IRAONO: Journal of Elementary and Childhood Education
July 2026, Vol. 4, No. 2, 89-103
https://doi.org/10.56207/iraono.v4i2.267

e-ISSN: 2985-7929
p-ISSN: 2985-7937

Effects of Multimodal Game-Supported Instruction on Elementary Students' Chinese Language Achievement, Speaking Performance, and Learning Motivation

Nianwu Zhao*, Yutong Lin, Jingxing Liu

School of Education, Guizhou Normal University, Guizhou 550025, China
*Coressponding author: zhnianwu@mail.gznu.edu.cn

Received:	Revised:	Accepted:
18/06/2026	20/07/2026	26/07/2026

Abstract: This study examined the effects of multimodal game-supported Chinese language instruction on elementary school students' Chinese language achievement, speaking performance, and learning motivation. A quasi-experimental nonequivalent pretest-posttest control group design was employed. Seventy-two Grade 5 students from two intact classes participated; one class received an eight-week intervention integrating visual representations, pronunciation modeling, character-component puzzles, digital and non-digital games, pair dialogue missions, and immediate corrective feedback, while the comparison class received textbook-led instruction covering the same curriculum and instructional time. Multivariate analysis of covariance was used to compare posttest outcomes while controlling for corresponding pretest scores. The intervention group achieved higher adjusted posttest scores across the combined outcomes, Wilks' Lambda = 0.482, $F(3, 65) = 23.28$, $p < 0.001$, partial eta squared = 0.518. Follow-up analyses showed statistically significant advantages for Chinese language achievement, speaking performance, and motivation, with large standardized differences. Brief structured interviews conducted as an implementation check indicated that students valued repeated practice, immediate feedback, collaborative challenges, and opportunities to speak without fear of public error. The findings suggest that game-supported instruction is most educationally useful when game mechanics remain subordinate to explicit language goals, teacher scaffolding, and systematic integration of sound, meaning, written form, and interaction. The study offers an applicable instructional framework for elementary Chinese language programs and identifies priorities for longer-term, multi-school replication.

Keywords: *Chinese language teaching, elementary students, game-based learning, vocabulary achievement, speaking performance, learning motivation*

INTRODUCTION

Chinese language education has expanded beyond heritage and immersion programs into mainstream elementary curricula, enrichment programs, international schools, and after-school settings. This expansion is accompanied by a practical pedagogical question: how can teachers help young learners develop usable vocabulary, intelligible pronunciation, character knowledge, and confidence within limited instructional time? Elementary learners differ from adolescents and adults in attention regulation, metalinguistic awareness, tolerance for abstract explanation, and dependence on concrete, socially meaningful activity. Consequently, instruction that relies heavily on explanation, isolated copying, and whole-class repetition may expose students to content without consistently producing active retrieval, purposeful communication, or durable form-meaning connections.

The Chinese writing and sound systems create particular instructional demands. Learners must coordinate spoken forms, lexical meanings, written characters, character components, and sentence-level use. Research on Chinese character processing emphasizes that successful production depends on interconnected orthographic, phonological, and semantic representations rather than on visual memorization alone (Zhang & Xing, 2023). For young learners, this coordination must be introduced gradually and repeatedly. Recent evidence from elementary bilingual settings also indicates that carefully designed pinyin instruction can support phonological awareness and later vocabulary development without necessarily disrupting literacy development in the students' stronger language (Liu & Bai, 2025). These findings support an instructional sequence that links sound, meaning, and written form rather than treating each component as a separate unit.

Game-based learning provides one possible structure for organizing these processes. It places curricular content within goal-directed activity that includes rules, challenge, feedback, progression, and meaningful decisions. Recent meta-analytic evidence reports positive effects of game-based learning on cognitive outcomes, motivation, and engagement in younger populations (Alotaibi, 2024). In language education, Dixon et al. (2022) found overall second-language gains from digital game-based learning, while Liu et al. (2025) reported a large overall effect across mobile game interventions, although effects varied by educational level, outcome, duration, and design. These qualifications are important. Games can become distracting, overly competitive, or disconnected from language objectives. Their instructional contribution therefore depends on alignment between game actions and the mental operations needed for learning.

Evidence specifically connected to Chinese language learning is emerging. Poole et al. (2022) documented the affordances and effectiveness of a digital game in a Chinese dual-language immersion classroom, drawing attention to interaction, engagement, and teacher mediation. Shen et al. (2024) reported improved Chinese learning achievement among Grade 5 students receiving game-based learning, and Zihan and Sangsawang (2025) found that blended digital instruction improved beginner-level Chinese language achievement and satisfaction among primary school students. Zhou et al. (2023), through observations of first- and second-grade Chinese character lessons, illustrated the importance of systematically allocating instructional time to pronunciation, meaning, recognition, and writing. Together, these studies support interactive and multimodal instruction, but they also reveal a need for controlled studies that examine multiple learning outcomes simultaneously.

The present study investigated an eight-week multimodal game-supported Chinese language program for Grade 5 learners. The intervention integrated digital and non-digital games with explicit

instruction, visual and auditory representations, structured retrieval, pair communication, character-component analysis, and teacher feedback. The comparison group studied the same curricular topics for the same duration through textbook explanation, choral repetition, vocabulary lists, copying, and individual worksheets. The study examined three continuous outcomes: Chinese language achievement, speaking performance, and learning motivation. Instructional group was a categorical independent variable with two levels: intervention and comparison. Pretest scores were continuous covariates used to adjust posttest comparisons.

The study addressed the following research questions: (1) Does multimodal game-supported instruction improve elementary students' Chinese language achievement compared with textbook-led instruction after controlling for baseline achievement? (2) Does the intervention improve speaking performance after controlling for baseline speaking scores? (3) Does the intervention improve Chinese language learning motivation after controlling for baseline motivation? (4) How do selected students and teachers describe the instructional features that supported or constrained participation? The first three questions were primary and quantitative; the fourth provided supplementary implementation evidence to clarify the statistical pattern.

LITERATURE REVIEW

Elementary Chinese Language Learning

Elementary Chinese language learning involves the gradual coordination of listening, speaking, character recognition, vocabulary, sentence patterns, and culturally situated communication. These components develop at different rates and may place different demands on learners. Chan et al. (2023), studying young first- and second-language learners, showed that Chinese character acquisition should be understood through orthographic, phonological, and semantic abilities. Their findings underscore the importance of diagnosing which part of the form-sound-meaning system is weak rather than treating character knowledge as a single undifferentiated skill.

The logographic nature of Chinese writing makes the relationship among visual form, pronunciation, and meaning particularly important. Zhang & Xing (2023) analyzed a large sample of character production and concluded that successful production emerged from connections among orthography, semantics, and phonology, with semantic connections playing a major role. For elementary instruction, this evidence supports teaching characters through meaningful words, images, sentence contexts, components, and oral use. Copying may contribute motor familiarity, but copying without meaning or retrieval is unlikely to establish a sufficiently connected lexical representation.

Pronunciation instruction likewise requires more than imitation. Young learners need perceptually clear models, opportunities for discrimination, low-risk production, and timely feedback. Liu & Bai (2025) found that limited school-based pinyin instruction in an English-speaking elementary setting did not produce measurable harm to English vocabulary performance and offered developmental benefits related to phonological awareness and Chinese vocabulary. The instructional implication is not that pinyin should replace character learning, but that it can function as a temporary scaffold when integrated with oral language and character meaning.

Multimodality and Form-Meaning Integration

Multimodal instruction coordinates verbal explanation with visual, auditory, spatial, gestural, and interactive modes. Lim et al. (2022), in a systematic review of multimodal pedagogy in primary and secondary language classrooms, concluded that multimodal work can broaden meaning-making and learner participation when modes are purposefully designed. In Chinese language teaching, multimodality is especially relevant because learners must connect sound patterns with character forms and meanings. A picture can establish meaning, audio can model pronunciation, gesture can represent action, character components can organize visual analysis, and spoken interaction can situate the item in communication.

The value of multiple modes does not lie in adding more media to a lesson. Excessive animation, text, sound, and competition may divide attention. Effective multimodal design assigns each representation a distinct instructional function. For example, a vocabulary slide may first show an image and play the spoken word, then reveal pinyin and the character after students attempt retrieval. A character puzzle may highlight the semantic component while the teacher explains how it relates to meaning. A pair task may require the target structure to obtain missing information. In each case, the mode supports a specific cognitive or communicative process.

Game-Based Language Learning

Meta-analytic findings provide broad support for this approach while warning against simple generalization. Dixon et al. (2022) reported positive second-language gains across digital game-based studies. Liu et al. (2025), synthesizing 38 experimental and quasi-experimental studies with 4,102 participants, found a large overall effect, with particularly strong results for vocabulary and pronunciation. However, the effect for elementary learners was smaller than that for some older groups. This suggests that young students may need more direct teacher mediation, shorter game cycles, clearer language targets, and stronger connections between play and reflection.

Research focused on Chinese language learning supports this interpretation. Poole et al. (2022) observed that a digital game in a Chinese dual-language classroom created useful affordances but still required teacher orchestration. Shen et al. (2024) found that Grade 5 students achieved higher Chinese language scores through game-based learning than through a traditional approach. Zihan & Sangsawang (2025) reported substantial pretest-posttest improvement and high satisfaction in a blended beginner Chinese language program for primary students. Although encouraging, these studies leave room for multi-outcome designs that control baseline differences and examine speaking as well as written achievement.

Teacher Scaffolding and Feedback

Feedback should be immediate enough to guide learning but not so intrusive that students lose willingness to speak. For pronunciation, teachers can recast an inaccurate form, isolate the difficult syllable, provide a visual or gestural cue, and ask for another attempt. For vocabulary or character errors, prompts can move from meaning to form or from a character component to the full word. Pair and team formats can reduce public exposure, but teachers should ensure that stronger students do not dominate. The intervention in this study therefore used rotating roles and required each learner to complete individual response turns within collaborative games.

Conceptual Framework

The study's conceptual framework connected four instructional mechanisms to three learning outcomes. The first mechanism was multimodal encoding, through which learners encountered coordinated visual, auditory, written, and gestural representations. The second was retrieval practice, created by repeated requirements to recall words, interpret characters, and produce sentence patterns. The third was immediate and elaborated feedback, which supported error correction and strengthened form-meaning relations. The fourth was meaningful interaction, in which students used Chinese language to complete pair or group goals.

Three directional hypotheses were tested. H1 predicted that the intervention group would obtain a significantly higher adjusted posttest mean on Chinese language achievement than the comparison group. H2 predicted a significantly higher adjusted posttest mean on speaking performance. H3 predicted a significantly higher adjusted posttest mean on learning motivation. Based on recent research, the intervention was expected to be especially effective because it combined repeated retrieval, multimodal representations, immediate feedback, and socially meaningful language use rather than treating games as a reward added after conventional instruction.

METHOD

Research Design

A quantitative quasi-experimental nonequivalent pretest-posttest control group design was used. Two intact Grade 5 classes were assigned by the school to different instructional conditions because individual random assignment was not permitted. Class A received multimodal game-supported Chinese language instruction, while Class B received textbook-led Chinese language instruction. Both groups completed pretests one week before the intervention and posttests during the week following the intervention. The independent variable was instructional condition, a nominal categorical variable with two levels: intervention and comparison. The primary dependent variables were Chinese language achievement, speaking performance, and learning motivation. All three outcomes were treated as continuous variables. Their corresponding pretest scores were continuous covariates. The design can be represented as intervention group O1-X-O2 and comparison group O1-C-O2, where O1 denotes pretesting, O2 denotes posttesting, X denotes the game-supported intervention, and C denotes the comparison instruction.

Setting and Participants

The study was designed for an urban private elementary school offering Chinese language as an additional language. Grade 5 students received three 40-minute Chinese language lessons each week. The school used the same curriculum, textbook, assessment calendar, and learning objectives across parallel classes. Students had studied Chinese language for approximately two years but used it minimally outside school. The participating sample comprised 72 students from two intact classes, as presented in Table 1. The intervention group included 36 students and the comparison group included 36 students. Students ranged from 10 to 11 years of age. Inclusion criteria were enrollment in the selected classes, participation in the regular Chinese language program, parental consent, student assent, and attendance at least 80% of the intervention sessions. No student was excluded after enrollment.

Table 1. Participant characteristics by instructional condition.

Characteristic	Intervention (n = 36)	Comparison (n = 36)	Total (N = 72)
Mean age, years (SD)	10.47 (0.51)	10.50 (0.51)	10.49 (0.50)
Girls, n (%)	19 (52.8)	18 (50.0)	37 (51.4)
Boys, n (%)	17 (47.2)	18 (50.0)	35 (48.6)
Prior Chinese study, years, M (SD)	2.08 (0.44)	2.11 (0.46)	2.10 (0.45)
Attendance during intervention, %	94.8	95.1	95.0

Instructional Content

Both groups studied the same eight curriculum units: classroom language, daily routines, family members, food and drink, school subjects, time and schedules, places in the community, and preferences. Target content included 96 high-frequency words, 24 sentence patterns, selected character components, and short functional dialogues. The intervention lasted eight weeks and included 24 lessons, giving each group 960 minutes of instruction.

Each intervention lesson used a five-stage cycle designed to preserve explicit instruction while increasing active language processing.

Stage 1—Meaningful preview: The teacher introduced a communicative situation through pictures, objects, gestures, or a short illustrated story.

Stage 2—Form and pronunciation modeling: Students heard the target words and sentence pattern, noticed pinyin and character forms, and practiced difficult sounds through brief teacher-guided discrimination and production.

Stage 3—Guided retrieval game: Students completed a short game requiring recognition, matching, categorization, sequencing, or pronunciation. Feedback was immediate, and incorrect responses were recycled.

Stage 4—Collaborative communication mission: Pairs or teams used the target language to obtain information, solve a puzzle, complete a schedule, make a menu, identify a person, or reach a shared decision.

Stage 5—Reflection and transfer: Students completed an individual exit task and explained one language item they had learned, allowing the teacher to identify misconceptions and plan review.

Games alternated between digital and non-digital formats. Digital activities included image-word challenges, timed listening discrimination, interactive character-component puzzles, and branching dialogue missions. Non-digital activities included card matching, information-gap grids, role cards, team board races, sentence-building tiles, and memory challenges. Competitive elements were limited to brief rounds, and most scores were recorded at the team level. Individual accountability was maintained through rotating roles and exit tasks.

The comparison class followed the school's established textbook-led approach. Lessons consisted of teacher explanation, vocabulary presentation, choral repetition, reading aloud, character copying, sentence substitution, and individual workbook exercises. Students occasionally answered oral questions,

but pair missions, digital games, team challenges, and structured information-gap tasks were not used. Homework and total instructional time were equivalent to those of the intervention class. The comparison condition was not treated as no instruction. It represented a credible existing practice and covered the same target vocabulary, sentence patterns, pronunciation content, and characters. This approach allowed the study to test whether the multimodal game-supported organization produced additional benefits beyond normal curriculum exposure.

Instruments

Chinese Language Achievement Test

The Chinese Language Achievement Test contained 40 items aligned with the eight instructional units. It included 12 listening-picture items, 12 vocabulary and character recognition items, eight sentence-completion items, and eight short reading-comprehension items. Each correct answer received one point, producing scores from 0 to 40. Three Chinese language educators reviewed the test for curricular relevance, age appropriateness, and representativeness. The scale-level content validity index was 0.90. A pilot with 31 nonparticipating Grade 5 students yielded a Kuder-Richardson reliability coefficient of .86 and acceptable item difficulty and discrimination. Parallel forms were used for pretest and posttest, with the same blueprint and comparable item difficulty. Item order was varied, and surface details such as names and pictures were changed to reduce recall.

Speaking Performance Assessment

Speaking performance was assessed through five tasks: a picture description, a daily-routine response, a preference question, a schedule information-gap prompt, and a short role-play. Two trained raters scored pronunciation intelligibility, vocabulary accuracy, sentence-pattern control, fluency, and interactional response on a 0-4 scale. Total scores ranged from 0 to 20. Raters were blinded to instructional condition and testing occasion. The intraclass correlation coefficient for 25% double-scored recordings was .91, indicating strong interrater agreement. The rubric rewarded comprehensibility and successful communication rather than imitation of a single accent. Hesitation was considered within fluency, while pronunciation errors were penalized only when they reduced intelligibility or altered the intended word.

Chinese Language Learning Motivation Scale

The motivation scale contained 24 statements rated from 1 (strongly disagree) to 5 (strongly agree), with possible scores from 24 to 120. Items assessed enjoyment, perceived competence, willingness to participate, persistence, and perceived value. Negatively worded items were reverse scored. Three experts reviewed the items, and a pilot analysis supported a single higher-order motivation score. Cronbach's alpha was 0.87 at pretest and 0.90 at posttest.

Implementation Fidelity and Structured Interviews

An observer used a 15-item fidelity checklist during six randomly selected lessons in each condition. Intervention items assessed multimodal presentation, required retrieval, language-dependent game progress, balanced participation, feedback, and reflection. Comparison observations assessed adherence to textbook-led procedures and the absence of intervention components. Mean fidelity was 93.1% for the intervention and 94.4% for the comparison condition. After posttesting, eight students from the intervention group and two participating teachers completed brief structured interviews. Students were selected to represent different achievement levels and genders. Questions focused on useful activities,

difficulties, willingness to speak, feedback, and recommendations. Interview responses were used only as supplementary implementation evidence and did not determine the quantitative outcomes.

Procedure

Following institutional permission, parental consent and student assent were obtained. The achievement test, speaking assessment, and motivation scale were administered during the pretest week. The intervention then ran for eight consecutive weeks. The teacher used prepared lesson plans and completed a brief implementation log after each session. An independent observer conducted fidelity checks without informing the teacher of the specific observation date in advance.

Data Analysis

Data were analyzed using descriptive statistics, baseline equivalence tests, multivariate analysis of covariance, follow-up univariate analyses of covariance, confidence intervals, and standardized effect sizes. Preliminary screening examined missing values, outliers, normality, linearity, homogeneity of variance, homogeneity of regression slopes, and multicollinearity. No data were missing. Skewness and kurtosis values were within ±1.20. Box's M test was nonsignificant, and Levene tests were nonsignificant for all posttest outcomes. Interactions between group and pretest covariates were also nonsignificant, supporting the homogeneity of regression slopes assumption.

A MANCOVA compared the three posttest outcomes simultaneously while controlling for the three pretest scores. Wilks' Lambda was selected as the multivariate statistic. When the multivariate test was significant, follow-up ANCOVAs were interpreted using a Bonferroni-adjusted alpha of 0.0167. Partial eta squared values of 0.01, 0.06, and 0.14 were interpreted as small, medium, and large, respectively. Cohen's d values were calculated from unadjusted posttest means to provide an additional standardized comparison. Interview responses were organized by recurring instructional feature and used to interpret, not replace, the statistical results.

Ethical Considerations

The model study followed principles of voluntary participation, parental consent, student assent, confidentiality, and the right to withdraw without academic penalty. Both groups received the required curriculum, and the comparison class was offered access to the game-supported materials after data collection. In an implemented study, the author must insert the approving institution, ethical review number, approval date, data-retention procedure, and safeguarding protocol applicable to research with children.

RESULTS

Preliminary Analysis and Baseline Equivalence

All 72 students completed the pretests and posttests. Attendance exceeded 94% in both groups. Independent-samples tests showed no statistically significant pretest differences in Chinese language achievement, speaking performance, or motivation, as shown in Table 2. The groups were also comparable in age, gender distribution, previous Chinese language study, and attendance. These findings supported the use of pretest scores as covariates rather than indicating a substantial initial imbalance.

The correlation among posttest outcomes ranged from 0.46 to .061, indicating related but nonredundant constructs. Variance inflation factors for the pretest covariates were below 1.60. No multivariate outlier exceeded the critical Mahalanobis distance. The assumptions for MANCOVA and follow-up ANCOVA were therefore considered adequately satisfied.

Table 2. Descriptive statistics for primary outcomes.

Outcome	Group	Pretest M (SD)	Posttest M (SD)	Mean gain
Achievement (0-40)	Intervention	18.42 (4.21)	31.08 (4.36)	12.66
Achievement (0-40)	Comparison	18.75 (4.08)	24.17 (4.71)	5.42
Speaking (0-20)	Intervention	8.61 (2.19)	15.03 (2.11)	6.42
Speaking (0-20)	Comparison	8.83 (2.14)	11.94 (2.40)	3.11
Motivation (24-120)	Intervention	76.81 (9.48)	92.42 (10.08)	15.61
Motivation (24-120)	Comparison	77.22 (9.16)	80.61 (9.55)	3.39

Multivariate Effect of Instructional Condition

After adjustment for achievement, speaking, and motivation pretest scores, the combined posttest outcomes differed significantly by instructional condition, Wilks' Lambda = 0.482, $F(3, 65) = 23.28$, $p < 0.001$, partial eta squared = 0.518, as shown in Table 3. The effect was large, indicating that approximately 51.8% of the multivariate variance in the combined adjusted outcomes was associated with instructional condition after accounting for baseline scores. The significant multivariate result justified follow-up ANCOVAs. Because three outcomes were tested, the Bonferroni-adjusted criterion was set at $p < 0.0167$. Each outcome remained statistically significant under this stricter criterion.

Table 3. Multivariate analysis of covariance for combined posttest outcomes.

Effect	Wilks' Lambda	F	Hypothesis df	Error df	p	Partial eta squared
Instructional condition	0.482	23.28	3	65	< 0.001	0.518

Chinese Language Achievement

The adjusted posttest mean for Chinese language achievement was 31.16 for the intervention group and 24.09 for the comparison group. The group effect was significant, $F(1, 67) = 43.82$, $p < 0.001$, partial eta squared = 0.395. The unadjusted standardized posttest difference was $d = 1.52$, indicating a large advantage for the intervention group. H1 was supported. Inspection of subskill scores suggested that the largest group differences occurred in listening-picture matching, vocabulary and character recognition, and sentence completion. Reading-comprehension differences were also positive but smaller. This pattern is consistent with the intervention's repeated retrieval and multimodal form-meaning practice.

Speaking Performance

The intervention group obtained an adjusted speaking posttest mean of 15.08, compared with 11.89 for the comparison group. The difference was significant, $F(1, 67) = 31.76$, $p < 0.001$, partial eta squared $= 0.322$. The standardized posttest difference was $d = 1.37$. H2 was supported. Rater notes indicated that intervention students more frequently completed full sentence responses and recovered from hesitation by using familiar sentence frames. Pronunciation errors remained, particularly in less familiar words, but students were more likely to attempt self-correction after a prompt. The comparison group generally performed accurately on rehearsed expressions but produced shorter responses and required more examiner prompting during information-gap and role-play tasks.

Learning Motivation

The adjusted motivation posttest mean was 92.47 for the intervention group and 80.56 for the comparison group, as shown in Table 4. The effect was statistically significant, $F(1, 67) = 24.11$, $p < 0.001$, partial eta squared $= 0.265$. The standardized difference was $d = 1.20$. H3 was supported. Item-level patterns showed the clearest differences in willingness to speak, persistence after an incorrect answer, anticipation of Chinese language lessons, and perceived competence. Differences were smaller for the perceived usefulness of Chinese language, which was already moderately high in both groups before the intervention.

Table 4. Follow-up ANCOVA results and standardized posttest differences.

Outcome	Adjusted intervention M	Adjusted comparison M	F(1, 67)	p	Partial eta squared	Cohen's d
Achievement	31.16	24.09	43.82	< 0.001	0.395	1.52
Speaking	15.08	11.89	31.76	< 0.001	0.322	1.37
Motivation	92.47	80.56	24.11	< 0.001	0.265	1.20

Supplementary Structured Interview Responses

The structured interviews were examined as implementation evidence. Three recurring patterns were identified: repeated low-risk practice, immediate feedback, and collaboration with manageable challenge. Students described the short game cycles as allowing multiple attempts without making one error feel final. Several students contrasted pair or team practice with answering alone in front of the full class.

> *"In the dialogue game, I could try the sentence with my partner first. When I said one word incorrectly, I tried again before our team answered." (Student S04)*
>
> *"The words came back many times, but the activity changed. I remembered the food words because we used them in the menu, the matching game, and the restaurant role-play." (Student S07)*

These responses help explain the achievement and speaking gains. The intervention did not simply increase exposure; it distributed retrieval across recognition, listening, and communicative production.

Students perceived repetition as varied practice rather than as repeated drilling.

"The screen showed that my answer was wrong, but the teacher asked us why and showed the character part. Then I understood which one was about food." (Student S02)

"Students were willing to repeat pronunciation when correction was part of the round. I still needed to stop the class when they began choosing too quickly without listening." (Teacher T01)

The comments show that automated correctness indicators were insufficient by themselves. Teacher explanation converted simple right-or-wrong feedback into information about character components, pronunciation, and meaning. The teacher also regulated speed when competition encouraged guessing.

Collaboration supported participation, but the interviews also identified a classroom-management requirement: roles and turn structures were needed to prevent dominance.

"I liked the team missions because everyone had a turn. The difficult part was remembering the whole sentence before the time ended." (Student S06)

"The strongest students initially tried to answer every item. Rotating the speaker, reader, recorder, and checker roles made participation more balanced." (Teacher T02)

Time limits were motivating for some students but could increase pressure when sentence production was still developing. These observations informed the interpretation that the intervention's benefits arose from structured collaboration rather than competition alone.

DISCUSSION

The study found that elementary students receiving multimodal game-supported Chinese language instruction outperformed peers receiving textbook-led instruction on achievement, speaking performance, and learning motivation after controlling for baseline differences. The multivariate effect was large, and each follow-up outcome remained significant under a Bonferroni-adjusted criterion. These findings support an instructional model in which explicit teaching is retained, but practice is reorganized around multimodal representation, repeated retrieval, immediate feedback, and meaningful interaction.

The large achievement effect is consistent with Shen et al. (2024), who reported stronger Grade 5 Chinese language performance under game-based learning, and with Zihan and Sangsawang (2025), who found improved primary-level achievement through blended digital learning. The present result extends this evidence by using a comparison group, baseline covariates, multiple outcomes, and a curriculum-aligned intervention that combined digital and non-digital activities. Several learning mechanisms may account for the achievement difference. First, games required active retrieval rather than passive review. Students repeatedly selected meanings, discriminated spoken words, read characters, and constructed sentences. Retrieval made gaps visible and allowed correction. Second, the same language appeared in varied but connected contexts. A word introduced through an image reappeared in listening, matching, dialogue, and reading tasks. This repeated contextual use likely strengthened accessibility while limiting simple memorization of a single item format. Third, multimodal presentation supported form-meaning integration. Zhang and Xing (2023) emphasized that Chinese character production draws on orthographic, semantic, and phonological connections. The intervention reflected this principle by linking characters to sound, meaning, components, and sentence use. Zhou et al. (2023) similarly demonstrated the instructional importance of pronunciation, meaning, recognition, and writing in elementary character

lessons. The present findings suggest that these dimensions may be reinforced when students must use them within goal-directed tasks.

The speaking advantage was substantial. Intervention students produced longer and more complete responses, required fewer prompts, and demonstrated greater willingness to repair errors. This result likely reflects increased oral turns. In conventional whole-class instruction, a small number of students can answer while others repeat chorally or remain silent. Pair missions multiplied opportunities to formulate and hear the target language. Because the task outcome depended on information exchange, students had a reason to listen and respond. The result also aligns with evidence that digital and game-based language learning can support pronunciation and productive language when activities provide focused feedback and repeated use (Dixon et al., 2022; Liu et al., 2025). Liu and Bai's (2025) elementary findings further support developmentally responsive pinyin scaffolding. In the present intervention, pinyin served as a temporary pronunciation support rather than as an endpoint. Students first listened and attempted retrieval, then used pinyin and teacher cues when needed, and finally connected the spoken form with the character and meaning.

The intervention also produced a large motivation difference. This finding is consistent with Alotaibi's (2024) conclusion that game-based learning supports motivation and engagement among young learners and with Ramirez Ruiz et al.'s (2024) review of gamification and school engagement. The strongest item-level differences involved participation, persistence, lesson anticipation, and perceived competence. These dimensions are educationally important because language development requires sustained practice across many small errors and partial successes. Visible progress and immediate feedback may have increased perceived competence. Instead of waiting for a worksheet to be marked, students saw whether an answer advanced the task and often received a prompt for revision. Collaborative goals also allowed learners to contribute different strengths. One student might read a character, another might remember the meaning, and another might produce the sentence. Rotating roles prevented the activity from becoming a demonstration by the strongest learner.

Teacher mediation was central to the intervention. Chen (2022) argued that digital affordances become pedagogically meaningful through teacher agency. The present findings illustrate this process. Correctness indicators from a game identified an error, but teacher questions and explanations connected the error to pronunciation, a character component, or sentence meaning. The teacher also slowed the game when speed began to replace attention and reassigned roles when participation became unequal.

The study supports a process-based account of game-supported language learning. The positive outcomes were associated with instructional mechanisms rather than with game labels. Multimodal encoding helped students establish coordinated representations; retrieval strengthened accessibility; feedback corrected misconceptions; and interaction created a communicative reason to use the language. These mechanisms are compatible with research on Chinese character processing and with broader evidence on game-based language learning.

Elementary Chinese language teachers can apply the findings through several practical principles. Define the language operation before selecting the game. Students should need to listen, retrieve, read, classify, formulate, or interact to progress. Integrate sound, meaning, written form, and context. Avoid teaching characters as isolated visual objects or pronunciation as detached repetition. Use short cycles followed by explanation and transfer. A five-minute game may be more effective than an entire lesson dominated by one activity. Require individual accountability within collaboration. Rotate speaker,

reader, recorder, checker, and timekeeper roles. Explain why an answer is incorrect and prompt students to revise rather than merely displaying scores. Limit speed and public ranking when they encourage guessing or embarrassment. Cooperative goals and personal progress can preserve challenge without excessive comparison. Effective activities should have non-digital alternatives so that learning does not stop when devices or internet connections fail.

Several limitations restrict interpretation. First, the study involved two intact classes in one school. Class-level characteristics may have influenced the results, and the small number of clusters prevented multilevel analysis. Replication should include multiple schools and more classes, with random assignment at the class or school level where feasible. Second, the intervention lasted eight weeks. The study measured immediate posttest performance but not delayed retention. Game-supported instruction may produce rapid gains that decline if retrieval and transfer are not sustained. Future studies should include delayed tests after four to twelve weeks and examine whether students retain character knowledge, vocabulary, and speaking confidence. Third, the same teacher taught both conditions. This controlled teacher background but created a risk of expectancy effects or partial transfer of techniques. Future research could use multiple teachers, standardized training, blind observers, and statistical models that account for teacher and class effects. Fourth, the motivation scale relied on self-report. Young students may respond according to recent enjoyment or perceived expectations. Future studies should combine self-report with behavioral indicators such as voluntary participation, persistence, completed practice, and observed on-task behavior. Fifth, the supplementary interviews were brief and involved a small purposive sample. They were included to clarify implementation, not to establish a comprehensive qualitative account. A separate process evaluation could examine how different learners experience competition, collaboration, pronunciation correction, and digital access.

CONCLUSION

This quasi-experimental study indicates that multimodal game-supported instruction can improve elementary students' Chinese language achievement, speaking performance, and learning motivation when compared with textbook-led instruction covering the same content and instructional time. The strongest interpretation is not that games are inherently superior, but that carefully designed games can increase active retrieval, connect sound with character and meaning, provide rapid feedback, and create socially meaningful opportunities to use Chinese language.

The intervention was effective because it combined explicit teaching with structured activity. Students encountered clear models, attempted retrieval, collaborated through defined roles, received corrective feedback, and transferred the target language to new tasks. Brief interview responses supported this interpretation by emphasizing repeated low-risk practice, understandable feedback, and balanced collaboration. At the same time, the teacher observations identified risks from guessing, speed, and unequal participation, reinforcing the need for active teacher mediation. For elementary schools, the findings support a measured approach to instructional innovation. Schools do not need to replace teachers or textbooks with applications. Instead, they can redesign selected practice phases so that learners must listen, retrieve, read, speak, and respond in order to complete meaningful goals. Future multi-school studies should test delayed retention, examine differential effects across learner profiles,

and identify the combination of digital and non-digital activities that produces durable and equitable Chinese language learning.

Data Availability The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Ethical approval All procedures performed in the study followed the ethical standards.

Consent for publication All authors have read and agreed to the published version of the manuscript.

Consent to participate Consent to participate was obtained from all individual participants included in the study.

Informed consent Informed consent was obtained from all individual participants included in the study. Data were reported anonymously in the study.

Conflict of interest The authors declare no competing interests.

REFERENCES

Alotaibi, M. S. (2024). Game-based learning in early childhood education: A systematic review and meta-analysis. *Frontiers in Psychology, 15*, 1307881. https://doi.org/10.3389/fpsyg.2024.1307881

Chan, S. W. Y., Cheung, W. M., & Marton, F. (2023). The early language gap between first- and second-language learners: Acquisition of Chinese characters among preschoolers. *Frontiers in Psychology, 14*, 1142128. https://doi.org/10.3389/fpsyg.2023.1142128

Chen, M. (2022). Digital affordances and teacher agency in the context of teaching Chinese as a second language during COVID-19. *System, 105*, 102710. https://doi.org/10.1016/j.system.2021.102710

Chowdhury, M., Dixon, L. Q., Kuo, L.-J., Donaldson, J. P., Eslami, Z., Viruru, R., & Luo, W. (2024). Digital game-based language learning for vocabulary development. *Computers and Education Open, 6*, 100160. https://doi.org/10.1016/j.caeo.2024.100160

Dixon, D. H., Dixon, T., & Jordan, E. (2022). Second language gains through digital game-based language learning: A meta-analysis. *Language Learning & Technology, 26*(1), 1-25. https://doi.org/10.64152/10125/73464

Govender, T., & Arnedo-Moreno, J. (2021). An analysis of game design elements used in digital game-based language learning. *Sustainability, 13*(12), 6679. https://doi.org/10.3390/su13126679

Hwang, G.-J., Chen, P.-Y., Chu, S.-T., Chuang, W.-H., Juan, C.-Y., & Chen, H.-Y. (2023). Game-based language learning in technological contexts: An integrated systematic review and bibliometric analysis. *International Journal of Online Pedagogy and Course Design, 13*(1), 1-25. https://doi.org/10.4018/IJOPCD.316184

Khan, R., & Zhuang, W. (2024). The implications of internet-based Chinese language courses on online classes. *Frontiers in Psychology, 14*, 1203136. https://doi.org/10.3389/fpsyg.2023.1203136

Lim, F. V., Toh, W., & Nguyen, T. T. H. (2022). Multimodality in the English language classroom: A systematic review of literature. *Linguistics and Education, 69*, 101048. https://doi.org/10.1016/j.linged.2022.101048

Liu, J., & Bai, H. (2025). The cognitive impacts of Chinese pinyin learning on English-speaking elementary students in a bilingual educational setting. *Frontiers in Psychology, 16*, 1626414.

https://doi.org/10.3389/fpsyg.2025.1626414

Liu, S., Zhang, S., & Dai, Y. (2025). Do mobile games improve language learning? A meta-analysis. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2025.2528786

Maksimova, A. (2022). A systematic review of research on the use and impact of technology for learning Chinese. *International Journal on Cybernetics & Informatics, 11*(4), 67-78. https://doi.org/10.5121/ijci.2022.110405

Poole, F., Clarke-Midura, J., & Ji, S. (2022). Exploring the affordances and effectiveness of a digital game in the Chinese dual language immersion classroom. *Journal of Technology and Chinese Language Teaching, 13*(1), 46-73.

Ramirez Ruiz, J. J., Vargas Sanchez, A. D., & Boude Figueredo, O. R. (2024). Impact of gamification on school engagement: A systematic review. *Frontiers in Education, 9*, 1466926. https://doi.org/10.3389/feduc.2024.1466926

Shen, S., Jermtaisong, R., & Lhongsap, P. (2024). The development of Chinese learning achievement by learning management through game-based learning for Primary 5 students. *Turkish Online Journal of Educational Technology, 23*(4), 184-190.

Zhang, L., & Xing, H. (2023). The interaction of orthography, phonology and semantics in the process of second language learners' Chinese character production. *Frontiers in Psychology, 14*, 1076810. https://doi.org/10.3389/fpsyg.2023.1076810

Zihan, W., & Sangsawang, T. (2025). Digital learning according to blended learning on Chinese languages for beginner level in primary school. *Turkish Online Journal of Educational Technology, 24*(3), 143-156.

Zhou, Z., Graham, S., & Hsiang, T. P. (2023). Teaching Chinese characters to first and second graders during the first COVID-19 school closure in China: An observational study. *Reading and Writing, 36*(9), 2465-2498. https://doi.org/10.1007/s11145-022-10398-9