DOI: 10.53469/jerp.2025.07(11).24

Constructing and Implementing a "Tri-Phase-Tri-Dimensional" Digital Assessment System for Junior High School English Academic Quality

Yuwen Sun[1], Ming Yin[2,*]

School of Foreign Languages and Cultures, Nanjing Normal University Taizhou College, Taizhou 225300, China
[1]05220241@nnutc.edu.cn, [2]20061004@nnutc.edu.cn
Correspondence Author

Abstract: ***Junior high school English plays a critical role in fostering students' comprehensive development and enhancing teaching quality in China. Traditional evaluation methods struggle to meet the core competency requirements outlined in the English Curriculum Standards for Compulsory Education. This study, grounded in the New National English Curriculum Standards and the China Standards of English Language Ability, proposes a "Tri-Phase-Tri-Dimensional" digital assessment system that integrates formative, process-oriented, and summative evaluations. By leveraging intelligent tools, the system enables multidimensional data collection, real-time feedback, and personalized interventions. A case study demonstrated the system's efficacy, with data revealing a 15.5% improvement in language skills and a 20.8% enhancement in learning competencies, a 48.2% reduction in teachers' grading workload, and 92.1% of students achieving the targeted speaking proficiency. The study concludes that the "Tri-Phase-Tri-Dimensional" digital assessment system represents a systematic and theoretically robust model that significantly enhances the scientific validity and effectiveness of English academic quality evaluation, offering a practical solution for improving teaching quality and a scalable framework for assessment reform.***

Keywords: Junior High School English, Academic Quality, "Tri-Phase-Tri-Dimensional" Digital Assessment System, Teaching Implementation**.**

1. Introduction

In the era of educational digital transformation, the assessment of academic quality in junior high school English, a cornerstone of compulsory education in China, is pivotal for fostering students' holistic development and enhancing instructional quality. However, traditional assessment methods are often criticized for their subjectivity, delayed feedback, and lack of personalization, failing to meet the core competency requirements outlined in the *New National English Curriculum Standards* (*New Curriculum Standards, NCS*) [1]. In the context of educational digital transformation, this global wave is profoundly reshaping educational ecologies, with teaching and assessment at its core [2]. As a key subject in fundamental education, the assessment of junior high school English academic quality directly impacts students' language proficiency and comprehensive competency development [3], [4], [5]. Yet, traditional paper-based tests prioritize linguistic knowledge, often neglecting the four core competencies—language ability, cultural awareness, thinking quality, and learning ability [1]. Additionally, lengthy feedback cycles hinder teaching optimization and fail to address personalized learning needs.

Digital Assessment, powered by technologies such as artificial intelligence (AI), natural language processing (NLP), and learning analytics, emerges as a potent solution to this challenge [6], [7]. Unlike their static, analog predecessors, digital tools can offer immediate, personalized, and actionable feedback. For instance, AI-driven speech recognition can evaluate pronunciation in real-time, while learning management systems (LMS) can track engagement patterns, providing a granular, multi-faceted view of student progress. By benchmarking performance against established frameworks like *China's Standards of English Language Ability* (CSE) [8], these technologies can translate raw data into meaningful pedagogical insights.

Despite the promise, much of the existing research and practice in digital English assessment remains fragmented. Many studies focus on the efficacy of a single tool (e.g., an automated writing evaluation system) or a single skill, failing to address the holistic nature of language competency [9]. This study addresses this critical gap by proposing and validating a comprehensive, theoretically-grounded framework: the "Tri-Phase-Tri-Dimensional" Digital Assessment System. This system aims to construct a coherent assessment ecology that systematically integrates pedagogical theory with a suite of digital tools. Its purpose is to enhance the scientific rigor, diagnostic precision, and formative power of junior high school English assessment, ultimately providing robust support for students' holistic development and fostering a culture of continuous instructional improvement.

2. Theoretical Framework and Literature Review

The design of the proposed system is not technologically deterministic; rather, it is deeply rooted in a synthesis of established and contemporary assessment theories. This section elaborates on the four pillars of our theoretical framework and reviews relevant literature.

2.1 Evidence-Centered Design (ECD)

Proposed by Mislevy, Steinberg, and Almond [10], ECD is a robust framework for designing and developing assessments. It conceives of assessment as a process of reasoning from evidence. The core idea is to make explicit the chain of inference from what students say, do, or create to claims about their competencies. ECD comprises three primary models:

1) **Student Model:** Defines the knowledge, skills, and abilities (KSAs) to be assessed. In our study, this model is populated by the core competencies outlined in the NCS.

2) **Evidence Model:** Specifies the behaviors or performances that constitute evidence of the KSAs defined in the student model. It links observations to competency claims, defining what evidence to collect and how to interpret it.

3) **Task Model:** Describes the tasks or situations that will be presented to students to elicit the required evidence. By adopting ECD, we ensure that our assessment design is principled, coherent, and transparent. Every task is designed to elicit specific evidence linked to a defined competency, strengthening the validity of our assessment claims.

2.2 Formative Assessment and Self-Regulated Learning

The foundational work of Black and Wiliam [11] established formative assessment as a powerful lever for improving student learning. It involves the use of evidence to adapt teaching and learning activities to meet student needs. A key mechanism through which formative assessment works is by promoting Self-Regulated Learning (SRL) [12]. SRL refers to the process whereby learners actively manage their own learning by setting goals, monitoring their progress, and adjusting their strategies. Digital technologies supercharge this process. Immediate feedback from an AI tutor, for example, allows a student to instantly "monitor" their performance and "adjust" their pronunciation strategy. A digital portfolio that visualizes progress over time helps a student "set goals" for the next learning phase. Our system is designed to provide the timely, specific feedback necessary to foster this cycle of self-regulation, moving students towards greater learner autonomy.

2.3 Learning Analytics

Learning Analytics (LA) is "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" [13]. LA allows us to move beyond simple right or wrong scores to analyze the process of learning. By capturing digital traces—such as time spent on a task, number of attempts, resources accessed, and interaction patterns in a collaborative forum—we can identify learning strategies, detect potential difficulties, and predict future performance. Our system leverages LA as its analytical engine. It aggregates multi-modal data from various digital tools to create a dynamic, longitudinal profile of each student. This enables not only summative judgments but also diagnostic and predictive insights, allowing for proactive, data-informed interventions.

2.4 Ecological Assessment

Traditional assessment often occurs in isolated, inauthentic "test" environments. An ecological perspective, inspired by Bronfenbrenner's [14] ecological systems theory, argues that assessment should be situated within the complex, interconnected systems where learning naturally occurs (e.g., the classroom, home, online communities). Ecological assessment is therefore continuous, context-sensitive, and involves multiple sources of evidence collected across different environments. Our "Tri-Phase" design (pre-, process-, and summative assessment) is a direct application of this principle. It captures student learning not as a single snapshot, but as a continuous developmental trajectory across various contexts: individual preparatory work, in-class collaboration, and post-class reinforcement. This provides a more authentic and holistic picture of a student's true capabilities.

3. Constructing the "Tri - Phase - Tri - Dimensional" Digital Assessment System

Following the principles of ECD, this section details the architecture and operationalization of the digital assessment system.

3.1 Research Design and Participants

This study employed a quasi-experimental, single-group pre-test/post-test design embedded within a descriptive case study. The primary goal was to explore the implementation process and evaluate the efficacy of the digital assessment system in a naturalistic classroom setting. The participants were 45 eighth-grade students (22 male, 23 female, average age 13.5 years) from an urban public school in Nanjing, China. The participating teacher had 10 years of experience and received a two-week training workshop on the use of the digital tools and the underlying pedagogical framework. The study was conducted over a four-week period, covering the *Good Manners* unit of the Oxford English textbook (Yilin Press).

3.2 System Architecture: The Three Dimensions and Three Phases

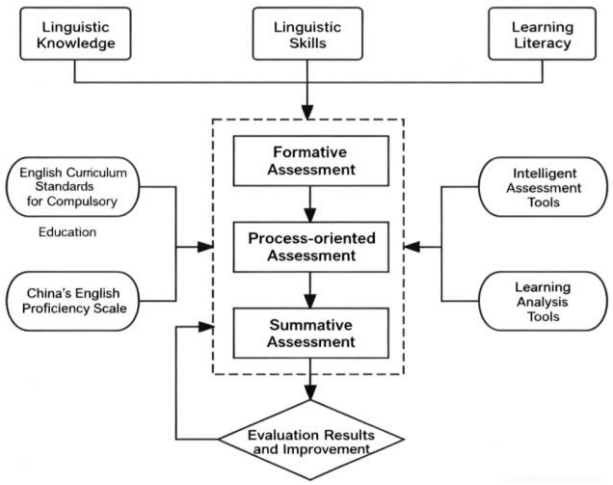

Figure 1: "Tri-Phase-Tri-Dimensional" Digital Assessment System

To align with the "Double Reduction" policy, this study composes the "Tri-Phase-Tri-Dimensional" Digital Assessment system (see Figure 1). The system's architecture is defined by its two core components: the dimensions of assessment (the *what*) and the phases of implementation (the *when* and *how*).

3.2.1 The Three Dimensions (The Student Model)

1) Linguistic Knowledge: This foundational dimension covers the explicit "declarative knowledge" of the language. Aligned with Level 3 of the curriculum standards, it includes a core vocabulary of 1,600 words, key sentence patterns, and grammatical rules (e.g., the use of modal verbs for giving advice).

2) Linguistic Skills: This dimension focuses on the practical application of knowledge ("procedural knowledge"). It is operationalized using the "can-do" statements from the CSE. For this unit, target skills included:

- Listening: Understand conversations about social etiquette.
- Speaking: Give and respond to advice on polite behavior.
- Reading: Identify specific rules and general ideas in a text about etiquette.
- Writing: Write a short email offering advice on proper conduct.

3) Learning Literacy: This dimension assesses transversal skills essential for lifelong learning.

- Metacognition & Learning Strategies: Measured through student self-assessment checklists and analysis of their planning and revision behaviors on digital platforms.

- Cultural Awareness: Assessed via collaborative tasks requiring comparison of cultural norms (e.g., table manners in China vs. the West).

- Digital Literacy & Collaboration: Evaluated through the production of digital artifacts (e.g., collaborative mind maps, video presentations) and analysis of interaction logs on platforms like Padlet.

3.2.2 The Three Phases (The Assembly and Task Models)

1) Phase 1: Formative Assessment (Diagnostic): Before the unit began, a diagnostic pre-test was administered via Wenjuanxing (a popular online survey/quiz tool in China). It consisted of 10 multiple-choice questions targeting prerequisite vocabulary and grammar. The goal was to establish a baseline and identify potential learning gaps to inform initial instructional design.

2) Phase 2: Process-oriented Assessment (Formative): This is the core of the system, involving continuous data collection during the unit. A suite of digital tools was used:

- ClassIn (a virtual classroom platform) was used for live lessons. Its features allowed tracking of student participation (e.g., number of times raising a hand, speaking duration) and conducting instant polls.

- Padlet (a collaborative online whiteboard) was used for a brainstorming task where students co-created mind maps comparing etiquette rules. The platform's contribution logs provided data on collaborative dynamics.

- Yike (a mobile learning app) delivered after-class practice exercises. Its adaptive engine provided targeted drills based on individual performance. All attempts and scores were logged.

- EAP Talk (an AI-powered oral practice tool) was used for speaking homework. Students recorded themselves reading dialogues, and the AI provided instant feedback on pronunciation, fluency, and intonation, generating a detailed report.

3) Phase 3: Summative Assessment (Evaluative): At the end of the unit, assessment was based on two performance tasks rather than a traditional test:

- A collaborative infographic on "Do's and Don'ts of Digital Etiquette," created using Canva.

- An individual 2-minute video presentation on Flipgrid, where students gave advice to a foreign visitor about etiquette in China. These artifacts were evaluated using a detailed rubric that integrated all three dimensions (linguistic accuracy, clarity of communication, cultural appropriateness, digital creativity). The final output of the system was a holistic, multi-dimensional diagnostic report for each student, visualized using a radar chart.

3.3 Data Collection and Analysis

A mixed-methods approach was used for data analysis.

- Quantitative Data: Pre- and post-test scores, accuracy rates from Yike exercises, fluency scores from EAP Talk, and participation metrics from ClassIn were collected. Paired-samples t-tests were used to analyze significant changes in performance.

- Qualitative Data: Student-created artifacts (mind maps, videos), teacher's observational notes, and responses from a post-unit student satisfaction survey (including open-ended questions) were collected. This data was analyzed thematically to provide context and depth to the quantitative findings.

4. Results

The implementation of the system yielded rich, multi-faceted data, revealing significant improvements in student learning and engagement.

4.1 Improvement in Linguistic Knowledge and Skills

Quantitative analysis showed statistically significant gains in both linguistic knowledge and language skills. The class average on the pre-test was 70.4%, which increased to 87.2% on a structurally equivalent post-test ($t(44) = 8.12$, $p < .001$),

confirming formative evaluation's efficacy [15].

- Speaking Fluency: Data from EAP Talk showed a 23.5% average increase in words per minute and a 17.8% increase in pronunciation accuracy scores over four practice sessions. In the final video task, 92.1% of students were rated as meeting or exceeding the CSE Level 3 standard for speaking on familiar topics.

- Listening and Reading: On the Yike platform, the average accuracy for listening comprehension exercises increased from 75.2% in the first week to 83.6% in the final week. Reading comprehension accuracy similarly rose from 78.1% to 85.3%.

- Writing: Analysis of the writing components in the final tasks showed a significant reduction in grammatical errors related to modal verbs (a key focus of the unit), with the error rate dropping from an average of 3.2 errors per 100 words to 1.1.

4.2 Development of Learning Literacy

The system was particularly effective in fostering competencies beyond pure language skills.

- Metacognition: The post-unit survey revealed that 88% of students "agreed" or "strongly agreed" that the instant feedback from digital tools helped them understand their own mistakes and know what to improve. One student wrote, "The AI report for my speaking showed my fluency was low. I realized I was pausing too much, so I practiced speaking in longer chunks. It really helped." This demonstrates a clear cycle of monitoring and strategic adjustment.

- Collaboration and Digital Literacy: The Padlet mind maps and Canva infographics served as direct evidence of these skills. A content analysis of the mind maps showed that 86.1% were logically structured and demonstrated a synthesis of ideas rather than a simple list. Interaction logs from Padlet indicated a relatively balanced contribution pattern in most groups after the teacher introduced structured roles.

- Cultural Awareness: The quality of comparisons in the mind maps and the appropriateness of advice in the Flipgrid videos were key indicators. 85.2% of the final video presentations were rated as "culturally appropriate" or "highly culturally appropriate" by the rubric.

4.3 Impact on Teaching and Learning Environment

The system also had a profound impact on classroom dynamics and teacher workload.

- Teacher Efficiency: The teacher reported that automated grading for quizzes and speaking practice via Yike and EAP Talk reduced her marking workload by an estimated 48.2%, freeing up approximately 5 hours per week. She reallocated this time to designing more engaging activities and providing one-on-one coaching to struggling students.

- Student Engagement: Data from Class. In showed a 35% increase in voluntary student participation (e.g., raising hands, responding in the chatbox) compared to previous, non-digitally-enhanced units. The post-unit survey showed a student satisfaction rate of 95% with the new learning method.

5. Discussion

The results of this study are promising, demonstrating that a systematically designed digital assessment system can significantly enhance English language learning. This section interprets these findings in relation to the theoretical framework and discusses their broader implications.

5.1 Alignment with Theoretical Promises

The findings largely affirm the theoretical underpinnings of the system. The observed improvement in student performance, coupled with their self-reported growth in metacognitive awareness, strongly supports the synergistic relationship between formative assessment and self-regulated learning [11], [12], [16], [17]. The digital tools did not just provide scores; they provided immediate, specific feedback that empowered students to take ownership of their learning, acting as agents in their own educational journey.

The rich, process-oriented data captured by the system underscores the value of Learning Analytics [12]. Instead of relying on a single test score, the teacher had access to a dashboard visualizing a student's engagement, effort (e.g., number of practice attempts), and progress over time. This aligns with the principles of Ecological Assessment [13], as it painted a holistic picture of learning as it unfolded across different contexts and modalities. The system's structured design, guided by ECD, ensured that these rich data points were not just noise, but meaningful evidence systematically linked to specific competency claims.

5.2 Challenges and Mitigation in a Real-World Context

While successful, the implementation was not without its challenges. The initial accuracy of AI-driven emotion recognition in ClassIn was found to be unreliable (10.4% misjudgment rate), highlighting the current limitations of affective computing and the irreplaceable role of human teachers in interpreting emotional cues. This was mitigated by training the teacher to treat AI data as a supplementary "second opinion" rather than an absolute truth.

Furthermore, a "digital divide" in terms of technical skill was observed. An initial 15% of students struggled with the video creation task on Flipgrid. This finding cautions against assuming universal digital literacy among "digital natives." The successful mitigation via a brief, targeted tutorial (which raised the completion rate to 88.1%) suggests that scaffolding for digital skills must be an integral part of such systems. This experience underscores the socio-technical nature of educational technology implementation [2].

5.3 Deeper Implications: Ethics, Equity, and the Evolving Teacher Role

This study prompts critical reflection on the broader implications of data-intensive assessment.

- Ethics and Privacy: The comprehensive collection of student data, while pedagogically powerful, raises significant ethical questions. It is imperative to move towards an "ethics-by-design" approach, incorporating principles of data minimization, transparency, and student consent. Students should not merely be subjects of datafication but should have agency over their own data.

- Algorithmic Fairness: The reliance on AI for scoring brings the risk of algorithmic bias. Does an AI speaking assessment tool penalize students with regional accents? Does an automated writing evaluator favor simple, formulaic sentences over creative but slightly flawed expressions? Ensuring equity requires continuous auditing of these algorithms for bias and maintaining a "human-in-the-loop" for contested or nuanced evaluations.

- The Future of the Teacher: This system does not render the teacher obsolete; it redefines their role. Freed from the drudgery of rote marking, the teacher becomes a "learning architect," a "data-informed diagnostician," and a "human mentor." Their expertise shifts from information delivery to instructional design, data interpretation, and providing socio-emotional support. This transformation necessitates a fundamental rethinking of teacher professional development, focusing on data literacy, pedagogical design, and ethical awareness.

6. Conclusion and Future Directions

This research designed, implemented, and evaluated the "Tri-Phase-Tri-Dimensional" digital assessment system. By integrating the principles of ECD, formative assessment, learning analytics, and ecological assessment, the system provides a holistic, robust, and theoretically grounded model for evaluating junior high school English academic quality. It moves beyond fragmented tool use to create a coherent assessment ecology. The case study demonstrated the system's effectiveness in promoting not only language proficiency but also crucial future-ready competencies, while simultaneously enhancing teaching efficiency.

This study has its limitations, though. The findings of this study should be interpreted with caution. The single-group, small-sample design limits generalizability. The study's short duration also precludes any claims about long-term effects. Furthermore, the assessment of complex competencies like "cultural awareness" remains a challenge and could benefit from more refined metrics. Finally, the successful implementation was contingent on a level of technological infrastructure and teacher support that may not be available in all contexts, highlighting the issue of scalability and equity.

Future research should aim to address these limitations.

1) Scale and Generalizability: Conduct larger-scale, longitudinal studies across diverse school contexts (e.g., rural vs. urban) to validate the model's robustness and identify contextual factors influencing its effectiveness.

2) Advanced Analytics: Leverage more sophisticated LA techniques. Predictive models could be developed to identify at-risk students proactively. Cluster analysis could be used to group students based on their learning behavior patterns, enabling more differentiated instruction.

3) Human-AI Symbiosis: Explore optimal models for human-AI collaboration in assessment. Research could focus on designing interfaces that help teachers interpret AI-generated feedback and make more informed pedagogical decisions.

4) Student Agency in Assessment: Investigate the impact of giving students more control over their learning data, such as co-designing assessment rubrics or choosing which data to share.

In conclusion, the digitalization of education offers an unprecedented opportunity to reinvent assessment. The system proposed here represents one step towards a future where assessment is no longer a final judgment but a continuous, insightful, and empowering dialogue that nurtures lifelong learners.

References

[1] Ministry of Education of the People's Republic of China, English Curriculum Standards for Compulsory Education, Beijing Normal University Press, Beijing, 2022.
[2] Selwyn N, Is Technology Good for Education?, John Wiley & Sons, Hoboken, 2016.
[3] Chen, Y. A study on English teaching evaluation in junior middle school, Northwest Normal University, Lanzhou, 2010.
[4] Wu, Z., Yang, Y., Ren, Z, "Student Academic Achievement Analysis, Feedback and Guidance System: A Diagnostic Evaluation of Middle School Students' English Learning Based on Curriculum Standards", Foreign Language Teaching Theory and Practice, (4), pp. 15-22, 2011.
[5] Zheng X R, Li Q, Chen X W, "A Study on the Application of Formative Assessment in Junior High School English Teaching", New Curriculum Learning, (3), pp. 8-9., 2014.
[6] Wang T. P., Pan Q. Q., Li Z., "Functional Logic and Practical Approach of Educational Digitalization in Promoting the Construction of a Powerful Education Country", Journal of Distance Education (1672-0008), (2), 2024.
[7] Wu, Z., Xiong, Z., Yang, Y., "The Only Way to Build a Strong Country in Education: A Study on The Path And Mechanism of Digital Education to Promote The Construction of a Learning-Oriented Country", Journal of Distance Education, (1), 2024.
[8] National Education Examinations Authority, China's Standards of English Language Ability, Shanghai Foreign Language Education Press, Shanghai, 2024.

[9] Chapelle C A, Douglas D, “The technology thread,” Assessing language through computer technology, pp.1-19, 2006.
[10] Almond R, Mislevy R J, Steinberg L S, “On the Structure of Educational Assessments”, Measurement: Interdisciplinary research and perspectives, 1(1), pp. 3-62, 2003.
[11] Black P, Wiliam D, “Assessment and classroom learning”, Assessment in Education: principles, policy & practice, 5(1), pp. 7-74, 1998.
[12] Zimmerman B J. Becoming a self-regulated learner: An overview [J]. Theory into practice, 2002, 41(2): 64-70.
[13] Siemens G, Gasevic D, “Guest Editorial-Learning and Knowledge Analytics”, Journal of Educational Technology & Society, 15(3), pp. 1-2, 2012.
[14] Bronfenbrenner U, “Contexts of Child Rearing: Problems and Prospects”, American psychologist, 1979, 34(10), pp. 844, 1979.
[15] Bennett R E. “Formative Assessment: A Critical Review”, Assessment in Education: Principles, Policy & Practice, 18(1), pp. 5-25, 2011.
[16] Stiggins R., “From formative assessment to assessment for learning: A path to success in standards-based schools”, Phi Delta Kappan, 87, pp. 324-328, 2005.
[17] Gao L., "Current Situation Survey and Reflection on Summative Assessment and Formative Assessment in Middle School English Teaching", Science & Technology Information, (11), pp. 18-18, 2013.

Author Profile

Yuwen Sun is studying at the School of Foreign Languages and Cultures of Nanjing Normal University Taizhou College under the guidance of his mentor Ming Yin. As a senior English major student, he will obtain his bachelor’s degree in the coming June.

Ming Yin received his master’s degree in English Language and Literature from Nanjing Normal University in 2015. He is now studying PhD courses on English Linguistics at Nanjing Normal University. As associate professor, he is now teaching and tutoring undergraduates at the School of Foreign Languages and Cultures, Nanjing Normal University Taizhou College. His major research field is corpus linguistics and English teaching.