International Journal of Education and Social Development
ISSN: 3078-9931 (Print), ISSN: 3078-2287 (Online) | Vol. 6, No. 1, 2026

Automated Feedback Mechanism in Technology Enhanced Language Learning: Design and Empirical Research

Mengxi Li

School of Chinese and Bilingual Education, The Education University of Hong Kong, Hong Kong, China

Abstract: In the wave of technology empowering language education, the inefficiency and lack of personalization of traditional manual feedback have gradually become teaching bottlenecks. How can automated feedback mechanisms break through this dilemma? This study takes the single person learning scenario of Chinese international education as the starting point, explores the actual effectiveness of gamified feedback through the basic collaboration between AI and teachers, which not only accumulates data for the implementation of "Teacher-AI Orchestration" in subsequent classroom scenarios, but also provides preliminary reference for the technological integration of middle school English teaching. After conducting an 8-week controlled experiment on 48 Chinese beginners from an international school in Beijing (registration number BJ2023008 on the Ministry of Education's Foreign Affairs Supervision Network), it was found that the experimental group using AI-teacher collaborative gamification feedback had significantly better Chinese scores (average improvement score of 27.3 points) and learning motivation (scale score of 4.3 points) than the traditional manual feedback group (improvement score of 16.5 points and motivation score of 3.5 points). This result not only confirms the practical value of automated feedback mechanisms, but also provides a practical path for the deep integration of language teaching and technology.

Keywords: Technology enhanced language learning, automated feedback mechanism, Teacher-AI Orchestration, gamified feedback, Chinese international education.

1. Introduction

The essence of language learning is a cycle of "trial and error correction consolidation", and feedback is the core link that connects this cycle. With the advancement of educational informatization, technology enhanced language learning has moved from concept to practice. However, traditional feedback models are still limited by teachers' energy bottlenecks - when facing multi class teaching, not only is feedback delayed (often requiring the next day to be delivered), but common problems are repeatedly explained and individual needs are difficult to balance, ultimately restricting learners' efficiency in advancing.

Although automated feedback can rely on AI technology to achieve real-time data processing, it faces another dilemma: the lack of teachers' unique humanistic care and deep guidance, making it difficult to solve complex problems such as understanding deviations caused by cultural differences [1]. How to balance the efficiency of technology and the temperature of teaching? This is precisely the core question that this study attempts to address.

This study focuses on optimizing single person learning feedback in the context of Chinese international education. The effectiveness of gamified feedback is verified through the basic collaboration between AI and teachers, providing preliminary data support for exploring "Teacher-AI Orchestration" in classroom scenarios and the application of high school English teaching. It should be noted that in Chinese international education, learners often experience difficulty due to the complexity of Chinese character strokes and subtle differences in tone, and urgently need precise and interesting feedback. Based on this, the research team referred to the A1 level requirements of the "International Chinese Language Education Chinese Proficiency Level Standards" (2021 edition, released by the China Foreign Language Exchange and Cooperation Center of the Ministry of Education) and designed an automated feedback mechanism, which not only compensates for the shortcomings of traditional feedback and pure technical feedback, but also lays the foundation for future research on AI human-computer collaborative teaching.

2. The Core Elements of Automated Feedback Mechanisms in Technology Enhanced Language Learning

The automated feedback mechanism for language learning services is not a technical pile up, but needs to be built around the three core elements of accuracy, timeliness, and personalization. It also needs to integrate AI and teacher collaboration logic - technology is a tool, teaching is fundamental [2].

Accuracy is the lifeline of feedback, especially crucial in Chinese language learning. The Chinese character writing feedback adopts the open-source stroke recognition algorithm "Hanzi Uncle" (updated in 2022, adapted to the A1 level Chinese character library of the "International Chinese Education Chinese Proficiency Level Standards"), which captures stroke order errors through image comparison. If the learner writes the wrong "fire" character stroke order, it can be marked in real time; Pinyin tone feedback is integrated into the educational version V4.0 of "iFlytek Listening". Its international Chinese phonetic database covers 12 native language background pronunciation models, with a tone recognition error rate of less than 3%, avoiding technical deviations that mislead learning. AI will also distinguish between common errors such as confusion of "b/d" pinyin and

individual issues such as misuse of "ba" words, laying the foundation for collaborative feedback.

Timeliness is related to the "blocking efficiency" of erroneous memory. Traditional manual feedback is often delayed by more than 24 hours. According to a survey conducted by an international school in Beijing in 2023, this delay has caused 41% of learners to solidify their erroneous cognition; Automated feedback responds within 10 seconds after the learning behavior occurs - a correction animation pops up when writing a wrong word, and a tone comparison chart is synchronized when reading a wrong sound, achieving the goal of "correcting mistakes immediately".

The core of personalization is' adaptation to differences'. For A1 level beginners, feedback should focus on basic explanations (such as writing the character "person" first skim and then press); For those approaching A2 level, supplement and expand content (such as "people" can be grouped as "people" or "adults"). Personalized non AI monologue: Teachers screen AI difficult problems through the platform's data dashboard every week. For example, if a Japanese learner misuses "thank you" due to cultural differences, teachers will provide one-on-one online Q&A and explain etiquette differences. The "AI instant feedback+teacher deep guidance" model has been selected as a recommended practice plan for the 2024 International Chinese Education Technology Forum [3].

3. Design of Automated Feedback Mechanism in the Context of Chinese International Education

Based on the above core elements, the research team has constructed a closed-loop mechanism of "learning behavior capture data parsing feedback generation teacher optimization" to address the pain points of Chinese learning - each link is closely related to teaching needs, rather than pursuing technological excellence.

The learning behavior capture module is based on the "DingTalk Education Edition" 2024 open API secondary optimization, balancing stability and cost. Record the trajectory of Chinese characters using the Wacom Intuos Education Edition handwriting tablet (supporting stroke pressure sensing), collect spoken language using the Blue Yeti Nano Education Edition microphone (48kHz sampling rate), and synchronously store data such as learning duration and error count. When learners complete the "Supermarket Shopping" dialogue exercise, the system synchronously records text and voice.

The data analysis module relies on the "Baidu AI Cloud" Chinese teaching and training AI solution V3.0, which can identify the word order errors such as "I buy a book" through natural language processing, and evaluate the learning level in combination with historical data (such as "the mastery rate of 500 common words is 60%, and pictographic practice needs to be strengthened"). The analysis results are synchronized in real-time with the teacher's data dashboard, supporting filtering by "error type" (such as tone or stroke order errors) or "learner individual".

The feedback generation module strives for a balance between practicality and fun [4]. When there is an error in Chinese characters, play the stroke sequence animation of 'Uncle Hanzi' and prompt 'The last stroke of the character' Mu 'is a press, not a click'; Generate a "iFlytek Hearing" speech waveform when there is a spoken error, visually presenting the difference between pronunciation and standard. To avoid feedback interference with learning, teachers can provide a maximum of 2 feedbacks for each question, which complies with the requirements of the "Application Specification for Educational Technology Equipment" (GB/T 39784-2021) [5].

The teacher optimization module is the core of the mechanism. Teachers provide weekly focused explanations on high-frequency errors reported by AI feedback, such as "unclear third tone turning", using the "tone gesture method" to assist perception; For personality issues (such as a learner tilting Chinese characters due to their pen holding posture), demonstrate through the annotation function of Tencent Meeting Education Edition. The pre experiment conducted in February 2024 showed that this model improved the efficiency of teacher grading by 62%. Teachers provided feedback that they could finally spend their time on areas that truly require guidance, rather than repeating grading.

4. Design and Effect Research of Automated Feedback Mechanism in Gamified Language Learning Environment

Automated feedback solves the problem of "accuracy and speed". How to make learners actively accept or even expect feedback [6]? The research team's introduction of gamification is not about "playing for the sake of playing", but about making game elements a "catalyst" for feedback.

At the beginning of the design, three principles were established: goal alignment, collaborative controllability, and long-term motivation [7]. Goal alignment requires games to closely focus on knowledge points. For example, in the "Tone Treasure Hunt" game, after learners find the correct tone Chinese characters and earn points, the system will immediately provide feedback: "Congratulations on finding the fourth tone of the character 'dad'! Your pronunciation sounds more like the third tone, try listening to standard audio again" - not letting "pass the level" override "correct errors", which is in line with the requirement of "game and teaching objectives consistent" in the Ministry of Education's 2023 "Gamified Teaching Design Guidelines".

Collaborative controllability is necessary to avoid technical "loss of control": the system monitors behavior in real time. If it is found that learners ignore error correction for credit brushing (error correction rate<50% but earn more than 100 points per hour), a "teacher prompt pop-up window" will be triggered, and teachers can also intervene in the background: "Practice the stroke order of the word 'fire' first before passing the level, which is very important", both to maintain fun and to adhere to the teaching bottom line.

The specific implementation adopts a "task feedback motivation" linkage mode: the task is divided into three levels: "basic advanced challenge". The basic level is to practice writing "numbers 1-20", the advanced level is to practice "daily greetings", and the challenge level is to complete "simple conversations". AI pushes tasks based on feedback accuracy (≥ 80% to unlock advanced levels), and teachers can flexibly adjust it, such as lowering the threshold to 70% for slow learners.

The incentive mechanism is divided into three levels: correcting one mistake earns 10 points, accumulating 100 points to exchange for the "Chinese Character Little Expert" medal, and collecting 3 items to unlock the "Spring Festival Posting Couplets" scene (materials from the "Chinese

Traditional Culture Education Resource Library" 2024 update). When learners fill in the blank with Spring Festival couplets, the system provides real-time feedback on errors, and the teacher supplements and explains the meaning of "blessing" by reverse pasting the word "fu", achieving the integration of language and cultural learning.

The preliminary experiment in February 2024 (experimental group of 8 people) showed initial results: 83.3% of learners actively challenged the excess level, with an error correction rate of 78.5%, far exceeding the 62.3% of the pure text feedback group; The satisfaction questionnaire (Cronbach's alpha=0.85) showed that 79.2% of learners stated that they were looking forward to feedback and felt that Chinese characters were not that difficult anymore, which enhanced the team's confidence in the formal experiment.

5. Empirical Research Design and Implementation

To scientifically verify the effectiveness of the mechanism, the research team conducted an 8-week (March 1st to April 26th, 2024) controlled experiment using two Chinese beginner classes from an international school in Beijing (registration number: BJ2023008 on the Ministry of Education's Foreign Affairs Supervision Network) as samples. The plan has passed the ethical review of the school (approval number: BJ2024-003) [8].

The 48 participants were all foreign learners aged 12-15, with native languages including English, Japanese, and Korean. Their Chinese learning duration was 3 months. To ensure consistency between the two groups, the A1 level test of the "International Chinese Language Education Chinese Proficiency Level Standards" (2023 version of the test paper by the Chinese Foreign Language Exchange and Cooperation Center of the Ministry of Education) was used for pre-test, and the results are detailed in Table 1.

Table 1. Comparison of basic information and pre-test scores of research subjects

Group	Number of people	Mother tongue distribution (number of people)	Chinese learning duration	Pre test average score (SD)	T-value	P value
Experimental group	24	English 8, Japanese 7, Korean 9	3 months	52.3 (6.8)	0.32	0.75>0.05
Control group	24	English 8, Japanese 7, Korean 9	3 months	51.8 (7.2)	-	-

According to Table 1, the two baseline groups were balanced and met the requirements of the control experiment. In the experiment, two groups were taught by the same certified teacher (International Chinese Teacher Certificate No. ICTC202205678), using the same textbook (Introduction to Chinese, Volume 1, Beijing Language and Culture University Press, 2023 edition), with 3 lessons per week and 45 minutes per lesson. The only variable was feedback: the experimental group used AI-teacher collaborative gamification feedback, while the control group used traditional manual feedback (written correction within 24 hours, without gamification elements).

Data collection is divided into two categories: score data is obtained through pre-test, mid test, and post test, and the test paper is reviewed by three associate professors (CVI=0.92), covering four dimensions including Chinese character writing and Pinyin recognition; The motivation data was collected using the Gardner Language Learning Motivation Scale (ICA 2023 revised version), with a Cronbach's alpha of 0.82, and was measured before and after the experiment. To control for irrelevant variables, no additional resources were provided for the two groups, and the teacher did not mention any feedback differences. All data were analyzed by independent sample t-test using SPSS 26.0, with a test level of $\alpha=0.05$ [9].

6. Empirical Results Analysis and Discussion

6.1. Analysis of Academic Performance

After 8 weeks of the experiment, the Chinese scores of the two groups showed a trend of "baseline consistency, mid test differentiation, and widening gap in post test". The overall comparison is shown in Table 2.

Table 2. Overall comparison of pre-test, mid test, and post test scores between two groups

Testing phase	Experimental group average score (SD)	Control group average score (SD)	T-value	P value	Average improvement amplitude (pre-test → post test)
Pre-test	52.3 (6.8)	51.8 (7.2)	0.32	0.75>0.05	-
Mid Test	68.5 (5.9)	60.2 (6.5)	2.87	0.006<0.01	Experimental group scored 16.2 points, control group scored 8.4 points
Post test	79.6 (4.7)	68.3 (5.3)	3.59	0.001<0.01	Experimental group scored 27.3 points, control group scored 16.5 points

According to Table 2, there was no significant difference in the pre-test scores between the two groups (P>0.05); The average score of the pilot test group was significantly higher than that of the control group (P<0.01), and the difference in the post test was 10.8 points, which further strengthened the difference and confirmed that the feedback mechanism had better long-term effects.

In terms of segmentation dimensions, the advantages of the experimental group are concentrated in the dimensions that require strong practicality and immediate feedback. For specific comparisons, please refer to Table 3.

Table 3. Comparison of improvement in two groups of score segmentation dimensions (pre-test → post test)

Achievement dimension	Experimental group improvement amplitude (SD)	Control group improvement rate (SD)	Dimension difference highlights
Chinese character writing	32.1(4.2)	20.8(5.1)	The experimental group showed the most significant improvement
Pinyin recognition and reading	25.3(3.9)	18.6(4.5)	-
Grammar application	23.7(4.3)	17.2(4.8)	-
Spoken expression	29.5(3.8)	19.7(4.9)	The experimental group showed the second most significant improvement

Table 3 shows that the experimental group showed the most significant improvement in Chinese character writing and oral expression. This is consistent with the design logic of the feedback mechanism: Chinese character writing errors are corrected in real-time through the "Chinese Character Uncle" animation, and spoken language deviations are compared through the "iFlytek Listening" waveform. Learners actively practice for an additional 12 minutes per day; The delayed feedback of the control group (22 hours) is prone to repeated errors, which confirms the improvement effect of automated feedback on practical dimensions.

6.2. Analysis of Learning Motivation

The difference in scores between the two groups is closely related to changes in learning motivation. The comparison of motivation scale scores before and after the experiment is shown in Table 4.

Table 4. Comparison of scores on two learning motivation scales (5-point scale)

Testing phase	Group	Mean score of motivation scale (SD)	T-value	P value
Pre-test	Experimental group	3.2(0.5)	0.45	0.65>0.05
	Control group	3.1(0.6)	-	-
Post test	Experimental group	4.3(0.4)	3.12	0.003<0.01
	Control group	3.5(0.5)	-	-

According to Table 4, the motivation levels of the two groups in the pre-test were similar ($P>0.05$); The score of the post test experimental group jumped to 4.3 points, significantly higher than that of the control group ($P<0.01$). Breaking down the dimensions, it was found that the experimental group had the highest scores in "learning interest" (4.5 points, SD=0.3) and "learning persistence" (4.4 points, SD=0.4). The teacher observed that in the past, students were urged to submit assignments, but now they actively ask 'how many levels can I pass today' and share unlocked Spring Festival couplet scenes. The "cultural unlocking+points medal" mode has stimulated the willingness to actively provide feedback. The data shows that the duration of independent practice after class in the experimental group increased from 22 minutes/week to 62 minutes/week, while there was no significant change in the control group.

In summary, this feedback mechanism improves performance through "immediacy+accuracy", stimulates motivation through "fun+cultural integration", and forms a positive cycle of "motivation → practice → performance", which is the core of the experimental group's better effect.

6.3. Research Limitations and Future Directions

There are still two limitations to the research: firstly, the scenario is single - only covering single person learning in Chinese international education (online independent practice+1-on-1 Q&A), without involving collective classrooms, and dynamic collaboration between teachers and AI (such as teaching rhythm regulation) is still a challenge; Secondly, there is insufficient depth of collaboration - the feedback mechanism focuses on "error correction+motivation incentives", and the collaborative logic between teachers and AI in terms of control and role division is not deep enough, which is the core issue of Teacher-AI Orchestration.

In the future, expanding to English listening and speaking teaching in Hong Kong secondary schools, following the 2023 curriculum guidelines of the Hong Kong Education Bureau, exploring the Teacher-AI Orchestration model. Hong Kong secondary school English classes have a large capacity (35-40 students/class) [10], and AI can provide BBC Learning English graded news. Teachers can adjust the classroom based on AI data (such as past tense pronunciation error rate), verify the cross scenario applicability of the mechanism, and provide solutions for the implementation of Teacher-AI Orchestration.

7. Conclusion

The 8-week empirical study has shown us the potential of automated feedback mechanisms in technology enhanced language learning - it is not meant to replace teachers, but to break through the bottleneck of traditional feedback through the synergy of "AI efficiency+teacher temperature". In the single player scenario of Chinese international education, the gamified automated feedback designed according to the A1 level of the "International Chinese Education Chinese Proficiency Level Standards" not only advanced learners' Chinese grades by an average of 27.3 points (far exceeding the control group's 16.5 points), but also significantly improved their learning motivation, and many people changed from "afraid of Chinese characters" to "love to pass levels". This change is the best proof of technology serving teaching.

Of course, the limitations of the research also clearly indicate the future direction: expanding from individual scenarios to collective classrooms, deepening from simple collaboration to the core logic of Teacher-AI Orchestration, especially in new scenarios such as English teaching in Hong Kong secondary schools, exploring models such as "AI assisted task push+personalized teacher guidance" and "AI

data parsing+teacher classroom regulation" will be the next focus.

Ultimately, the core of technology enhanced language learning has always been 'learner centered'. The value of automated feedback mechanisms lies not only in improving grades, but also in making learning more interesting and efficient, allowing teachers to devote more energy to warm teaching interactions - perhaps this is the true meaning of AI human-machine collaborative teaching, and also a reflection that this study hopes to contribute to educational practice.

References

[1] Shi H, Aryadoust V. A systematic review of AI-based automated written feedback research [J]. ReCALL, 2024, 36(2): 187-209.

[2] VanPatten B, Williams J. Theories in second language acquisition [M]. Oxon: Routledge, 2014.

[3] Dehghanzadeh H, Fardanesh H, Hatami J, et al. Using gamification to support learning English as a second language: a systematic review [J]. Computer Assisted Language Learning, 2021, 34(7): 934-957.

[4] Zichermann G, Cunningham C. Gamification by design: Implementing game mechanics in web and mobile apps [M]. " O'Reilly Media, Inc.", 2011.

[5] Dicheva D, Dichev C, Agre G, et al. Gamification in education: A systematic mapping study [J]. Journal of educational technology & society, 2015, 18(3): 75-88.

[6] Venkatesh V, Davis F D. A theoretical extension of the technology acceptance model: Four longitudinal field studies [J]. Management science, 2000, 46(2): 186-204.

[7] Kalmpourtzis G, Romero M. Constructive alignment of learning mechanics and game mechanics in Serious Game design in Higher Education [J]. International Journal of Serious Games, 2020, 7(4): 75-88.

[8] Shadish W R, Cook T D, Campbell D T. Experimental and quasi-experimental designs for generalized causal inference [M]. Houghton, Mifflin and Company, 2002.

[9] Larson-Hall J. A guide to doing statistics in second language research using SPSS and R [M]. Routledge, 2015.

[10] Zhan Z, He G, Li T, et al. Effect of groups size on students' learning achievement, Motivation, cognitive load, collaborative problem-solving quality, and in-class interaction in an introductory ai course [J]. Journal of Computer Assisted Learning, 2022, 38(6): 1807-1818.