Journal of Contemporary Educational Research, 2025, Volume 9, Issue 11
http://ojs.bbwpublisher.com/index.php/JCER
ISSN Online: 2208-8474
ISSN Print: 2208-8466

Bridging the Gap: Aligning Communicative Language Testing Principles with AI-Driven Assessment in Civil Aviation Ground Service English

Jinmei Fan*

Guangzhou Civil Aviation College, Guangzhou 510000, Guangdong, China

**Author to whom correspondence should be addressed.*

Abstract: The integration of Communicative Language Testing (CLT) principles with AI-driven automated assessment poses a significant challenge in professional language testing. Addressing this issue within the specific context of Civil Aviation Ground Service English, this study explores pathways for their logical reconciliation. Through conceptual analysis and theoretical deduction, with a focus on human-AI interaction scenarios, we demonstrate that the synergy between CLT and AI stems from a shared focus on competency measurement. Key findings reveal that: (1) standardized competency dimensions in CLT can be operationalized into data-processable formats for AI; (2) within professional contexts, AI algorithms can be tailored using authentic service corpora to meet CLT's demand for situational authenticity; and (3) a division of labor based on competency level—where AI handles standardized scoring of lower-order competencies and human-AI collaboration assesses higher-order competencies—effectively resolves the tension between CLT's dynamic communication and AI's static algorithms. Ultimately, the study constructs a three-dimensional integration framework encompassing "professional register," "competency level," and "human-AI division of labor," offering a theoretical model for CLT-AI integration and a practical blueprint for innovating Civil Aviation Ground Service English assessment.

Keywords: Civil Aviation Ground Service English; Communicative Language Testing (CLT); AI automated assessment; Human-AI interaction scenario; Logical reconciliation

Online publication: December 8, 2025

1. Introduction

The theory of Communicative Language Testing (CLT), rooted in Dell Hymes' concept of "communicative competence," was systematically developed through the frameworks of Canale and Swain (linguistic and strategic competence) and Bachman (communicative language ability). This evolution established an assessment tradition prioritizing authenticity, interactivity, and contextual fidelity, with the ultimate goal of ensuring test performance generalizes to real-world communication by simulating target language use domains [1,2]. In contrast, AI-driven assessment, grounded in computational linguistics and machine learning, operates on the core principles of algorithmic standardization and scalability. It seeks to evaluate language proficiency efficiently through data-driven model training [3,4]. The integration of AI into "human-AI interaction" language testing contexts, such as the Aviation Ground Service English Test, has brought to the fore a fundamental tension: the CLT emphasis on dynamic communication appears to conflict with the static, pre-defined logic of algorithmic assessment. While CLT requires test tasks to be contextually authentic, AI assessment often relies on standardized corpora, potentially undermining the dynamic nature of communication. Furthermore, CLT values interpersonal collaboration inherent in interactivity, raising the question of whether AI-mediated "human-AI interaction" can fulfill the requirements of genuine communication.

This tension not only challenges the contemporary relevance of CLT theory but also questions the theoretical legitimacy of AI assessment within professional language testing. Current academic debate reflects two opposing views: CLT proponents question AI's capacity to replicate the dynamism of real communication through "algorithmic simulation," while AI researchers contend that large-scale corpus training can enable "communication-like" assessment, though they have yet to adequately address CLT's core requirement of contextual fidelity.

Against this backdrop, this study investigates the following core research questions: How can the principles of authenticity and interactivity in CLT be conceptually reconciled with the standardization and algorithmic logic of AI assessment within the specific context of civil aviation ground service English testing? What are the theoretical foundations and boundaries of such a reconciliation? By addressing these questions, this study aims to contribute a "professional contextualization" dimension to the "human-AI collaboration" framework in language testing. Practically, it seeks to provide a theoretical basis for the informed application of AI in civil aviation English testing, thereby avoiding technologically deterministic approaches.

2. The evolution of CLT: From theoretical competence models to contextualized assessment

The theoretical basis of CLT lies in the concept of "communicative competence," initially proposed by Dell Hymes in 1967 in response to the limitations of Noam Chomsky's focus on grammatical competence. Hymes argued that linguistic knowledge must encompass "pragmatic appropriateness," thereby defining communicative competence as the ability to use language effectively in specific social contexts. This framework was later formalized by Michael Canale and Merrill Swain, who elaborated it into a multi-component model comprising grammatical, sociolinguistic, and strategic competencies [5]. "Strategic competence" was brought into the perspective of measurement for the first time. Building upon this foundation, Lyle Bachman advanced the model further in the 1990s with his framework of "Communicative Language Ability" (CLA). This model organized language competence into "language knowledge," "strategic competence," and "psychophysiological mechanisms," and established the core principles of CLT: authenticity, interactivity, and practicality (often related to consistency in specific contexts) [3,4]. The evolution of CLT represents a fundamental paradigm shift from assessing static linguistic knowledge to evaluating dynamic communicative competence. Unlike traditional language testing, which prioritizes grammatical and lexical knowledge in isolation, CLT is defined by its focus on authentic language use in real-world contexts. This orientation is characterized by three core principles:

(1) Authenticity, requiring test tasks to simulate real-life language use; (2) Interactivity, demanding dynamic engagement between the test-taker and the task or interlocutor; and (3) Functionality, emphasizing the practical purposes of communication. Consequently, the central value of CLT lies in its commitment to measuring the ability to communicate effectively in context, which fundamentally distinguishes it from context-independent knowledge assessment.

3. AI-driven language assessment: A paradigm shift from rule matching to contextual awareness

The concept of artificial intelligence (AI) was first proposed in 1956 [6]. The essence of automated AI assessment lies in its data-driven, model-based approach to evaluating language proficiency, which is premised on technologies from computational linguistics and machine learning. Historically, this field has evolved through three key phases, characterized by a progression from computational and perceptual to cognitive intelligence [7]. The technological evolution has progressed from computational intelligence, which facilitates rule-based scoring, to perceptual intelligence, which leverages statistical modeling, and finally to cognitive intelligence, which achieves contextual interpretation through deep learning.

AI has the potential to fundamentally reshape key aspects of assessment, such as test design, psychometric methods, and security [8]. In terms of its functional characteristics, AI-powered automated assessment offers three key advantages: namely, automation minimizing human intervention, scalability handling massive volumes of test data, and real-time performance providing immediate feedback to test-takers [9]. However, there exists an inherent tension in its value orientation: on one hand, AI pursues “standardization,” as the consistency of algorithms can reduce scoring errors; on the other hand, the “dynamic nature” of language communication requires assessment to focus on context and pragmatics, a factor that early AI models often struggled to capture [10]. AI automated assessment is not meant to replace humans but to assist humans. An ideal assessment system should be “human-machine collaborative,” where AI undertakes automated scoring of low-level competencies such as grammar and vocabulary, while humans are responsible for evaluating high-level competencies like pragmatics and strategies [11].

4. Civil Aviation Ground Service English: Human-AI interaction at the intersection of professional context and technical media

Compared with human-to-human dialogue, the human-AI dialogue assessment modality offers greater standardization and, consequently, enhanced scoring reliability [12]. The human-AI interaction scenario in Aviation Ground Service English represents a convergence of civil aviation professional contexts, AI interaction technology, and English communication. It involves test-takers interacting in English with an AI system within a simulated ground service environment. A focus on test “authenticity” is a cornerstone of communicative language testing, distinguishing it fundamentally from traditional approaches [13]. Therefore, the functional design of these scenarios is guided by two criteria: professional authenticity, demanding high-fidelity simulation of service workflows with authentic language data, and technical interactivity, necessitating advanced natural language processing for fluid dialogue. The value proposition, therefore, lies in balancing two objectives: the accurate measurement of professional communicative competence—assessing the clarity, appropriateness, and effectiveness of English used in service tasks—and the achievement of technical efficiency through AI, which enhances the test’s scalability and practicality.

5. Further clarification of concept boundaries

Communicative Language Testing (CLT) targets communicative competence, emphasizing practical language use, whereas traditional assessments focus on linguistic knowledge, prioritizing grammatical and lexical accuracy. A parallel dichotomy exists in assessment methods: algorithm-centered AI prioritizes the standardized scoring of lower-order skills, whereas human assessors excel at the pragmatic evaluation of higher-order abilities through subjective judgment. The context of civil aviation ground service English constitutes a form of professional communication, requiring that both corpora and scenarios adhere to strict industry standards. This section analyzes the logical tensions between CLT and AI assessment within this specific context and explores potential paths for reconciliation.

5.1. Deriving core propositions: A framework for logical argumentation

The logical connection between CLT and AI-automated assessment rests on a fundamental convergence between the objectives of ability measurement and the paths for their technical implementation. This convergence is dictated by the very nature of communicative competence, which CLT seeks to foster, and the data-driven paradigm that underpins AI. Consequently, the design of any new test must prioritize the principle that testing behaviors should authentically mirror real-world language use [14]. The capacity to use language appropriately in authentic contexts, known as communicative competence, is not merely the mastery of discrete grammatical or lexical items but constitutes an integrated synthesis of linguistic, pragmatic, and strategic competencies. To measure this comprehensive ability, CLT must rely on situational interaction, requiring test-takers to demonstrate their abilities through interactions with interlocutors in real or simulated scenarios. In contrast, the core of AI automated assessment lies in data-driven standardization. By processing large volumes of structured or semi-structured data, algorithms identify measurable dimensions of language ability such as grammatical accuracy and vocabulary complexity, and generate consistent assessment results. Therefore, the logical connection between CLT and AI automated assessment must be built on the premise that the ability dimensions of CLT can be converted into data processable by AI. While dynamic competencies like "pragmatic competence" and "strategic competence" were once considered beyond the reach of static algorithms, advances in Large Language Models (LLMs) have enabled a breakthrough. LLMs can now transform this dynamism into analyzable data through contextual semantic understanding, thereby creating a foundation for alignment. In the specific context of civil aviation ground service English, the "target domain usage behavior" is defined as "completing service procedures using standard-compliant English," the core of which is governed by "professional pragmatic rules." For example, if a test-taker responds to a passenger's complaint in a harsh tone, violating the "politeness principle" by saying, "Your bag was lost because you didn't label it properly," AI can use the context of "the passenger is making a complaint" to identify that this response fails to meet the requirement of "calming the passenger's emotions." Conversely, if the test-taker uses a reassuring tone and says, "I'm very sorry about your lost bag. Let's check the tracking system together," AI can accurately determine the appropriateness of this response. Similarly, in a flight delay scenario, an AI model trained on authentic dialogue corpora can learn to distinguish appropriate from inappropriate responses based on the presence or absence of key informational components. An appropriate response typically includes: the reason for the delay (e.g., "due to air traffic control"), the expected duration (e.g., "approximately 90 minutes"), and actionable advice (e.g., "please monitor the airport display boards for updates"). An inappropriate response, by contrast, lacks these elements—for example, a vague statement like, "Your flight is delayed. Wait here." Since such communicative norms are relatively fixed and explicitly codified within the domain, they can be effectively translated into algorithmic logic through training on real-world civil aviation service dialogues.

For the assessment of more complex and dynamic communicative abilities, logical consistency can be established through an "ability-level division of labor" that integrates human calibration. This approach is necessary because evaluating higher-order strategic competencies, such as the overall "reasonableness" of a response, inherently requires human judgment grounded in social and contextual experience. Within this framework, lower-level competencies, including grammatical accuracy, vocabulary range, and basic pragmatic conventions, are efficiently evaluated by AI to ensure objectivity and scoring consistency. In contrast, higher-level abilities such as the application of complex communication strategies and nuanced emotional recognition are assessed through human-AI collaboration, thereby addressing the demands of dynamic interaction. This division is not adversarial but fundamentally complementary, with both components operating synergistically within a unified assessment system to serve the overarching goal of measuring communicative competence. The resultant logical reconciliation between CLT and AI-driven assessment is predicated on an "alignment of ability objectives and technical pathways." This model effectively resolves the inherent tension between dynamic communication and static algorithms through the aforementioned ability-level division of labor. Furthermore, it ensures construct validity by bounding the AI's scope with "professional register rules" and "professional corpus training." This coherent logical chain provides a rigorous theoretical foundation for the subsequent analysis of implementation pathways within human-AI interaction scenarios for civil aviation ground service English.

5.2. Bridging theoretical divides: Reconciling CLT and AI-driven assessment for innovative assessment

From the CLT perspective, Lyle Bachman's "Model of Language Communicative Competence" argues that assessment should cover the complete chain of "knowledge application–strategy selection–situational interaction." This aligns closely with the proposition of this study that "the competence dimensions of CLT need to be converted into AI-processable data." Both jointly point to the core logic that "competence assessment must balance comprehensiveness and operability," and serve as the direct theoretical basis for the design of the human-AI interaction scenario for civil aviation ground service English in this study. Specifically, we define "professional authenticity" as "scenarios that replicate core processes and corpora derived from real dialogues." In essence, this translates the principle of "consistency" into the professional field of civil aviation, thereby enabling the assessment of test-takers' ability to use English in specific professional contexts [15]—a continuation of CLT's core tradition of "situational adaptation." From the perspective of AI-enabled assessment, this study's framework resonates with the "digital-first assessment ecology" advocated by Burstein *et al.*, which underscores the importance of human-AI collaboration. This alignment strongly supports our proposed model of "ability-level division of labor." Both perspectives converge on an assessment logic wherein technology assists rather than replaces human judgment. This consensus indicates that the present study does not subvert classical theories, but rather deepens and specifies classical logic within a professional context.

Regarding the core debate in the field, whether CLT's dynamic communication is compatible with AI's static algorithms, this study, through proposition deduction, finds that the root of the debate lies in the neglect of "differences in ability levels." In fact, "dynamic communication" can be decomposed into two levels: "low-level pragmatic adaptation" and "high-level strategic decision-making." Low-level adaptation refers to routine adjustments based on context, which can be achieved through the contextual semantic understanding capability of large language models (LLMs). For instance, AI can recognize the context of "passenger complaints" and judge whether a response complies with the "politeness principle." High-level decision-making, by contrast, involves "complex experience-based choices" that require human calibration. When the condition of "differences in ability levels" is incorporated, the "dynamic nature" of CLT and the "static nature" of AI are no longer opposing forces. Instead, they form a complementary relationship within the same assessment system, where technology covers standardized low-level abilities, and human judgment oversees non-standardized high-level abilities. This resolution not only addresses the long-standing debate but also provides an operable logical bridge for the integration of CLT and AI.

6. Conclusion

Centering on the core issue of logical reconciliation between CLT and AI automated assessment in the context of human-AI interaction for civil aviation ground service English, this study integrates the particularity of professional contexts, hierarchical differences in competence, and functional division of labor between humans and AI into a unified analytical framework through conceptual deconstruction, proposition deduction, and theoretical dialogue. It not only retains the core essence of CLT, i.e., the comprehensiveness of communicative competence, but also responds to the development needs of AI technology in terms of standardization and efficiency. Furthermore, it provides a practical theoretical tool tailored to the characteristics of professional fields such as civil aviation. Admittedly, this study has limitations: it does not conduct a more in-depth refinement of indicators for the manual calibration standards of high-level competence. In the future, this aspect can be further improved by combining specific cases of civil aviation ground service. However, on the whole, this study clarifies the logical connection between CLT and AI automated assessment through theoretical deliberation. It offers a new professional contextualization perspective for the continuation of classical theories in the technological era, and also provides an operable theoretical tool for the intelligent development of language testing in professional fields. This not only represents adherence to the core spirit of CLT, but also reflects the rational utilization of the value of AI technology, and further responds to the contemporary demand of technology empowering education.

Funding

This work was supported by the Vocational Education Research Project from China Commercial Technicians Association: "Research on the Construction and Application of AI-Enabled English Testing System for Civil Aviation Ground Services" (20ZSJYB20250420); the Education Science Planning Projects (Higher Education Special Program) from Guangdong Provincial Department of Education: "Research on the Evaluation System of Digital Competence in Curriculum Ideology and Politics for Higher Vocational Teachers in Guangdong under the Background of Educational Digitalization" (2024GXJK877) and "Digital Empowerment for High-Quality Development in Guangdong: An Innovative Study on Cultivating Interdisciplinary Foreign Language Talents" (2023GXJK691).

Disclosure statement

The author declares no conflict of interest.

References

[1] Chen X, 2010, The Shift of "Authenticity" Criteria under Bachman's Communicative Language Testing Theoretical Model. Journal of Jiangsu University (Social Science Edition), (02): 81–84.

[2] Xue R, 2008, Communicative Language Testing: Theoretical Models and Assessment Criteria. Foreign Language Teaching, (03): 68–76.

[3] Burstein J, LaFlair, GT, Von Davier AA, 2025, A Theoretical Assessment Ecosystem for Digital-First Language Assessment: The Duolingo English Test, Duolingo Research Report DRR, 4–16.

[4] Wang H, Wu H, He Z, et al., 2022, Progress in Machine Translation. Engineering, 18(11): 143–153.

[5] Canale MS, Swain M, 1980, Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing, Applied Linguistics, 31–47.

[6] Russell S, Norvig P, 2020, Artificial Intelligence: A Modern Approach (4th Edition), Pearson, Hoboken.

[7] Jin Y, Xu MJ, 2024, Ethical Reflections on Intelligent Technology Empowering Language Assessment. Foreign Languages and Their Teaching, (03): 1–10 + 145.

[8] Hao J, von Davier A, Yaneva V, et al., 2024, Transforming Assessment: The Impacts and Implications of Large Language Models and Generative AI. Educational Measurement: Issues and Practice, (43): 16–29.

[9] Alam S, Usama M, Alam MM, et al., 2023, Artificial Intelligence in the Global World: A Case Study of Grammarly as an E-Tool on ESL Learners' Writing at Darul Uloom Nadwa. International Journal of Educational Technology in Higher Education, 13(11): 1741–1747.

[10] Schaefer E, Martin J, 2023, Language Testing in Changing Times: An Interview with Professor Daniel Isbell, TEVAL: Technology-Enhanced Language Assessment Review, 1–5.

[11] Mahmoud RH, 2022, Implementing AI-Based Conversational Chatbots in EFL Speaking Classes: An Evolutionary Perspective. Journal of Technology-Enhanced Language Learning, 1–21.

[12] Ockey G, Chukharev-Hudilainen E, 2021, Human Versus Computer Partner in the Paired Oral Discussion Test. Applied Linguistics, 42(5): 924–944.

[13] Spolsky B, 1985, The Limits of Authenticity in Language Testing. Language Testing, 31–40.

[14] Bachman LF, Palmer AS, 1996, Language Testing in Practice, Oxford University Press, Oxford.

[15] Douglas D, 2000, Assessing Language for Specific Purposes, Cambridge University Press, New York.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.