

A Review on the Application of Language Proficiency Scales in Speaking Tests

Tang Jing ^{1,*}

¹ School of Foreign Languages and Literature, Beijing Normal University, Beijing, China

* Correspondence: Tang Jing, School of Foreign Languages and Literature, Beijing Normal University, Beijing, China

Abstract: This article begins with a review of Language Proficiency Scales (LPS) and established speaking-test scales, placing particular emphasis on how these frameworks conceptualize oral proficiency and how they are used to support consistent assessment across diverse learning contexts. The discussion covers major theoretical foundations related to the description of communicative language ability, including the functional interpretation of speaking performance, principles of rating-scale development, the hierarchical subdivision of speaking skills, and the increasing need for contextualized task design in oral assessment. Building on this theoretical background, the article then examines how proficiency scales can be systematically linked to speaking tests, using the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR) and China's Standards of English Language Ability (CSE) as practical cases that demonstrate alignment processes, scale interpretation, and challenges in operationalizing descriptors for real testing situations. In addition, the article provides an overview of emerging trends in research on speaking assessment, such as fine-grained performance profiling, the integration of multimodal evidence, and advances in technology-enhanced rating methods, and discusses how these developments may inform future applications, refinement, and innovation of the CSE in both pedagogical and assessment settings.

Keywords: language proficiency scales; speaking tests; Common European Framework of Reference; China's Standards of English Language Ability

1. Introduction

A Language Proficiency Scale (LPS) is a structured set of descriptors that outline a language user's ability to use a given language, typically organized into hierarchical levels that represent progressive stages of proficiency development. Such scales function as essential tools for interpreting language attainment and for guiding curriculum design, instructional practice, and assessment development. Studies have indicated that a national foreign language proficiency scale provides a multi-level description of English language ability and contributes to constructing a coherent theoretical basis for English learning, teaching, and evaluation [1,2]. Research on major international foreign language proficiency scales offers valuable insights for the establishment of China's Standards of English Language Ability (CSE) and for the development of corresponding assessment instruments.

By drawing on international experience in linking proficiency scales to testing systems, researchers in China aim to align the CSE with both international and domestic English examinations. This alignment supports more direct interpretation of test results through a unified descriptive framework, thereby enhancing the transparency, comparability, and practicality of English assessments. Such efforts also contribute to

Published: 22 November 2025



Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

improving the coherence of testing practices within China and promoting greater compatibility between the CSE and widely recognized proficiency frameworks.

2. Overview of Proficiency Scales and Proficiency Description Theories

2.1. Language Proficiency Scales and Oral Proficiency Scales

Studies of language proficiency scales began in the early 1950s, when the United States developed a series of oral proficiency standards to assess overseas military personnel. The FSI scale categorizes speaking performance into basic, limited working, basic vocational, vocational, and native or bilingual levels [3]. It was the first to use face-to-face interviews to evaluate oral communication and to describe real-life oral communicative competence. However, distinctions between low and medium levels were insufficient, transitions between levels were not smooth, and the scale only itemized language skills without an overall assessment.

Around the same period, teaching-oriented standards were developed in the form of ACTFL, which refined the low and medium levels to reflect minor progress in language learning. It covered four skills-listening, speaking, reading, writing-and considered the learner's education level.

Speaking Rating Scales typically divide oral proficiency into four levels-superior, advanced, intermediate, and novice-with sub-levels for finer differentiation. These scales describe linguistic characteristics in specific contexts using criteria such as global tasks and functions, context, accuracy, and text type.

The Canadian Language Benchmarks (CLB) scale, issued in 1996, evaluates language proficiency of immigrants. It divides proficiency into three stages and twelve benchmarks, with speaking assessed through ability, knowledge, and strategy, further subdivided by communicative tasks in social interaction, instruction, persuasion, and information sharing. The benchmark levels form a clear continuum of knowledge and skills.

In Europe, the ALTE and CEFR scales are highly influential. The ALTE scale uses "Can-do Statements" to describe language ability at each level, comparing with the previous level, and differentiates productive (speaking, writing) and receptive (listening, reading) skills across social, academic, and professional contexts. Only typical tasks are listed due to the impossibility of covering all language use situations.

The CEFR comprehensively describes knowledge, skills, and cultural background required for effective communication, establishing competence levels for learner assessment [4]. Oral competence is described in terms of range, accuracy, fluency, interaction, and coherence, while communicative competence is represented through communicative activities, strategies, and overall competence. Its application in teaching is limited by readability and lack of scale-related glossary [5]. Both ALTE and CEFR draw on the theory of communicative language ability, with CEFR adopting an action-oriented approach that categorizes language actions into output, reception, and mediation [5].

Early scales primarily served evaluation purposes, gradually evolving into standards for language teaching and assessment, influencing curricula, policies, and syllabus design. These frameworks are valuable references for formulating unified language proficiency standards in China.

The China Standards of English Language Ability (CSE), implemented in 2018, adopts a use-oriented approach based on communicative language ability [6]. Language proficiency is divided into comprehension (listening, reading), expression (oral, written), and mediation (translation, interpreting), scaled into nine levels grouped into elementary (CSE 1-3), intermediate (CSE 4-6), and advanced (CSE 7-9). Each descriptor includes performance, criteria, and conditions, emphasizing the strategies learners use to accomplish communicative tasks.

CSE comprehensively defines English proficiency standards across listening, speaking, reading, writing, and translation. Descriptors include typical activities, knowledge, and strategies for each subskill, supporting scientific evaluation and sub-scale

selection. Descriptor sources include Chinese curriculum standards, syllabi, examination materials, and foreign proficiency frameworks. Indicators of each descriptor were tracked, forming a quality hierarchy to optimize classification, understanding, and application. A Chinese-characteristics descriptor repertoire ensures scientific validity and sustainability.

The CSE Oral Scale comprises oral expression skills and strategies, describing communication activities and the corresponding competencies required for learning and teaching spoken English. Some subscales (oral explanation, exposition, and interaction) address field-specific and academic competencies, though alignment with the total scale is sometimes inconsistent, and transitions between levels may be abrupt. The CSE aims to describe general English proficiency progression regardless of learner age or background, requiring ongoing evaluation and refinement.

2.2. Theories Relating to Oral Proficiency Description

The formulation of language proficiency scales reflects evolving understanding of language ability and represents the development of language teaching and learning theories. Continuous advancements in these theories drive ongoing revisions and improvements of proficiency scales. Both CEFR and CSE are based on communicative language ability theory. Recent years have seen significant progress in oral proficiency description, including refined rating scales, subdivision of speaking skills-particularly pragmatic and interactive skills-and greater contextualization of speaking tests. These developments support more precise evaluation and the enhancement of CSE-based assessments.

2.2.1. Bachman's Mode of Communicative Language Ability (CLA)

Bachman's CLA model comprises language competence, strategic competence, and psychophysiological mechanisms. It expands earlier models by distinguishing components of knowledge from skills and by describing how these components interact with each other and with the context of language use [7].

Language competence is divided into organizational and pragmatic competence. Organizational competence involves controlling grammatical structures, understanding propositional content, and arranging sentences into coherent texts. Pragmatic competence includes illocutionary and sociolinguistic competence. Illocutionary competence enables expression and interpretation of ideational, manipulative, heuristic, and imaginative functions according to sociocultural norms and discourse features. Sociolinguistic competence ensures sensitivity to dialect, register, cultural references, and conventions, allowing appropriate performance in varied language contexts.

Strategic competence encompasses assessment, planning, and execution, enabling language users to apply their competences effectively in context. Psychophysiological mechanisms involve neurological and physiological processes used during language execution, including auditory and visual skills for receptive use and articulatory skills for productive use, such as correct pronunciation, stress, and intonation.

2.2.2. Cohen's Rating Scale for Pragmatic Speaking

Cohen's pragmatic speaking scale evaluates oral proficiency across two dimensions: form and content. Form examines naturalness of discourse, stylistic appropriateness, and clarity of expression. Content assesses suitability of expression, accuracy of conveyed information, and the completeness of information provided. This multidimensional approach enables comprehensive evaluation of both linguistic and pragmatic competence [8].

2.2.3. Nunn's Rating Scales for Small Group Interaction

Nunn's small group interaction scales include four ratings: turn-taking and negotiation, content of contributions, pronunciation and intonation, and grammar and

vocabulary. The first two assess interactive skills, such as maintaining conversation and exchanging information. The latter two evaluate intelligibility through microlinguistic accuracy. These scales cover both organizational and pragmatic competence, emphasizing the role of negotiation and interactive dynamics in oral communication [9].

2.2.4. Fulcher's Performance Decision Trees (PDT)

Performance Decision Trees (PDT) prioritize performance effects by integrating data from specific communicative contexts. PDT combines detailed performance descriptors with simplified decision-making procedures. Interactional competence in service encounters is assessed through discourse competence, discourse management, and pragmatic competence. Discourse competence supports structuring questions and responses coherently. Discourse management focuses on managing transitions, clarifying purposes, defining roles, closing interactions, and using backchannel signals. Pragmatic competence ensures appropriate interaction and rapport through relational communication [10].

While these frameworks provide foundational guidance for oral proficiency assessment, several challenges remain. These include scaling tacit knowledge such as cultural awareness, intercultural communication, and critical thinking; accurately describing specific contexts like classrooms, social situations, and workplaces; and ensuring descriptor quality, readability, classification rationale, and differentiation. Addressing these challenges is essential for advancing oral proficiency scale development and improving their scientific validity and practical application.

3. Cases of Linking Proficiency Scales to Speaking Tests

Analyzing current practices of designing domestic speaking tests based on CEFR and linking foreign speaking tests to the CSE provides valuable insights for aligning China's English tests with international standards and for designing domestic tests guided by the CSE. Such alignment enhances the international recognition and acceptance of China's tests while promoting the practical application of CSE within the national English testing system.

3.1. Cases of Designing Domestic Speaking Tests Based on the CEFR

1) Test of English Proficiency Oral (TEP)

The TEP, part of the comprehensive reform project for College English in municipal universities, is guided theoretically by the College English Curriculum Requirements and the CEFR [11]. It emphasizes oral communicative interaction, flexibility, and authenticity. Oral proficiency for non-English majors is divided into three stages: elementary, intermediate, and advanced. Assessment focuses on four aspects: communication effect, content organization, pronunciation and intonation, and grammar and vocabulary.

Elementary stage: Can discuss daily topics in English, deliver short prepared speeches on familiar topics, and engage in simple discussions.

Intermediate stage: Can communicate fluently on general topics, express personal opinions and feelings, and state facts, reasons, or describe events clearly.

Advanced stage: Can engage in fluent and accurate communication on general or professional topics, summarize complex oral or written materials, and provide evaluations.

2) Test of Oral Proficiency in English (TOPE)

TOPE was designed to assess the oral communicative ability of college students [12]. The rating scale levels are aligned with CEFR reference levels. Researchers analyzed students' language functions with video-recorded examples and used Conversation Analysis to identify speaking features differentiating proficiency levels. Although fully linking language functions to proficiency levels proved challenging, TOPE highlighted

task processes, contextual factors, and scoring features, offering guidance for future development of rating scales.

3.2. Cases of Linking Foreign Speaking Tests to The CSE

Several international tests have been aligned to the CSE to facilitate comparability with China's proficiency standards.

1) IELTS Speaking Test

The IELTS Speaking test covers CSE levels 3 to 7, with most descriptors drawn from levels 5 and 6 [13]. Section 1 targets level 4, Section 2 mainly aligns with level 5, and Section 3 includes descriptors from levels 6 and 7. The test addresses a range of text function categories, such as description, exposition, argumentation, and interaction. Interactional and managing interaction functions are emphasized in Section 3, reflecting the need for abstract and argument-based responses.

2) Aptis Speaking Test

Aptis spans four CSE levels, from 3 to 6. Each part of the test shows progression in difficulty while allowing some overlap between adjacent parts [13]. Part 1 targets CSE 3, Part 2 CSE 3-4, Part 3 CSE 4-5, and Part 4 CSE 4-6. The test is designed to evaluate communicative ability across multiple contexts, excluding instruction and interaction functions, which are addressed elsewhere.

3) TOEFL Speaking Test

The TOEFL iBT test engages multiple skills to simulate academic language use [6]. Alignment with the CSE identified 32 relevant scales, with speaking cut scores ranging from CSE 4 to CSE 8. Except for CSE 5, cut scores were rounded down to align more closely with established TOEFL-iBT-IELTS concordance mappings. CSE descriptors cover and extend beyond CEFR and other curriculum standards, supporting their transformation into courses and assessment tools for diverse learners.

3.3. He Lianzhen's Linking Use Argument

The Linking Use Argument framework, derived from the Assessment Use Argument, systematically connects test development, score interpretation, and result application [14]. LUA emphasizes evidence collection and data analysis to ensure linking outcomes benefit all stakeholders. Key principles include:

- 1) Decisions based on linking results should be beneficial and equitable.
- 2) They must align with educational, social, and legal norms.
- 3) Information used for decision-making should be relevant, sufficient, meaningful, generalizable, and impartial.
- 4) Linking processes, experts, and standard-setting methods should produce consistent results.

Linking tests to proficiency scales establishes relationships between scores and standards, defines cut scores, and clarifies the abilities required for different tasks. The benefits include promoting systematic, modernized, and internationally aligned language education, improving evaluation and selection by schools, and supporting learners' autonomous development and teachers' assessment competence. Risks such as misinterpretation of linking purposes, oversimplified results, and unscientific implementation can be mitigated through public education, clear guidance, and standardized, expert-led linking procedures.

4. New Trends of Speaking Test in the Era of Digital Humanities

Recent developments in speaking test research, in terms of testing methods and research perspectives, offer both technical support and innovative insights for enhancing the CSE framework and designing CSE-based tests. These trends reflect the integration of digital technologies and data-driven approaches into language assessment, facilitating more precise, efficient, and contextually valid evaluation of oral proficiency.

4.1. New Practice Methods of Speaking Test

1) Computerization and Automation

The integration of computer technologies into English teaching has led to the development of computerized oral proficiency assessments, commonly referred to as Computerized Oral Proficiency Tests (COPT) [15]. These tests focus on optimizing question types, scoring methods, test administration procedures, and the reliability and validity of spoken language assessments. The automation of scoring and test delivery provides increased efficiency and objectivity in evaluating speaking performance. Automated scoring systems, such as EduRater, illustrate practical approaches to designing reliable scoring mechanisms, integrating algorithmic analysis of speech features with structured rating criteria. This digital approach allows for large-scale testing and consistent assessment standards, enhancing comparability and fairness in oral proficiency evaluation [16].

2) English Learners' Spoken Corpora

Learner corpora play a crucial role in identifying linguistic features that distinguish proficiency levels, forming the basis for scoring and proficiency criteria. Notable examples in China include the College Learners Spoken English Corpus (COLSEC) and the Spoken English Corpus of the Public English Test System (SECOPETS). COLSEC covers 39 spoken test topics categorized into personal life and study, social concerns, and campus life, providing a comprehensive resource for systematic investigation of EFL learners' spoken performance. SECOPETS, on the other hand, is designed to evaluate the current oral English abilities of Chinese learners across different backgrounds and to explore effective strategies for enhancing oral proficiency [17]. These corpora support evidence-based test design, offer authentic examples of learner language, and contribute to the development of valid, context-sensitive scoring rubrics [18].

4.2. New Perspectives of Speaking Test

Cognitive diagnosis has emerged as a powerful approach in speaking assessment, drawing on advances in cognitive psychology, item response theory, and statistical modeling [19]. This method infers students' unobservable cognitive states from their observable test responses, allowing precise measurement of specific knowledge structures and processing skills. In speaking tests, cognitive diagnosis enables detailed analysis of individual performance, highlighting strengths and weaknesses in specific competencies. This information provides test stakeholders with richer, more actionable feedback on learners' oral proficiency. Recent efforts have focused on developing computer-adaptive language testing systems based on cognitive diagnosis principles (CD-CALT), which dynamically adjust item selection according to real-time diagnostic results. Such systems enhance the efficiency, fairness, and granularity of oral proficiency assessment, representing a significant shift toward data-driven, personalized evaluation in the digital humanities era [20].

5. Conclusion

Language proficiency scales articulate the complex system of language development in a clear and accessible manner, providing benchmarks, references, and standardized measures for comparing different language proficiency systems. Such scales hold considerable significance for both language teaching and assessment. The China's Standards of English Language Ability (CSE) facilitates the alignment and mutual recognition of test scores, degrees, and other qualifications with international standards, promoting greater comparability and transparency in English language evaluation.

The CSE addresses several longstanding challenges in China's foreign language education system, including the disconnect between teaching and testing objectives, inconsistencies in examination standards, inadequate cohesion across learning stages, and vague classification of language proficiency levels. Through these efforts, the CSE has

been formally incorporated into the regulations and standards system of the State Language Commission for the first time, providing an official framework for the assessment of English proficiency.

The development and implementation of the CSE offer a systematic framework and practical guidance for designing and comparing tests across different educational and assessment contexts. Nonetheless, challenges remain in its practical application to speaking tests, particularly regarding how to integrate specific professional language skills required by learners' majors or future careers into the proficiency descriptors. To address these challenges, future research can draw upon the successful experience of CEFR, constructing a scientifically grounded and operationally practical proficiency description system. By leveraging advances in technical methods and research perspectives, the CSE's usability and effectiveness can be further enhanced, enabling it to play a central role in the ongoing reform and modernization of China's English testing system.

References

1. B. North, "Scales of language proficiency," *Melbourne Papers in Language Testing*, vol. 4, no. 2, pp. 60-111, 1995.
2. P. Leeming, and J. Harris, "Measuring speaking proficiency growth in the language classroom: An investigation of practical approaches for teachers," *Language Teaching Research*, vol. 29, no. 7, pp. 2963-2985, 2025.
3. J. L. Clark, and R. T. Clifford, "The FSI/ILR/ACTFL proficiency scales and testing techniques: Development, current status, and needed research," *Studies in Second Language Acquisition*, vol. 10, no. 2, pp. 129-147, 1988.
4. Council of Europe. Council for Cultural Co-operation. Education Committee. Modern Languages Division, "Common European framework of reference for languages: Learning, teaching, assessment," *Cambridge University Press*, 2001.
5. N. H. De Jong, "Assessing second language speaking proficiency," *Annual Review of Linguistics*, vol. 9, no. 1, pp. 541-560, 2023.
6. S. Papageorgiou, S. Wu, C. N. Hsieh, R. J. Tannenbaum, and M. Cheng, "Mapping the TOEFL iBT test scores to China's Standards of English language ability: Implications for score interpretation and use," *ETS Research Report Series*, vol. 2019, no. 1, pp. 1-49, 2019.
7. L. F. Bachman, "Fundamental considerations in language testing," *Oxford university press*, 1990. doi: 10.2307/329499
8. A. D. Cohen, "Testing Language Ability in the Classroom," 1980. doi: 10.2307/324160
9. R. Nunn, "Designing rating scales for small-group interaction," *ELT Journal*, vol. 54, no. 2, pp. 169-178, 2000. doi: 10.1093/elt/54.2.169
10. G. Fulcher, F. Davidson, and J. Kemp, "Effective rating scale development for speaking tests: Performance decision trees," *Language Testing*, vol. 28, no. 1, pp. 5-29, 2011. doi: 10.1177/0265532209359514
11. Z. Xihua, "The Study of Test of English Proficiency Oral for Non-English Majors," *Journal of Beijing International Studies University*, vol. 36, no. 8, p. 70, 2014.
12. L. Liu, and G. Jia, "Looking beyond scores: validating a CEFR-based university speaking assessment in Mainland China," *Language Testing in Asia*, vol. 7, no. 1, p. 2, 2017. doi: 10.1186/s40468-017-0034-3
13. Q. Zhang, "Impacts of World Englishes on local standardized language proficiency testing in the Expanding Circle: A study on the College English Test (CET) in China," *English Today*, vol. 38, no. 4, pp. 254-270, 2022.
14. C. J. Weir, "Language testing and validation," *Hampshire: Palgrave MacMillan*, vol. 10, p. 9780230514577, 2005. doi: 10.1057/9780230514577
15. L. Zuo, "Computer network assisted test of spoken English," *Computer Systems Science and Engineering*, vol. 34, no. 6, pp. 319-323, 2019. doi: 10.32604/csse.2019.34.319
16. W. Hu, "A study on the scoring method of oral English test in college English online computer test," In *EAI International Conference, BigIoT-EDU*, July, 2022, pp. 25-36. doi: 10.1007/978-3-031-23950-2_4
17. D. Man, "A Longitudinal Corpus Investigation of Grammatical Stance Markers in the English Argumentative Writing of Chinese Students (Doctoral dissertation, University of Malaya (Malaysia)),", 2020.
18. Y. Yu, and G. Zeng, "Lesser relevance markers in Chinese academic spoken English corpus: a cross-disciplinary study on pragmatic features," *Frontiers in Psychology*, vol. 14, p. 1297038, 2023. doi: 10.3389/fpsyg.2023.1297038
19. Y. W. Lee, and Y. Sawaki, "Cognitive diagnosis approaches to language assessment: An overview," *Language Assessment Quarterly*, vol. 6, no. 3, pp. 172-189, 2009.
20. M. Heidari Vincheh, A. Mirzaei, and A. Roohani, "A cognitive diagnostic approach to IELTS speaking test: unveiling the subskills and test-takers' perceptions," *Language Testing in Asia*, vol. 14, no. 1, pp. 1-21, 2024. doi: 10.1186/s40468-024-00311-2

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the publisher and/or the editor(s). The publisher and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.