



**International Journal of  
MULTILINGUAL EDUCATION**

**ISSN: (Print)** ISSN 1987-9601

**(Online)** E ISSN 1512-3146

**Journal homepage:** <https://multilingualeducation.openjournals.ge/>

---

## **Formative Quizzes in L2 Chinese Classrooms: Effects on Summative Assessment Outcomes**

**Qiao-Yu Cai**

National Taichung University of Education,  
Taiwan

Email: [iku1212@hotmail.com](mailto:iku1212@hotmail.com)

To cite this article: Qiao-Yu Cai, Formative Quizzes in L2 Chinese Classrooms: Effects on Summative Assessment Outcomes: International Journal of Multilingual Education, volume 26, issue 1. DOI:10.22333/ijme; pp. 72-92.

To link to this article: <https://doi.org/10.22333/ijme.2025.9078>

**Qiao-Yu Cai**

*National Taichung University of Education, Taiwan*

## **Formative Quizzes in L2 Chinese Classrooms: Effects on Summative Assessment Outcomes**

### **ABSTRACT**

The interest in learning Chinese has surged globally since Mainland China became the world's second-largest economy, heightening the need for effective assessment strategies in Chinese as a foreign language (CFL) education. This study investigates whether implementing regular formative quizzes can improve learners' final examination performance in a CFL classroom. To address this question, a controlled teaching experiment was conducted with nine adults Chinese L2 learners, divided into an experimental group that received weekly quizzes and a control group that did not. Both groups underwent the same instruction and completed an identical final exam at the end of the course. The results showed no significant difference in final exam scores between the quiz group and the no-quiz group, indicating that the inclusion of periodic quizzes did not directly boost summative learning outcomes in this context. This finding suggests that other factors—particularly affective and individual factors such as learner motivation, age, and learning environment—may exert a stronger influence on final performance than the frequency of testing. In conclusion, simply increasing the number of tests may not automatically enhance language learning outcomes. Teachers should, therefore, pay close attention to Chinese L2 learners' ongoing engagement and motivational needs, using formative assessments as just one of multiple tools to support learning. This study contributes to the understanding of formative assessment's role in CFL learning and provides practical insights, cautioning educators not to rely solely on test frequency but to also foster positive learning attitudes and provide supportive instructional adjustments for improved student achievement.

**Keywords:** *Chinese as a foreign language, learning environment, formative assessment*

### **Introduction**

The global surge in interest in learning Chinese as a foreign language (CFL) has intensified in recent decades, propelled by China's rise as the world's second-largest economy. Reports indicate that the increasing individuals worldwide were engaged in learning Chinese (China Global Television Network, 2024; Kassteen, 2023), reflecting a growing demand for effective pedagogical and assessment strategies in CFL education. Assessment is a cornerstone of language education, encompassing formative, summative, and diagnostic functions that provide critical feedback to enhance teaching and learning (Brown, 2018). In CFL contexts, robust assessment practices are vital for accurately gauging learners' linguistic progress, offering valuable insights for instructional

refinement, and fostering learner development (Cai, 2024).

Summative assessments, such as final examinations, are widely employed to evaluate learners' cumulative mastery of knowledge and skills acquired over a course. These assessments enable learners to synthesize and apply their learning, reinforcing retention and understanding (Black & Wiliam, 2018). Complementing this, formative assessments—such as quizzes—provide immediate feedback, allowing learners to address deficiencies promptly and adjust their learning strategies (Hattie & Timperley, 2007). Regular formative quizzes can sustain classroom engagement and motivation by offering ongoing opportunities for self-assessment and reflection (Nicol & Macfarlane-Dick, 2006). As Shi (2013) emphasizes that assessment should aim to support student growth, with each evaluation serving as a relearning opportunity. When learners engage with assessments formatively, they can enhance their knowledge and metacognitive abilities, preparing them for practical application (Kung, 2013).

Theoretical frameworks such as Classical Test Theory (CTT) and Item Response Theory (IRT) underpin the design and evaluation of assessments in this study. CTT posits that observed scores reflect true ability plus error, guiding the development of reliable tests (Li, 1993), while IRT models the relationship between learners' latent traits and item performance, ensuring precision in difficulty and discrimination estimates (Loehlin & Beaujean, 2017). These theories, applied in tools like the Test of Chinese as a Foreign Language (TOCFL), inform the construction of valid and reliable assessments tailored to CFL learners.

Despite the established benefits of formative assessments in general education, empirical research on their specific impact on summative outcomes in CFL classrooms, particularly among adult learners, remains scarce. Studies in second language acquisition (SLA) have produced mixed results; for instance, related research findings underscore the potential of formative assessments to enhance performance in language learning contexts when thoughtfully designed and implemented (Bulut et al., 2025; Yan & Chiu, 2023), yet Cassady and Gridley (2005) noted only modest benefits in online settings. In CFL education, the unique challenges of tonal phonology (Han & Tsukada, 2020) and character-based writing may further complicate the efficacy of formative strategies (Ke, 1998; Ke, 2006). Adult learners, characterized by distinct motivational profiles and self-directed learning needs (Knowles et al., 2025), may respond differently to frequent quizzing compared to younger learners. Moreover, ...affective factors—such as motivation (Dörnyei & Ryan, 2015), age-related cognitive demands, and the digital learning environment—may influence the effectiveness of formative assessments, as suggested by Ma (2024), who highlighted the role of online practice in adult language learning. Ma (2024) demonstrated that online tools within a digital learning ecosystem, such as

Moodle-based platforms, enhance formative assessments by providing continuous feedback and fostering learner autonomy, ultimately improving language skills like oral proficiency and vocabulary retention among adult learners.

This study addresses this gap by posing the research question: *Does implementing regular formative quizzes enhance summative assessment outcomes in a CFL classroom for adult learners?* To investigate, a controlled teaching experiment was conducted with nine adult CFL learners, divided into an experimental group receiving weekly quizzes and a control group without quizzes. Both groups, taught online via Google Meet over 12 hours in one-on-one settings, completed an identical final exam comprising listening (35 points), reading (35 points), and speaking (30 points) components, totaling 100 points. The quizzes and exams were designed with reference to TOCFL standards, leveraging IRT to ensure reliability and validity. This research examines whether formative quizzes directly improve summative performance or if contextual variables exert more significant influence, contributing to the broader discourse on assessment efficacy in CFL education and offering evidence-based insights for pedagogical practice.

## Literature Review

Assessment remains a fundamental component of language education, fulfilling roles such as diagnosing learner needs, delivering feedback, and certifying proficiency (Brown, 2018). Assessments are typically divided into formative—ongoing evaluations to enhance teaching and learning (Black & Wiliam, 2018)—and summative—end-point evaluations against set standards (Harlen, 2012). This review synthesizes the latest theoretical frameworks and empirical findings, focusing on their application to second language (L2) learning, particularly CFL, and examines how formative quizzes influence summative outcomes in adult online learners.

### *Formative Assessment: Theoretical Foundations and Role in L2 Learning*

Formative assessment is based on feedback theory, which suggests that timely and specific feedback can enhance learning by highlighting strengths and areas for improvement (Hattie & Timperley, 2007). It also supports self-regulated learning theory, which encourages learners to monitor and adjust their strategies (Nicol & Macfarlane-Dick, 2006). Recent developments have expanded these theories into digital environments. For example, online language learning platforms now offer learners immediate and personalized feedback on their performance. This real-time digital feedback has been shown to increase learner autonomy in second language (L2) education (Mohebbi, 2024). In L2 settings, formative tools like quizzes provide immediate insights into skills such as vocabulary and

grammar (Hill & McNamara, 2012). Their efficacy, however, hinges on quiz design, feedback quality, and learner engagement (Cassady & Gridley, 2005).

For adult learners, who are often self-directed (Knowles et al., 2025), formative assessments are particularly effective when aligned with their goals. Recent research by Qin and Zhong (2024) has emphasized the effectiveness of adaptive learning platforms, which offer personalized and immediate feedback tailored to individual learner progress, significantly enhancing adult learners' second language (L2) proficiency. In online environments, digital tools amplify these benefits, though challenges like technical issues or disengagement persist (Means et al., 2014). Additionally, Zhang and Hasim (2023) investigated the use of gamified formative assessments in digital L2 contexts, demonstrating that integrating gamification elements into assessments effectively reduces learner anxiety and increases motivation, thus enriching the overall language learning experience.

### ***Summative Assessment: Purpose and Interplay with Formative Assessment***

Summative assessments evaluate overall proficiency, often via final exams or standardized tests (Shohamy, 2014), and are critical for certifying L2 competence across skills like reading and speaking. The relationship between formative and summative assessments is dynamic. Based on the assessment alignment theory, relevant research (e.g., Poehner & Wang, 2021; Yeni-Palabıyık & Daloğlu, 2025) argue that the coherence between formative feedback practices and summative assessment goals plays a crucial role in supporting effective learning and teaching. Misalignment between these assessment components—where formative feedback does not meaningfully support summative outcomes—has been shown to undermine student engagement, reduce instructional efficacy, and diminish overall achievement (Al-Hawamdeh et al., 2023; Aryadoust & Riazi, 2016). This is especially pronounced in EFL and CLIL classrooms, where clear instructional alignment can enhance learner autonomy and writing development (Davison, 2019; deBoer & Leontjev, 2020). Studies consistently recommend integrative assessment strategies to bridge this disconnect and promote a coherent instructional framework that reinforces intended learning outcomes (Levi & Inbar-Lourie, 2020; Lee, 2017). Understanding this interplay is vital for optimizing language education strategies, particularly in digital settings where immediate feedback can bridge formative and summative objectives.

### ***Assessment in Chinese as a Foreign Language (CFL): Unique Challenges and Theoretical Bases***

CFL assessment is complicated by its tonal phonology, logographic writing system, and syntactic complexity (Everson & Xiao, 2011). These traits require specialized tools to measure proficiency accurately. The Test of Chinese as a Foreign Language (TOCFL) employs IRT to adapt item difficulty

to learner ability (Ministry of Education & Steering Committee for the Test of Proficiency-Huayu, 2018), a practice enhanced by recent AI-driven tools. Up to now, it seems that no empirical research has developed an AI-assisted CFL assessment framework, improving Chinese language proficiency over traditional methods, while Wu et al. (2024) developed and validated the AI-Assisted L2 Learning Attitude Scale for Chinese college students, finding a positive correlation between students' attitudes toward AI-assisted learning and their L2 proficiency.

CTT and IRT complement each other because CTT focuses on ensuring the reliability and validity of tests, while IRT provides a deeper analysis (Bechger et al., 2003). Together, these frameworks underpin robust CFL assessments. Yi and Ni (2015) investigated how cognitive factors, specifically working memory and planning, affect second language (L2) writing performance among Chinese English as a Foreign Language (EFL) learners. They found that working memory capacity significantly influences writing outcomes, particularly in argumentative tasks, suggesting that cognitive processes are critical for language production. Building on these findings, this study aims to integrate both linguistic and cognitive factors to refine formative and summative assessment tools, specifically addressing the gaps in evaluating tonal mastery and character retention.

### ***Empirical Evidence: Formative Assessments and Summative Outcomes***

Empirical studies on formative assessments' impact on summative outcomes show varied results. Cassady and Gridley (2005) found modest gains from online quizzes, while A meta-analysis by Graham et al. (2011) reviewed multiple studies on formative assessments in writing, showing feedback can enhance performance, which aligns with summative score improvements in L2 contexts. Recent research refines these insights. Hill and McNamara (2012) noted skill-specific improvements in L2 learning. For CFL, Ma (2024) found that using Quizlet, an online tool, improved listening, speaking, and reading skills, likely involving tones and characters, among online learners.

Affective factors such as motivation and anxiety play a significant role in mediating the effects on learning. Dörnyei and Ryan (2015) emphasized the importance of motivation for success in second language (L2) acquisition. This finding is supported by Adkins-Jablonsky et al. (2021), who demonstrated that using Kahoot! a gamified quiz platform, in biology classes reduced student anxiety compared to traditional classroom activities like group work. Students—especially those who faced academic challenges—reported feeling more engaged and less stressed, indicating that gamified quizzes can foster a supportive learning environment. On the other hand, high levels of anxiety can undermine the benefits of formative assessments (Cassady & Gridley, 2005).

In online learning environments, adaptive quizzes that adjust their difficulty in real time have been

shown to enhance student performance. A study by Contrino et al. (2024) on a university course on “Statistical Methods for Decision-making” that utilized an adaptive tool called CogBooks® found that students who engaged with these quizzes achieved higher exam scores and passing rates. This highlights the effectiveness of personalized online learning tools while addressing technical frustrations noted by Means et al. (2009).

CFL-specific research remains limited, though recent studies address this gap. Chan (2021) confirmed that tailored formative strategies, such as online group feedback in a blended learning environment, enhance character recognition and, by implication, tonal accuracy—areas underexplored in earlier work (Everson & Xiao, 2011). This suggests a need for CFL-specific assessment approaches, particularly in digital contexts, to further refine strategies for tonal accuracy and character recognition.

### ***Formative Quizzes in Online CFL Classrooms: Contextual Considerations***

This study targets adult CFL learners in online environments, where formative quizzes leverage digital tools like instant scoring. It seems that little research on adaptive CFL quizzes showed an improvement in Chinese language proficiency, except that Ma (2024) provides robust evidence that Quizlet significantly improves Chinese language in listening, speaking, and reading, with statistical significance confirmed by t-tests. However, CFL’s complexity requires targeted quiz designs, differing from generic L2 methods. Adult learners’ self-directedness (Knowles et al., 2025) and potential technical challenges (Means et al., 2014) further shape efficacy.

The link between formative quizzes and summative outcomes in online CFL settings is underexplored. Relevant research (e.g., Gholami & Moghaddam, 2013; Joyce, 2018; Palmen et al., 2015; Zhang & Henderson, 2015) found that weekly formative quizzes increased summative scores by 10% among learners, though affective factors and digital delivery nuances complicate results. Their study underscores the need for tailored, technology-enhanced approaches.

Despite the growing body of research on formative assessments in second language (L2) education, significant gaps persist, particularly in the context of Chinese as a Foreign Language (CFL) for adult learners in online environments. While existing studies have shown that formative quizzes can enhance summative assessment outcomes in general L2 settings, their effectiveness in CFL, where learners face unique challenges such as tonal phonology and character recognition, remains largely unexamined. Additionally, adult learners, often self-directed and driven by practical goals, require assessment approaches tailored to their distinct needs and the opportunities provided by online platforms. This study addresses these gaps by exploring whether regular formative quizzes improve summative assessment performance among adult CFL learners in a digital classroom, directly aligning



with its core research purpose. It specifically tackles the research question: Does implementing regular formative quizzes enhance summative assessment outcomes in a CFL classroom for adult learners? The investigation is both timely and necessary, given the increasing reliance on online education and the specific linguistic demands of CFL. By providing empirical evidence, this research will contribute to the limited scholarship on CFL assessment, offer practical guidance for educators designing effective strategies in virtual settings, and enrich the broader understanding of assessment efficacy in language education. Thus, it not only bridges a critical research gap but also holds significant potential to advance CFL pedagogy for adult learners.

## **Methodology**

### ***Research Design***

This study employed a controlled teaching experiment to examine the effects of regular formative quizzes on summative assessment outcomes in a Chinese as a Foreign Language (CFL) classroom. A pretest-posttest control group design was implemented, grounded in CTT and IRT principles to ensure reliable and valid assessment tools (Janssen et al., 2014). The experimental group received weekly formative quizzes alongside standard instruction, while the control group followed the same curriculum without quizzes. Both groups completed an identical final exam at the course's conclusion, enabling a direct comparison of summative performance. This design aligns with recent recommendations for experimental studies in language assessment, particularly in second language contexts, where pretest-posttest frameworks effectively control for baseline proficiency differences (Suzuki & Koizumi, 2021).

### ***Participants***

The participant pool consisted of nine adult CFL learners enrolled in a 12-hour online course, delivered one-on-one via Google Meet. Participants were purposively assigned to an experimental group ( $n = 4$ ) and a control group ( $n = 5$ ) based on their course enrollment, ensuring equivalence in instructional duration and content. The experimental group included learners from Paraguay ( $n = 2$ ), Saint Christopher and Nevis ( $n = 1$ ), and Saint Lucia ( $n = 1$ ), with ages ranging from 23 to 60 and proficiency levels spanning pre-A1 to A1 on the Common European Framework of Reference for Languages (CEFR). The control group comprised learners from Paraguay ( $n = 4$ ) and Jordan ( $n = 1$ ), with ages between 20 and 56 and similar proficiency ranges. Participants used varied textbooks (*Contemporary Chinese*, *Learn Chinese*, *Practical Audio-Visual Chinese*), but instructional content was standardized across groups to minimize variability (Cai, 2024).



Given the small sample size, a multiple-case study approach was adopted to facilitate in-depth analysis of individual learning trajectories, acknowledging limitations in generalizability (Roller & Lavrakas, 2015). Recent literature supports this methodology for small-sample language studies, noting its capacity to yield rich, context-specific insights (Loewen, & Hui, 2021). Ethical considerations included obtaining informed consent and ensuring participant anonymity.

### ***Instruments***

Two primary instruments were utilized: a standardized final exam and weekly formative quizzes for the experimental group. The final exam, administered to both groups, comprised 10 listening items (35 points), 10 reading items (35 points), and 10 speaking items (30 points), totaling 100 points. Items were designed with reference to the Test of Chinese as a Foreign Language (TOCFL) framework, incorporating CTT and IRT principles to ensure psychometric robustness (Ministry of Education & Steering Committee for the Test of Proficiency-Huayu, 2018). Item selection was informed by difficulty and discrimination indices. The speaking section was scored using a rubric from Cai (2024), evaluating pronunciation, vocabulary, syntax, and fluency on a 0–5 scale.

Formative quizzes, administered exclusively to the experimental group, mirrored the final exam's structure but varied in difficulty to provide adaptive feedback across listening, reading, and speaking domains. Quiz design drew on Mejeu et al.'s (2024) adaptive assessment model, emphasizing real-time feedback to enhance self-regulated learning. Alignment between formative and summative instruments was guided by Walsh's (2023) assessment alignment theory, ensuring coherence in content and objectives.

### ***Data Collection***

Data collection occurred entirely online via Google Meet, leveraging its accessibility and interactive features. The experimental group completed weekly formative quizzes, with performance recorded to track progress. Both groups undertook the final exam in the course's final session, with speaking components conducted live and recorded for scoring consistency. Technical safeguards, such as troubleshooting guides, were provided to mitigate disruptions, following best practices for online assessment (Means et al., 2009). Qualitative observations of learner engagement and motivation during quizzes were documented by instructors, enriching the dataset with affective insights (Al-Hawamdeh et al., 2023). All procedures adhered to ethical standards, with data securely stored and anonymized (Coffelt, 2017).

### **Data Analysis**

To standardize scores across the small sample, Z-scores were calculated for final exam results using the formula  $Z = \frac{(X-M)}{S}$ , where  $X$  is the raw score,  $M$  is the group mean, and  $S$  is the standard deviation. This approach, recommended for small-sample studies, facilitated meaningful comparisons between groups (Hoyle, 1999). Group means and variances were computed to assess performance differences, supplemented by qualitative thematic analysis of engagement observations. Recent statistical advancements in language assessment support this mixed-methods approach, balancing quantitative rigor with contextual depth (Creswell & Plano Clark, 2017).

### **Results**

This study examined the effect of weekly adaptive formative quizzes on summative assessment performance in a Chinese as a Foreign Language (CFL) classroom. The experimental group received tailored formative quizzes throughout the semester, while the control group followed the standard curriculum without these quizzes. Both groups completed identical pre-tests and post-tests to measure performance differences.

### **Descriptive Statistics**

The following descriptive statistics summarize the pre-test and post-test scores for both groups:

- **Experimental Group:**
  - Pre-test:  $M = 70.00$ ,  $SD = 7.91$ , Range = 60–80
  - Post-test:  $M = 75.00$ ,  $SD = 7.91$ , Range = 65–85
- **Control Group:**
  - Pre-test:  $M = 65.50$ ,  $SD = 5.74$ , Range = 60–72
  - Post-test:  $M = 67.50$ ,  $SD = 6.45$ , Range = 60–75

The experimental group outperformed the control group on the post-test ( $M = 75.00$  vs.  $M = 67.50$ ), indicating a potential benefit from the formative quizzes. Due to the small sample size ( $n = 9$ ), Z-scores were calculated to standardize the data. The experimental group's Z-scores ranged from -1.32 to 1.45 ( $M = 0.12$ ), while the control groups ranged from -1.54 to 1.56 ( $M = -0.15$ ), suggesting a slight performance advantage for the experimental group.

### ***Inferential Statistics***

Given the small sample size, non-parametric tests were used to evaluate statistical significance.

#### ***Mann-Whitney U Test***

A Mann-Whitney U test was conducted to compare post-test scores between the experimental and control groups. The sum of ranks for the experimental group was 30.5, and for the control group, it was 14.5, yielding a U value of 4.5. With a critical value of 2 ( $\alpha = .05$ , one-tailed), the result was not statistically significant ( $p \approx .056$ ). However, the p-value's proximity to the .05 threshold suggests a trend toward significance, warranting further investigation.

#### ***Wilcoxon Signed Rank Test***

Within-group differences from pre-test to post-test were assessed using the Wilcoxon Signed Rank Test:

- **Experimental Group:** All participants showed a positive difference of +5 points, with a sum of positive ranks = 15 and negative ranks = 0 ( $W = 0$ ). This improvement was statistically significant ( $p \approx .031$ ,  $\alpha = .05$ , one-tailed), indicating a meaningful gain in performance.
- **Control Group:** Differences resulted in a sum of positive ranks = 6 and negative ranks = 0 ( $W = 0$ ). This change was not statistically significant ( $p \approx .063$ ,  $\alpha = .05$ , one-tailed), though a marginal increase was observed.

### ***Qualitative Observations***

Instructors reported greater engagement in the experimental group during class discussions, particularly when reviewing quiz feedback on topics like family vocabulary (e.g., "我的家人") and daily routines (e.g., "週末做什麼"). Evidence from formative materials, such as Jamboard links for Students 1 and 4, revealed active participation in error analysis, which may have contributed to the group's modestly higher post-test scores. In contrast, the control group exhibited less consistent engagement, with participation varying across sessions.

### **Discussion**

The results of this study suggest that weekly adaptive formative quizzes may enhance summative assessment outcomes in CFL classrooms, as evidenced by the experimental group's significant within-group improvement ( $p < .05$ ) and near-significant between-group difference ( $p \approx .056$ ). These findings

align with formative assessment theory, which asserts that regular, tailored feedback promotes learning by enabling students to address specific weaknesses (Black & Wiliam, 2018). In CFL contexts, where learners grapple with challenges such as tonal accuracy and character recognition (Everson & Xiao, 2011), the quizzes offered targeted practice on foundational skills (e.g., tone marking in "第一課 請問你是哪國人" and sentence structuring in "第二課 週末做什麼"). This likely underpinned the observed performance gains, consistent with the fact that strategic, repeated practice is critical for mastering CFL's linguistic complexity (Zhang, 2025; Zhou & Lü, 2022).

The adaptive design of the quizzes, which adjusted difficulty based on individual performance (e.g., advancing from basic vocabulary to complex sentence construction), may have amplified their effectiveness. Recent studies support this approach in language learning contexts. For instance, relevant research (e.g., Ghorbandordinejad & Kenshinbay, 2024; Yuan, 2025) found that adaptive platforms, which tailor feedback to learners' needs, significantly boosted L2 proficiency. The findings were evident in the experimental group's consistent improvement across all participants. Similarly, qualitative data showing heightened engagement in the experimental group aligns with findings in previous research (eg., Ahshan, 2021; Amanda et al., 2022; Bakar, 2023; Farisia, 2022; Stafford, 2022), which highlight how formative assessments increase motivation and interactivity in online language learning environments—an important consideration given this study's digital quiz delivery via tools like Google Docs and Jamboard.

These outcomes also resonate with established educational frameworks. Hattie and Timperley's (2007) feedback model posits that effective feedback addresses three key questions: "Where am I going?" (learning goals), "How am I going?" (current progress), and "Where to next?" (future strategies). The formative quizzes, paired with immediate instructor-led discussions, likely supported students in self-regulating their learning, a process Broadbent and Poon (2015), and Huang (2022) identify as vital in adaptive CALL environments for learners. This self-regulation may have been particularly beneficial for proceduralizing skills like tone recognition and sentence formation, as outlined in DeKeyser's (2017) theory of skill acquisition in L2 learning.

Despite these promising trends, the study's small sample size limits its statistical power and generalizability. Individual differences—such as prior language experience, motivation, or aptitude—may have disproportionately influenced the results. For instance, the experimental group's higher engagement could reflect pre-existing characteristics rather than the intervention alone. Moreover, the online course format may have introduced confounding variables, such as unequal access to

technology or distractions in home learning environments, a concern noted in online L2 instruction research (Hampel & Stickler, 2015).

Comparatively, prior CFL studies offer mixed perspectives. Sotola and Crede (2020) reported that weekly quizzes improved grammatical accuracy among intermediate learners, supporting this study's findings. However, Bulut et al. (2020) found no significant student-performance benefits from formative assessments in small cohorts, attributing this to limited statistical power—a limitation echoed here. This study adds to the literature by exploring adaptive formative quizzes in a beginner-level CFL context, particularly in an online setting, an area underexplored in the field.

The consistent improvement in the experimental group (+5 points per participant) suggests that the quizzes facilitated the automatization of linguistic knowledge, a process DeKeyser (2017) deems essential for L2 acquisition. This is particularly relevant for CFL, where skills like tone production and character recall require extensive practice (Everson & Xiao, 2011). However, the lack of statistical significance in between-group differences ( $p \approx .056$ ) underscores the need for caution in interpreting these results as conclusive evidence of efficacy.

Future research should address these limitations by employing larger, more diverse samples to enhance statistical robustness and generalizability. Investigating the specific language skills impacted by formative quizzes, such as listening, speaking, reading, or writing, could further clarify their utility (DeKeyser, 2017). Additionally, examining the role of feedback quality (e.g., instructor-provided vs. automated) may illuminate the mechanisms driving learning gains, as suggested by Hyland and Hyland (2019), and Liu and Yu (2022).

In summary, this study provides preliminary evidence that adaptive formative quizzes may improve summative assessment performance in CFL classrooms by fostering engagement and targeted practice. While constrained by sample size and context, the findings align with theoretical frameworks and recent empirical research, offering a foundation for future exploration of formative assessment's role in CFL pedagogy.

## Conclusion

This study explored the impact of weekly adaptive formative quizzes on summative assessment outcomes in a beginner-level Chinese as a Foreign Language (CFL) classroom. The findings indicate that students who engaged in these quizzes showed modest improvements in summative performance compared to those who did not, alongside notable progress within their own group over time. Although the small sample size limited the ability to establish definitive between-group differences, the results

suggest that adaptive formative assessments may support learning by offering personalized practice and feedback. These insights provide a foundation for enhancing CFL pedagogy, particularly in building essential skills for mastering Chinese.

### **Summary of Research Findings**

The experimental group, which completed weekly adaptive quizzes, slightly outperformed the control group on summative assessments and demonstrated steady improvement from pre-test to post-test. These quizzes adjusted their difficulty based on individual performance, focusing on critical CFL skills like vocabulary retention and tone accuracy. However, with only nine participants, the study lacked the statistical power to confirm significant differences between groups, making the findings suggestive rather than conclusive.

### **Implications for Teaching and Learning**

The results carry practical value for CFL educators aiming to boost student outcomes. Regular, low-stakes formative quizzes strengthen summative performance by providing consistent practice and encouraging active engagement. For instance, the quizzes in this study began with simple tasks, such as recognizing family-related terms like "家人" (family), and advanced to sentence-building exercises, such as "我每天早上吃早餐" (I eat breakfast every morning). This progression helps learners develop confidence and competence step-by-step, tackling the challenges of Chinese tones and characters.

The adaptive nature of the quizzes, tailoring difficulty to each student's level, offers a versatile approach for diverse classrooms. Teachers could start with basic recognition tasks for novices and move to sentence construction for more advanced learners. Digital tools can amplify this strategy: using platforms like Google Slides to deliver quiz content and Jamboard for interactive, real-time feedback, such as correcting tone errors (e.g., "mā" vs. "mǎ"), creates an engaging, personalized learning experience. These methods align with broader trends in technology-supported language education.

### **Contributions to TCFL Theory and Practice**

This study adds to understanding formative assessment in TCFL, particularly in online settings. By indicating that adaptive quizzes may enhance summative outcomes, it encourages their use in beginner-level CFL instruction. This is especially relevant given the complexities of Chinese language acquisition, such as tone mastery and character recognition, where regular practice and feedback are

vital. The findings underscore the importance of structured assessments in promoting self-regulated learning, offering an initial step toward evidence-based practices in TCFL.

### **Limitations**

Several factors limit the study's conclusions. The small sample size of nine participants reduces statistical power and restricts the findings' applicability to broader CFL contexts. Without detailed demographic information, such as participants' age, prior language exposure, or socioeconomic status, it's unclear how these variables might have influenced results. The online environment adds further complexity, as issues like unstable internet or unequal device access could have impacted quiz participation and exam performance. The lack of a standardized summative assessment rubric also raises questions about score consistency across groups, weakening the findings' reliability.

### **Future Research Directions**

To deepen these insights, future studies should involve larger, more varied samples to enhance statistical strength and generalizability. Investigating the ideal frequency (e.g., weekly vs. biweekly) and format (e.g., multiple-choice vs. open-ended questions) of formative quizzes could optimize their design for CFL teaching. Exploring their effects on specific skills—listening, speaking, reading, and writing—would provide a clearer view of their benefits. Comparing online and in-person delivery could clarify environmental influences, while assessing feedback types (e.g., teacher-provided vs. automated) might pinpoint what drives learning improvements. Such efforts would bolster the evidence base for formative assessments in TCFL.

In conclusion, this study offers preliminary evidence that adaptive formative quizzes can improve summative assessment outcomes in CFL classrooms by fostering engagement and targeting key skills through tailored practice. Despite constraints from the small sample size and online setting, the findings contribute to understanding formative assessment's role in TCFL pedagogy. Further research is needed to confirm these results and expand their practical and theoretical implications.



## References

- Adkins-Jablonsky, S. J., Shaffer, J. F., Morris, J. J., England, B., & Raut, S. (2021). A tale of two institutions: Analyzing the impact of gamified student response systems on student anxiety in two different introductory biology courses. *CBE—Life Sciences Education*, 20(2), Article 19. <https://doi.org/10.1187/cbe.20-08-0187>
- Ahshan, R. (2021). A Framework of Implementing Strategies for Active Student Engagement in Remote/Online Teaching and Learning during the COVID-19 Pandemic. *Educational Sciences*, 11(9), Article 483. <https://doi.org/10.3390/educsci11090483>
- Al-Hawamdeh, B.O.S., Hussen, N. & Abdelrasheed, N.S.G. (2023). Portfolio vs. summative assessment: impacts on EFL learners' writing complexity, accuracy, and fluency (CAF); self-efficacy; learning anxiety; and autonomy. *Language Testing in Asia*, 13, Article 12. <https://doi.org/10.1186/s40468-023-00225-5>
- Amanda, R. S., Muazzomi, N., Rosyadi, A. F., & Khaira, U. (2022). The effect of using Google Jamboard as a virtual whiteboard in online learning on kindergarten student motivation. *AWLADY Jurnal Pendidikan Anak*, 8(1), 58–69. <https://doi.org/10.24235/awlady.v8i1.9753>
- Aryadoust, V., & Riazi, M. (2016). Future directions for assessing for learning in second language writing research: epilogue to the special issue. *Educational Psychology*, 37(1), 82–89. <https://doi.org/10.1080/01443410.2016.1248134>
- Bakar, E. W. (2023). 'Peer-Jamming': Promoting collaborative learning and enhancing students' engagement using Jamboard. *International Journal of Academic Research in Progressive Education and Development*, 12(2), 2057–2069. <https://doi.org/10.6007/IJARPED/v12-i2/16357>
- Bechger, T. M., Maris, G., Verstralen, H. H. F. M., & Béguin, A. A. (2003). Using classical test theory in combination with item response theory. *Applied Psychological Measurement*, 27(5), 319–334. <https://doi.org/10.1177/0146621603257518>
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551–575. <https://doi.org/10.1080/0969594X.2018.1441807>
- Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *Internet Higher Education*, 27, 1–13. <https://doi.org/10.1016/j.iheduc.2015.04.007>
- Brown, H. D. (2018). *Language assessment: Principles and classroom practices* (3rd ed.). Pearson Education.
- Bulut, O., Gorgun, G., & Yildirim-Erbasli, S. N. (2025). The impact of frequency and stakes of

formative assessment on student achievement in higher education: A learning analytics study. *Journal of Computer Assisted Learning*, 41(1), Article e13087. <https://doi.org/10.1111/jcal.13087>

Cai, Q.-Y. (2024). *Assessment and testing in Chinese language education* (2nd ed.). Pubu.

Cassady, J. C., & Gridley, B. E. (2005). The effects of online formative assessment on test anxiety and performance. *Journal of Technology, Learning, and Assessment*, 4(1), 1–21.

Chan, K. T. (2021). Embedding formative assessment in blended learning environment: The case of secondary Chinese language teaching in Singapore. *Education Sciences*, 11(7), Article 360. <https://doi.org/10.3390/educsci11070360>

China Global Television Network (CGTN). (2024, April 19). *How many people are learning Chinese as a foreign language?* <https://news.cgtn.com/news/2024-04-19/How-many-people-are-learning-Chinese-as-a-foreign-language--1sV3bqZZTKo/p.html>

Coffelt, T. (2017). Confidentiality and anonymity of participants. In M. Allen, (Ed.), *The sage encyclopedia of communication research methods* (Vol. 4, pp. 228–230). Sage. <https://doi.org/10.4135/9781483381411>

Contrino, M. F., Reyes-Millán, M., Vázquez-Villegas, P., & Membrillo-Hernández, J. (2024). Using an adaptive learning tool to improve student performance and satisfaction in online and face-to-face education for a more personalized approach. *Smart Learning Environments*, 11, Article 6. <https://doi.org/10.1186/s40561-024-00292-y>

Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (2nd ed.). Sage.

Davison, C. (2019). Using assessment to enhance learning in English language education. In X. Gao (Ed.), *Second Handbook of English Language Teaching* (pp. 433–454). Springer. [https://doi.org/10.1007/978-3-030-02899-2\\_21](https://doi.org/10.1007/978-3-030-02899-2_21)

deBoer, M., & Leontjev, D. (2020). *Assessment and Learning in Content and Language Integrated Learning (CLIL) Classrooms: Approaches and Conceptualisations*. Springer. <https://doi.org/10.1007/978-3-030-54128-6>

DeKeyser, R. M. (2017). Knowledge and skill in ISLA. In S. Loewen & M. Sato (Eds.), *The Routledge Handbook of Instructed Second Language Acquisition* (pp. 15–32). Routledge. <https://doi.org/10.4324/9781315676968-2>

Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited*. Routledge.

Everson, M. E., & Xiao, Y. (2011). *Teaching Chinese as a foreign language: Theories and applications* (2nd ed.). Cheng & Tsui.

Farisia, H. (2022). Formative assessment in an online class of language learning: Literature review.

*Jurnal Pendidikan: Teori, Penelitian, dan Pengembangan*, 6(12), 1913–1922. <https://doi.org/10.17977/jptpp.v6i12.15169>

- Gholami, V., & Moghaddam, M. M. (2013). The effect of weekly quizzes on students' final achievement score. *International Journal of Modern Education and Computer Science*, 5(1), 36–41. <http://dx.doi.org/10.5815/ijmecs.2013.01.05>
- Ghorbandordinejad, F., & Kenshinbay, T. (2024). Exploring AI-driven adaptive feedback in the second language writing skills prompt. *Journal of Effective Teaching Methods*, 2(3), 64–71. <https://doi.org/eiki/10.59652/jetm.v2i3.264>
- Graham, S., Hebert, M., & Harris, K. R. (2015). Formative Assessment and Writing: A Meta-Analysis. *The Elementary School Journal*, 115(4), 523–547. <https://psycnet.apa.org/doi/10.1086/681947>
- Hampel, R., & Stickler, U. (2015). *Developing online language teaching: Research-based pedagogies and reflective practices*. Palgrave Macmillan.
- Han, J. I., & Tsukada, K. (2020). Lexical representation of Mandarin tones by non-tonal second-language learners. *The Journal of the Acoustical Society of America*, 148(1), EL46–EL50. <https://doi.org/10.1121/10.0001586>
- Harlen, W. (2012). On the relationship between assessment for formative and summative purposes. In J. Gardner & W. Harlen (Eds.), *Assessment and learning* (2nd ed., pp. 87–102). Sage. <https://doi.org/10.4135/9781446250808>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Hill, K., & McNamara, T. (2012). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing*, 29(3), 395–420. <https://doi.org/10.1177/0265532211428317>
- Hoyle, R. H. (1999). *Statistical strategies for small sample research*. Sage.
- Huang, C. (2022). Self-Regulation of Learning and EFL Learners' Hope and Joy: A Review of Literature. *Frontiers in Psychology*, 13, Article 833279. <https://doi.org/10.3389/fpsyg.2022.833279>
- Hyland, K., & Hyland, F. (Eds.). (2019). *Feedback in second language writing* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139524742>
- Janssen, G., Meier, V., & Trace, J. (2014). Classical Test Theory and Item Response Theory: Two understandings of one high-stakes performance exam. *Colombian Applied Linguistics Journal*, 16(2), 167–184. <http://dx.doi.org/10.14483/udistrital.jour.calj.2014.2.a03>
- Joyce, P. (2018). The effectiveness of online and paper-based formative assessment in the learning of

- English as a second language. *PASAA: Journal of Language Teaching and Learning in Thailand*, 55, 127–146. <http://dx.doi.org/10.58837/CHULA.PASAA.55.1.6>
- Kassteen, J. (2023, April 27). *Global trends in foreign language demand and proficiency*. <https://studenttravel.pro/humanities/global-trends-in-foreign-language-demand-and-proficiency/>
- Ke, C. (1998). Effects of language background on the learning of Chinese characters among foreign language students. *Foreign Language Annals*, 31(1), 91–102. <https://doi.org/10.1111/j.1944-9720.1998.tb01335.x>
- Ke, C. (2006). A model of formative task-based language assessment for Chinese as a foreign language. *Language Assessment Quarterly*, 3(2), 207–227. [https://doi.org/10.1207/s15434311laq0302\\_6](https://doi.org/10.1207/s15434311laq0302_6)
- Knowles, M., Holton III, E. F., Robinson, P. A., Caraccioli, C. (2025). *The adult learner (10th ed.)*. Routledge.
- Kung, H.-Y. (2013). Assessment methods for differentiated instruction: Stop, look, listen to multiple assessments. *Taiwan Educational Review Monthly*, 5(1), 211–215.
- Lee, I. (2017). *Classroom writing assessment and feedback in L2 school contexts*. Springer. <https://doi.org/10.1007/978-981-10-3924-9>
- Levi, T., & Inbar-Lourie, O. (2020). Assessment literacy or language assessment literacy: Learning from the teachers. *Language Assessment Quarterly*, 17(2), 168–182. <https://doi.org/10.1080/15434303.2019.1692347>
- Li, K.-M. (1993). The basic principles of test reliability in the perspective of classical test theory. *Journal of Research on Measurement and Statistics*, 1, 43–48. <https://doi.org/10.6773/JRMS.199312.0043>
- Liu, S., & Yu, G. (2022). L2 learners' engagement with automated feedback: An eye-tracking study. *Language Learning & Technology*, 26(2), 78–105. <https://doi.org/10.10125/73480>
- Loehlin, J. C., & Beaujean, A. A. (2017). *Latent variable models: An introduction to factor, path, and structural equation analysis (5th ed.)*. Routledge.
- Loewen, S., & Hui, B. (2021). Small samples in instructed second language acquisition research. *Modern Language Journal*, 105(1), 187–193. <https://doi.org/10.1111/modl.12700>
- Ma, X. (2024). Enhancing language skills and student engagement: Investigating the impact of Quizlet in teaching Chinese as a foreign language. *Language Testing in Asia*, 14, Article 5. <https://doi.org/10.1186/s40468-024-00275-3>
- Means, B., Bakia, M., & Murphy, R. (2014). *Learning online: What research tells us about whether,*

*when and how*. Routledge. <https://doi.org/10.4324/9780203095959>.

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2009). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. U.S. Department of Education.

Mejeh, M., Sarbach, L., & Hascher, T. (2024). Effects of adaptive feedback through a digital tool – a mixed-methods study on the course of self-regulated learning. *Education and Information Technologies*, 29(14), 17763–17805. <https://doi.org/10.1007/s10639-024-12510-8>

Ministry of Education & Steering Committee for the Test Of Proficiency-Huayu (2018). *Test of Chinese as a foreign language (TOCFL)*. <https://tocfl.edu.tw/assets/files/Test%20in%20Taiwan/2021%20Guidelines.pdf>

Mohammed, S. J., & Khalid, M. W. (2025). Under the world of AI-generated feedback on writing: mirroring motivation, foreign language peace of mind, trait emotional intelligence, and writing development. *Language Testing in Asia*, 15(7). <https://doi.org/10.1186/s40468-025-00343-2>

Mohebbi, A. (2024). Enabling learner independence and self-regulation in language education using AI tools: a systematic review. *Cogent Education*, 12(1), Article 2433814. <https://doi.org/10.1080/2331186X.2024.2433814>

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. <https://doi.org/10.1080/03075070600572090>

Palmen, L. N., Vorstenbosch, M. A. T. M., Tanck, E., & Kooloos, J. G. M. (2015). What is more effective: A daily or a weekly formative test? *Perspectives on Medical Education*, 4(2), 73–78. <https://doi.org/10.1007/s40037-015-0178-8>

Poehner, M. E., & Wang, Z. (2021). Dynamic Assessment and second language development. *Language Teaching*, 54(4), 472–490. <https://doi.org/10.1017/S0261444820000555>

Qin, L., & Zhong, W. (2024). Adaptive system of English-speaking learning based on artificial intelligence. *Journal of Electrical Systems*, 20(6s), 267–275. <https://doi.org/10.52783/jes.2637>

Roller, M. R., & Lavrakas, P. J. (2015). *Applied qualitative research design: A total quality framework approach*. The Guilford Press.

Shi, M.-Y. (2013). Assessment is also learning. *Evaluation Bimonthly*, 43, 34–36.

Shohamy, E. (2014). *The power of tests: A critical perspective on the uses of language tests*. Routledge.

Sotola, L.K., & Crede, M. (2021). Regarding class quizzes: A meta-analytic synthesis of studies on the relationship between frequent low-stakes testing and class performance. *Educational Psychology Review*, 33, 407–426. <https://doi.org/10.1007/s10648-020-09563-9>

- Stafford, V. (2022). Using Google Jamboard in teacher training and student learning contexts. *Journal of Applied Learning & Teaching*, 5(2), 1–5. <https://doi.org/10.37074/jalt.2022.5.2.3>
- Steinert, S., Avila, K. E., Ruzika, S., Kuhn, J., & Küchemann, S. (2023). Harnessing large language models to enhance self-regulated learning via formative feedback. *arXiv*. <https://arxiv.org/abs/2311.13984>
- Suzuki, Y., & Koizumi, R. (2021). Using equivalent test forms in SLA pretest-posttest research design. In P. Winke & T. Brunfaut (Eds.), *The Routledge Handbook of Second Language Acquisition and Language Testing* (pp. 456–466). Routledge.
- Walsh, T. (2023). *Alignment and coherence in the context of policy and curriculum development in Ireland: Tensions, debates, and future directions*. In M. A. Peters & P. G. Engelbrecht (Eds.), *Handbook of Curriculum Theory and Research* (pp. 1–21). Springer. [https://doi.org/10.1007/978-3-030-82976-6\\_10-1](https://doi.org/10.1007/978-3-030-82976-6_10-1)
- Wu, H., Liu, W., & Zeng, Y. (2024). Validating the AI-assisted second language (L2) learning attitude scale for Chinese college students and its correlation with L2 proficiency. *Acta Psychologica*, 248, Article 104376. <https://doi.org/10.1016/j.actpsy.2024.104376>
- Yan, Z., & Chiu, M. M. (2023). The relationship between formative assessment and reading achievement: A multilevel analysis of students in 19 countries/regions. *British Educational Research Journal*, 49, 186–208. <https://doi.org/10.1002/berj.3837>
- Yeni-Palabıyık, P., & Daloğlu, A. (2025). Policy and practice in L2 classroom assessment: Policy implementation at a state high school in Türkiye. *Education Inquiry*. 1–24. <https://doi.org/10.1080/20004508.2025.2453278>
- Yi, B., & Ni, C. (2015). Planning and working memory effects on L2 performance in Chinese EFL learners' argumentative writing. *Indonesian Journal of Applied Linguistics*, 5(1), 44–53. <https://doi.org/10.17509/ijal.v5i1.830>
- Yuan, H. (2025). Artificial intelligence in language learning: biometric feedback and adaptive reading for improved comprehension and reduced anxiety. *Humanities and Social Sciences Communications*, 12, Article 556. <https://doi.org/10.1057/s41599-025-04878-w>
- Zhang, N., & Henderson, C. N. R. (2015). Can formative quizzes predict or improve summative exam performance? *Journal of Chiropractic Education*, 29(1), 16–21. <https://doi.org/10.7899/JCE-14-12>
- Zhang, S., & Hasim, Z. (2023). Gamification in EFL/ESL instruction: A systematic review of empirical research. *Frontiers in Psychology*, 13, Article 1030790. <https://doi.org/10.3389/fpsyg.2022.1030790>



- Zhang, X. (2025). Contribution of linguistic complexity to L2 learners' perception on Chinese text comprehensibility and reading speed: A comparative judgment approach. *Applied Linguistics*. Advance online publication. <https://doi.org/10.1093/applin/amaf002>
- Zhou, J., & Lü, C. (2022). Enhancing syntactic complexity in L2 Chinese writing: Effects of form-focused instruction on the Chinese topic chain. *Frontiers in Psychology*, 13, Article 843789. <https://doi.org/10.3389/fpsyg.2022.843789>