Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 26, JUNE 2026

DOI: 10.37943/LHFX3759

Aigerim Kereibayeva
PhD, Senior lecturer, Department of Languages
linakei87@gmail.com, orchid.org/0000-0002-0330-9033
International Information Technology University, Kazakhstan
Marco Angelo Vasquez
Master's Degree, Assistant Professor, Department of Languages
m-a.vasquez@iitu.edu.kz, orchid.org/0000-0003-2609-3009
International Information Technology University, Kazakhstan

TEACHER-CURATED AI FOR COMMUNICATIVE CHINESE ACQUISITION: A CASE STUDY AT KAZAKH ABLAIKHAN UNIVERSITY

Abstract: Mastering the Chinese logographic system remains a major challenge in Teaching Chinese as a Second Language (TCSL), often leading to a "communicative gap," in which structural inaccuracies hinder effective language use. This study evaluates the pedagogical effectiveness of a teacher-curated GPT-4o interface designed to transform character acquisition from rote memorization into active semantic deconstruction. Using a quasi-experimental mixed-methods design, the research involved 29 first-year students (N = 29) at Kazakh Ablai Khan University of International Relations and World Languages during the 2025–2026 academic year. The intervention employed an eight-week "Free Chat" protocol guided by Socratic "Prime Prompts" to mitigate "complexity paralysis," defined as the cognitive overload caused by characters with high stroke density. Quantitative analysis, utilizing a Weighted Orthographic Accuracy Model (WOAM) and a 2×2 mixed-design ANOVA, revealed a significant interaction effect, $F(1, 27) = 15.82$, $p < .001$, $\eta_p^2 = .37$. The experimental group achieved a post-test mean accuracy of 84.20%, with a large effect size (Cohen's d = 1.69). Qualitative thematic analysis of GPT-4o interaction logs and interviews indicated that the model's Liushu-based (六书) interpretations functioned as a cognitive scaffold, facilitating a "leveling effect" through L1-mediated (Russian/Kazakh) semantic narratives. However, the emergence of a "Digital-Analog Gap" suggests that while digital scaffolding enhances orthographic logic, it does not automatically develop the haptic–motor memory required for manual calligraphy. The findings therefore indicate that teacher-curated AI can serve as a powerful "communicative bridge," provided pedagogy incorporates a deliberate "fading" of the scaffold to prevent epistemic outsourcing.

Keywords: Artificial Intelligence; Chinese character acquisition; Teaching Chinese as a Second Language; communicative competence; semantic awareness; digital transformation; Kazakhstan higher education;

Introduction

The global educational landscape is undergoing a systemic transformation driven by the rapid maturation of Generative Artificial Intelligence (AI). In the Republic of Kazakhstan, this transformation has been framed as a key pillar of national competitiveness. It is emphasized by President Kassym-Jomart Tokayev's strategic mandate to achieve total AI literacy by 2027 [1]. Within the field of Teaching Chinese as a Second Language (TCSL), this shift necessitates a transition from traditional rote-based pedagogical approaches toward human-centered digital learning frameworks. Despite these developments, a significant research gap remains. While existing scholarship has extensively examined the use of AI in English as a Second Language instruction, the specific orthographic and semantic challenges associated with the Chinese logographic writing system among Central Asian learners remain insufficiently explored. To address this gap, this paper presents an institutional case study focusing on undergraduate learners within the specialized language curriculum at Kazakh Ablai Khan University of International Relations and World languages.

Received: 25.02.2026 Accepted: 17.07.2026 Published: 30.06.2026

DOI: 10.37943/LHFX3759

To ensure empirical rigor, this research anchors its constructs in established SLA frameworks while engaging with critical debates on technological mediation. Teacher-curated AI is operationalized as "pedagogical scaffolding" within Vygotsky's Sociocultural Theory, functioning as a controlled environmental variable within the Zone of Proximal Development (ZPD). This curation is a theoretical necessity to mitigate algorithmic bias the tendency of Large Language Models (LLMs) to prioritize standardized or culturally skewed linguistic outputs and to prevent the "unreliable answers" typical of unconstrained models [2-4].

Furthermore, this study conceptualizes complexity paralysis as a manifestation of extraneous cognitive load within the framework of Cognitive Load Theory. However, recent scholarship increasingly warns about the risk of epistemic outsourcing, a phenomenon in which learners rely on the generative capacity of AI rather than internalizing underlying linguistic structures, potentially leading to dependency [5-8]. To address this issue, the intervention employs AI to deconstruct Chinese characters into semantic narratives. This approach aims to promote germane cognitive load. The mental effort required to construct durable cognitive schemas while simultaneously mitigating the risk of cognitive offloading, whereby individuals delegate memory processes to digital devices, potentially undermining long-term retention [9-11].

Historically, character acquisition in TCSL has relied on rote memorization, an approach that often fails to translate orthographic knowledge into functional communicative competence [12-13]. While digital tools can produce a "leveling effect," a critical concern remains regarding the Digital-Analog Gap. This study operationalizes this gap as the performance divergence between digital character input and autonomous handwriting recall, situating the research within debates on the "haptic-motor" loss in digital learning [14-15]. The Communicative Bridge is thus defined through the lens of Willingness to Communicate (WTC) as a measurable reduction in the discrepancy between intent and structural accuracy [16].

Methods and Materials

The study employed a quasi-experimental, mixed-methods design to evaluate the impact of Pedagogical Scaffolding through teacher-curated AI on Chinese character acquisition. Conducted during the 2025-2026 academic year at the Kazakh Ablai Khan University of International Relations and World Languages (KazUIR&WL), the eight-week intervention targeted the "Digital-Analog Gap"-defined here as the cognitive disconnect between digital character recognition and autonomous orthographic recall. The purposive sample consisted of 29 first-year students ($N = 29$) with pre-intermediate proficiency, divided into an experimental group ($n = 15$) and a control group ($n = 14$). To control for extraneous variables, including prior language exposure and potential teacher effects, all participants were verified to have received an equivalent duration of prior Chinese instruction (four months). In addition, the same instructor taught both groups to ensure pedagogical consistency Baseline equivalence between the two groups was established through a pre-test essay assessment, which indicated no statistically significant differences in orthographic proficiency ($p = .812$) or initial motivation levels ($M = 4.2/5$, $p = .92$). Ethical approval for the study was granted by the Institutional Review Board of Kazakh Ablai Khan University of International Relations and World Languages (Protocol No. 3, 28 January 2025). All participants provided written informed consent for the analysis of diagnostic work and encrypted, anonymized digital chat logs in accordance with established data privacy protocols.

To identify and interpret symbolic representations of Chinese characters, the intervention utilized a specialized Custom GPT interface built upon the GPT-4o (Omni) multimodal architecture. This model employs Vision-Transformer (ViT) layers and Multi-Modal Tokenization, allowing it to process both UTF-8 encoded logographs and student-uploaded images as discrete semantic units. The recognition of handwritten inputs is facilitated by the model's native Computer Vision (CV) capabilities, while symbolic interpretation is governed by a Socratic Scaffolding Model. This model was operationalized through system-level 'Prime Prompts' that constrained the AI to apply the Liushu (六书) traditional classification system, specifically deconstructing characters through xiangxing (象形) and huiyi (会意) logic. As illustrated in Fig. 1, the AI

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 26, JUNE 2026

generated radical-based "semantic narratives" in the participants' L1 (Russian/Kazakh), facilitating systematic morphemic deconstruction and reducing extraneous cognitive load.

To ensure instructional fidelity, both groups followed a homogeneous curriculum based on the Road to Success (成功之路) series. To isolate the treatment effect from a simple practice effect, time-on-task was strictly controlled at 15 minutes per session. The Control Group engaged in traditional manual radical tracing, while the Experimental Group utilized the curated AI interface for "Free Chat" sessions [17]. The intervention was designed to shift "Complexity Paralysis" (the cognitive overwhelm caused by high-stroke density) from extraneous to germane cognitive load. Students used the AI to verify these semantic stories before finalizing their handwritten work (*Fig. 2*). This approach, supported by recent studies on diagnostic error corpora [18], addressed the "Communicative Bridge" construct by linking graphemic-semantic logic directly to active manual production.

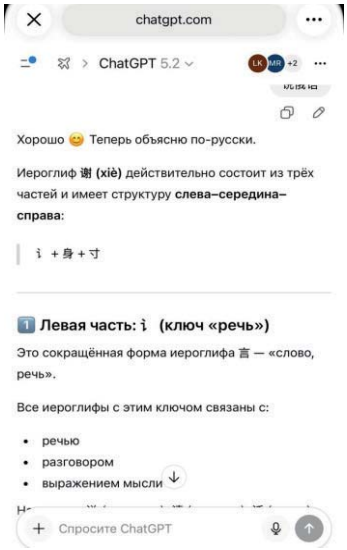


Figure 1. Structural and radical deconstruction

Figure 2. Radical story for *cí* (慈), handwriting input interface

Data collection followed a three-stage protocol. The primary outcome variable, character accuracy, was operationalized through a Weighted Orthographic Accuracy Model (WOAM). A 50-character sample was extracted from identical pre-test and post-test handwritten essays. Accuracy (A) was quantified using the following formal mathematical model:

$$A = \left(\frac{\sum s_i}{n}\right) \times 100 \quad (1)$$

Where s_i denotes the weighted recognition score for each character i, determined by a tripartite qualitative-to-quantitative scale: 1.0 (full mastery), 0.5 (minor stroke/spatial error), and 0.0 (radical substitution/omission). n represents the total number of characters in the evaluation sample (n = 50). To eliminate expectancy bias, essays were anonymized and graded by two independent instructors blinded to the study conditions. Inter-rater reliability was calculated at Cohen's $\kappa = 0.86$, indicating high agreement.

Quantitative analysis utilized a 2×2 mixed-design ANOVA in SPSS to test the interaction effect, with assumptions verified (Shapiro-Wilk $p>.05$; Levene's $p=.34$). Effect sizes were reported using Cohen's d and partial eta-squared η_p^2. Qualitative data from logs and interviews (n=15) were

DOI: 10.37943/LHFX3759

analyzed via thematic analysis. To mitigate confirmation bias, researchers conducted "negative case analysis," specifically coding for "epistemic outsourcing" (dependency risks) where AI mediation may have bypassed active cognitive retrieval.

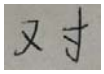
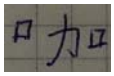
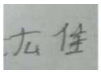
Results

Pre-Intervention Diagnostic Results (Initial Error Corpus)

To establish a pedagogical baseline, a diagnostic error corpus comprising over 120 student assignments was analyzed in early 2025. This comprehensive corpus provided the categorical framework employed to evaluate the experimental (n = 15) and control (n = 14) groups. The analysis revealed that students struggled primarily with the orthographic-semantic logic of Chinese characters rather than simple rote recall.

The frequency of baseline errors was categorized into four taxonomies, which informed the scoring rubric for the subsequent intervention:

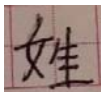
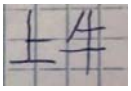
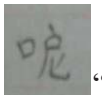
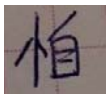
1. Structural Fragmentation (35% frequency).

 “对”  “咖”  “雄”  “斑”  “舍”  “最”

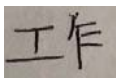
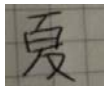
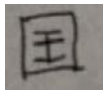
The most prevalent error observed was the spatial fragmentation of character components. Students exhibited a limited gestalt perception, often writing left–right or top-bottom radicals with excessive spacing. For example, components that should form a single cohesive unit were rendered as separate elements (e.g., splitting “**斑**” into three distinct parts). This pattern suggests cross-linguistic interference from alphabetic, linear writing habits, in which characters are perceived as a sequence of individual shapes rather than as an integrated block-frame.

2. Stroke Redundancy and Omission (28% frequency)

Stroke Redundancy (Additions)

 “姓”  “买”  “上午”  “呢”  “怕”

Stroke Omission (Deletions)

 “工作”  “然”  “说”  “夏”  “国”

During the initial stages of Chinese language acquisition, one of the most common errors among students is the incorrect addition or omission of strokes during character production. This issue primarily stems from the high density of orthographically similar characters (near-homoglyphs) within the Chinese writing system. Novice learners often lack the visual-orthographic sensitivity necessary to recognize that altering a single stroke can transform a character into a completely different lexical unit, with a distinct pronunciation and meaning. For example, the subtle distinctions between “**午**” (noon) and “**牛**” (cow), or between “**王**” (king) and “**玉**” (jade), illustrate how a single stroke can serve as a critical phonemic and semantic differentiator.

3. Graphemic-Semantic Substitution (Component Displacement) (22% frequency)

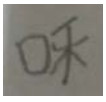
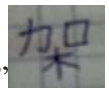
 “题”  “悔”  “物”  “楼”  “觉”  “顺”

This category of error involves students replacing a correct grapheme or radical with a similar-looking but semantically different component. This stems from a lack of morphemic awareness regarding radical properties. Students often perceive radicals as purely visual shapes rather than meaningful building blocks. A prime example observed in the corpus is the character “悔” (huǐ – "to regret"). Students frequently substitute the "heart" radical (“忄”), which signifies emotion, with the "wood" radical (“木”), resulting in the character “梅” (méi – "plum"). This substitution completely alters the lexical meaning and the phonetic value of the character. To

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 26, JUNE 2026

mitigate such errors, pedagogical intervention must emphasize the semantic differentiation of frequently confused radicals. Instructors should provide comparative analyses of graphemes with high visual similarity but distinct meanings, such as: “彳” (step) vs. “亻” (person); “广” (shelter) vs. “疒” (sickness); “氵” (water) vs. “冫” (ice); “寸” (inch) vs. “扌” (hand);“又” (again) vs. “攵” (tap/action).

4. Structural Conuration: Positional and Proportional Errors (15% frequency).

 “男”  “路”  “太”  “部”  “和”  “架”

The final category of identified errors involved the spatial configuration and positional order of character components. Student work revealed significant inconsistencies in the relative size, placement, and internal proportions of radicals, disrupting character symmetry and visual balance. A prevalent error was the lateral inversion of components, such as reversing left-right positions in “部” and “和,” indicating a failure to internalize the "fixed-position" nature of Chinese graphemes. To mitigate these errors, instructors must provide comparative structural analyses, guiding students to adhere to standardized "writing architecture" and hierarchical component relationships. Diagnostic data confirmed no statistically significant difference in error profiles between cohorts at the study's onset. Both the Experimental Group (n = 15) and Control Group (n = 14) shared a reliance on linear, alphabetic habits and lacked semantic awareness of radical functions. This systemic "Communicative Gap" the inability to translate recognition into accurate production established a balanced baseline for the eight-week intervention.

Intervention Outcomes: Quantitative Analysis

Following the eight-week intervention, identical handwritten essay assignments (200-character tasks) were administered as both the Pre-test and Post-test to ensure task consistency. A standardized 50-character sample was extracted from each essay to evaluate the "Digital-Analog Gap". To determine the efficacy of the teacher-curated AI scaffolding, a 2 × 2 mixed-design ANOVA was conducted. The primary outcome variable, character accuracy, was calculated using the Weighted Orthographic Accuracy Model (WOAM).

The ANOVA revealed a significant main effect for Time, F (1,27) = 42.15, p <.001, η_p^2=.61, indicating that both groups improved over the eight weeks. More importantly, a significant Time × Group interaction effect was observed, F(1,27) = 15.82, p <.001, η_p^2= .37. This interaction confirms that the improvement in character accuracy was significantly greater for the Experimental Group, supported by the GPT-4o Socratic scaffold, than for the Control Group. These results are detailed in Table 1.

Table 1. Descriptive Statistics for Character accuracy

Period	Group	*N*	Mean (*M*)	*SD*	*t* (27)	*p*	Cohen’s *d*
Pre_test	Experimental	15	68.50	7.20	-0.24	.812	0.08
	Control	14	69.10	6.80			
Post_test	Experimental	15	84.20	5.40	4.68	<.001	1.69
	Control	14	71.80	8.90			

Qualitative Analysis of AI-Student Interaction Logs

To elucidate the cognitive mechanisms underlying the Experimental Group’s quantitative gains, a thematic analysis was conducted on the digital transcripts from the "Free Chat" sessions. Following Braun and Clarke’s (2006) six-phase framework, two researchers independently coded the logs using an inductive-deductive approach. An initial inter-coder agreement was calculated at κ = 0.88, with discrepancies resolved through iterative consensus meetings. This analysis aimed to identify how learners leveraged real-time feedback from the GPT-4o scaffold to mitigate the

DOI: 10.37943/LHFX3759

baseline errors identified in the pre-intervention corpus. The distribution of these interactions is summarized in Table 2.

Table 2. Frequency of AI-Assisted Corrections in Chat Logs

Interaction Theme	Frequency	Primary Error Addressed
Immediate Orthographic Correction	25%	Structural Fragmentation / Stroke Accuracy
Semantic/Radical Clarification	23%	Graphemic-Semantic Substitution
Cognitive Load Management	18%	Complexity Paralysis / Stroke Omission
Structural Architecture Rules	17%	Directional & Mirror Errors
Contextual/Grammar Refinement	17%	Contextual Misuse

Orthographic Reinforcement and Spatial Discipline (25%): As shown in Table 2, the most frequently observed theme involved real-time spatial feedback. The AI functioned as a digital "grid," reinforcing the cohesive block-frame structure of Chinese logographs. For fragmented characters, such as the three-part bān (斑) or the vertically integrated bí (鼻), the AI provided immediate structural deconstruction. This repeated reinforcement corresponded directly with the significant reduction in structural fragmentation observed in the post-test essays.

Morphemic Scaffolding and Semantic Logic (23%): A primary affordance of the teacher-curated AI was its ability to operationalize the Liushu (六书) system. Rather than providing rote definitions, the GPT-4o model interpreted the radical logic behind character components. For instance, students were prompted to distinguish the "eye" radical (目) in jīng (睛) from the "sun" radical (日) in qíng (晴), shifting the acquisition process from rote memorization to logical deduction within the learner's Zone of Proximal Development (ZPD).

Cognitive Load Management through Narrative Deconstruction (18%): The GPT-4o interface mitigated "Complexity Paralysis" by interpreting and deconstructing high-stroke characters into "semantic stories." For complex characters such as xiè (谢) or fán (繁), the AI explained the relationships between components. By transforming a daunting visual task into a predictable narrative, the AI lowered the intrinsic cognitive load, enabling students to attempt characters they had previously avoided due to high stroke density.

Fixed-Position Logic and Structural Architecture (17%): Logs demonstrated that the AI enforced standardized writing architecture by targeting mirror errors. For asymmetrical radicals such as fù (阝), the AI clarified that placement on the left (e.g., duì 队) relates to terrain, while the right (e.g., bù 部) relates to administration. This precision assisted students in internalizing the hierarchical relationship between components rather than treating radicals as isolated shapes.

Transition to Functional Communicative Competence (17%): The "Free Chat" environment appeared to lower the learners' affective filter, facilitating an increase in Willingness to Communicate (WTC). Transcripts showed a marked increase in the usage of abstract characters-such as dǒng (懂) and shūfu (舒服)-within original dialogue. This suggests a transition from mechanical reproduction to functional communication.

Negative Case Analysis: Risks of Epistemic Outsourcing

To mitigate confirmation bias, researchers analyzed instances in which the intervention failed to elicit cognitive engagement. In approximately 8% of the logs, signs of epistemic

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 26, JUNE 2026

outsourcing were observed, with students relying on the GPT-4o multimodal output for immediate translation rather than engaging with the Liushu-based deconstruction logic. One participant remarked, "It was easier to let the AI explain the parts than to try to remember the radical myself." These findings suggest that without a deliberate teacher-led "fading" of the scaffold, there is a risk of digital dependency.

Qualitative Perspectives: Student Perceptions of AI Integration

To capture the subjective experience of the "communicative bridge," semi-structured interviews were conducted with the 15 participants in the experimental group. These interviews aimed to move beyond numerical measures of accuracy and to explore the cognitive and emotional shifts facilitated by the teacher-curated AI. The resulting themes were analyzed to identify the tool's affordances and limitations, as summarized in Table 3.

Table 3. Thematic Analysis of Student Interview Results

Theme	**Student Perspective Summary / Illustrative Quote**	**Pedagogical Implication**
I. Affordances		
Bilingual Scaffolding	When I didn't understand the complex radical logic in Chinese, the AI explained the 'story' in Russian first, which made the internal logic clear before I tried to use it.	Russian-language support lowers the affective filter and ensures conceptual accuracy in radical deconstruction.
Cognitive Load Reduction	Before, characters like cān (餐) felt like a mess of 16 lines. The AI broke it into 'evening,' 'food,' and 'hand.' It turned a 'drawing' into a 'story' I could remember.	AI facilitates the "chunking" of complex graphemes into manageable semantic units.
Low-Anxiety Environment	I felt safer making mistakes with the AI, words like 舒服 without fear of embarrassment	Higher frequency of "Target Character Usage" in chat logs.
Real-Time Spatial Correction	The AI is like a digital 'grid.' It caught my 'drifting' radicals immediately, before I practiced it wrong	Eliminates "fossilized" structural fragmentation errors.
II. Challenges		
Immediate Feedback Loop	In class, I wait a week for my homework to be graded. By then, I've forgotten why I made the mistake. The AI tells me why my radical is wrong the second I type it.	Prevents the "fossilization" of incorrect writing habits through real-time intervention.
Technical Constraint	Sometimes the AI didn't recognize my stroke order if I uploaded a photo, only the final result. It corrected the 'look' but not the 'process'."	AI may prioritize orthographic output over the calligraphy process.
Correction Fatigue	Sometimes the AI is too 'perfectionist.' I was trying to tell a story about my weekend, but it kept stopping me to explain a	Constant pedagogical intervention can interrupt the "Communicative Bridge" and discourage fluency.

DOI: 10.37943/LHFX3759

	radical. It broke my flow of thought.	
Input Method Bias	The AI is great at fixing my typing errors, but it doesn't know that my handwriting is messy or that I'm writing the 'water' radical (氵) with four strokes instead of three."	AI primarily corrects the result, often ignoring the stroke-order process essential for character balance.
Digital Dependency	"I felt very smart while the chat was open, but during the paper-and-pen post-test, I felt a bit 'naked.'	Students may become reliant on the "Semantic Decoding" prompts.
Over-Standardization	I tried to use a more creative way to say 'I'm tired,' but the AI kept pushing me back to the textbook phrases from Road to Success.	Pre-conditioned semantic logic can limit linguistic "play" and natural idiomatic expression.

The feedback presented in Table 3 highlights a critical duality in the use of teacher-curated AI. On one hand, the "advantages" emphasize a fundamental shift in orthographic processing. By transforming characters from visual puzzles into "semantic stories," the AI effectively reduced the intrinsic cognitive load associated with complex characters. Additionally, the real-time spatial correction functioned as a preventative measure, ensuring that the "fixed-position" nature of radicals was mastered before errors could become habitual.

However, the challenges reported by students-particularly the sensation of being "naked" during the post-test-underscore the digital–analog gap. This finding suggests that, while AI strengthens character recognition and structural logic, it does not inherently guarantee autonomous recall. Additionally, the phenomenon of "correction fatigue" indicates that, for AI to function as an effective scaffold, its curation must balance orthographic rigor with communicative flow. Collectively, these insights suggest that the intervention should be considered a cognitive scaffolding tool rather than a complete replacement for traditional pedagogical methods.

Discussion

The findings of this study indicate that the teacher-curated GPT-4o interface functions as a significant cognitive mediator in the second-language (L2) Chinese classroom. The statistically significant interaction effect revealed by the 2×2 mixed-design ANOVA (F (1, 27) = 15.82, p < .001, η_p^2= .37) demonstrates that the transition from a "communicative gap" to a "communicative bridge" is more than a metaphorical construct; it reflects a measurable alteration in cognitive processing. While the control group remained reliant on rote memorization, the experimental group successfully adopted a logic-deduction system. This shift can be interpreted through the lens of Cognitive Load Theory, wherein the GPT-4o scaffold actively mitigated "complexity paralysis" by transforming the high stroke density of Chinese logographs from an extraneous to a germane cognitive load. By deconstructing complex characters into manageable "semantic stories," the AI functioned as a pedagogical scaffold within the learners' Zone of Proximal Development. This deconstruction allowed students to bypass the cognitive overwhelm typical of early-stage character acquisition, thereby lowering their affective filters and facilitating an increase in Willingness to Communicate through the use of more abstract lexical units.

Furthermore, the Socratic tutoring paradigm implemented through system-level "prime prompts" effectively circumvented the limitations of general-purpose AI by preventing passive, direct translation. This approach compelled students to participate in a "negotiation of meaning," wherein the GPT-4o's Liushu-based (象形 and 会意) feedback reinforced the Liushu system as an underlying cognitive framework. By operationalizing semantic logic rather than rote recall, the

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 26, JUNE 2026

intervention addressed the "fixed-position" nature of radicals and corrected "fossilized" structural fragmentation before it could solidify. However, the qualitative emergence of the "Digital-Analog Gap" vividly described by students who felt "naked" during paper-based post-tests identifies a critical integration borderline. These findings suggest that, while digital scaffolding effectively reinforces character recognition and structural logic, it does not inherently ensure the haptic–motor memory necessary for autonomous handwriting- an observation consistent with recent research on the neurological divergence between typing and writing [19-20]. The observed 8% frequency of "epistemic outsourcing" reinforces this concern, suggesting that without a deliberate teacher-led "fading" of the scaffold [20], a risk of digital dependency exists. Consequently, curated AI should be positioned as an augmentative cognitive tool rather than a replacement for traditional orthographic practice. While the large effect size (Cohen's $d = 1.69$) provides promising evidence for this exploratory framework, the small sample size and single-institution scope suggest that these findings serve as a preliminary proof-of-concept rather than a final national blueprint. Future research must incorporate delayed post-tests to evaluate the long-term sustainability of this transition from digital scaffolding to autonomous recall.

Conclusion

The integration of teacher-curated artificial intelligence at Kazakh Ablai Khan University of International Relations and World Languages (2025-2026) presents a compelling preliminary framework for transforming Chinese logographic pedagogy. This study moves beyond the traditional binary of "rote memorization versus digital translation" by demonstrating that a Socratic GPT-4o scaffold can effectively foster Liushu-based morphemic awareness and Willingness to Communicate. By transferring the cognitive burden of complex character deconstruction from the student to a teacher-governed digital agent, the research provides a viable pathway for mitigating "complexity paralysis" among early-stage learners.

However, the identified "Digital-Analog Gap" highlights a fundamental pedagogical boundary. The study concludes that, while the GPT-4o model's interpretation of orthographic logic enhances recognition, it does not yet substitute for the haptic–motor memory required for autonomous handwriting. This finding underscores the need for "hybrid" pedagogical models that balance digital scaffolding with traditional calligraphy practice. Moreover, the observed risks of "epistemic outsourcing" serve as a critical reminder that AI integration must remain human-centric and teacher-guided to prevent cognitive dependency.

While the small sample size and single-institution scope (N =29) mean these results serve as a proof-of-concept rather than a final national blueprint, the significant effect sizes ($d = 1.69$) underscore the potential for this model to support Kazakhstan's digital modernization goals. Future research should utilize delayed post-tests to measure long-term retention and explore the use of haptic stylus interfaces to bridge the gap between digital recognition and manual production. Ultimately, this study positions curated AI as an augmentative bridge not a replacement for the cognitive processes underlying true linguistic sovereignty.

Limitations

Despite the significant statistical effects observed, several limitations must be acknowledged. First, the empirical evaluation was conducted within a single institutional context Kazakh Ablai Khan University of International Relations and World Languages and involved a relatively small sample size ($N = 29$). Consequently, the findings may not fully capture the diverse demographic and pedagogical variables present across all higher education institutions in Kazakhstan. Second, the eight-week duration of the intervention restricts the ability to evaluate long-term retention of Chinese characters or the sustainability of the "communicative bridge" without active digital scaffolding. Future research should implement longitudinal designs, delayed post-tests, and larger multi-institutional cohorts to validate the generalizability of these preliminary results.

DOI: 10.37943/LHFX3759

References

[1] Tokayev K.-J. (2024). Poslanie Glavy gosudarstva Kasym-Zhomarta Tokaeva narodu Kazahstana «Kazahstan v epohu iskusstvennogo intellekta: aktual'nye zadachi i ih resheniya cherez cifrovuyu transformaciyu» [Address of the Head of State Kassym-Jomart Tokayev to the people of Kazakhstan "Kazakhstan in the era of artificial intelligence: Urgent tasks and their solutions through digital transformation"]. Official Site of the President of the Republic of Kazakhstan (Akorda). URL: https://www.akorda.kz/ru/poslanie-glavy-gosudarstva-kasym-zhomarta-tokaeva-narodu-kazahstana-kazahstan-v-epohu-iskusstvennogo-intellekta-aktualnye-zadachi-i-ih-resheniya-cherez-cifrovuyu-transformaciyu-885145

[2] Panagiotidis P. (2024). LLM-based chatbots in language learning, *European Journal of Education*, vol. 7, no. 1, pp. 102-123. https://doi.org/10.26417/919pei96z. Available at: https://revistia.com/ejed/article/view/3011

[3] Amaro I., Della Greca A., Francese R., Tortora G., & Tucci C. (2023). AI unreliable answers: A case study on ChatGPT, *Artificial Intelligence in HCI,* pp. 23–40. https://doi.org/10.1007/978-3-031-35894-4_2

[4] Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, *32*, 1052–1092. https://doi.org/10.1007/s40593-021-00285-9

[5] Zhang Z. & Huang X. (2024). The impact of chatbots based on large language models on second language vocabulary acquisition, *Heliyon*, vol. 10, no. 3, e25370, https://doi.org/10.1016/j.heliyon.2024.e25370

[6] Franjic M. & Bozinovski B. (2025). Exploring ChatGPT's impact on ESL writing proficiency: An experimental approach, *Teaching English with Technology*, vol. 25, no. 2, pp. 53–81. https://doi.org/10.56297/vaca6841/ZDDD4208/LUZG7476

[7] Bearman M. & Ajjawi R. (2023). Learning to work with the black box: Pedagogy for a world with generative AI, *British Journal of Educational Technology*, vol. 54, no. 5, pp. 1160–1173. https://doi.org/10.1111/bjet.13337

[8] Baidoo-Anu D. & Ansah L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning, *Journal of AI*, vol. 7, no. 1, pp. 52–62. https://doi.org/10.2139/ssrn.4337484

[9] Sila C., William C., Yunus M., & Rafiq, K. (2023). Exploring students' perception of using ChatGPT in higher education, *International Journal of Academic Research in Business and Social Sciences*, vol. 13, no. 12, pp. 4044–4054. https://doi.org/10.6007/IJARBSS/v13-i12/20250

[10] Kereibayeva A., Sharshova R., Dyussenova D., Aliyeva D., & Toktarova T. (2025). The effectiveness of VR glasses as a tool for immersive language education: Insights from questionnaires and interviews, *Procedia Computer Science*, vol. 272, pp. 528–533. https://doi.org/10.1016/j.procs.2025.10.242

[11] Risko E. F. & Gilbert S. J. (2016). Cognitive offloading, *Trends in Cognitive Sciences*, vol. 20, no. 9, pp. 676–688. https://doi.org/10.1016/j.tics.2016.07.002

[12] Chen C., & Gong Y. (2025). The role of AI-assisted learning in academic writing: A mixed-methods study on Chinese as a second language students, *Education Sciences*, vol. 15, no. 2, pp. 141. https://doi.org/10.3390/educsci15020141

[13] Cong Y. (2024). AI language models: An opportunity to enhance language learning, *Informatics,* vol. 11, no. 3, pp.49. https://doi.org/10.3390/informatics11030049

[14] Behforouz B., Al-Gaithi A., & Al Balushi Z. (2025). Revolutionizing Vocabulary Acquisition through AI Chatbots, International Journal of Learning, *Teaching and Educational Research*, vol. 24, no. 7, pp. 1-22. https://doi.org/10.26803/ijlter.24.7.1

[15] Kohnke L., Moorhouse B. L., & Zou D. (2023). ChatGPT for language teaching and learning", *RELC Journal*, vol. 54, no. 2, pp. 537–550. https://doi.org/10.1177/00336882231162868

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 26, JUNE 2026

[16] Baskara F. X. R. & Mukarto F. X. (2023). Exploring the implications of ChatGPT for language learning in higher education, *Indonesian Journal of English Language Teaching and Applied Linguistics*, vol. 7, no. 2, pp. 343–358. https://doi.org/10.21093/ijeltal.v7i2.1447

[17] Guo J., Lyu B., Zhang Z., & Qi X. (2025). The impact of error diagnosis in handwriting practice on learners' Chinese character production and orthographic awareness, *Computer Assisted Language Learning*, vol. 28, no. 1, pp. 1–31. https://doi.org/10.1080/09588221.2025.2505917

[18] Wiley R. W. & Rapp B. (2021). The effects of handwriting experience on literacy learning, *Psychological Science*, vol. 32, no. 7, pp. 1086–1103. https://doi.org/10.1177/0956797621993111

[19] Park S. & Kim O. (2025). Scaffolding with generative AI in language classrooms: A critical thinking framework, *Studies in Linguistics*, vol. 77, pp. 351–374. https://doi.org/10.17002/sil..77.202510.351

[20] Lee H. Y., Chen P. H., Wang W. S., Huang Y. M., & Wu T. T. (2024). Empowering ChatGPT with guidance mechanism in blended learning: Effect of self-regulated learning, higher-order thinking skills, and knowledge construction, International Journal of Educational Technology in Higher Education, vol. 21, no. 16. https://doi.org/10.1186/s41239-024-00447-4