkin et al., 2025; Ali et al., 2025; Heung & Chiu, 2025). Writing gains support claims that ChatGPT improves writing quality, linguistic accuracy, and performance while facilitating feedback and idea generation (Bai, 2024; Cipta et al., 2025; Teng & Huang, 2026; Bauer et al., 2026; Su et al., 2023; Stockwell, 2024). Engagement gains echo findings that ChatGPT boosts participation, interaction, motivation, and feedback uptake (Lo et al., 2024; Heung & Chiu, 2025; Meşe et al., 2025; Ali et al., 2023; Yildiz, 2023; Wu & Ho, 2025; Zhan & Yan, 2025). These results corroborate the literature and indicate that ChatGPT-supported feedback promotes learning outcomes and sustained engagement.

RQ2.

The RQ2 findings reveal a distinct triadic correlation pattern. In Group 1's post-test, the strengthened associations among engagement, reading, and writing, alongside the high Reading–Writing correlation (.782), strongly align with the reciprocal literacy model, which conceptualizes reading and writing as mutually reinforcing processes relying on shared linguistic and cognitive resources (Graham & Hebert, 2011; Graham et al., 2018; Fitzgerald & Shanahan, 2000; Berninger et al., 2016; Duke & Cartwright, 2021). Reading provides vocabulary and discourse knowledge supporting writing (Graham et al., 2018; Graham & Sandmel, 2011), while writing activities are positively related to reading comprehension gains (Graham & Hebert, 2011; Hebert et al., 2012), echoing integrated Chinese literacy perspectives (Zhu, 2008, 2010b, 2019). The strengthened engagement–literacy associations corroborate that engagement scores are positively related to reading comprehension (Connor et al., 2009), and that effort and belongingness are

significantly associated with reading performance (Lee, 2014). For writing, these associations support that engagement shapes participation in feedback activities (Fredricks et al., 2004; Philp & Duchesne, 2016), engagement strategies are positively related to writing performance (Ndruru et al., 2025), and active participation correlates with greater linguistic proficiency, fluency, accuracy, and complexity (Rad et al., 2024). Conversely, Group 2's consistently weak associations suggest that without engagement-oriented instructional approaches (e.g., gamification; Lee & Baek, 2023), this triadic synergy does not emerge.

RQ3.

The RQ3 findings offer support for engagement as a mediator and the holistic reading-writing-engagement pathways. After confirming the baseline between group1 and 2's pre-test, in Group 1's post-test, the significant within-time indirect effects (reading → engagement → writing) align with the view that engagement mediates instructional effects on literacy outcomes (Alvermann, 2002; Guthrie et al., 1998). This corroborates claims that cognitive and behavioral engagement predict achievement (Zhu et al., 2023; Lee, 2014; Connor et al., 2009), and that engagement strategies positively predict writing performance (Ndruru et al., 2025; Rad et al., 2024; Fredricks et al., 2004; Philp & Duchesne, 2016), with engagement depth mediating writing improvement (Zhang & Hyland, 2018). This synergy echoes that integrated literacy activities foster sustained engagement (Park, 2016; Bock & Erickson, 2015; Roe, 1997; Kletzien & Hushion, 1992; de Milliano, 2013; Alexander, 2018; Ng & Graham, 2018).

Conversely, Group 2's post-test weak indirect effects align with the notion that insufficient engagement limits reading–writing connections (Hidi & Anderson, 2014; Mason et al., 2023; Aktaş et al., 2023), especially without engagement-oriented instruction (Lee & Baek, 2023).

Importantly, the non-significant cross-lagged mediation in both groups echoes that the engagement–outcome relationship is developmental rather than fixed (Li & Lerner, 2011), and reciprocal reading–writing connections unfold nonlinearly over an extended period to fully manifest across lexical and text levels (Kim et al., 2018), so the generally non-significant cross-lagged effect may be caused by a semester interval between pretest and posttest.

Conclusion

This study investigated the impact of ChatGPT-supported feedback on Chinese L1 students' reading, writing, and engagement. Findings revealed significant Time × Group effects, demonstrating that ChatGPT substantially enhanced all three outcomes with large effect sizes. Correlation and path analyses confirmed a triadic synergy among these constructs and validated engagement as a within-time mediator. Within the broader context of AI-integrated literacy education, these results highlight the pedagogical value of ChatGPT in fostering immediate literacy gains. It is recommended that educators integrate such tools while implementing longitudinal tracking to sustain long-term development.

Limitations

First, participants were drawn from a single middle school, limiting generalizability. Second, the no more than four-month interval was relatively short; given that engagement–outcome relationships are developmental and reading–writing connections unfold nonlinearly over extended periods as we mentioned in discussion, this timeframe may underestimate longitudinal mediation. Third, the large effect sizes (Cohen's $d > 1.5$) and post-test construct fusion (Reading–Writing $r = .782$) may raise concerns regarding common method bias, although the reliability and validity are quite good.

Future research should recruit multi-site, diverse samples to enhance generalizability. Longitudinal designs spanning multiple semesters with multiple waves are needed to fully capture the nonlinear, developmental trajectories of reading–writing reciprocity and engagement. To mitigate common method bias concerns, future studies should incorporate external observations.

Declarations

Ethics Statement

Ethical approval for this study was granted by the Ethics Committee of Minnan University of Science and Technology (No. KYLL2025-08-26-1). 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.

Consent for publication

Consent for publication of anonymized data was obtained from all participants and, where applicable, their guardians.

Declaration of generative AI use

Generative AI was used solely to assist with language editing and improve the clarity and readability of the manuscript. It was not used for study design, literature synthesis, methodology development, data collection, data analysis, or the generation of research findings. The author takes full responsibility for the accuracy, integrity, and originality of the manuscript.

Data availability statement

The dataset supporting the findings of this study has been deposited in Zenodo (https://doi.org/10.5281/zenodo.21412506) under restricted access to protect participant confidentiality. Access will be granted where appropriate and in accordance with ethical and institutional requirements.

Competing interests

The author declares no competing interests.

Funding

This research received no external funding.

Author contributions

Dr. Zhu Siyi Augustine (first author) was responsible for all aspects of the study except data collection and the ethical approval application, including conceptualization, data analysis, and manuscript drafting and revision. https://orcid.org/0009-0000-8193-3192 Prof. Zhu JianJun (second author) was responsible for data collection and the ethical approval application. https://orcid.org/0009-0002-5114-2626

Acknowledgements

The author sincerely thanks the participating school, teachers, and students for their support and cooperation throughout the study. The author also gratefully acknowledges the Taiwan Association for Educational Communication and Technology (TAECT) for the opportunity to present an earlier version of this work at the 2026 TAECT Annual International Conference. The constructive peer-review and academic feedback received during the conference submission process are also gratefully acknowledged.

References

1. Ali JKM, Shamsan MAA, Hezam TA, Mohammed AAQ (2023) Impact of chatgpt on learning motivation. J Engl Stud Arabia Felix 2(1):41–49. https://doi.org/10.56540/jesaf.v2i1.51
2. Ali MM, Alaa AM, Wael Alharbi, AlQurashi I (2025) Investigating the Impact of ChatGPT on Reading Proficiency in EFL Context. Int J Technol Educ 8(4):1222–1244. https://doi.org/10.46328/ijte.1203
3. Alvizo JB (2025) The Impact of Chatgpt on Students' Academic Writing Performance: A Meta-Analytic Study. J Educational Analytics 4(4):877–892. https://doi.org/10.55927/jeda.v4i4.381
4. Aktaş N, Özdemir Çetinkaya, E., Kaya Özgül B (2023) The Relationship between Reading Comprehension and Writing Skills of Fourth Grade Students and Their Motivation to Read and Write. Int J Educational Res Rev 8(2):122–136. https://doi.org/10.24331/ijere.1202499
5. Alvermann DE (2002) Effective literacy instruction for adolescents. J Lit Res 34(2):189–208. https://doi.org/10.1207/s15548430jlr3402_4
6. Alexander PA (2018) Engagement and literacy: reading between the lines. J Res Reading 41(4):732–739. https://doi.org/10.1111/1467-9817.12262
7. BATEKİN Y (2025) Improving students' reading, comprehension, and reading motivation through texts generated from their prompts using ChatGPT-4: a mixed-method study. Humanit Social Sci Commun 13(1). https://doi.org/10.1057/s41599-025-06265-x

8. Bai X (2024) Effectiveness of Utilizing ChatGPT as AI Assistance Tool on Essay Writing Performance of Chinese College Students. Light Explorer 9(3). https://doi.org/10.57180/vwqx4610
9. Bauer N, Fütterer T, Brucker B, Gerjets P (2026) Leveraging ChatGPT in Academic Writing: ChatGPT Enhances Students' Writing Quality, Writing Experience, and Ownership. Computers Educ Open 100351. https://doi.org/10.1016/j.caeo.2026.100351
10. Berninger VW, Abbott RD, Cook CR, Nagy W (2016) Relationships of attention and executive functions to oral language, reading, and writing skills and systems in middle childhood and early adolescence. J Learn Disabil 50(4):434–449. https://doi.org/10.1177/0022219415617167
11. Bock AK, Erickson KA (2015) The influence of teacher epistemology and practice on student engagement in literacy learning. Res Pract Persons Severe Disabil 40(2):138–153. https://doi.org/10.1177/1540796915591987
12. Cheong CM, Liao X, Zhu X (2017) Improving teachers' questioning in teaching of reading texts in Chinese based on using the six types of reading comprehension processes. Int J Chin Lang Educ 1(1):73–98
13. Connor CM, Lara J, Crowe J, E. C., Meadows JG (2009) Instruction, student engagement, and reading skill growth in reading first classrooms. Elementary School J 109(3):221–250. https://doi.org/10.1086/592305
14. Culham L (2003) 6 + 1 Traits of Writing: The Complete Guide (Grades 3 and Up). *Scholastic Professional Books*. ISBN: 0439280389
15. Culham R (2006) The trait lady speaks up. Educational Leadersh 64(2):53–57
16. Cipta YA, Rasuki M, Fatkurochman H (2025) IMPROVING EFL STUDENTS' WRITING PERFORMANCE THROUGH CHATGPT. Jurnal Penelitian Humaniora 58–66. https://doi.org/10.23917/humaniora.v26i2.11540
17. de Milliano I (2013) Literacy development of low-achieving adolescents: The role of engagement in academic reading and writing. *Universiteit van Amsterdam* [Host]. ISBN: 9789090274454
18. Duke NK, Cartwright KB (2021) The science of reading progresses: Communicating advances beyond the simple view of reading. Reading Res Q 56(S1):S25–S44. https://doi.org/10.1002/rrq.411
19. Etkin HK, Etkin KJ, Carter RJ, Rolle CE (2025) Differential effects of GPT-based tools on comprehension of standardized passages. Front Educ 10. https://doi.org/10.3389/feduc.2025.1506752
20. Fredricks JA, Blumenfeld PC, Paris AH (2004) School engagement: Potential of the concept, state of the evidence. Rev Educ Res 74(1):59–109. https://doi.org/10.3102/00346543074001059
21. Fitzgerald J, Shanahan T (2000) Reading and writing relations and their development. Educational Psychol 35(1):39–50. https://doi.org/10.1207/S15326985EP3501_5
22. Ge S (2016) 6 + 1 [Improving students' Chinese writing ability through the 6 + 1 elements training approach]. , (29), 70–71. https://doi.org/10.16657/j.cnki.issn1673-9132.2016.29.042

23. Graham S, Hebert M (2010) Writing to read: Evidence for how writing can improve reading. *Alliance for Excellent Education*. https://media.carnegie.org/filer_public/9d/e2/9de20604-a055-42da-bc00-77da949b29d7/ccny_report_2010_writing.pdf
24. Graham S, Hebert M (2011) Writing to read: A meta-analysis of the impact of writing and writing instruction on reading. Harv Educational Rev 81(4):710–744. https://doi.org/10.17763/haer.81.4.t2k0m13756113566
25. Graham S, Sandmel K (2011) The process writing approach: A meta-analysis. J Educational Res 104(6):396–407. https://doi.org/10.1080/00220671.2010.488703
26. Graham S, McKeown D, Kiuhara S, Harris KR (2012) A meta-analysis of writing instruction for students in the elementary grades. J Educ Psychol 104(4):879–896. https://doi.org/10.1037/a0029185
27. Graham S, Liu X, Bartlett B, Ng C, Harris KR, Aitken A, Barkel A, Kavanaugh C, Talukdar J (2018) Reading for writing: A meta-analysis of the impact of reading interventions on writing. Rev Educ Res 88(2):243–284. https://doi.org/10.3102/0034654317746927
28. Guthrie JT, Cox KE, Anderson E, Harris K, Mazzoni S, Rach L (1998) Principles of integrated instruction for engagement in reading. Educational Psychol review 10(2):177–199. https://www.jstor.org/stable/23359424
29. Hebert M, Simpson A, Graham S (2012) Comparing effects of different writing activities on reading comprehension: A meta-analysis. Read Writ 26(1):111–138. https://doi.org/10.1007/s11145-012-9386-3
30. HEUNG YME, Chiu TKF (2025) How ChatGPT impacts student engagement from a systematic review and meta-analysis study. Computers Education: Artif Intell 8. https://doi.org/10.1016/j.caeai.2025.100361
31. Hidi S, Anderson V (2014) Situational interest and its impact on reading and expository writing. In The role of interest in learning and development (pp. 215–238). *Psychology Press*. https://doi.org/10.4324/9781315807430
32. Kim Y-SG, Petscher Y, Wanzek J, Al Otaiba S (2018) Relations between reading and writing: A longitudinal examination from Grades 3 to 6. Read Writing: Interdisciplinary J 31(7):1591–1618. https://doi.org/10.1007/s11145-018-9855-4
33. Kletzien SB, Hushion BC (1992) Reading workshop: Reading, writing, thinking. J Read 35(6):444–451. https://www.jstor.org/stable/40007556
34. Kuehner AV, Hurley J (2019) How integrating reading and writing supports student success. J Dev Educ, 20–26. https://www.jstor.org/stable/45221759
35. Kozlow M, Bellamy P (2004) Experimental study on the impact of the 6 + 1 trait writing model on student achievement in writing. Portland, OR: *Northwest Regional Educational Laboratory*. https://educationnorthwest.org/sites/default/files/resources/Student_Achievement_in_Writing.pdf
36. Lee JS (2014) The relationship between student engagement and academic performance: Is it a myth or reality? J Educational Res 107(3):177–185. https://doi.org/10.1080/00220671.2013.807491

37. Lee JY, Baek M (2023) Effects of gamification on students' English language proficiency: A meta-analysis on research in South Korea. *Sustainability*, 15(14), 11325. https://doi.org/10.3390/su151411325?urlappend=%3Futm_source%3Dresearchgate.net%26utm_medium%3Darticle
38. Li Y, Lerner RM (2011) Trajectories of school engagement during adolescence: implications for grades, depression, delinquency, and substance use. Dev Psychol 47(1):233
39. Lo CK, Hew KF, Jong MS (2024) The influence of ChatGPT on student engagement: A systematic review and future research agenda. Comput Educ 219:105100. https://doi.org/10.1016/j.compedu.2024.105100
40. Mason LH, Meadan H, Hedin LR, Cramer AM (2023) Avoiding the struggle: Instruction that supports students' motivation in reading and writing about content material. *In Motivating Writers in Class* (pp. 70–96). Routledge. https://doi.org/10.4324/9781003416814-5
41. Meşe E, Kuru Ö, Atay D (2025) Facilitating student engagement through ChatGPT. J Educational Res 1–11. https://doi.org/10.1080/00220671.2025.2510394
42. Ndruru F, Cahyono BY, Rovikasari M, Mulati DF (2025) Unpacking the impact of writing task complexity, use of digital tools, and engagement strategies on university students' academic writing performance. J Inform Technol Education: Res 24(1):1–23. https://doi.org/10.28945/5609
43. Ng C, Graham S (2018) Improving literacy engagement: enablers, challenges and catering for students from disadvantaged backgrounds. J Res Reading 41(4):615–624. https://doi.org/10.1111/1467-9817.12252
44. Olson DR (1996) The world on paper: The conceptual and cognitive implications of writing and reading. Cambridge University Press
45. Pan X (1997) History of modern writing education in China [□□□□□□□□□□]. *Fujian People's Publishing House*. ISBN: 9787211028832
46. Pan X (2013) Writing: Life oriented toward self-actualization [□□: □□□□□□□□□□]. *Fujian People's Publishing House*. ISBN: 7030076117
47. Pan X (2017) Chinese language: Expression and existence [□□: □□□□□] (Vols. 1–4). *Fujian People's Publishing House.* ISBN: 9787211076765
48. Park J (2016) Integrating reading and writing through extensive reading. Elt J 70(3):287–295. https://doi.org/10.1093/elt/ccv049
49. Philp J, Duchesne S (2016) Exploring engagement in tasks in the language classroom. Annu Rev Appl Linguist 36:50–72. https://doi.org/10.1017/S0267190515000094
50. Rad HS, Alipour R, Jafarpour A (2024) Using artificial intelligence to foster students' writing feedback literacy, engagement, and outcome: A case of Wordtune application. Interact Learn Environ 32(9):5020–5040. https://doi.org/10.1080/10494820.2023.2208170
51. Reeve J (2013) How students create motivationally supportive learning environments for themselves: The concept of agentic engagement. J Educ Psychol 105(3):579–595. https://doi.org/10.1037/a0032690

52. Roe MF (1997) Combining enablement and engagement to assist students who do not read and write well. Middle School J 28(3):35–41. https://doi.org/10.1080/00940771.1997.11494450
53. Stockwell G (2024) ChatGPT in language teaching and learning: Exploring the road we're travelling. Technol Lang Teach Learn 6(1):2273. https://doi.org/10.29140/tltl.v6n3.2273
54. Su Y, Lin Y, Lai C (2023) Collaborating with ChatGPT in argumentative writing classrooms. Assess Writ 57:100752. https://doi.org/10.1016/j.asw.2023.100752
55. Teng MF, Huang J (2026) Assessing ChatGPT Feedback for EFL Learners' Engagement and Writing Performance. Int J Appl Linguistics. https://doi.org/10.1111/ijal.70155
56. Wang D (2016) 6 + 1 [Improving students' writing ability through the 6 + 1 elements]. , (18), 80
57. Wang M-T, Fredricks J, Ye F, Hofkens T, Linn JS (2019) Conceptualization and assessment of adolescents' engagement and disengagement in school: A Multidimensional School Engagement Scale. Eur J Psychol Assess 35(4):592–606. https://doi.org/10.1027/1015-5759/a000431
58. Wu C-H, Ho VT (2025) Critical factors for why ChatGPT enhances learning engagement and outcomes. Educ Inform Technol. https://doi.org/10.1007/s10639-025-13346-6
59. Yildiz TA (2023) The Impact of ChatGPT on Language Learners' Motivation. J Teacher Educ Lifelong Learn 5(2):582–597. https://doi.org/10.51535/tell.1314355
60. Zhan Y, Yan Z (2025) Students' engagement with ChatGPT feedback: implications for student feedback literacy in the context of generative artificial intelligence. Assess Evaluation High Educ 1–14. https://doi.org/10.1080/02602938.2025.2471821
61. Zhang ZV, Hyland K (2018) Student engagement with teacher and automated feedback on L2 writing. Assess Writ 36:90–102. https://doi.org/10.1016/j.asw.2018.02.004
62. Zheng G (2007) Listening to Ms. Zheng Guihua's classroom teaching []. *East China Normal University Press*. ISBN: 9787561754047
63. Zheng G (2012) Reflection and construction of Chinese teaching []. *Commercial Press*. ISBN: 9787100094603
64. Zheng G (2014) Research on writing instruction []. *Guangxi Education Press*. ISBN: 9787543584310
65. Zhu X (2005) Further development of the model of cognitive abilities and related questions on reading test. J Chin Lang Educ 2(6):18–39
66. Zhu X, Cheong CM, Li GY, Wu J (2020) Primary school teachers' conceptions of reading comprehension processes and its formulation. Front Psychol 11:615. https://doi.org/10.3389/fpsyg.2020.00615
67. Zhu A, Mofreh SAM, Salem S, Li Z, Yao M (2023) A Review of the Effect of Reading Engagement on Reading Achievement. Encyclopaideia 27(67):17–28. https://doi.org/10.6092/issn.1825-8670/16180
68. Zhu JJ (2008) On audience awareness in students' writing and its cultivation []. Global Education [], (9), 81–85

69. Zhu JJ (2009a) A century-long debate: Reading or writing? Re-examining the conflict and integration between Ye Shengtao's and Li Jinxi's views of writing [□□□□: □□, □□□□:□□□□□□□□□□□□□□□□□□□□]. Chin Lang Secondary Schools [□□□□], (7–8), 12–16
70. Zhu JJ (2009b) Exploring the reading–writing relationship in traditional Chinese language education [□□□□□□□□□□□□□□]. *The Communication of Chinese Teaching [□□□□□□□]*, (4B), 48–50
71. Zhu JJ (2010a) Constructing writing communities: An effective strategy for secondary-school writing instruction [□□□□□□□□: □□□□□□□□□□□□□□]. Global Educ [□□□□□□□], (3), 89–92
72. Zhu JJ (2010b) Evaluative reading and instructional strategies: A comparison of textbook adaptations of The Necklace in China and the United States [□□□□□□□□□□□□□□□: □□□□□□□□□□□□□□□□□□□□]. Chin Teach Middle School [□□□□□□□□□], (12), 13–16
73. Zhu JJ (2019) Research on reading–writing integration: Constrained by experience, aspiring beyond it [□□□□□□□: □□□□□, □□□□□]. Foreign Lang Teach Schools [□□□□□□□], (5), 12–15

Figures

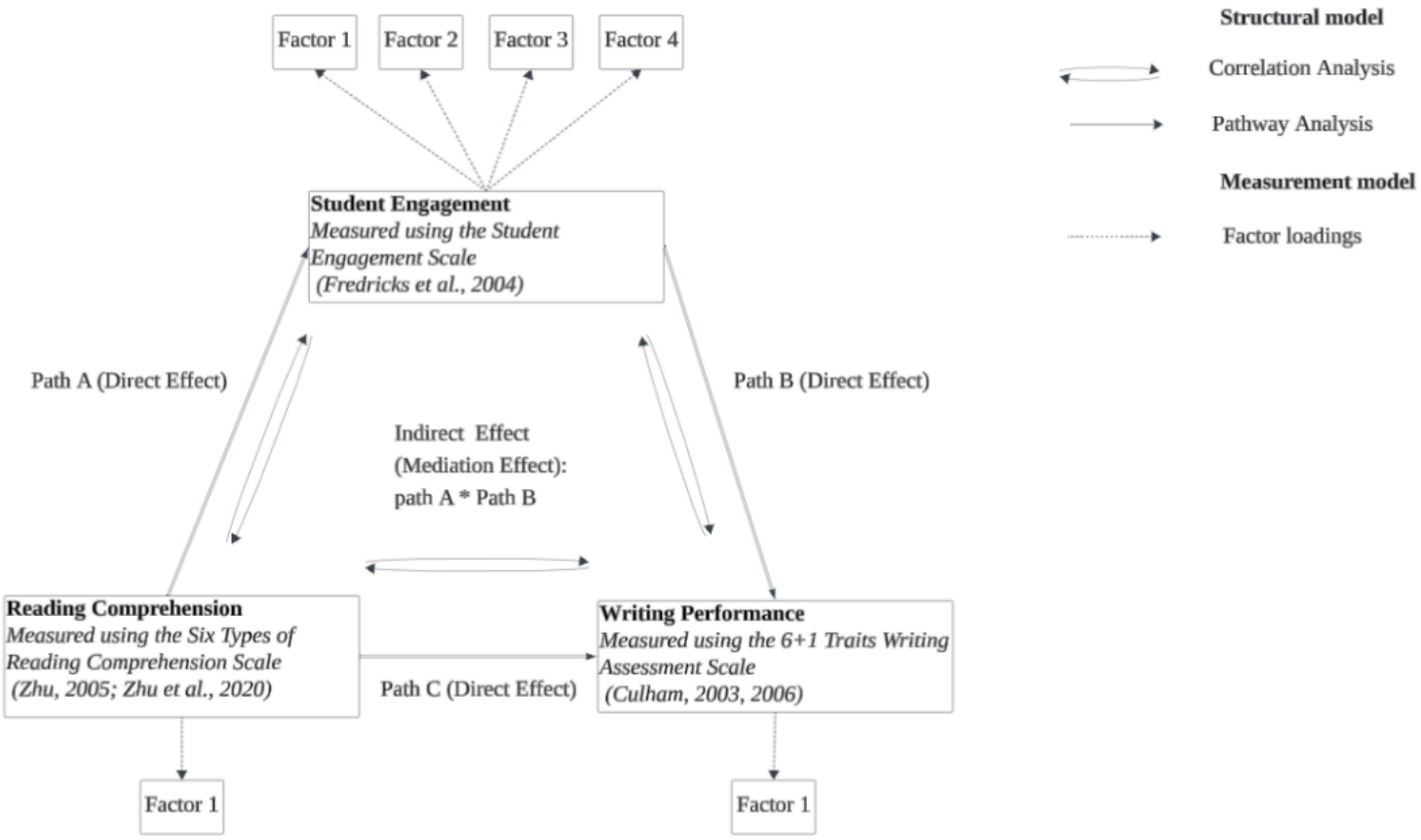


Figure 1

Conceptual Framework of Reading–Writing Integration and Engagement.

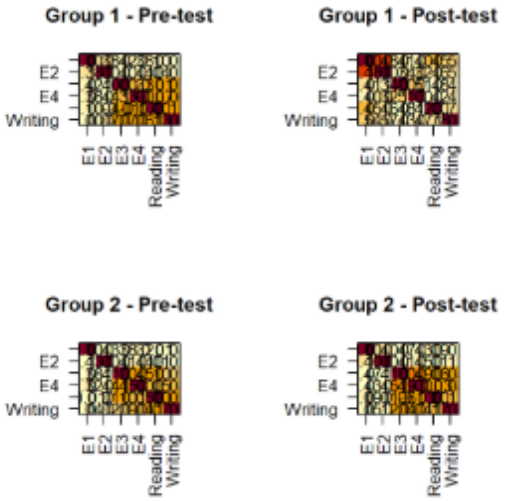


Figure 2

Correlation heatmaps across variables and times