

Constructing a Quality Evaluation Framework for Chinese Language Learning: A Focus on International Students in China

Xinyi Tian ¹, Pingping Zhang ^{2,*}

¹ Beijing Language and Culture University, Beijing 100083, China

² School of Education, Zhengzhou University, Zhengzhou 450001, China

*** Correspondence:**

Pingping Zhang

zhangpingping@zzu.edu.cn

Received: 15 June 2026 / Accepted: 24 June 2026 / Published online: 28 June 2026

Abstract

Constructing a rigorous quality evaluation system for Chinese language learning is essential for advancing international education in Chinese universities. After defining the concept of learning quality, we delineate an evaluation framework for international undergraduates as a systematic, outcome-oriented structure that integrates multidimensional, longitudinal, and multilevel influences on the learning process. Based on a review of existing systemic deficiencies and globally recognized educational evaluation models, we propose four pathways to operationalize this framework: instituting a comprehensive evaluation policy, cultivating an effective multi-stakeholder evaluation mechanism, developing a hierarchically designed indicator system, and creating a dynamic, interactive feedback loop to drive continuous educational improvement.

Keywords: International Students in China; Quality of Chinese Language Learning; Evaluation System; Educational Evaluation Reform

1. Introduction

In recent years, the enrollment of inbound international undergraduates in China has expanded significantly. According to data from the Ministry of Education (MOE), the total international student population surged from 61,869 (including 11,797 undergraduates) in 2001 to 492,185 (including 160,783 undergraduates) in 2018 (MOE, 2011, 2019). In response to this rapid quantitative growth, the General Office of the Chinese Communist Party Central Committee and the State Council jointly issued the Opinions on Enhancing the Opening-up of Education in the New Era in 2016, explicitly establishing the goal of elevating the quality of international education (China, 2016). Consequently, China's international education sector has shifted from a scale-driven model toward a quality-oriented development paradigm. Current priorities center on

systematically improving the professionalism, methodological rigor, standardization, and institutionalization of international education programs (Daily, 2023).

Given the diverse countries of origin among international undergraduate students, the high heterogeneity in their academic readiness upon admission, and the substantial disparities in institutional management capacities across universities, improving educational quality requires a comprehensive approach. This approach must encompass multiple stages, including admissions, entrance assessments, academic training, administration, and evaluation. Specifically, it is crucial to construct an effective quality evaluation system for international undergraduate education, one that meets the demands of contemporary quality-oriented development and aligns with China's national and educational contexts. Such a system would enable a comprehensive assessment and diagnosis of institutional teaching and learning processes, thereby serving as a vital mechanism for enhancing the overall quality and efficacy of international education in China.

Educational policymakers worldwide recognize that the paradigm for evaluating teaching effectiveness has shifted from assessing instructional delivery to measuring student learning outcomes, positioning the latter as the core component of educational quality assessment. Evaluating the actualized learning outcomes of international students provides the most direct and accurate reflection of institutional teaching efficacy and student development processes. Consequently, such evaluation serves as a crucial prerequisite for designing targeted interventions to enhance the overall quality of international education. Within this context, as the primary medium of instruction for inbound international undergraduate students, Chinese language proficiency functions as an indispensable tool for both their academic success and daily sociocultural integration.

Students' Chinese language proficiency directly impacts their disciplinary learning and, consequently, the overall quality of international education at the institutional level. Given the substantial variance in pre-admission Chinese proficiency among inbound undergraduates, universities typically offer remedial or supplementary Chinese language courses post-enrollment. These courses are designed to mitigate academic barriers in students' major studies stemming from language deficiencies.

Currently, the literature concerning international undergraduate education in China predominantly focuses on macro-level issues, such as admission criteria and programmatic requirements (Cui, 2014; Xu *et al.*, 2018). This macro-level focus has left a relative dearth of evaluative research on students' holistic academic development and its underlying determinants. While some scholars have pursued micro-level analyses, these studies tend to be narrowly confined to specific linguistic competencies or unidimensional variables affecting language acquisition. For instance, Hao (2018) demonstrated that Chinese character usage frequency was the most critical determinant in character recognition tasks. Similarly, Fan and Li (2018) employed a learning difficulties inventory to investigate time allocation, instructional quality, and learning barriers among international students in Chinese language foundation programs. Furthermore, drawing on pandemic-era online Chinese courses as a case study, Han *et al.* (2021) revealed that students' adaptability to online environments significantly impacted their satisfaction with language acquisition.

Overall, the existing body of evaluative research remains fragmented, overly reliant on anecdotal and experiential observations, and lacking in comprehensive correlational analyses. Consequently, there is a compelling need to conduct a systematic evaluation of the quality of Chinese language acquisition among inbound international undergraduate students.

The development and implementation of a quality evaluation system for Chinese language acquisition among inbound international undergraduates are instrumental in standardizing quality benchmarks and enhancing the overall educational outcomes of international programs in China. First, the establishment of this evaluation system aligns with the imperative to pursue high-quality international education. Conducting such evaluations provides a clear diagnostic overview of current teaching and learning dynamics, thereby generating comprehensive, evidence-based data to inform targeted quality enhancement interventions. Second, the implementation of this system addresses the pressing need to refine and optimize the existing Quality Assurance (QA) framework for international education. The proposed evaluation framework emphasizes the active involvement of both internal and external stakeholders throughout the assessment process. By operationalizing continuous evaluation, it cultivates a virtuous cycle comprising "evidence-based monitoring, problem identification, targeted remediation, and re-evaluation" (Li, 2021). This collaborative approach fosters dialogue and consensus-building between internal and external assessors, enabling joint pedagogical analyses, facilitating the formulation of quality improvement strategies, and ultimately driving the high-quality development of international education in China.

Building upon a comprehensive review of the existing literature on Chinese language evaluation for inbound undergraduates, and drawing upon well-established theoretical evaluation paradigms, this study constructs a multidimensional quality evaluation framework. This framework is conceptualized across four core dimensions: institutionalizing evaluation mechanisms, coordinating diverse evaluation stakeholders, delineating a robust system of evaluation indicators, and optimizing the utilization of evaluation outcomes.

2. Target Population and Scope of the Evaluation System

2.1. Inbound International Undergraduate Students

In accordance with the definition of international students outlined in the Statistical Indicator System for Educational Monitoring and Evaluation in China (2020 Edition), inbound international undergraduate students are defined as degree-seeking undergraduates who hold foreign passports, are officially enrolled in Chinese higher education institutions (HEIs), and undertake their academic programs with Chinese as the medium of instruction (MOE, 2020). This study specifically focuses on this degree-seeking cohort, as they serve as a critical indicator of the internationalization of higher education and reflect the depth of global engagement and institutional cooperation within Chinese HEIs. Notably, the target population explicitly excludes international undergraduates majoring in the Chinese language or related disciplines at Chinese HEIs.

2.2. Quality of Chinese Language Learning

Stakeholders, including educational policymakers, researchers, institutions, students, and parents, often hold diverse perceptions of "quality." Because the conceptualization of the quality of Chinese language learning among inbound international undergraduates directly determines the architecture of the evaluation system, defining its operational definition serves as a necessary point of departure. According to the Dictionary of Education, educational quality is defined as an evaluation of the level and effectiveness of education, which is ultimately manifested in the competencies of the individuals cultivated (Gu, 1998).

Interpretations of quality vary across international standards and reports issued by global organizations and various nations. For instance, the Indicators of Quality in Brazilian Education: Measuring and Stimulating Improvement places greater emphasis on educational outcomes and descriptive assessments of quality, aiming to establish a foundation for educational accountability (Buchmann & Neri, 2008). Conversely, the European Report on the Quality of School Education focuses centrally on educational outcomes while highlighting institutional processes, such as school resources, infrastructure, curriculum, teaching staff, and administration, yet it largely overlooks the influence of individual-level student factors on these outcomes (Communities, 2001). Meanwhile, the OECD Education Indicators encompass a broader range of dimensions, including educational and learning outcomes, financial resources, educational opportunities, and learning environments, thereby systematically examining the combined influence of both institutional and individual student-level factors on educational outcomes (Bottani, 1996).

Additionally, Li and Xin (2021) synthesized domestic and international perspectives, proposing that educational quality is a multidimensional, multi-tiered, and value-laden concept that reflects societal expectations for high-quality education. Consequently, the conceptualization of educational quality should transcend current metrics to adopt a future-oriented perspective, thereby reflecting the trajectory of educational reform and development. Evidently, educational quality is a multifaceted and systemic concept wherein various interacting components collectively constitute a cohesive framework.

Within this systemic framework, the quality of learning serves as the core. It refers to the evaluation of the processes and outcomes of students' academic progression and development at a specific stage relative to predetermined educational objectives; this evaluation entails both the assessment of actualized learning outcomes and the appraisal of various contextual factors that influence academic achievement (Jia, 2009). Consequently, conceptualizing the quality of Chinese language learning for inbound international undergraduates should encompass the entire trajectory of student development rather than being confined solely to summative academic grades. This necessitates the establishment of an evaluation system grounded in a holistic view of quality.

Specifically, the quality of Chinese language learning for these students is defined as a systematic process anchored in predetermined educational goals. It involves the continuous collection of data regarding students' Chinese language learning outcomes and their corresponding determinants throughout their undergraduate studies. This process enables

stakeholders to describe, evaluate, and adjust language-learning activities to ensure that educational objectives are effectively realized.

2.3. The Evaluation System

Guided by specific value orientations, an evaluation system is defined as a systematic construct comprising two or more organically linked and interacting elements. This system assesses educational quality and its determinants against established standards to provide feedback and establish regulatory mechanisms, thereby optimizing educational decision-making and driving quality enhancement (Pan, 2004).

As the conceptualization of educational quality continuously evolves, its evaluative dimensions have become increasingly comprehensive and systemic. Consequently, the evaluation system for the quality of Chinese language learning among inbound international undergraduates during their studies in China must address four fundamental questions: why to evaluate (purpose), who to involve in evaluation (assessment agents), what to evaluate (content), and how to evaluate (methodology).

From a systems theory perspective, the evaluation system can be structured into three core components: input, process, and output evaluation. In terms of its constituent elements, this system encompasses evaluation objectives, assessment agents, evaluation standards, evaluative content, evaluation methodologies, and the utilization of evaluation outcomes.

3. Current Status and Challenges of the Evaluation System

Evaluating the Chinese language learning outcomes of inbound international undergraduates necessitates the development of valid measurement instruments, the innovation of quality assessment paradigms, and the adoption of comprehensive, outcome-based evaluations as pivotal benchmarks of educational quality in China. However, in practice, a substantial gap persists between these objective evaluative imperatives and the current state of Chinese language learning quality assessment.

3.1. Lack of Policy and Institutional Mechanisms for Chinese Language Learning Evaluation

Promulgated in 2018, the Quality Standards for Higher Education of International Students in China (Trial) (hereafter referred to as the Standards) represent the Ministry of Education's first regulatory document dedicated to international education. While the Standards stipulate admission thresholds and graduation criteria regarding the Chinese language proficiency of international students enrolled in Chinese-medium programs, they fail to articulate explicit requirements for the instructional process (MOE, 2018). Consequently, they do not provide a sufficient baseline for evaluating the quality of students' Chinese language acquisition throughout their degree programs.

Furthermore, in the 2021 Audit and Evaluation Indicator System for Undergraduate Education and Teaching in Regular Higher Education Institutions (Trial), "the development of international

student education" was embedded as an audit focus under the primary indicator "Student Development" (MOE, 2021). However, because international education was not established as an independent, standalone indicator within this audit framework, the system lacks the capacity to comprehensively reflect the actual pedagogical realities of international programs.

At the national level, a systematic and robust, government-mandated institutional framework for evaluating international students' learning quality remains absent. At present, optimizing macro-level policy architecture, formulating highly targeted and actionable guidelines grounded in institutional realities, and formally establishing a quality evaluation system for inbound international undergraduates constitute the most pressing priorities in the quality assurance of international education in China.

3.2. Inadequate Mechanisms for Multi-Stakeholder Participation

Although the Standards provide a foundational basis for guiding international higher education activities and facilitating institutional self-evaluation, the policy fails to clearly demarcate the specific responsibilities and division of labor among key actors, namely, the government, third-party social organizations, and HEIs themselves, in conducting learning quality evaluations.

At present, host HEIs remain the primary agents responsible for evaluating the quality of Chinese language learning. However, the prevailing practice of "lenient admission and graduation" (often colloquially characterized as "easy in, easy out") significantly dampens these institutions' initiative and motivation to undertake rigorous self-evaluations (Wang, 2014). From an organizational perspective, HEIs predominantly regulate international education through administrative measures, such as admissions, curriculum delivery, infrastructure, and security mechanisms. Crucially, they lack dedicated, independent departments tasked with evaluating international student learning outcomes. Consequently, institutional self-evaluation and self-directed development mechanisms remain underdeveloped.

Furthermore, government oversight regarding the quality evaluation of Chinese language acquisition among international undergraduates remains deficient. Specifically, state authorities have failed to establish regulatory guidelines for overarching strategic areas, including the value orientations, core objectives, and assessment standards of the evaluation framework. There is also a lack of standardized licensing procedures and industry benchmarks for professional evaluation agencies, practitioners, and social organizations. Currently, research by specialized agencies remains heavily skewed toward developing high-stakes entrance examinations for international students. Consequently, there is a distinct shortage of scientifically rigorous, impartial, and professional third-party evaluation mechanisms that specialize in the formative (process-oriented) assessment of international students' Chinese language acquisition.

3.3. Absence of Validated Chinese Academic Proficiency Tests and Comprehensive Evaluation Indicator Systems

International undergraduates' Chinese language proficiency, represented by their academic achievement in Chinese, lies at the core of the quality evaluation system. A foundational task of this system is to describe the current status and developmental trajectory of Chinese language education quality, thereby identifying systemic strengths and weaknesses. However, to date,

China has not established standardized Chinese academic proficiency assessments specifically tailored for the undergraduate stage. Consequently, boundaries remain blurred between general language proficiency tests and academic achievement tests, leading to unclear role positioning for both. Furthermore, existing research rarely addresses the formative assessment of language learning outcomes, suffering from a chronic imbalance that prioritizes summative outcomes over developmental processes.

The most widely recognized examination, the Chinese Proficiency Test (hereafter referred to as the New HSK), is a standardized international assessment of Chinese language competence, evaluating the communicative abilities of non-native speakers across daily, academic, and professional contexts. The Standards mandate that international undergraduates enrolled in Chinese-medium programs must achieve at least Level 4 under the International Curriculum for Chinese Language Education (MOE, 2018), making the New HSK score a prerequisite for admission. To accommodate the rapid expansion of Confucius Institutes and the global promotion of Chinese language education, the measurement construct and utility of the New HSK have undergone significant shifts, emphasizing a "washback effect" aimed at promoting teaching and learning through assessment (Chai, 2015). Notably, the target populations for this washback effect primarily comprise overseas Confucius Institute learners, non-degree-seeking short-term students, and general language learners. However, the pre-admission New HSK score cannot diagnose whether a student's linguistic competence is sufficient to meet the rigorous demands of specialized undergraduate courses, nor can it identify specific pedagogical issues within the classroom.

Another well-established language assessment is the Putonghua Proficiency Test (hereafter referred to as the PSC). The PSC primarily assesses candidates' spoken proficiency by measuring the standardization of their pronunciation, vocabulary, and grammar through an oral examination (Commission). Crucially, the PSC does not evaluate the core skills of listening, reading, and writing, which are essential for academic success among international undergraduates. Although the newly revised Guidelines for the Putonghua Proficiency Test adapted their reading passages and conversational topics to accommodate international test-takers, the validity of the PSC for international students remains to be rigorously verified due to the demographic complexity, varying proficiency levels, and diverse native language backgrounds of this population (Zhang, 2023). Consequently, the PSC is currently ill-suited for large-scale academic assessments of international undergraduates.

Established in 2009, the Unified Chinese Proficiency Examination for Pre-University International Undergraduates under the Chinese Government Scholarship assesses students' receptive and productive language skills (listening, speaking, reading, and writing) while balancing the development of both General Chinese and Chinese for Academic Purposes (MOE, 2011). However, it is critical to note that this pre-university exit exam is inherently a summative assessment. Its outcomes are primarily utilized to certify whether pre-university students have met the minimum language proficiency threshold required for undergraduate matriculation. In contrast, evaluating the quality of Chinese language learning during the undergraduate stage requires continuous monitoring and evaluation of students' academic outcomes post-matriculation,

prioritizing formative assessment and emphasizing the diagnostic and developmental functions of evaluation.

In sum, the target populations, test constructs, and scopes of the New HSK, the PSC, and the pre-university exit examination do not align with the requirements of a quality evaluation system for registered international undergraduates. Furthermore, these three assessments are inherently summative, administered only at the culmination of specific instructional phases. This functional positioning diverges from the continuous, longitudinal evaluation required for quality assurance. Because none of these tests capture the formative and longitudinal nature of language acquisition, they fail to evaluate the ongoing efficacy of teaching and learning or to diagnose classroom deficiencies, thereby inhibiting the regulatory and formative potential of assessment. Consequently, there is an urgent need to refine the integration of formative and summative assessments for matriculated international undergraduates and to establish a robust framework for the formative evaluation of their learning outcomes (Chai, 2018).

Another foundational task in evaluating Chinese language learning quality is to identify key contributing factors and their underlying mechanisms of influence, thereby providing robust empirical evidence to optimize pedagogical practices. Existing research on the determinants of international students' language learning quality lacks a multi-level research design that systematically accounts for student-level, teacher-level, and institution-level variables. Moreover, while contemporary large-scale educational assessments globally utilize advanced psychometric and statistical methodologies, such as multilevel modeling, to partition variance in student performance and explore the interactive mechanisms of contributing factors, research in China remains methodologically homogeneous. Current domestic studies rely on rudimentary analytical techniques and lack rigorous relational or causal analyses of evaluative data.

4. Theoretical Frameworks and Models for System Construction

The evolution of educational evaluation models and their widespread application in student development assessments provide a solid theoretical foundation for integrating Chinese language proficiency assessment with its underlying contributing factors. To this end, this section reviews three educational evaluation models frequently utilized in both national and international educational quality assessment programs.

4.1. Management-Oriented Educational Evaluation Models

Among management-oriented evaluation models, the most prominent is the Context-Input-Process-Product (CIPP) model proposed by Stufflebeam. Stufflebeam argued that educational evaluation should transcend merely diagnosing problems and current conditions; instead, it should prioritize the formative and improvement-oriented functions of assessment. In this view, evaluation must be decision-oriented, serving to inform decision-makers and thus enhance teaching and learning. Compared to Tyler's objective-oriented evaluation model, the CIPP model critically addresses the formulation of objectives and their underlying rationale. It represents a comprehensive, full-cycle evaluation spanning from program conceptualization and implementation to ultimate outcomes, thereby facilitating continuous programmatic and

pedagogical improvement (Jiang, 2007). When constructing an indicator system based on the CIPP model, researchers typically delineate sub-dimensions under its four primary domains, context, input, process, and product, and subsequently select specific, representative indicators within these sub-dimensions.

At the practical level, the OECD drew upon the CIPP framework to establish an "Input-Process-Outcome" indicator model. Under this framework, the OECD aligned the PISA context questionnaires across four nested levels: individual participants, instructional contexts, educational institutions, and the broader education system (OECD, 2019). By structuring indicators of educational context, input, and outcomes across macro, meso, and micro levels, this framework offers a valuable blueprint for constructing a multidimensional quality evaluation indicator system for Chinese language learning among inbound international undergraduates.

4.2. Learning Outcome Evaluation Models Based on the Value-Added Concept

Astin (1993) posited that learning outcomes in higher education can be broadly categorized into cognitive and affective domains. Cognitive outcomes encompass the knowledge and skills acquired through general and specialized education, whereas affective outcomes include dimensions such as student satisfaction and learning motivation. Rooted in the value-added paradigm of student development, Astin proposed the Input-Environment-Outcome (IEO) model. This model emphasizes that students' prior knowledge, pre-entry academic experiences, and demographic or family backgrounds constitute critical "input" factors. The instructional processes, curricular designs, faculty interactions, institutional climate, and students' own involvement and effort during their undergraduate studies serve as "environmental" factors. Consequently, the resulting knowledge, skills, attitudes, values, and behaviors represent the ultimate "outcomes" of higher education (OECD, 2012).

By focusing on the net gains in learning outcomes, essentially controlling for students' initial inputs, the IEO model embodies the core principles of value-added evaluation. This approach offers a more precise and equitable reflection of institutional effectiveness and the true quality of student development. Consequently, this model provides a vital theoretical foundation and methodological reference for selecting indicators to construct a multidimensional quality evaluation system for international undergraduates' Chinese language acquisition.

4.3. The Opportunity to Learn Model Based on Process Equity

The Opportunity to Learn (OTL) framework, originally conceptualized by Carroll, is a theoretical model that encompasses the school-level inputs and instructional processes necessary for students to achieve desired educational outcomes. Analyses of results from international large-scale educational assessments indicate that the constituent components of the OTL model serve as significant predictors of student academic achievement (Carroll, 1963). Unlike explicit input indicators (such as educational funding or physical resources), OTL focuses on micro-level heterogeneity, such as curricular content, pedagogical organization, and student learning behaviors, thereby providing a more authentic reflection of the active instructional process.

Practical applications of OTL assessment results demonstrate that this model enables institutions to diagnose common characteristics of student learning and identify systemic

disparities, thereby offering empirical references to design targeted interventions for specific student cohorts. Furthermore, it meticulously documents institutional instructional processes, providing concrete leverage points for educational improvement. Longitudinal testing over time generates cyclical databases that yield vital empirical evidence for implementing and validating institutional reform initiatives. Compared with the CIPP and IEO models, the OTL framework operates at a profoundly micro-level. As an actionable and quantifiable theoretical framework for evaluating equity in the educational process, it provides a robust foundation for screening key influencing factors in the quality evaluation of international undergraduates' Chinese language learning.

Synthesis of the Evaluation Models. A comparative analysis of these frameworks reveals distinct complementary strengths. The CIPP and IEO models share fundamental similarities; in terms of output, both emphasize not only quantifiable academic performance but also students' intrinsic affective indicators, such as learning interest, self-efficacy, and motivation. Typically, the influencing factors identified by these two models span the macro, meso, and micro levels, encompassing the educational system, institutions, teachers, and individual students, thereby enabling a comprehensive evaluation of educational quality, equity, and efficiency. Specifically within this broader scope, the IEO model focuses heavily on factors that facilitate academic gains from the perspectives of students, instructors, and institutions during the academic progression, thereby prioritizing the value-added nature of evaluation.

In contrast to the multi-level approach of CIPP and IEO, the OTL model posits that the institutional instructional factors enabling students to achieve desired outcomes are predominantly shaped by teacher pedagogy or the broader educational system. Consequently, the influencing factors designed under the OTL model are concentrated explicitly at the institutional and teacher levels. This targeted focus allows for the rigorous identification of subtle variations in institutional management and pedagogical behaviors across different regions and schools in the short term. It provides multidimensional feedback on student learning, curricular design, teacher quality, resource allocation, and school climate. Ultimately, these three evaluation models are not inherently superior to one another; rather, different educational evaluation initiatives should select the model, or a hybrid thereof, that best aligns with their specific evaluative objectives.

5. Pathways for Developing the Evaluation System

Drawing upon the aforementioned theoretical models of educational evaluation and addressing the current status and practical demands of inbound international education in China, this section outlines the fundamental pathways for developing a quality evaluation system for international undergraduates' Chinese language acquisition. This framework is conceptualized through the lens of the evaluation system's core constituent elements.

5.1. Establishing a Systematic and Comprehensive Institutional Framework

The institutionalization of evaluation addresses the fundamental question of why the quality of Chinese language learning among international undergraduates must be assessed. First, this institutional framework must clearly define the value orientation, objectives, and functions of the

evaluation, thereby guiding the proper trajectory of educational assessment while ensuring alignment with broader policy mandates and administrative directives. Designed with a formative, student-centered focus, the framework should serve as a comprehensive, cyclical quality assurance system established at the national policy level. It aims to guide the comprehensive, longitudinal tracking of international undergraduates' learning outcomes throughout their entire academic lifecycle.

Second, the institutional framework must delineate the responsibilities of various evaluation stakeholders across different levels and clarify key tasks, thereby providing actionable guidelines and robust operational frameworks. Specifically, the framework should encompass guiding philosophies, fundamental principles, core mandates, and implementation protocols to facilitate the systematic and phased execution of evaluation efforts. Additionally, a suite of supporting policies must be refined. These include establishing Chinese proficiency testing standards tailored to the specific developmental phases of international degree programs, enhancing the professionalization of evaluation agencies, and instituting robust accreditation and supervisory mechanisms led by independent third-party organizations, such as industry associations and professional academic societies.

5.2. Refining the Operational Mechanisms for Diverse Evaluative Stakeholders

The composition, rights, and responsibilities of evaluative stakeholders address the fundamental question of who conducts the evaluation of Chinese language learning quality among international undergraduates. The evaluative ecosystem should encompass government bodies, professional agencies, independent third-party organizations, and HEIs, thereby establishing a multi-actor governance system characterized by mutual coordination and accountability. The government assumes primary responsibility for formulating evaluation policies and institutional frameworks, exercising administrative and supervisory functions over educational quality assessment. Additionally, government entities must streamline relations among stakeholders across various levels to prevent redundant or overlapping evaluations, thereby fostering a synergistic and data-sharing operational mechanism.

As highlighted in the Overall Plan for Deepening the Reform of Educational Evaluation in the New Era (China, 2020), there is a pressing need to strengthen professional capacity and leverage the expertise of professional agencies and social organizations. Evaluating the quality of international undergraduates' Chinese language acquisition is a complex and systematic endeavor. Every phase, including the development of language proficiency testing instruments, the selection of influencing variables, logistical implementation, and the analysis of results, requires profound professional expertise. Commissioning independent third-party organizations and professional agencies to conduct monitoring and assessment represents a highly efficient and viable approach. Concurrently, the government should accelerate the formulation of accreditation standards for evaluation agencies and professional certification for evaluators, thereby ensuring the methodological rigor and institutional credibility of these external bodies.

Typically, assessments conducted by governments and macro-level agencies tend to be broad and comprehensive. Consequently, HEIs are encouraged to design their own internal evaluation

frameworks, aligning with external indicator systems while incorporating their unique institutional characteristics. These internal evaluations can target specific instructional activities, particular phases of the academic progression, or comprehensive institutional performance. Such endeavors will contribute to a continuously evolving mechanism of self-regulation, intrinsic motivation, and continuous institutional enhancement in the realm of inbound international education.

5.3. Constructing a Hierarchically Designed Evaluation Indicator System

Constructing a comprehensive evaluation indicator system that encompasses both students' Chinese language proficiency and its influencing factors addresses the fundamental question of what to evaluate. First, it is essential to clarify that this evaluation should function as a diagnostic and formative assessment during the students' undergraduate trajectory. The primary objectives are to evaluate and enhance the current state of language acquisition and institutional instruction while identifying key contributing variables.

In terms of content, the evaluation should encompass not only academic achievement but also learning attitudes and values. For inbound international students, components such as sociocultural orientation to China, cross-cultural adaptation, and the development of positive host-country affinity (commonly conceptualized in Chinese policy as *Zhihua Youhua*, or "knowing and befriending China") must be integrated into the quality assessment scope. Structurally, the framework should incorporate curricular standards and pedagogical processes, spanning multiple hierarchical levels: student learning, faculty instruction, institutional management, and macro-educational policy. Methodologically, a mixed-methods approach combining quantitative and qualitative analyses is warranted. Furthermore, since assessment and improvement are intrinsically linked, the framework must inherently generate actionable, data-driven solutions to enhance instructional quality.

Second, the core dimensions of the evaluation must be systematically integrated. Based on varying assessment objectives and contexts, a multi-level framework of influencing factors should be established. This framework must account for the applicability of different evaluation models across systemic, institutional, classroom, and individual levels. Developing these indicators requires prioritizing the stability of key constructs and survey items, thereby facilitating the collection of reliable longitudinal data for trend comparisons.

In the nascent stages of system development, the indicator framework should prioritize parsimony over exhaustive complexity. Researchers can adopt the Opportunity to Learn model to collect micro-level data, gradually narrowing the focus to isolate key determinants. Simultaneously, the dynamic nature of these indicators must be maintained. Drawing on theoretical models such as CIPP and IEO, the system should be periodically recalibrated to accurately reflect the evolving realities and challenges of international undergraduates' language acquisition.

To address how to evaluate, robust and feasible methodologies must be rigorously selected. Specifically, techniques such as matrix sampling and test equating designs are recommended for the longitudinal tracking of language acquisition trajectories. Additionally, Value-Added Models

and Hierarchical Linear Modeling can be employed to unpack the complex mechanisms through which various multi-level factors influence academic achievement, thereby offering robust empirical evidence to optimize both teaching and learning.

5.4. Developing a Dynamic and Interconnected Feedback and Enhancement Mechanism for Evaluation Results

The feedback and enhancement mechanism addresses the critical question of how to effectively operationalize assessment data regarding international undergraduates' Chinese language acquisition. Establishing robust evaluation databases, information-sharing platforms, and regular reporting channels based on large-scale surveys is indispensable to this system. To maximize the formative and reformative functions of educational evaluation, empirical data must be rigorously analyzed to provide educational authorities and HEIs with precise insights that support evidence-based policymaking and pedagogical improvement. By investigating the sources of variance within the influencing factors model, researchers can elucidate and validate underlying operational mechanisms, thereby establishing a cohesive framework for enhancing language learning quality. Concurrently, longitudinal and cyclical assessment results provide forward-looking evidence for calibrating Chinese proficiency tests, updating indicator frameworks, and selecting appropriate technical instruments, ultimately driving the dynamic evolution of the evaluation system itself.

Conducting continuous monitoring across the entire educational pipeline to collect and analyze Input, Process, and Output data represents a fundamental paradigm shift in assessing language learning quality. Through this mechanism, students can reflect on their linguistic strengths and weaknesses, diagnose underlying causes, and subsequently enhance their learning outcomes. Meanwhile, HEIs, instructors, and administrators can draw upon these results to construct comprehensive learning databases, using assessment data as a catalyst to drive systemic reforms in pedagogy, curricula, and institutional administration. At the macro level, educational authorities can harness the positive washback effects of formative evaluation to optimize educational quality and efficacy. This alignment seamlessly supports the strategic transition of inbound international education in China from rapid scale expansion to substantive, quality-driven development (Neihan Fazhan), thereby accelerating the establishment of the "Study in China" brand and elevating the global prestige of Chinese higher education.

Author Contributions:

Conceptualization, X.T.; methodology, X.T.; resources, P.Z.; writing—original draft preparation, X.T.; writing—review and editing, P.Z. All authors have read and agreed to the published version of the manuscript.

Funding:

This research project is funded by the Graduate Innovation Fund of Beijing Language and Culture University (Approval Number: 26YCX049).

Institutional Review Board Statement:

Not applicable.

Informed Consent Statement:

Not applicable.

Data Availability Statement:

Not applicable.

Conflict of Interest:

The authors declare no conflict of interest.

Reference

- Astin, A. W. (1993). Principles of Good Practice for Assessing Student Learning. In *Leadership Abstracts*, 6, 4.
- Bottani, N. (1996). OECD international education indicators. *International Journal of Educational Research*, 25(3), 279-288.
- Buchmann, G., & Neri, M. C. (2008). The brazilian education quality index (Ideb): measurement and incentives upgrades. EPGE Brazilian School of Economics and Finance-FGV EPGE (Brazil). 686.
- Carroll, J. B. (1963). A model for school learning. *Teachers College Record*, 64, 723-732.
- Chai, X. (2015). A comparative study on the enrollment and examination systems of international students between China and the United States. *Journal of China Examinations*, 5, 39-46.
- Chai, X. (2018). *Research on second language testing issues*. University of International Business and Economics Press, 212-213.
- China, G. O. o. t. S. C. o. t. P. s. R. o. (2016). Opinions on doing a good job in education opening-up to the outside world in the new era. Ministry of Education of the People's Republic of China. Retrieved from http://www.moe.gov.cn/jyb_xwfb/s6052/moe_838/201605/t20160503_241658.html
- China, G. O. o. t. S. C. o. t. P. s. R. o. (2020). Overall plan for deepening the reform of education evaluation in the new era. Retrieved from http://www.moe.gov.cn/jyb_xxgk/moe_1777/moe_1778/202010/t20201013_494381.html
- Commission, S. L. Putonghua proficiency test syllabus. National Putonghua Proficiency Test. Retrieved from <https://www.cltt.org/#/articleDetail?id=5835a5a7-ceb5-4c1a-b2d-bddc91960f99&type=work>
- Communities, E. (2001). European report on the quality of school education. Office for Official Publications of the European Communities, 5-7.
- Cui, S. (2014). Discussion on the orientation mode for international students in colleges and universities. *Higher Education Exploration*, 3, 154-158.
- Daily, C. E. (2023). Building a globally influential and important education center: The seventh commentary on studying and implementing the spirit of General Secretary Xi Jinping's

- important speech at the fifth collective study session of the Political Bureau of the CPC Central Committee. 1.
- Fan, W., & Li, X. (2018). An empirical study on the influencing factors of Chinese learning difficulties of international preparatory students in colleges and universities. *Applied Linguistics*, 4, 132-139.
- Gu, M. (1998). *Dictionary of education (Revised and combined edition)*. Shanghai Education Press, 798.
- Han, S., et al. (2021). A survey on online education satisfaction of international students in China: A case study of Teaching Chinese as a Foreign Language courses. *Journal of Ningbo Vocational and Technical College*, 25(3), 55-59.
- Hao, M. (2018). Research on the influencing factors of Chinese character recognition and reading of advanced Chinese language learners. *Language Teaching and Linguistic Studies*, 5, 1-12.
- Jia, K. (2009). Research on the evaluation of learning quality of higher vocational students. Doctoral dissertation, East China Normal University.
- Jiang, G. (2007). Theory and practice of higher education evaluation based on CIPP. *China Higher Education Research*, 8, 10-12.
- Li. (2021). Evaluation and monitoring of basic education: New fields and new challenges of education supervision system construction. *Journal of China Examinations*, 5, 48-55.
- Li, & Xin, T. (2021). Connotation and theoretical model of monitoring and evaluation of basic education quality. *Journal of East China Normal University (Educational Sciences)*, 39(4), 15-29.
- MOE. (2011). Brief statistics of international students in China 2001. 35-36.
- MOE. (2018). Notice of the Ministry of Education on issuing the "Quality Code for Higher Education of International Students in China (Trial)". Retrieved from http://www.moe.gov.cn/srcsite/A20/moe_850/201810/t20181012_351302.html
- MOE. (2019). Brief statistics of international students in China 2018. 35.
- MOE. (2020). Notice of the Ministry of Education on issuing the "Statistical Indicator System of China Education Monitoring and Evaluation (2020 Edition)". Retrieved from http://www.moe.gov.cn/srcsite/A03/s182/202101/t20210113_509619.html
- MOE. (2021). Notice of the Ministry of Education on issuing the "Implementation Scheme for Undergraduate Education and Teaching Audit and Evaluation of Regular Colleges and Universities (2021-2025)". Retrieved from http://www.moe.gov.cn/srcsite/A11/ss7057/202102/t20210205_512709.html
- OECD. (2012). AHELO feasibility study report: Volume 1. 1, 99-100.
- OECD. (2019). PISA 2018 assessment and analytical framework. 105-107.
- Pan, W. (2004). Research on the quality evaluation system of postgraduate education in China. Doctoral dissertation, East China Normal University.
- Wang, J. (2014). The basic positioning and coping strategies of education for international students in China. *China Higher Education Research*, 8, 88-92.

- Xu, X., et al. (2018). A preliminary study on the cultivation of international students in colleges and universities. *Heilongjiang Education (Research and Evaluation of Higher Education)*, 8, 71-73.
- Zhang, C. (2023). Applicability analysis and improvement strategies of Putonghua Proficiency Test for international students. *Journal of Yunnan Normal University (Teaching and Research on Chinese as a Foreign Language Edition)*, 21(2), 79-92.

License: Copyright (c) 2026 Author.

All articles published in this journal are licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0). This license permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and source are properly credited. Authors retain copyright of their work, and readers are free to copy, share, adapt, and build upon the material for any purpose, including commercial use, as long as appropriate attribution is given.