Arab World English Journal
INTERNATIONAL PEER REVIEWED JOURNAL ISSN: 2229-9327

Arab World English Journal (AWEJ) Volume 16. Number 2, June 2025 Pp.276-310
DOI: https://dx.doi.org/10.24093/awej/vol16no2.16

Cognitive Diagnostic Assessment for IELTS Speaking Skills: A Chinese High School Case Study

Mengjia Liu
Faculty of Education, Universiti Kebangsaan Malaysia,Bangi, Malaysia
Corresponding Email: p126174@siswa.ukm.edu.my

Harwati Hashim
Faculty of Education, Universiti Kebangsaan Malaysia, Bangi, Malaysia
Email: harwati@ukm.edu.my

Nur Ainil Sulaiman
Faculty of Education, Universiti Kebangsaan Malaysia,Bangi, Malaysia
Email: nurainil@ukm.edu.my

Received:02/25/2025 Accepted:06/18/2025 Published:06/26/2025

Abstract
This qualitative case study investigated how a Cognitive Diagnostic Assessment framework enhanced International English Language Testing System speaking skills among Chinese high school English as a Foreign Language learners. The significance of this research lies in addressing the critical gap between diagnostic assessment theory and practical implementation in examination-oriented educational contexts. The study examined three key questions: how cognitive subskills develop through diagnostic assessment, what feedback processing strategies students employ, and what pedagogical challenges teachers face during implementation. The 16-week intervention at an international school in southwest China involved five Grade 11 students at the Common European Framework of Reference A2 level and one experienced teacher. Data collection included semi-structured interviews, classroom observations, mock test recordings, and document analysis. Findings revealed significant improvements across five subskills: linguistic knowledge, syntactic complexity, pronunciation mastery, discourse management, and speech structure awareness. Students evolved from surface-level feedback engagement to sophisticated metacognitive processing strategies, particularly in pattern recognition and cross-contextual transfer. The teacher's implementation emphasized culturally sensitive adaptations, addressing face-saving concerns and reconciling diagnostic assessment with examination traditions. The study demonstrates that integrating diagnostic assessment principles with culturally responsive pedagogy enhances speaking development for high-stakes examinations, with implications for curriculum design and teacher education. The single-site nature and small sample size limit generalizability, warranting broader implementation studies across diverse Chinese educational contexts.

Keywords: Chinese English learners, cognitive diagnostic assessment, diagnostic feedback, English speaking assessment, second language speaking developmen

How to cite: Liu, M., Hashim, H., & Sulaiman, N. A. (2025). Cognitive Diagnostic Assessment for IELTS Speaking Skills: A Chinese High School Case Study. *Arab World English Journal, 16*(2):276-310. https://dx.doi.org/10.24093/awej/vol16no2.16

Introduction

The International English Language Testing System (IELTS) has emerged as a crucial gateway for millions of students pursuing international education and professional opportunities worldwide, with speaking proficiency being one of its most challenging yet essential components (Bogachevska, 2022; Fleming, & Shinjee, 2024; Karim & Haq, 2014; Shiny & Karthikeyan, 2022). In China particularly, where over 27 million university students are engaged in English language learning and assessment, IELTS speaking scores play a decisive role in academic and career advancement (Cheng & Curtis, 2010). Moreover, the significance of the IELTS speaking assessment has intensified as China continues to be one of the largest sources of international students globally, with Chinese test-takers comprising a notable proportion of IELTS candidates around the world (Hu, 2019; Zhang, 2024). Validation studies highlight the generalizability and consistency of linking IELTS speaking tests to China's Standards of English Language Ability, underscoring the role of standardized frameworks in bridging diagnostic feedback and instructional practices (Chen & Wu, 2022).

The current IELTS speaking assessment framework has demonstrated relatively strong reliability and validity in evaluating candidates' oral proficiency through its standardized band scoring system (Brown, 2006a, 2006b; Read, 2022; Suss, 2023). However, it presents significant limitations in providing detailed diagnostic information about learners' specific strengths and weaknesses across these dimensions (Dashti & Razmjoo, 2020; Heidari Vincheh et al., 2024; Kafipour, & Khoshnood, 2023; Quaid, 2018). As Kuo et al. (2016) emphasize, traditional evaluation methods that rely solely on total test scores provide substantially less diagnostic information than comprehensive assessment approaches, making it challenging for learners to identify specific areas requiring focused improvement.

Furthermore, research consistently indicates that Chinese test-takers lag behind global averages in speaking scores, despite demonstrating strong performance in other test components (Heidari Vincheh et al., 2024; Hu, 2019). Traditional assessment methods may not adequately capture or address the specific cognitive and linguistic challenges faced by Chinese learners in oral communication tasks (Huang, 2024; Wang & Wright, 2018). These challenges often stem from fundamental differences between Chinese and English language systems, including phonological patterns, tonal features, syntactic structures, and pragmatic conventions. Nevertheless, as demonstrated by Csapó and Molnár (2019), technology-based assessment systems can effectively monitor students' cognitive development and make the psychological dimension of learning more visible, suggesting promising avenues for improvement.

In response to these challenges, Cognitive Diagnostic Assessment (CDA) has emerged as a particularly promising framework for enhancing language assessment practices. CDA represents a psychometric approach that provides fine-grained diagnostic feedback by identifying specific cognitive attributes or skills that learners have mastered or need to develop (Jang, 2008; Javidanmehr, & Anani, 2017; Leighton & Gierl, 2007; Leighton et al., 2010). Unlike traditional assessment methods that produce single proficiency scores, CDA decomposes performance into

multiple cognitive subskills, enabling attribute-level diagnosis that reveals learners' strengths and weaknesses across distinct competency areas (Tang & Zhan, 2021). This diagnostic granularity is achieved through sophisticated statistical modeling that maps test items to underlying cognitive attributes, thereby generating detailed learner profiles for targeted instructional interventions (Li, 2019; Zhai et al., 2023).

Notably, the CDA framework presented by Heidari Vincheh et al. (2024) offers an innovative approach to understanding the complex cognitive processes involved in language production, emphasizing diagnostic feedback for learner development through systematic analysis of speaking performance patterns. By focusing on the detailed mapping of cognitive skills underlying successful task performance, CDA offers potential solutions to the limitations of traditional scoring methods for IELTS assessment (Minakova, 2020). Recent research has provided compelling evidence for the efficacy of cognitive diagnostic approaches in language testing contexts, demonstrating significant improvements in learners' speaking performance through systematic diagnostic assessment and targeted feedback mechanisms (Ravand & Robitzsch, 2018; Souzandehfar, 2024; Nakatsuhara et al., 2017).

While previous CDA applications have primarily focused on receptive skills such as reading and listening (Jang, 2008; Leighton & Gierl, 2007), the framework's application to productive skills like speaking remains largely unexplored, particularly in examination-oriented contexts such as Chinese educational settings. The highly valued examination culture in Chinese society has created a testing environment that heavily emphasizes summative evaluation over formative assessment (Cheng & Curtis, 2010; Minakova, 2020). This cultural context presents both challenges and opportunities for implementing cognitive diagnostic frameworks in addressing the specific needs of Chinese learners.

Building upon the Cognitive Diagnostic Assessment framework established by Heidari Vincheh et al. (2024), this study examines the implementation and adaptation of this diagnostic approach within the context of Chinese high school EFL learners' IELTS speaking skills development. The significance of this investigation lies in its potential to provide empirical evidence for the practical application of CDA in examination-oriented educational settings, offering insights into how diagnostic assessment can be culturally adapted while maintaining its theoretical integrity. Furthermore, this study addresses the critical gap in qualitative research examining the cognitive processes underlying speaking skill development through diagnostic assessment.

Given the complex, context-dependent nature of implementing diagnostic assessment in real classroom settings, this investigation employs a qualitative case study methodology. This methodological approach enables deep exploration of participants' experiences, cognitive processes, and pedagogical adaptations, capturing the nuanced interactions between diagnostic feedback and high-stakes test preparation. Through systematic investigation aligned with the qualitative paradigm, this research examines the following three research questions:

1. How does the implementation of a Cognitive Diagnostic Assessment (CDA) framework impact the development of specific speaking subskills among Chinese high school students?
2. What cognitive processes do Chinese EFL learners employ when engaging with diagnostic feedback during their IELTS speaking preparation?
3. How does the Chinese EFL teacher implement the CDA framework in pedagogical practice, and what challenges or modifications does the teacher encounter during the process?

These research questions are designed to capture the complexity of implementing diagnostic assessment through qualitative inquiry, enabling rich descriptions of learner development trajectories, cognitive processing strategies, and pedagogical adaptations within the specific cultural and educational context of Chinese IELTS speaking preparation.

Literature Review

IELTS Speaking Assessment

The IELTS speaking assessment, a globally recognized benchmark for evaluating oral language proficiency, has been integral to academic and professional contexts since its introduction in 1989 (Souzandehfar, 2024). The speaking test is structured across three sections—an introduction and interview, a long turn, and a discussion—providing candidates with varied opportunities to demonstrate their speaking skills. It employs a band scoring system ranging from 1 to 9 that evaluates four primary criteria: fluency and coherence, lexical resource, grammatical range and accuracy, and pronunciation (Brown, 2006a, 2006b; Read, 2022). The standardized nature of the framework ensures consistency and comparability across diverse testing environments. However, its predominantly summative orientation limits its ability to provide nuanced, diagnostic feedback for learners aiming to enhance specific aspects of their performance (Dashti & Razmjoo, 2020; Panahi, & Mohebbi, 2022).

One of the strengths of the IELTS speaking assessment lies in its reliability and robust design, which has been extensively validated through empirical research. Ginting et al. (2023) affirm that the test reliably measures speaking proficiency, ensuring fairness and accuracy in evaluating candidates' abilities. However, reliability can sometimes be undermined by examiner variability. Hajiabdorrasouli & Ahmadi, (2023) and Mehraban (2022) highlight that subjective differences in examiner styles and interpretations of band descriptors can influence scoring, impacting inter-rater reliability. While standardized rubrics mitigate discrepancies, absolute consistency across examiners remains a challenge.

Concerns regarding the validity of the IELTS speaking test have also been raised. Gagen, & Faez (2024) critique the predictive validity of the test, questioning its effectiveness in gauging real-world communicative competence. While the test effectively assesses structured speaking tasks, it may not fully capture the dynamic, interactive nature of authentic communication. Similarly, Heidari Vincheh et al. (2024) question the content validity of the assessment, pointing

out gaps in its ability to evaluate a comprehensive range of speaking skills, including pragmatic and sociolinguistic competencies essential for academic and professional success. These critiques emphasize the need for more holistic evaluation criteria that encompass both linguistic accuracy and communicative effectiveness.

A notable limitation of the IELTS speaking framework is its inability to provide detailed diagnostic feedback, crucial for identifying and addressing specific weaknesses. The band descriptors, while offering a broad overview of performance, lack the granularity required for targeted skill development (Huang, 2024; Kafipour & Khoshnood, 2023). This is especially problematic for non-native English-speaking learners, such as Chinese EFL students, who face unique linguistic and cognitive challenges in mastering speaking skills (Hu, 2019; Wang & Wright, 2018). For instance, phonological differences, such as the absence of certain English phonemes in Mandarin, contribute to pronunciation difficulties. Similarly, syntactic and pragmatic divergences impede fluency and discourse management, while cultural factors, including an emphasis on rote learning, limit authentic communicative practice (Huang et al. 2022).

There are also some disadvantages for certain learner groups in terms of examiner perceptions. Research suggests that cultural or linguistic biases may inadvertently affect scoring outcomes, particularly when learners display non-native-like pronunciation or discourse strategies that deviate from Anglo-centric norms (Borger, 2019; Woolridge et al. 2023). These biases underscore the need for inclusive assessment practices that acknowledge diverse linguistic and cultural identities while maintaining fairness and objectivity.

The potential of diagnostic approaches extends beyond individual skill development to broader pedagogical implications. Research supports the role of diagnostic assessment in informing curriculum design. Ismail et al. (2022) emphasize that detailed diagnostic information enables educators to align instructional strategies with learners' specific needs, enhancing language instruction effectiveness. Similarly, Barghi et al. (2025) demonstrate how diagnostic tools can identify granular weaknesses, such as grammatical accuracy or lexical depth, allowing for targeted interventions. Minakova (2020) highlights how dynamic assessment (DA) for IELTS speaking preparation promotes learners' ability to self-correct errors and enhances their use of complex linguistic structures like tense-aspect markers. DA, through systematic feedback, not only supports immediate test performance but also fosters transferable skills for broader language use.

Additionally, diagnostic assessments facilitate greater learner autonomy by providing actionable feedback that empowers students to take ownership of their learning journey. Souzandehfar (2024) underscores the importance of fostering metacognitive awareness through diagnostic feedback, helping learners reflect on their performance and set achievable goals for improvement. Csapó and Molnár (2019) demonstrate how technology-enhanced diagnostic systems make cognitive processes more visible, enabling learners to monitor progress and adapt strategies effectively. These findings align with the broader emphasis on formative assessment as a tool for promoting self-regulated learning and long-term skill development.

In the context of high-stakes tests like IELTS, the integration of diagnostic elements is particularly critical. Diagnostic frameworks bridge the gap between assessment and instruction, fostering a more learner-centered approach that prioritizes skill development alongside performance evaluation (Gebril, 2022). For instance, Nakatsuhara et al. (2017) demonstrate how systematic diagnostic assessments within speaking tests can provide detailed performance insights, enabling candidates to focus on areas requiring improvement. This alignment of diagnostic feedback with pedagogical objectives represents a transformative shift in language assessment practices.

While the IELTS speaking assessment has contributed substantially to language testing, its limitations highlight the need for innovation. Exploring diagnostic methodologies addresses these challenges, ensuring assessments evaluate proficiency while supporting learners in achieving communicative potential. Therefore, it could be argued that diagnostic frameworks complement traditional assessments, paving the way for comprehensive and equitable language testing practices.

Cognitive Diagnostic Assessment Implementation in China

Cognitive diagnostic assessment (CDA) has emerged as a transformative approach in language teaching, offering an alternative to traditional assessments by providing fine-grained diagnostic feedback. Grounded in cognitive psychology and psychometric theory, CDA allows educators to identify learners' strengths and weaknesses at a granular level, facilitating targeted instruction. Its application has evolved significantly, with early studies focusing on receptive skills like reading and listening (He et al., 2022; Min & He, 2022). However, in high-stakes speaking assessments such as IELTS, CDA remains underutilized, particularly in contexts like China, where the education system predominantly relies on summative evaluations.

The IELTS speaking test, while widely recognized for its reliability and validity, has long been criticized for its inability to provide detailed diagnostic feedback. Holistic scoring frameworks, as currently employed, obscure test-takers specific strengths and weaknesses, making it difficult for learners to focus on areas that require improvement (Heidari Vincheh et al., 2024). This limitation is especially problematic in China, where English learners face unique linguistic and cognitive challenges. Despite the growing demand for diagnostic assessments, few studies have systematically explored the application of CDA to speaking skills in Chinese educational contexts, leaving a critical gap in both research and practice.

A comprehensive CDA framework for IELTS speaking, proposed by Heidari Vincheh et al. (2024), could be used as a promising means to address these gaps within the EFL context. This framework identifies five subskills essential for speaking proficiency: linguistic knowledge, syntactic knowledge, pronunciation mastery, discourse management and fluency, and speech structure awareness. Each of these subskills represents a critical component of oral communication, offering a roadmap for diagnosing and developing speaking abilities in a structured and systematic manner. For instance, linguistic knowledge encompasses vocabulary and

grammatical accuracy, foundational elements for constructing meaningful sentences. Syntactic knowledge focuses on sentence structure, enabling learners to produce more complex and coherent speech. Pronunciation mastery addresses phonological accuracy, including stress, rhythm, and intonation, which are particularly challenging for Chinese learners due to the phonemic differences between English and Mandarin (Hu, 2019). Discourse management and fluency involve organizing ideas logically and maintaining a natural flow of speech, while speech structure awareness emphasizes the importance of cohesive and coherent responses.

It should be noted that the implementation of this framework in Chinese high schools presents both opportunities and challenges. One on hand, Chinese learners' emphasis on systematic learning aligns well with the structured approach of CDA. For example, the focus on speech structure awareness resonates with Chinese students' preference for clear frameworks and methodologies in learning. However, on the other hand, traditional teaching practices, which often prioritize rote memorization over interactive and communicative approaches, may hinder the development of subskills like discourse management and fluency (Tiu et al., 2023). EFL teachers in China often rely on prescriptive methods that emphasize correctness, which, while useful for building grammatical accuracy, can limit learners' ability to express ideas spontaneously and interactively.

Cultural and institutional factors further shape the implementation of CDA in Chinese educational settings. The hierarchical nature of teacher-student relationships and the cultural emphasis on face-saving can influence how diagnostic feedback is delivered and received. Teachers may hesitate to provide detailed feedback for fear of embarrassing students, while students may prefer indirect feedback that highlights their strengths rather than their weaknesses (Bui, 2024; Gao, 2021; Liu & Xu 2017). These dynamics necessitate culturally sensitive approaches to feedback, such as incorporating peer assessments or anonymous evaluations, to create a supportive environment for skill development.

Moreover, interactional competence, as highlighted by Lam et al. (2023), is an essential component of speaking proficiency that requires more attention within diagnostic frameworks. Interactional competence involves managing interactions effectively during speaking tasks, such as turn-taking, negotiating meaning, and responding appropriately to interlocutors. These aspects are particularly relevant in IELTS speaking tasks but are often overlooked in traditional assessments. Therefore, integrating diagnostic tools to address interactional competence can help Chinese learners develop the pragmatic and sociolinguistic skills necessary for real-world communication.

Despite these challenges, there are more significant opportunities for integrating CDA into Chinese high schools. Advances in technology-based assessments offer efficient ways to track and analyze cognitive skill development (Shi et al., 2024). These tools can be particularly valuable in large classrooms in China, where individualized diagnostic assessments are logistically challenging. Additionally, Shi et al. (2024) demonstrated the use of polytomous cognitive diagnostic models (CDMs) to diagnose Chinese EFL learners' writing performance. The study

highlighted how CDMs provide a granular understanding of learners' strengths and weaknesses, which could be adapted to speaking skills. These tools allow for the synthesis of diagnostic information into actionable feedback, further supporting personalized instruction. Moreover, case studies from other domains, such as the successful implementation of CDA in reading assessments in Changchun City, China, demonstrate the feasibility of adapting international frameworks to local contexts (Li et al., 2021). All in all, these experiences highlight the importance of aligning diagnostic assessments with curriculum goals, teacher training, and the use of localized materials to ensure cultural relevance.

Another essential factor in implementing CDA effectively is the adaptability of teachers and students to new assessment practices. Chinese EFL teachers have shown a growing willingness to adopt innovative methodologies, provided they align with both national educational standards and the practical realities of their classrooms (Mei & Chen, 2024). For instance, incorporating the CDA framework into existing preparatory programs for IELTS can provide students with actionable insights into their speaking abilities, enabling them to focus on specific areas for improvement. This approach not only enhances the effectiveness of test preparation but also supports broader language learning objectives.

Thus, the integration of Heidari Vincheh et al.'s CDA framework into Chinese high schools offers a promising pathway for addressing the limitations of traditional assessments in IELTS speaking. By providing detailed feedback across key subskills, this framework can bridge the gap between assessment and instruction, fostering a more learner-centered approach to language education. Moreover, it has the potential to transform how speaking skills are taught and evaluated, promoting a deeper understanding of learners' needs and empowering them to achieve their full communicative potential.

Future research should focus on longitudinal studies to evaluate the impact of CDA on learners' performance and perceptions in diverse Chinese contexts. Additionally, exploring the interplay between cultural factors and diagnostic feedback processes can provide valuable insights into how CDA can be effectively adapted to different educational settings. Through such efforts, the CDA framework can not only enhance the teaching and assessment of IELTS speaking skills but also contribute to the broader goal of improving English language education in China.

Methodology

This qualitative case study employed an interpretive research paradigm to investigate the implementation and impact of a Cognitive Diagnostic Assessment (CDA) framework on IELTS speaking skills development among Chinese high school EFL learners. The study utilized multiple data collection methods including semi-structured interviews, classroom observations, mock test recordings, and document analysis to capture the complexity of the intervention across 16 weeks. This multi-method approach enabled a comprehensive examination of student learning processes, teacher implementation strategies, and the cultural adaptations required for the successful integration of diagnostic assessment principles within the Chinese educational context.

Participants

This study employs a qualitative case study methodology to explore the implementation and impact of a Cognitive Diagnostic Assessment (CDA) framework on IELTS speaking skills development among Chinese high school EFL learners. This approach aligns with understanding complex educational phenomena in their natural settings (Creswell, 2013) and facilitates an in-depth examination of how the CDA framework influences speaking skill development across dimensions identified in the literature review. This qualitative approach also aligns with recent research demonstrating the effectiveness of in-depth exploratory methods in understanding learner experiences and educational interventions in language learning contexts (Abdulaal et al., 2022; Pang et al., 2024).

The qualitative research was conducted at a private international high school in southwest China, selected for its emphasis on preparing students for overseas higher education. Five Grade 11 students at the A2 CEFR proficiency level were purposefully selected through criterion-based sampling to ensure participants possessed the specific characteristics essential for investigating the CDA framework implementation. These students were preparing for undergraduate studies in Australia and represented a homogeneous group regarding their language proficiency needs. The study also involved an experienced Chinese EFL teacher with a Master's degree in TESOL from the UK and five years of IELTS training experience, selected for his expertise in assessment practices.

The purposeful sampling strategy was directly aligned with the research questions examining cognitive diagnostic assessment in IELTS preparation. As Palinkas et al. (2015) emphasize, this approach is appropriate for bounded systems where the population shares critical characteristics relevant to the research focus. The sampling ensures both the student cohort perspective while incorporating expert pedagogical insight, enhancing the study's credibility. According to Etikan et al. (2016), studying educational interventions in small, specialized populations provides rich, in-depth insights, though findings remain context-specific rather than generalizable.

Table One presents the demographic information of the five student participants involved in this study.

Table 1. *Demographic information of student participants*

Participants	Age	Gender	Years of English study	Proficiency level (CEFR)
Student 1	Female	16	7	A2
Student 2	Female	17	7	A2
Student 3	Male	16	8	A2
Student 4	Male	16	7	A2
Student 5	Male	16	7	A2

Table two provides the demographic profile of the teacher participants who implemented the CDA framework throughout the intervention period.

Table 2. *Demographic information of the teacher participant*

Participant	Age	Gender	Qualification	Teaching experience	Role
Chinese teacher	30	Male	Master's degree in TESOL (UK)	5 years	IELTS trainer and CDA implementer

Research Instruments

This study employs multiple carefully selected instruments to collect comprehensive data on implementing the Cognitive Diagnostic Assessment (CDA) framework for IELTS speaking skills development in the Chinese EFL context. Each instrument was chosen to ensure academic rigor while providing nuanced insights into both learners' and teachers' experiences with the CDA framework.

Semi-Structured Interview Protocols

The primary instrument in this study is the semi-structured interviews. This method was chosen due to its effectiveness in eliciting in-depth qualitative data while allowing flexibility to explore unanticipated themes (Kvale & Brinkmann, 2015). Semi-structured interviews are recognized for balancing the structure needed to maintain focus with openness that encourages rich, descriptive responses (Creswell & Poth, 2018).

The interview protocols underwent rigorous development through a systematic three-phase process. Initial drafts were grounded in the research questions mentioned above, followed by expert review from three TESOL professionals for content validation and clarity. The protocols were then refined through pilot testing with two students outside the main study cohort, allowing for adjustments in question sequencing and phrasing. The final protocols incorporated both structural and elaboration probes to encourage detailed responses while maintaining consistency across sessions. To capture distinct perspectives, separate protocols were developed for students and the EFL teacher, with each focusing on their unique experiences with the CDA framework. Interviews were conducted bi-weekly over 16 weeks (one academic semester), enabling the documentation of evolving perspectives and experiences throughout the implementation period. For the students, questions were designed to explore their perceptions of the feedback received and its impact on their speaking skills development. Sample questions include:

1. How do you feel the diagnostic feedback helps you improve specific aspects of your speaking skills?
2. What do you find most challenging about the speaking tasks in the CDA framework?
3. Can you provide an example of how the feedback has influenced your practice strategies?

For the EFL teacher, the questions focused on the instructional strategies and the teacher's experience with implementing the CDA framework. Examples include:

1. What challenges have you faced in integrating CDA into your teaching practices?
2. How do you think the CDA framework affects your students' learning progress?
3. Can you describe the feedback process and its influence on your teaching approach?

Classroom Observation Checklists

Classroom observations were conducted weekly using a structured checklist. Observations are widely used in qualitative studies to gather contextual data and triangulate findings (Merriam & Tisdell, 2016). The checklist development followed a systematic process grounded in educational research methodology. The final instrument included predefined categories examining student engagement through participation patterns, instructional strategies focusing on CDA implementation, and the integration of diagnostic feedback into classroom practice.

The observation protocol underwent multiple iterations of refinement, with each category carefully aligned to the research questions and CDA framework principles. This systematic approach to classroom observation provided real-time insights into the interaction between teaching and learning dynamics, particularly focusing on how diagnostic feedback influenced instructional practices and student responses.

Mock Speaking Test Recordings

Mock speaking test performances were assessed using a comprehensive measurement protocol that integrated the five cognitive subskills from the CDA framework with standard IELTS band descriptors. The assessment instrument consisted of a detailed analytical rubric with 25 performance indicators (five indicators per subskill) rated on a 1-9 scale. For linguistic knowledge, indicators included lexical precision, vocabulary range, idiomatic expression, collocation accuracy, and word formation skills. Syntactic complexity was measured through sentence variety, grammatical range, tense consistency, conditional structure usage, and subordinate clause implementation. The pronunciation mastery assessment focused on phonemic accuracy, stress patterns, intonation contours, connected speech features, and fluency markers. Discourse management evaluation examined coherence devices, topic maintenance, turn-taking strategies, self-correction mechanisms, and information organization. Speech structure awareness was assessed through introduction formulation, thesis development, supporting evidence provision, transitional element usage, and conclusion strategies. Inter-rater reliability was established by having both the EFL teacher and an external IELTS examiner independently score a sample of recordings (20%), yielding a Cohen's kappa coefficient of 0.87. Quantitative improvements were calculated as percentage increases from baseline to final assessments, with comprehensive score profiles documented for each student across all five subskills.

Audio recordings of mock IELTS speaking tests were adopted to track the student participants' progress in specific linguistic subskills, derived from the CDA framework. These subskills included linguistic knowledge, syntactic knowledge, pronunciation mastery, discourse management and fluency, and speech structure awareness (Heidari Vincheh et al., 2024). These recordings provided critical data for analyzing the effectiveness of the CDA framework in addressing targeted areas of speaking performance. This approach aligns with research emphasizing the importance of recording natural speech to analyze performance development over time (Nakatsuhara et al., 2017). Recordings were systematically reviewed to examine areas such as pronunciation accuracy, fluency, and discourse management, complementing findings from the diagnostic feedback.

Document Analysis

Document analysis involved reviewing lesson plans, diagnostic profiles, and feedback materials from the Chinese EFL teacher to examine how the CDA framework was operationalized. This method is commonly used in qualitative research to provide additional layers of evidence (Bowen, 2009). The analysis employed a systematic coding scheme developed specifically for this study, focusing on the alignment between diagnostic feedback and targeted subskills, integration of CDA principles into lesson planning, progression of feedback strategies, and modifications in teaching approaches based on diagnostic results.

The document analysis protocol ensured consistent examination of materials across the 16-week study period, enabling the researchers to track changes in instructional planning and feedback delivery. This systematic approach to document analysis provided valuable insights into how the CDA framework principles were translated into practical teaching strategies and how these strategies evolved in response to student needs and progress.

This comprehensive set of instruments, each carefully designed and implemented, ensured thorough data collection while maintaining methodological rigor. The combination of interviews, observations, recordings, and document analysis enabled robust triangulation of findings, strengthening the validity of the study's conclusions about the effectiveness of the CDA framework in developing IELTS speaking skills among Chinese high school EFL learners.

Research Procedures

This study followed a systematic approach to collect and analyze data on the implementation of the Cognitive Diagnostic Assessment (CDA) framework for IELTS speaking skills development among Chinese high school students. The procedures were meticulously designed to ensure academic rigor and to address the research questions accordingly.

The data collection was conducted over a 16-week period, spanning one academic semester. The 16-week implementation period was selected based on empirical and practical considerations. This timeframe aligns with research indicating that meaningful changes in speaking proficiency require 12-15 weeks of sustained intervention (Nakatsuhara et al., 2017). Corresponding with one academic semester in the Chinese educational system, the CDA framework implementation followed a structured three-phase protocol across the 16-week intervention. In the initial implementation phase (weeks 1-4), the EFL teacher conducted baseline diagnostic assessments using standardized IELTS speaking tasks across all three sections of the test. Students' performances were video-recorded and analyzed using the five-dimension cognitive diagnostic rubric adapted from Heidari Vincheh et al. (2024), with each subskill scored on a scale of 1-9 corresponding to IELTS band descriptors. The teacher then developed individualized diagnostic profiles for each student, highlighting specific strengths and weaknesses across the five cognitive subskills. During the core intervention phase (weeks 5-12), weekly 120-minute sessions integrated explicit instruction on each subskill with diagnostic feedback cycles. Each session typically comprised: (1) a 30-minute mini-lesson targeting a specific cognitive subskill; (2) a 40-minute practice component using authentic IELTS speaking tasks; (3) a 30-minute diagnostic

feedback session where students received individualized feedback on their performance using the diagnostic rubric; and (4) a 20-minute guided self-reflection activity where students documented their understanding of the feedback and planned targeted practice strategies. The final evaluation phase (weeks 13-16) focused on the integration of all subskills, with comprehensive mock tests administered in weeks 13 and 16 to assess overall progress.

What's more, the implementation of the CDA framework incorporated systematic cultural adaptations to enhance its effectiveness within the Chinese EFL context. These adaptations were informed by Hofstede's cultural dimensions theory (1986) and Jin's (2023) cultural learning approaches. Four primary modifications were implemented: (1) Feedback delivery was adjusted to incorporate indirect communication strategies aligned with face-saving cultural practices, such as using group-level feedback for common errors before addressing individual challenges; (2) Diagnostic profiles were framed within achievement-oriented terminology rather than deficit-based language, emphasizing progress potential rather than performance gaps; (3) Structured rubrics were developed with explicit criteria and exemplars to accommodate Chinese learners' preference for clear assessment frameworks; and (4) Metacognitive reflection activities were gradually introduced, beginning with teacher-guided protocols before transitioning to more autonomous reflection practices. These adaptations were systematically documented through teacher reflection logs and modified based on student responses throughout the implementation period.

Semi-structured interviews served as the primary instrument for gathering rich qualitative data about the CDA framework implementation. Following established qualitative principles (Kvale & Brinkmann, 2015), interviews were conducted bi-weekly over the 16-week period, resulting in eight interview sessions with each of the five Grade 11 students at A2 CEFR proficiency level and eight separate sessions with the experienced EFL teacher with IELTS training expertise. The interview protocols underwent rigorous development through expert review and pilot testing with two students outside the main study cohort, ensuring both validity and cultural appropriateness within the Chinese educational context. For student participants, questions explored their experiences with diagnostic feedback across five key cognitive subskills identified by Heidari Vincheh et al. (2024): linguistic knowledge, syntactic knowledge, pronunciation mastery, discourse management and fluency, and speech structure awareness. The protocols specifically examined how students processed and implemented feedback in relation to IELTS speaking task requirements. For the EFL teacher, interviews investigated instructional adaptations, implementation challenges within the Chinese high school setting, and perceptions of the framework's effectiveness in addressing students' specific linguistic and cognitive needs. Each interview session, lasting approximately 45-60 minutes, was conducted in a private setting to ensure confidentiality and encourage open dialogue. All interviews were audio-recorded with participant consent and transcribed verbatim for detailed analysis. The systematic documentation process included both immediate post-interview notes and reflective memos to capture contextual details and emerging insights.

To supplement the interviews, weekly classroom observations were conducted to document the integration of the CDA framework into regular instruction. Observations were guided by a structured checklist, which included categories such as student engagement, instructional strategies, and the interaction between diagnostic feedback and classroom practices. These observations also noted the emphasis on the five cognitive subskills as part of the instructional process, providing valuable contextual data and triangulating the findings from the interviews (Merriam & Tisdell, 2016). In addition to interviews and observations, mock IELTS speaking tests were administered periodically. The audio recordings from these tests were analyzed to track the participant's progress in the five key cognitive subskills: linguistic knowledge, syntactic knowledge, pronunciation mastery, discourse management and fluency, and speech structure awareness. This approach follows best practices in performance-based assessments (Nakatsuhara et al., 2017), providing longitudinal insights into the development of speaking proficiency. Document analysis also formed a key part of the data collection process. Lesson plans, diagnostic profiles, and feedback materials from the EFL teacher were examined to understand how the CDA framework was operationalized and aligned with instructional goals (Bowen, 2009). This analysis emphasized the inclusion of the five cognitive subskills as focal points in both planning and feedback processes, further substantiating the study's findings.

This study adhered to rigorous ethical protocols to protect participant rights and ensure research integrity. Written informed consent was obtained from all student participants, with additional parental/guardian consent secured for those under 18 years of age. The consent forms, available in both English and Mandarin, detailed the study's purpose, data collection methods, confidentiality measures, and participants' right to withdraw without consequences. To protect confidentiality, all data were stored in encrypted, password-protected files, with participants assigned pseudonyms in all documentation. Mock test recordings were securely stored on dedicated devices without internet connectivity. The teacher participant was fully briefed on all aspects of the research and actively involved in implementation decisions, ensuring professional agency throughout the process. Regular debriefing sessions were conducted to address emerging ethical considerations, particularly regarding the balance between research objectives and pedagogical responsibilities. Institutional approval was obtained from both the international school's administration and the researchers' university ethics committee. All these protocols aligned with best practices in qualitative research ethics (Creswell & Poth, 2018).

For data analysis, this research employed a systematic analytical approach following Braun and Clarke's (2006) six-phase thematic analysis framework. The thematic analysis process utilized a rigorous coding framework informed by both deductive and inductive approaches. First, researchers independently conducted multiple close readings of all transcripts and documentary evidence, generating initial codes. For calibration purposes, three researchers independently coded 20% of the dataset, then met to reconcile discrepancies and establish preliminary code definitions. This iterative process yielded a structured codebook containing 47 codes organized into six categories: Cognitive Subskill Development, Diagnostic Feedback Reception, Metacognitive

Processing Strategies, Implementation Challenges, Cultural Adaptations, and Pedagogical Modifications. Each code was defined using explicit criteria, example indicators, and exclusion parameters. The codebook was then validated through independent coding of another 20% of the data by two additional researchers, resulting in a Cohen's kappa coefficient of 0.87, indicating strong inter-rater reliability. For the mock test recordings analysis, a specialized analytical matrix was developed to document frequency counts of specific speaking features across the five cognitive subskills. Linguistic knowledge improvements were quantified by tracking advanced vocabulary usage rates, while syntactic complexity was measured by calculating the proportion of compound and complex sentences per response. Pronunciation accuracy was evaluated by documenting error reduction rates across specific phonemic categories. Discourse management progress was assessed through frequency analysis of cohesive device usage and topic maintenance markers, while speech structure improvement was determined through comparative analysis of organizational patterns against standardized templates for each speaking task type.

These practices, combined with systematic reflexive journaling, enhanced the trustworthiness of findings while providing an audit trail of analytical decisions (Creswell & Miller, 2000). This framework provided a robust foundation for examining CDA implementation outcomes, particularly regarding cognitive subskills critical to IELTS speaking performance among Chinese high school students.

Results

The findings address three research questions of the study: (1) the impact of CDA framework implementation on speaking subskill development among Chinese high school students; (2) the cognitive processes employed by Chinese EFL learners when engaging with diagnostic feedback; and (3) the EFL teacher's implementation of the CDA framework, including challenges and modifications. The results draw from multiple data sources collected over 16 weeks, including semi-structured interviews, classroom observations, mock test recordings, and document analysis.

Impact of CDA Framework on Specific Speaking Subskills Development

Overview of Subskill Development

The comprehensive analyses across multiple data sources revealed significant improvements in students' speaking proficiency through the CDA framework implementation. Mock test recordings demonstrated substantial gains across all five targeted subskills over the 16-week period: linguistic knowledge (40% increase in advanced vocabulary usage), syntactic complexity (35% improvement in complex sentence structures), pronunciation accuracy (45% reduction in phonemic errors), discourse management (50% enhancement in cohesive device usage), and speech structure organization (average improvement of 1.5 IELTS bands).

These improvements manifested in observable changes in students' speaking performance. Student 3 articulated this development in vocabulary sophistication:

"Before, I would just use simple words, but now I can express more complex ideas using words like substantial and feasible."

In the meantime, the EFL teacher corroborated this observation through classroom interaction notes:

"Students have become more conscious of word choice. They're not just replacing simple words with complex ones; they're learning to use vocabulary that precisely conveys their intended meaning."

Figure One below illustrates the progression of improvement across all five subskills, highlighting both initial and final assessment scores. The graph demonstrates consistent upward trends, with particularly notable gains in pronunciation mastery and discourse management. These improvements were reflected in both mock test performances and classroom interactions, as documented through the teacher observations and student self-reports. Appendix A provides a detailed summary of key improvements, evidence sources, and representative quotes for each subskill.

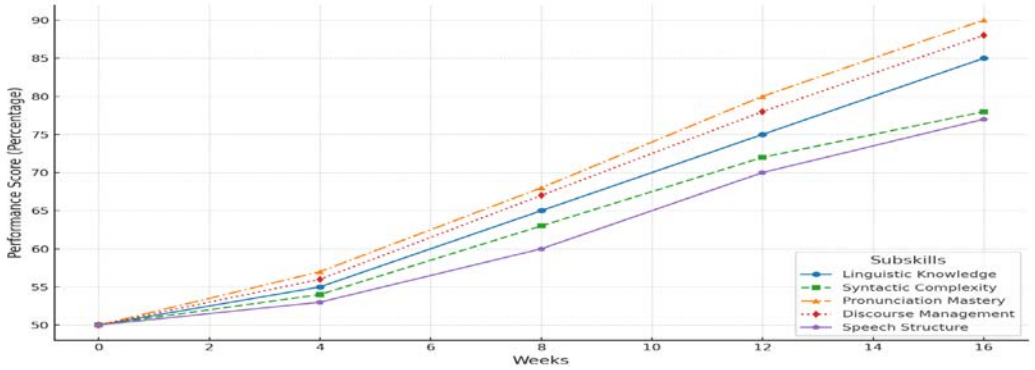


Figure 1. Development of speaking skills over 16-week period of the study

Detailed Analysis by Subskill

Linguistic Knowledge

The implementation of CDA revealed significant improvements in students' linguistic competence. Document analysis of mock test transcripts showed a 40% increase in the use of advanced vocabulary and idiomatic expressions over the 16-week period. As Student 3 reflected:

"The diagnostic feedback helped me identify overused words and phrases. Now I can express myself more precisely. For example, instead of just saying 'important', I can use words like 'crucial', 'essential,' or 'fundamental' depending on the context."

Teacher observation notes corroborated this development, documenting the increased frequency of academic vocabulary usage in classroom discussions. The EFL teacher noted:

"Students have become more conscious of word choice. They're not just replacing simple words with complex ones; they're learning to use vocabulary that precisely conveys their intended meaning."

ISSN: 2229-9327

Syntactic Knowledge

Analysis of mock test recordings demonstrated notable improvement in syntactic complexity. The frequency of compound and complex sentences increased by 35% from baseline assessments. Student 2 articulated this development:

"Previously, I would say 'Technology is good. It helps people. But it has problems.' Now I can say 'While technology offers numerous benefits to society, its rapid advancement also presents significant challenges that need careful consideration.' "

Pronunciation Mastery

Mock test recordings revealed substantial improvement in pronunciation accuracy, particularly in areas traditionally challenging for Chinese learners. Quantitative analysis of recordings showed a 45% reduction in phonemic errors related to sounds absent in Mandarin Chinese, such as /θ/ and /ð/. Student 4's reflection captured this development:

"The individualized feedback on my pronunciation helped me identify specific sound patterns I was struggling with. For instance, I learned to distinguish between 'think and 'sink' by practicing the tongue position the teacher showed me. Now I feel more confident with these sounds."

Classroom observation notes documented increased student willingness to attempt challenging pronunciations. The EFL teacher observed this change in confidence:

"Initially, students would avoid words containing difficult sounds. Now they actively seek opportunities to practice these pronunciations, showing remarkable improvement, particularly in stress patterns and intonation."

Figure Two illustrates the growing trends of improved pronunciation for students. In particular, confidence, stress patterns, phonemic awareness, and intonation emerged as the most impactful elements.

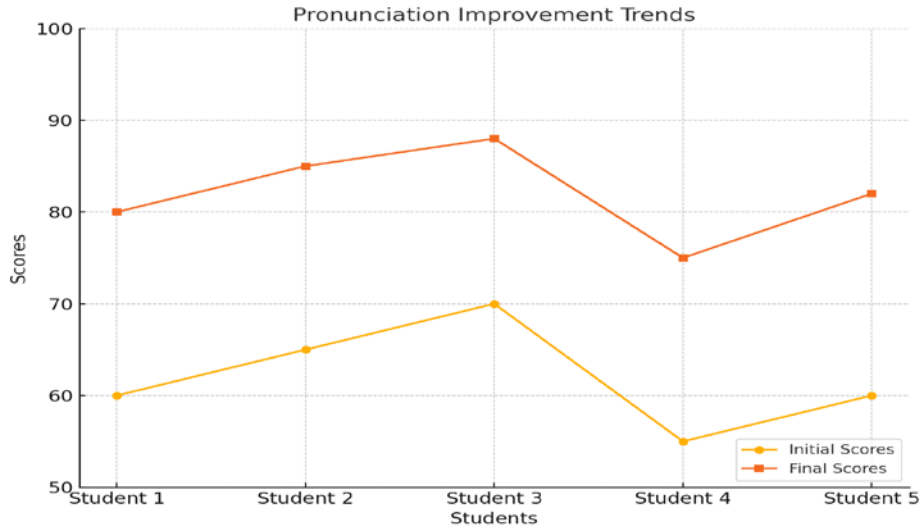


Figure 2. The growing trend of the improved pronunciation

ISSN: 2229-9327

Discourse Management and Fluency

Analysis of mock test transcripts demonstrated significant enhancement in discourse management capabilities. Students showed a 50% increase in the appropriate use of discourse markers and cohesive devices. Student 5's mock test responses exemplified this improvement:

"In my early notes, I would just list ideas without connection. Now I can say, 'While traditional education has its merits, there are compelling reasons to consider modern alternatives. Firstly... Additionally... Furthermore...' It helps me express my thoughts more clearly."

The teacher's assessment logs noted particular improvement in students' ability to maintain coherence during extended responses. As documented in observation notes:

"Students now demonstrate greater facility in connecting ideas and maintaining topic relevance throughout their responses. Their improved use of discourse markers has enhanced both the coherence and overall fluency of their speech."

Speech Structure Awareness

Document analysis of lesson plans and student performance records revealed systematic improvement in speech organization. Mock test recordings showed that by week 12, all five students consistently incorporated clear introductions, structured arguments, and coherent conclusions in their responses. Student 1 articulated this development:

"The CDA framework taught me to organize my thoughts before speaking. Now I always start with a clear position, support it with two or three main points, and end with a strong conclusion. This structure helps me feel more confident and makes my answers more convincing."

The teacher's reflective journal highlighted this improvement:

"The most striking development has been students' ability to structure their responses. They've moved from random, disconnected thoughts to well-organized arguments with clear progression and logical conclusion."

Data from mock test evaluations showed that student's scores in the 'coherence and cohesion' criterion of the IELTS speaking band descriptors improved by an average of 1.5 bands over the study period. This improvement was particularly evident in Part 2 (long turn) responses, where structured organization is crucial. Student 2 reflected on this progress:

"Before the CDA training, I would panic during the two-minute talk. Now I have a mental template: introduction, key points, examples, and conclusion. This helps me organize my thoughts quickly and speak more confidently."

Cognitive Processes in Engaging with Diagnostic Feedback

Analysis of interview data, classroom observations, and student reflection logs revealed complex patterns in how students processed and implemented diagnostic feedback throughout the 16-week study period. The cognitive engagement with feedback demonstrated a cyclical pattern of reception, processing, and implementation, with each phase building upon previous learning experiences.

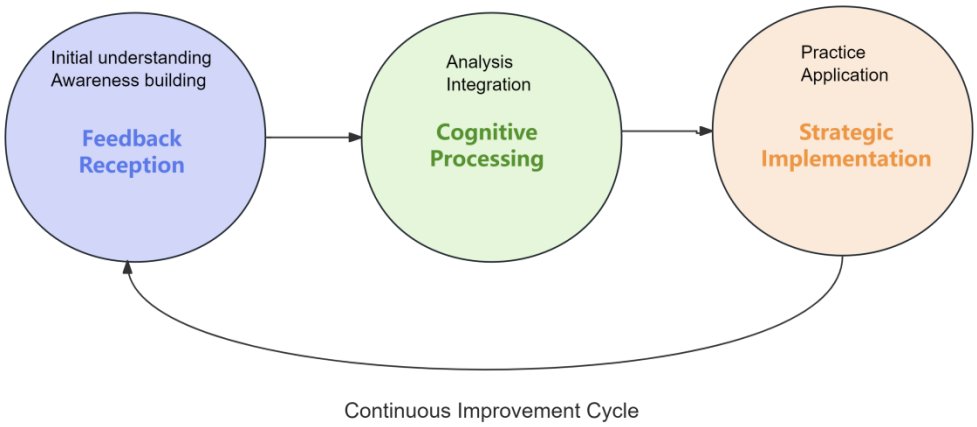


Figure 3. Student cognitive processing of diagnostic feedback

Figure Three showcases the interconnected nature of students' cognitive processing, indicating how feedback reception leads to cognitive analysis and subsequent implementation in a continuous improvement cycle. The diagram captures the recursive nature of learning observed in the study, where implementation experiences generated new insights that influenced how subsequent feedback was received and processed.

Student Engagement Patterns

Interview data revealed three distinct patterns in how students processed diagnostic feedback, each characterized by specific cognitive strategies and implementation behaviors. The first pattern, Immediate Application, was predominantly observed in the early stages of the study period. Students exhibiting this pattern would attempt to implement feedback directly without significant analytical processing. As Student 4 explained:

"Initially, when the teacher pointed out my pronunciation errors, I would just repeat the correct pronunciation immediately without understanding why I made the mistake in the first place."

The second pattern, Reflective Processing, emerged as students developed more sophisticated approaches to feedback engagement. This pattern involved deliberate analysis before implementation attempts. Student 1's reflection exemplified this approach:

"I started keeping a feedback journal where I would write down the teacher's comments and think about why certain aspects of my speaking needed improvement. This helped me understand the underlying patterns in my mistakes."

The third pattern, Integrated Adaptation, represented the most advanced level of feedback engagement. Students demonstrating this pattern showed an ability to synthesize multiple pieces of feedback and adapt them across different speaking contexts. The EFL teacher observed this development:

"By week 12, several students were able to take feedback from one speaking task and apply it creatively to different scenarios, showing a deep understanding of the principles behind the feedback."

Cognitive Processing Strategies

Analysis of classroom observations and interview data revealed specific cognitive strategies students employed when processing diagnostic feedback. Appendix B presents a summary of these strategies and their observed outcomes.

Evolution of Feedback Processing

The current study revealed a clear progression in how students processed and utilized diagnostic feedback over time. During the first four weeks, most students exhibited predominantly surface-level engagement with feedback, focusing on immediate corrections without deeper analysis. Student 2 reflected on this early phase:

"In the beginning, I was just trying to fix whatever the teacher pointed out without really thinking about the bigger picture."

By weeks 8-12, significant shifts in cognitive processing were observed. Students began developing more sophisticated strategies for engaging with feedback, as evidenced by their reflection logs and classroom participation. Student 5 described this transition:

"I started seeing connections between different pieces of feedback. For example, I realized that my pronunciation issues were often related to stress patterns, not just individual sounds."

The final weeks of the study (weeks 13-16) showed the emergence of advanced cognitive processing strategies. Students demonstrated the ability to:

- Anticipate potential areas for improvement before receiving feedback
- Connect feedback across different speaking tasks and contexts
- Develop personalized strategies for implementing feedback effectively

The EFL teacher noted this progression as well:

"The most significant change was in how students approached feedback. They moved from passive reception to active engagement, often initiating discussions about how to apply feedback in different situations."

Impact on Speaking Performance

The evolution in cognitive processing strategies showed a direct correlation with improvements in speaking performance. Analysis of mock test recordings revealed that students who demonstrated more sophisticated feedback processing patterns achieved greater improvements in their speaking scores. As Student 1 noted:

"Once I started really analyzing the feedback instead of just trying to implement it immediately, I saw much bigger improvements in my speaking ability."

Document analysis of teacher feedback logs supported this finding, showing that students who developed advanced cognitive processing strategies required progressively less explicit feedback to achieve improvements in their speaking performance. This suggests that enhanced cognitive processing of diagnostic feedback contributes to greater learner autonomy and more effective self-regulation in speaking skill development.

EFL teacher's Implementation of the CDA Framework

The analysis of teacher implementation data revealed three key dimensions of the CDA framework integration: pedagogical adaptations, implementation challenges, and strategic

responses. Document analysis of lesson plans, teaching logs, and interview transcripts demonstrated systematic development in the EFL teacher's approach throughout the 16-week period.

Pedagogical adaptations

The EFL teacher developed specific strategies to implement the CDA framework effectively in practice. Interview data revealed deliberate modification of instructional approaches:

"I restructured my feedback delivery to align with the five cognitive subskills, ensuring each student received targeted guidance in areas requiring development."

This systematic approach was evidenced in lesson planning documents, which showed explicit alignment between diagnostic outcomes and instructional interventions. The teacher's reflective journal documented the evolution of diagnostic practices:

"Initially, I focused on immediate corrections, but gradually developed more sophisticated diagnostic tools that helped identify patterns in students' cognitive processing of speaking tasks."

This progression was further supported by systematic documentation of student performance patterns and corresponding pedagogical responses.

Implementation Challenges and Solutions

Analysis of teacher interviews and observation data identified three primary implementation challenges:

- Differentiation of Feedback: "Balancing individualized diagnostic feedback with group instruction required careful planning and time management."
- Cultural Considerations: "Adapting feedback delivery to align with Chinese students' cultural preferences while maintaining diagnostic clarity was crucial."
- Resource Constraints: "Developing comprehensive diagnostic profiles within time constraints while maintaining instructional momentum presented ongoing challenges."

The teacher's strategic responses to these challenges demonstrated adaptive expertise, as evidenced in the modified assessment protocols and differentiated feedback mechanisms developed over the study period.

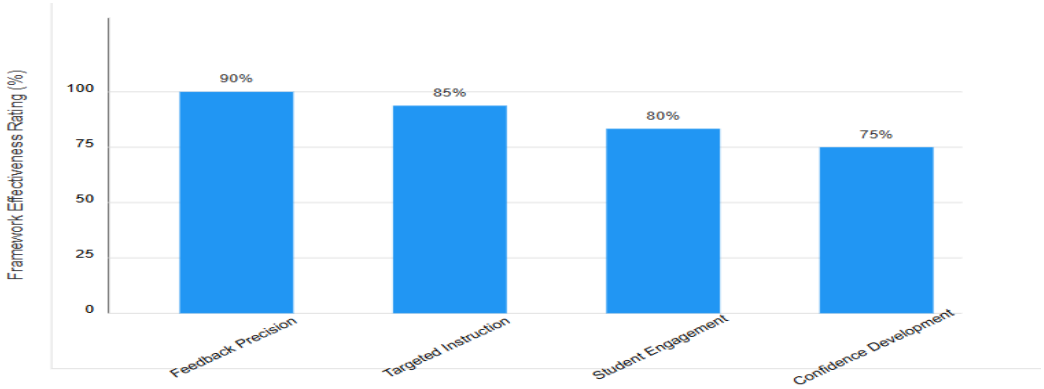


Figure 4. Teacher's perceptions of CDA framework effectiveness

Figure Four illustrates the teacher's evaluation of the CDA framework's effectiveness across key implementation dimensions. The data, derived from structured reflection logs and quantified assessment records, shows particularly strong outcomes in feedback precision (90%) and targeted instruction (85%), with notable gains also observed in student engagement (80%) and confidence development (75%).

Evidence-based Implementation Outcomes

Document analysis revealed systematic improvement in the teacher's diagnostic capabilities, evidenced by increasingly detailed and nuanced feedback records. The teacher noted:

"The framework provided a structured approach to identifying specific cognitive barriers in students' speaking development, enabling more precise instructional interventions."

Classroom observation data supported this finding, showing increased alignment between diagnostic assessments and subsequent teaching strategies. For example, pronunciation feedback evolved from generic correction to targeted analysis of cognitive processing patterns underlying specific pronunciation challenges.

Summary of Progress

Considering all the factors mentioned above, it is safe to say that the implementation of the CDA framework yielded significant improvements across multiple dimensions of speaking skill development. Appendix C presents a comprehensive summary of progress indicators across all five cognitive subskills, supported by multiple data sources.

The findings demonstrate that the CDA framework's implementation facilitated substantial improvements in students' speaking proficiency, particularly in areas traditionally challenging for Chinese EFL learners. The systematic integration of diagnostic assessment with targeted instruction provided a structured pathway for developing critical speaking subskills, addressing a significant gap in traditional EFL classroom practices.

Discussion

This qualitative case study investigated the implementation of a Cognitive Diagnostic Assessment (CDA) framework for developing IELTS speaking skills among Chinese high school EFL learners. The discussion synthesizes findings across three research questions, examining how the framework impacted speaking subskill development, cognitive processing strategies, and pedagogical implementation within the Chinese educational context.

RQ1: Development of Speaking Subskills

The first research question examined how CDA framework implementation impacts specific speaking subskill development among Chinese high school students. The findings revealed substantial improvements across five critical subskills: linguistic knowledge, syntactic complexity, pronunciation mastery, discourse management, and speech structure awareness. Linguistic knowledge showed marked advancement, with students progressing from basic vocabulary usage to sophisticated lexical choices. Students demonstrated increased grammatical accuracy and expanded vocabulary range, replacing generic terms with nuanced alternatives. For instance, learners evolved from using "important" to employing contextually appropriate variations such as "crucial" or "fundamental." This progression directly addresses the research question by demonstrating how systematic diagnostic feedback enables precise identification and

remediation of lexical weaknesses. These findings align with Jang's (2008) emphasis on explicit feedback's role in improving lexical choices. Heidari Vincheh et al. (2024) similarly found that targeted feedback fosters lexical sophistication. The CDA framework's detailed feedback reports encouraged students to systematically address specific weaknesses, ensuring consistent progress beyond what traditional instruction typically achieves.

Syntactic complexity exhibited parallel development, with students transitioning from simple to compound and complex sentence structures. The framework facilitated grammatical understanding through structured scaffolding that identified specific clause structure errors and prompted targeted practice. This finding supports Ellis's (2003) Input-Interaction Hypothesis, which emphasizes meaningful input's importance in promoting syntactic development. The ability to vary sentence types for rhetorical effect proved particularly valuable for IELTS speaking tasks, where sophisticated constructions contribute to higher band scores.

Pronunciation mastery, traditionally challenging for Chinese EFL learners, showed notable progress through systematic intervention. The framework addressed L1 phonological interference by focusing on problematic sounds, improving both phonemic accuracy and stress patterns. Students corrected persistent errors such as /r/ and /l/ confusion, while recordings revealed enhanced intonation patterns contributing to more natural speech production. This finding reinforces Wang and Wright's (2018) assertion that pronunciation difficulties stem from fundamental phonological differences between English and Mandarin. The improvement also aligns with Ma and Chong's (2022) findings that Chinese IELTS candidates strategically prioritize speaking skills due to perceived malleability and potential for rapid improvement. The CDA framework's structured approach appears to validate this perception while providing a systematic methodology for achieving pronunciation gains.

Discourse management and speech structure awareness demonstrated learners' enhanced ability to organize ideas coherently and articulate arguments effectively. Students developed a sophisticated understanding of introduction-body-conclusion formats and adapted responses for various task types. This development supports Mayer's (2014) Cognitive Load Theory, suggesting that clear instructional scaffolding reduces cognitive burden, enabling focus on higher-level organizational strategies.

RQ2:Cognitive Processing Patterns

The second research question explored cognitive processes Chinese EFL learners employ when engaging with diagnostic feedback during IELTS preparation. The findings revealed a developmental trajectory from surface-level feedback processing to sophisticated metacognitive strategies.

Initial engagement patterns showed students focusing on immediate corrections without deeper analysis. Learners repeated corrected phrases but struggled to transfer corrections across contexts. This mirrors Poehner and Wang's (2021) observations regarding initial diagnostic feedback integration stages. However, as the intervention progressed, students adopted increasingly reflective approaches characterized by meta-analytical review and contextual application. The evolution toward sophisticated processing strategies directly answers the research question by demonstrating how learners internalize diagnostic principles. Students progressed from mechanical repetition to analyzing error patterns, understanding root causes, and anticipating similar challenges. This transformation aligns with Zimmerman's (2002) Self-Regulated Learning Theory, highlighting how reflective practices foster autonomy and long-term improvement. The

framework encouraged students to understand error origins, similar to developmental patterns observed in hybrid computerized dynamic assessment studies (Silva Moreira et al., 2024). Feedback sessions prompted analysis of pronunciation and grammatical errors regarding L1 interference or task-specific demands, aligning with Jin and Liu's (2024) emphasis on contextualizing diagnostic feedback within learners' linguistic backgrounds. Advanced processing manifested as students began generalizing feedback principles across tasks, demonstrating what Poehner et al. (2015) term "cognitive modifiability." This higher-order cognitive process indicates genuine internalization of diagnostic principles. Students developed pattern recognition abilities, identifying recurring error types and preemptively addressing them during practice.

Despite progress, initial challenges emerged when transitioning from surface-level to deeper processing. Some students reported difficulty connecting specific feedback to broader patterns, echoing challenges noted in dynamic assessment literature (Poehner & Lantolf, 2022). The CDA framework addressed this through collaborative discussions where peers shared application strategies, fostering collective metacognitive growth consistent with Bakhoda and Shabani's (2019) findings on group-based dynamic assessment effectiveness.

The development of anticipatory error correction strategies represents sophisticated cognitive processing. Learners who mastered these processes reported increased confidence and preparation for varied speaking tasks, reflecting the profound shift from teacher-dependent feedback to self-regulated improvement. This aligns with Sun et al.'s (2023) assertions about dynamic assessment's relationship to learning potential and validates Csapó and Molnár's (2019) findings on metacognitive engagement's importance in diagnostic assessment.

RQ3:Teacher Implementation and Cultural Context

The third research question examined how the Chinese EFL teacher implemented the CDA framework in pedagogical practice and what challenges or modifications emerged. The findings revealed complex navigation between diagnostic assessment principles and cultural educational norms.

The teacher's implementation demonstrated sophisticated cultural calibration in feedback delivery methods. Recognizing students' face-saving preferences, the teacher employed indirect feedback strategies, anonymized error correction, and group discussions to minimize potential embarrassment. This adaptation directly addresses the research question by showing how cultural sensitivity shapes framework implementation without compromising diagnostic clarity. These modifications align with Hofstede's (1986) cultural dimensions theory emphasizing collectivist values in educational settings. Luo et al.'s (2020) argument that culturally calibrated feedback enhances learner receptivity found validation in increased student engagement when feedback honored face-saving concerns. Zhang (2021) similarly found indirect feedback strategies significantly increase engagement in high-context cultures, supporting the teacher's adaptive approach.

Examination-oriented educational traditions posed substantial implementation challenges. Students initially equated success with high scores rather than skill improvement, reflecting the dominance of summative assessment in Chinese contexts described by Huang et al. (2022). The teacher addressed this by explicitly linking diagnostic tasks to IELTS band score improvements, providing concrete examples of how subskill development translates to test performance. This framing reconciled diagnostic assessment with examination traditions, as Ismail et al. (2022) suggest when aligning formative and summative goals. The framework's structured nature

resonated with students' systematic learning preferences, addressing Tiu et al.'s (2023) findings on Chinese learners' methodological preferences. However, resource constraints emerged, particularly time requirements for individualized feedback with large classes. The teacher prioritized feedback for critical subskills and utilized technology for streamlined delivery, demonstrating pragmatic adaptation suggested by Ahmadi (2018) for scalable diagnostic practices.

Unexpected Outcomes

Beyond anticipated findings, several unexpected outcomes emerged, enriching the understanding of diagnostic assessment's broader impact. Peer feedback mechanisms developed organically as students internalized CDA principles and applied them collaboratively. This phenomenon, initially absent from the instructional design, created shared diagnostic language among learners. Souzandehfar's (2024) findings on diagnostic assessment's social dimensions support this collaborative feedback's role in fostering deeper cognitive engagement.

Substantial confidence increases emerged as an unanticipated byproduct. While not a primary framework focus, enhanced self-efficacy underscores diagnostic feedback's psychological benefits. Nakatsuhara et al. (2017) argue that understanding personal progress fosters self-efficacy and risk-taking willingness. Students reported increased participation in speaking activities both within and beyond classroom contexts, suggesting diagnostic assessment's motivational impact extends beyond skill development. Metacognitive awareness development further highlighted the framework's transformative potential. Through reflective practices, learners demonstrated increased self-regulation, recognizing and addressing weaknesses autonomously. This shift toward autonomous learning validates diagnostic frameworks' role in fostering lifelong language development capabilities.

Summary of Findings

The implementation of the CDA framework comprehensively addressed all three research questions, demonstrating significant enhancement of speaking subskills, promotion of sophisticated cognitive processing strategies, and successful cultural adaptation in pedagogical practice. These findings contribute to the theoretical understanding of diagnostic assessment in language education while providing practical insights for implementation in examination-oriented contexts. The framework's success in bridging formative assessment principles with high-stakes test preparation suggests broader applicability for similar educational environments where cultural sensitivity and examination success must coexist.

Implications

The findings of this study yield significant implications for pedagogical practices, policy development, and theoretical understanding of diagnostic assessment in language education. These implications extend beyond the immediate context of IELTS speaking preparation to broader applications in EFL instruction and assessment, offering substantial insights for various stakeholders in language education.

Pedagogical Implications

From a pedagogical perspective, the successful implementation of the CDA framework emphasizes the necessity of integrating diagnostic practices into IELTS preparation courses. Specifically, this study's findings showed notable improvements in cognitive subskills like

linguistic knowledge, syntactic complexity, and pronunciation mastery, underscoring the value of structured feedback tailored to these areas (Jin & Liu, 2024). For example, learners who received targeted pronunciation feedback demonstrated significant reductions in phonemic errors, indicating that diagnostic practices can address specific linguistic challenges effectively.

Teachers should adopt structured approaches to feedback delivery that explicitly align with subskill development. Practical strategies include using audio-visual tools to provide real-time corrective feedback on pronunciation and designing scaffolded speaking tasks to promote syntactic complexity. Moreover, integrating peer feedback sessions—guided by diagnostic rubrics—can enhance learners' ability to analyze and critique performance, as evidenced by the emergence of peer feedback mechanisms in this study. Such sessions not only reinforce diagnostic principles but also foster collaborative learning and shared accountability.

Teacher training programs are strongly recommended to incorporate modules focused on diagnostic assessment implementation, emphasizing strategies for creating diagnostic profiles, addressing large class sizes, and balancing individualized and group feedback. For example, training could include simulation activities where teachers practice delivering diagnostic feedback in diverse classroom settings, ensuring practical applicability (Zhorova et al., 2022).

Policy Recommendations

For policymakers and educational administrators, this study highlights the necessity of robust institutional support for implementing diagnostic assessment frameworks. As Csapó and Molnár (2019) highlight, effective diagnostic assessment requires substantial resource allocation, including dedicated time for teacher training, technology integration, and ongoing professional development. Institutions should prioritize investments in teacher readiness, such as training programs and scalable technologies like automated speech analysis platforms, which can complement teacher-led diagnostics in large classrooms by providing data-driven feedback.

Policies should integrate diagnostic elements into formative and summative systems to ensure continuous learner progress monitoring. For instance, mandating periodic diagnostic assessments during academic terms can generate actionable insights for instructional adjustments and targeted interventions (Ismail et al., 2022). Curriculum developers are highly recommended to embed diagnostic principles into IELTS preparation materials by incorporating features such as built-in diagnostic feedback prompts to guide learners in reflecting on their performance. Additionally, policies should address cultural influences to enhance framework effectiveness. For example, indirect feedback approaches, shown in this study to be effective in collectivist settings like China, should inform culturally responsive diagnostic designs (Cheng et al., 2021). Aligning diagnostic frameworks with cultural norms maximizes learner receptivity and fosters broader engagement.

Theoretical Contributions

This study contributes significantly to the theoretical understanding of diagnostic assessment by extending Heidari Vincheh et al.'s (2024) framework in the EFL context. Specifically, it highlights how diagnostic feedback fosters iterative cognitive engagement, enabling learners to move from surface-level corrections to reflective, autonomous adaptation. The findings on metacognitive awareness underscore the importance of integrating self-regulated learning theories, such as Zimmerman's (2002), into diagnostic frameworks to support long-term skill development.

The emergence of peer feedback mechanisms adds a social dimension to diagnostic theories, aligning with Vygotsky's (1978) principles of collaborative learning. This study demonstrates that learners internalize diagnostic principles through peer interactions, which can amplify cognitive engagement and foster shared responsibility for learning outcomes. However, future frameworks must account for the variability in peer feedback accuracy, emphasizing the need for teacher guidance to scaffold effective peer evaluation.

Cultural dimensions also emerge as critical to the effectiveness of diagnostic frameworks. In collectivist educational settings, group dynamics and indirect feedback approaches were found to significantly enhance learner engagement (Claramita et al., 2022). This underscores the need for diagnostic assessment theories to incorporate cultural adaptations explicitly, ensuring applicability across diverse contexts.

Finally, the study has implications for test developers, particularly those designing high-stakes assessments like IELTS. Incorporating diagnostic rubrics that align with cognitive subskills can enable precise identification of learner strengths and weaknesses while maintaining the reliability and validity of assessments (Nakatsuhara et al., 2017). Such innovations can bridge the gap between testing and instruction, providing actionable insights that benefit both educators and learners.

Conclusion

This qualitative case study aimed to investigate the implementation of a Cognitive Diagnostic Assessment (CDA) framework for enhancing IELTS speaking skills among Chinese high school EFL learners. The study examined how diagnostic feedback influenced learners' cognitive subskill development, their engagement with assessment data, and the cultural factors shaping implementation in the Chinese educational context. The findings revealed significant improvements across five cognitive subskills: linguistic knowledge, syntactic complexity, pronunciation mastery, discourse management, and speech structure awareness. Learners progressed from surface-level engagement with diagnostic feedback to employing sophisticated metacognitive strategies, demonstrating the framework's effectiveness in fostering self-regulated learning. The study also identified culturally specific adaptations, including the integration of traditional Chinese pedagogical values with diagnostic assessment principles, which enhanced learner receptiveness and engagement. This research contributes to the field by extending CDA frameworks to demonstrate how structured diagnostic feedback fosters metacognitive development in EFL contexts. The findings offer practical strategies for implementing culturally responsive diagnostic assessments in Chinese educational settings, with particular relevance for educators and test developers working in high-stakes examination environments. However, the study's small sample size and specific institutional context limit generalizability, pointing to the need for longitudinal research across diverse educational settings to validate these findings and explore the framework's sustained impact and broader applicability.

Funding: This research is not funded.
Acknowledgments: Not applicable.
Conflicts of Interest: The authors declare no conflict of interest.
Authenticity: This manuscript is an original work.
Artificial Intelligence Statement: AI and AI-assisted technologies were not used.

About the Authors:

Mengjia Liu is currently a Ph.D. candidate in TESL at Faculty of Education, Universiti Kebangsaan Malaysia. She holds an MA in Applied Linguistics and TESOL (Newcastle University, UK) and served as a university lecturer in China. Her research focuses on English speaking, writing, and language testing and assessment. ORCid ID: 0009-0002-2641-2639

Harwati Hashim, Ph.D, is an associate professor at the Centre for Innovation in Teaching and Learning, Faculty of Education, Universiti Kebangsaan Malaysia. Her areas of concentration are mobile learning, Mobile-assisted Language Learning (MALL), innovative pedagogy and technology in teaching and learning English as a Second Language (ESL). ORCid ID: 0000-0002-8817-427X

Nur Ainil Sulaiman, Ph.D, is a lecturer at the Faculty of Education, Universiti Kebangsaan Malaysia. Her research and teaching expertise includes second language acquisition, vocabulary development, and sociolinguistics within the broader domain of arts and applied arts education. ORCid ID: 0000-0001-6212-7494

References

Abdulaal, M.A.AD., Alenazi, M.H., Tajuddin, A.J.A. et al. Dynamic vs. diagnostic assessment: impacts on EFL learners' speaking fluency and accuracy, learning anxiety, and cognitive load. *Language Testing in Asia 12*, 32 (2022). https://doi.org/10.1186/s40468-022-00179-0

Ahmadi, D. M. R. (2018). The use of technology in English language learning: A literature review. *IJREE, 3*(2). Retrieved from http://ijreeonline.com/article-1-120-en.html

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.

Barghi, B., Afshinfar, J., Alavi, S. M., Jafarigohar, M., & Soleimani, H. (2025). An Exploratory Mixed Methods Study of Grammatical Range and Accuracy in IELTS: A True Diagnostic Approach to Cognitive Diagnostic Assessment. *International Journal of Language Testing, (3)*, 19-42. http://10.22034/ijlt.2024.481434.1374

Bogachevska, L. O. (2022). Passing of IELTS and preparation for its speaking section as an initial stage of students' integration in the global higher education area. *Transcarpathian Philological Studies, 1*(26), 87–92. Retrieved from http://zfs-journal.uzhnu.uz.ua/archive/26/part_1/17.pdf

Borger, L. (2019). Evaluating a high-stakes EFL speaking test: Teachers' practices and views. *Moderna Språk, 113*(1), 25–57. https://doi.org/10.58221/mosp.v113i1.7621

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal, 9*(2), 27–40. https://doi.org/10.3316/QRJ0902027

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative *Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brown, A. (2006a). Candidate discourse in the revised IELTS Speaking Test. *IELTS Research Reports* (Vol. 6, pp. 1–19). IELTS Australia and British Council. Retrieved from https://www.ielts.org/pdf/Volume%206,%20Report%203.pdf

Brown, A. (2006b). An examination of the rating process in the revised IELTS Speaking Test. *IELTS Research Reports* (Vol. 6, pp. 1–30). IELTS Australia and British Council. Retrieved from http://www.ielts.org/PDF/Vol6_Report2.pdf

Bui, H. P. (2024). Contextualizing synchronous online EFL classroom assessment: A face-saving perspective. *Education and Information Technologies, 29*, 24941–24961. https://doi.org/10.1007/s10639-024-12849-y

Cheng, L., & Curtis, A. (Eds.). (2010). *English language assessment and the Chinese learner (1st ed.)*. Routledge. https://doi.org/10.4324/9780203873045

Chen, Y., & Wu, X. (2022). Validation of the results of linking the speaking test of IELTS to

China\u2019s *Standards of English Language Ability. English Language Teaching, 15*(7), 75–85. https://doi.org/10.5539/elt.v15n7p75

Cheng, X., Zhang, L. J., & Yan, Q. (2021). Exploring teacher written feedback in EFL writing classrooms: Beliefs and practices in interaction. *Language Teaching Research, 29*(1), 136216882110576. https://doi.org/10.1177/13621688211057665

Claramita, M., Nurokhmanti, H., Qomariyah, N., Budiastuti, V. I., Utomo, P. S., & Findyartini, A. (2022). Facilitating student-centered learning: In the context of social hierarchies and collectivistic culture. In M. Claramita, A. Findyartini, D. D. Samarasekera, & H. Nishigori (Eds.), *Challenges and opportunities in health professions education*. Springer. https://doi.org/10.1007/978-981-16-7232-3_2

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment.* Cambridge University Press. https://www.coe.int/en/web/common-european-framework-reference-languages

Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches (3rd ed.).* SAGE Publications.

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory Into Practice, 39*(3), 124–130. https://doi.org/10.1207/s15430421tip3903_2

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches (4th ed.).* SAGE Publications.

Csapó, B., & Molnár, G. (2019). Online diagnostic assessment in support of personalized teaching and learning: The eDia system. *Frontiers in Psychology, 10*, 1522. https://doi.org/10.3389/fpsyg.2019.01522

Dashti, L., & Razmjoo, S. A. (2020). An examination of IELTS candidates' performances at different band scores of the speaking test: A quantitative and qualitative analysis. *Cogent Education, 7*(1). https://doi.org/10.1080/2331186X.2020.1770936

Ellis, R. (2003). *Task-based language learning and teaching.* Oxford University Press.

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics, 5*(1), 1–4. https://doi.org/10.11648/j.ajtas.20160501.11

Fleming, K., & Shinjee, B. (2024). English high-stakes testing and constructing the 'international' in Kazakhstan and Mongolia. *Applied Linguistics Review, 15*(4), 1253–1275. https://doi.org/10.1515/applirev-2022-0067

Gagen, T., & Faez, F. (2024). The predictive validity of IELTS scores: a meta-analysis. *Higher Education Research & Development, 43*(4), 873-888. https://doi.org/10.1080/07294360.2023.2280700

Gao, L. (2021). A review of multidimensional studies on the causes of classroom reticence. *IRA-International Journal of Education & Multidisciplinary Studies, 17*(2), 45–53. https://dx.doi.org/10.21013/jems.v17.n2.p1

Gebril, A. (Ed.). (2022). *Learning-oriented language assessment: Putting theory into practice. Studies in Language Assessment.* https://doi.org/10.58379/PBMV5990

Ginting, R. S., Dalimunthe, A., Dalimunthe, M. R., & Adelita, D. (2023). A critical review of IELTS speaking test. *Journal of Linguistics, Literature and Language Teaching, 9*(2), 138–155. https://doi.org/10.32505/jl3t.v9i2.7161

Hajiabdorrasouli, L., & Ahmadi, A. (2023). "Others have scored the same as me; do you want to change it"? Exploring rater dominance in negotiation rating sessions of EFL writing. *Iranian Journal of English for Academic Purposes, 12*(2), 87-106. https://doi.org/20.1001.1.24763187.2023.12.3.6.8

He, L., Xiong, L., & Min, S. (2022). Diagnosing listening and reading skills in the Chinese EFL context: Performance stability and variability across modalities and performance levels. *System, 106*, 102787. https://doi.org/10.1016/j.system.2022.102787

Heidari Vincheh, M., Mirzaei, A., & Roohani, A. (2024). A cognitive diagnostic approach to the IELTS

speaking test: Unveiling the subskills and test-takers' perceptions. *Language Testing in Asia, 14*, 42. https://doi.org/10.1186/s40468-024-00311-2

Hofstede, G. (1986). Cultural differences in teaching and learning. *International Journal of Intercultural Relations, 10*(3), 301–320. https://doi.org/10.1016/0147-1767(86)90015-5

Hu, M. (2019). The relationship between English phonological awareness of Chinese English learners and their English skills. *Theory and Practice in Language Studies, 9*(1), 45-51. https://doi.org/10.17507/tpls.0901.07

Huang, M. (2024). Empowering oral proficiency in a large-scale class: Video-recorded oral presentations and mobile-assisted peer assessment in a Chinese middle school. *Computer Assisted Language Learning, 1*(1), 1–21. https://doi.org/10.1080/09588221.2024.2382851

Huang, J., Li, J., Shu, T. and Zhang, Y. (2022) A mixed-methods national study investigating key challenges in learning English as a foreign language:A Chinese college student perspective. *Frontier Psychology,13*:1035819. https://doi.org/10.3389/fpsyg.2022.1035819

Ismail, S. M., Rahul, D. R., Patra, I., & Rezvani, E. (2022). Formative vs. summative assessment: Impacts on academic motivation, attitude toward learning, test anxiety, and self-regulation skill. *Language Testing in Asia, 12*(1), 40. https://doi.org/10.1186/s40468-022-00191-4

Jang, E. E. (2008). A framework for cognitive diagnostic assessment. In C. A. Chapelle, Y.-R. Chung, & J. Xu (Eds.), *Towards adaptive CALL: Natural language processing for diagnostic language assessment* (pp. 117–131). Ames, IA: Iowa State University.

Javidanmehr, Z., & Anani Sarab, M. R. (2017). Cognitive diagnostic assessment: Issues and considerations. *International Journal of Language Testing, 7*(2), 73–98.

Jin, C. (2023). The future is in the making: A review of literature on dynamic assessment in second language education. *Cambridge Educational Research e-Journal, 10*, 1–13. https://doi.org/10.17863/CAM.104605

Jin, C., & Liu, Y. (2024). Diagnosing and promoting learners' L2 inferential reading development through hybrid computerized dynamic assessment in the Chinese EFL classroom. *Computer Assisted Language Learning, 41*(1), 1–24. https://doi.org/10.1080/09588221.2024.2421521

Kafipour, R., & Khoshnood, A. (2023). Effect of feedback through dynamic assessment on EFL field-dependent and field-independent learners' speaking skill development. Frontiers *in Education, 8*, https://doi.org/10.3389/feduc.2023.1049680

Karim, S., & Haq, N. (2014). An assessment of IELTS speaking test. International *Journal of Evaluation and Research in Education, 3*(3), 152–157.

Kuo, B. C., Chen, C. H., Yang, C. W., & Mok, M. M. C. (2016). Cognitive diagnostic models for tests with multiple-choice and constructed-response items. Educational Psychology, 36(6), 1115–1133. https://doi.org/10.1080/01443410.2016.1166176

Kvale, S., & Brinkmann, S. (2015). *Interviews: Learning the craft of qualitative research interviewing* (3rd ed.). Sage Publications.

Lam, D. M. K., Galaczi, E., Nakatsuhara, F., & May, L. (2023). Assessing interactional competence: Exploring ratability challenges. *Applied Pragmatics, 5*(2), 208-233. https://doi.org/10.1075/ap.00014.lam

Leighton, J. P., & Gierl, M. J. (Eds.). (2007). *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge University Press. https://doi.org/10.1017/CBO9780511611186

Leighton, J. P., Gokiert, R. J., Cor, M. K., & Heffernan, C. (2010). Teacher beliefs about the cognitive diagnostic information of classroom- versus large-scale tests: Implications for assessment literacy. *Assessment in Education: Principles, Policy & Practice, 17*(1), 7–21. https://doi.org/10.1080/09695940903565362

Li, J. (2019). An evaluation of IELTS speaking test. *Open Access Library Journal, 6*, 1–17. https://doi.org/10.4236/oalib.1105935

Li, Y., Zhen, M. and Liu, J. (2021) Validating a Reading Assessment Within the Cognitive Diagnostic

Assessment Framework: Q-Matrix Construction and Model Comparisons for Different Primary Grades. *Frontier Psychology 12*:786612. https://10.3389/fpsyg.2021.786612

Liu, J., & Xu, Y. (2017). Assessment for learning in English language classrooms in China: Contexts, problems, and solutions. In H. Reinders, D. Nunan, & B. Zou (Eds.), *Innovation in language learning and teaching: New language learning and teaching environments* (pp. 125–145). Palgrave Macmillan. https://doi.org/10.1057/978-1-137-60092-9_2

Liu, Z. (2024). Effects of nonlinear dynamic online assessment model on language learners' learning outcomes and cognitive load. *Education and Information Technologies, 29*, 24255–24284. https://doi.org/10.1007/s10639-024-11592-7

Ma, H., & Chong, S. W. (2022). Predictability of IELTS in a high-stakes context: A mixed methods study of Chinese students' perspectives on test preparation. *Language Testing in Asia, 12*, 2. https://doi.org/10.1186/s40468-021-00152-3

Mayer, R. E. (Ed.). (2014). The Cambridge handbook of multimedia learning (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369

Mehraban, B. (2022). Assessing English Proficiency in the Age of Globalization : The Case of IELTS (Dissertation). Retrieved from https://urn.kb.se/resolve?urn=urn:nbn:se:lnu:diva-113738

Mei, H., & Chen, H. (2024). Cognitive diagnosis in language assessment: A thematic review. *RELC Journal*, *55*(2), 529–537. https://doi.org/10.1177/00336882221122357

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation (4th ed.)*. Jossey-Bass.

Minakova, V. (2020). Dynamic assessment of IELTS speaking: A learning-oriented approach to test preparation. *Language and Sociocultural Theory, 6*(2), 184-212.

Min, S., & He, L. (2021). Developing individualized feedback for listening assessment: Combining standard setting and cognitive diagnostic assessment approaches. *Language Testing*, *39*(1), 90-116. https://doi.org/10.1177/0265532221995475 (Original work published 2022)

Nakatsuhara, F., Inoue, C., Berry, V., & Galaczi, E. (2017). Exploring the use of video-conferencing technology in the assessment of spoken language: A mixed-methods study. *Language Assessment Quarterly, 14*(1), 1–18. https://doi.org/10.1080/15434303.2016.1263637

Panahi, A., & Mohebbi, H. (2022). Cognitive diagnostic assessment of IELTS listening: Providing feedback from its internal structure. *Language Teaching Research Quarterly, 28*, 147-165. https://doi.org/10.32038/ltrq.2022.29.10

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health, 42*(5), 533–544. https://doi.org/10.1007/s10488-013-0528-y

Poehner, M. E., & Lantolf, J. P. (2022). Advancing L2 Dynamic Assessment: Innovations in Chinese Contexts. *Language Assessment Quarterly, 20*(1), 1–19. https://doi.org/10.1080/15434303.2022.2158465

Poehner, M. E., & Wang, Z. (2020). Dynamic assessment and second language development. *Language Teaching, 54*(4), 1–19. https://doi.org/10.1017/S0261444820000555

Quaid, E. D. (2018). Reviewing the IELTS speaking test in East Asia: theoretical and practice-based insights. *Language Testing in Asia, 8*(2). https://doi.org/10.1186/s40468-018-0056-5

Ramalingam, S., Yunus, M. M., & Hashim, H. (2021). Exploring ESL learners' blended learning experiences and its' effectiveness through web-based technologies. *International Journal of Evaluation and Research in Education (IJERE), 10*(4), 1436-1445. https://doi.org/10.11591/ijere.v10i4.21465

Ravand, H., & Robitzsch, A. (2018). Cognitive diagnostic model of best choice: A study of reading comprehension. *Educational Psychology, 38*(10), 1255–1277. https://doi.org/10.1080/01443410.2018.1489524

Read, J. (2022). Test review: The International English Language Testing System (IELTS). *Language Testing*, *39*(4), 679–694. https://doi.org/10.1177/02655322221086211

Shi, X., Ma, X., Du, W., & Gao, X. (2024). Diagnosing Chinese EFL learners' writing ability using polytomous cognitive diagnostic models. *Language Testing*, *41*(1), 109–134. https://doi.org/10.1177/02655322231162840

Shiny, K. G., & Karthikeyan, J. (2022). Outlining international English language proficiency tests: A survey-based study on its need, purpose, and scope in the backdrop of migration. *Journal for Educators, Teachers and Trainers, 13*(4), 198–206. https://doi.org/10.47750/jett.2022.13.04.029

Silva Moreira J, Costa Ferreira P, Veiga Simão AM (2024) Fostering self-regulated learning in preschool through dynamic assessment methodologies. *PLoS ONE 19*(3): e0298759. https://doi.org/10.1371/journal.pone.0298759

Souzandehfar, M. (2024). New perspectives on IELTS authenticity: An evaluation of the speaking module. *International Journal of Language Testing*, *14*(1), 34–55. https://doi.org/10.22034/ijlt.2023.409599.1272

Suss, M. (2023). Examining the reliability of the International English Language Testing System design. *Bulletin of Advanced English Studies*, *8*(2), 72–92. https://doi.org/10.31559/BAES2023.8.2.4

Tang, F., & Zhan, P. (2021). Does diagnostic feedback promote learning? Evidence from a longitudinal cognitive diagnostic assessment. *AERA Open*, *7*, 1–10. https://doi.org/10.1177/23328584211060804

Templin, J. L., & Henson, R. A. (2009). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods, 11*(3), 287–305.

Tiu, J. S., Groenewald, E. S., Kilag, O. K. T., Balicoco, R. D., Wenceslao, S. B., & Asentado, D. E. (2023). Enhancing oral proficiency: Effective strategies for teaching speaking skills in communication classrooms. *Excellencia International Multi-disciplinary Journal of Education, 1*(6), 343-354.

Zhang, Z. (2021). Promoting student engagement with feedback: Insights from collaborative pedagogy and teacher feedback. Assessment & Evaluation in *Higher Education*, *47*(4), 540–555. https://doi.org/10.1080/02602938.2021.1933900

Wang, J., An, N., & Wright, C. (2018). Enhancing beginner learners' oral proficiency in a flipped Chinese foreign language classroom. *Computer Assisted Language Learning*, *31*(5–6), 490–521. https://doi.org/10.1080/09588221.2017.1417872

Woolridge, L. R., Leach, A.-M., Blake, C., & Elliott, E. (2023). Do Accents Speak Louder Than Words? Perceptions of Linguistic Speech Characteristics on Deception Detection. *Journal of Language and Social Psychology, 43*(2), 195-223. https://doi.org/10.1177/0261927X231209428

Xin, S., & Luxmon, H. (2023). Developing individualized feedback for listening assessment: Combining standard setting and cognitive diagnostic assessment approaches. *Educational Psychology, 29*(1). https://doi.org/10.1177/02655322199547

Pang, Y., Hashim, H., & Said, N. E. M. (2024). Learning Strategies for Chinese as Foreign Language Learners in College: A Qualitative Study. *Forum for Linguistic Studies*, *6*(4), 463–473. https://doi.org/10.30564/fls.v6i4.6795

Zhai, X., Haudek, K.C. & Ma, W. Assessing Argumentation Using Machine Learning and Cognitive Diagnostic Modeling. *Research Science in Education, 53*, 405–424 (2023). https://doi.org/10.1007/s11165-022-10062-w

Zhang, X. (2024). English language assessment practices in Chinese universities: A critical examination of testing methods and their implications for student learning. *SHS Web of Conferences*, *199*, Article 01005. https://doi.org/10.1051/shsconf/202419901005

Zhorova, I. Y., Kokhanovska, O. V., Khudenko, O. M., Osypova, N. V., & Kuzminska, O. H. (2022). Teachers' training for the use of digital tools of the formative assessment in the implementation of the concept of the New Ukrainian School. *Educational Technology Quarterly, 2022*(1), 56-72. https://doi.org/10.55056/etq.11

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, *41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

Appendices

Appendix A

Comprehensive Analysis of Speaking Subskill Development

Subskill	Key improvements	Evidence sources	Representative quotes
Linguistic Knowledge	Enhanced vocabulary range; Improved grammatical accuracy; Better idiomatic expression use.	Mock test recordings; Classroom observations; Interview data.	"Before, I would just use simple words, but now I can express more complex ideas using words like *substantial* and *feasible*." (Student 3)
Syntactic Knowledge	Increased complex sentence use; Reduced errors in tense/agreement; Better clause coordination.	Mock test recordings; Teacher feedback logs; Student interviews.	"I've learned to combine ideas using *while* and *whereas* instead of just using simple sentences." (Student 2)
Pronunciation Mastery	Improved stress patterns; Better intonation control; Enhanced phonemic accuracy.	Audio recordings; Teacher observations; Student self-reports.	"The targeted feedback helped me realize I was stressing the wrong syllables. Now my pronunciation sounds more natural." (Student 4)
Discourse Management	Better topic development; Improved coherence; Enhanced use of transitions.	Mock test recordings; Class discussions; Interview data.	"Now I can link my ideas smoothly using transitions like *furthermore* and *consequently*." (Student 1)
Speech Structure	Clear introduction/conclusion; Logical argument progression;	Test performances; Lesson plans; Student reflections.	"I always plan my response structure: introduction, main points, and conclusion."

	Better topic organization.		(Student 5)

Appendix B
Cognitive Processing Strategies and Outcomes

Strategy	Description	Examples of implementation	Impact on learning
Meta-analytical review	Systematic analysis of feedback patterns	"I began categorizing feedback based on different aspects of speaking - grammar, vocabulary, and pronunciation." (Student 3)	Enhanced ability to identify recurring patterns in speaking performance
Comparative analysis	Connecting new feedback with previous learning	"When I receive feedback now, I compare it with previous comments to see if I'm making progress." (Student 5)	Improved integration of feedback across learning experiences
Contextual application	Adapting feedback to different speaking contexts	Teacher observation notes showing students successfully transferring feedback across different speaking tasks.	Greater flexibility in applying learned skills

Appendix C

Progress Indicators across All Five Cognitive Subskills

Subskill	Observed development patterns	Supporting evidence	Implementation outcomes
Linguistic knowledge	Significant increase in advanced vocabulary range	"Students demonstrated broader lexical choices in spontaneous speech." (Teacher observation notes)	Progressive development in vocabulary sophistication
Syntactic complexity	Notable improvement in sentence structure variety	"I can now express complex ideas using different grammatical patterns." (Student 2)	Enhanced structural sophistication in responses
Pronunciation mastery	Substantial reduction in L1 interference	"Students showed marked improvement in challenging	Improved phonological awareness and control

		phonemic distinctions." (Teacher feedback logs)	
Discourse management	Enhanced use of cohesive devices	"Their responses flow much more naturally now." (Teacher interview)	More coherent extended discourse
Speech structure	Clear improvement in organizational patterns	"My answers have a clear beginning, middle, and end now." (Student 5)	Systematic approach to response organization