Construction of Smart Education Evaluation System: A Case Study of International Chinese Language Education

Ruiling Ma and Juan Xu*

***Abstract*—The enablement of intelligent technology and big data technology has provided an opportunity to transform and develop the evaluation of International Chinese Language Education. Based on the review of previous related research, this study constructs an overall framework for the International Chinese Language Smart Education evaluation system that contains five levels: user, application, information, data, and technology. The smart education evaluation model that contains four dimensions, namely smart learning environment, smart pedagogy, smart learning, and smart education support services, is constructed in this study. Meanwhile, some new features of evaluation on International Chinese Language Smart Education are assessed in this paper. Finally, taking an advanced Chinese writing course as a case, this paper adopts the Delphi and Analytic Hierarchy Process methods to determine the smart education evaluation index system, which includes 3 first-level indicators, 11 second-level indicators, 28 third-level indicators, and 120 fourth-level indicators.**

***Index Terms*—International Chinese language smart education, smart education evaluation, technology-assisted education evaluation reform, advanced Chinese writing evaluation.**

I. Introduction

Education evaluation is one of the essential components of the education system and it fulfills an irreplaceable function in improving instructional quality. Meanwhile, education evaluation is highly necessary for the comprehensive development of students and for improving the professional growth of teachers. In a word, scientific assessment activities also reflect the scientific nature of education [1–3].

As information technology is deeply integrated with education and teaching, digital technology has been widely applied in education reform and teaching practice, especially in education evaluation. Education evaluation is a dynamic development process, and the current trend in education evaluation reform is built based on objectivity, diversity, science, and openness [4, 5].

The International Achievement Evaluation Association advocates the implementation of multi-dimensional and comprehensive evaluation. Through the exploration and practice in countries around the world, several international education evaluation projects, such as “Program for International Student Assessment (PISA)”, “Progress in International Reading Literacy Study (PIRLS)”, and “Programme for the International Assessment of Adult Competencies (PIAAC)”, have been developed. Evaluation tools have transferred from the simplest pen-and-paper test to the combination of pen-an-paper based observation and computer-assisted testing, the evaluation function, on the other hand, closely follows diagnosis, development promotion, and decision-making services.

In 2020, the Chinese government promulgated the Overall Plan for *Deepening Educational Evaluation Reform in the New Era* to systematically promote and comprehensively deepen education evaluation reform [6]. Exploring new models of education evaluation has emerged as an inevitable trend and requirement for intensifying education evaluation reform [7].

With the advent of the smart era, technologies such as artificial intelligence, big data, and learning analytics are fully empowering teaching evaluation reform [8]. In the long past practice of education evaluation, various types of technologies have been applied to results evaluation, process evaluation, value-added evaluation, and comprehensive evaluation. However, there are objectively varying problems of different types and levels involved in each type of evaluation, which is often beyond the reach of traditional methods. Therefore, education evaluation calls for new technologies and methods to make breakthroughs.

International Chinese language Education means teaching Chinese to speakers of other languages (TCSOL). International Chinese language Smart Education is a novel type of education that harnesses emerging information technologies to promote accurate instruction for teachers and personalized learning for students. The intelligent education evaluation system measures and evaluates all the elements of the International Chinese Language Smart Education system, so that it can serve as a feedback adjustment to promote teaching and learning. With the support of intelligent technology, smart education evaluation enables educational judgment more accurate, comprehensive, and intelligent. Therefore, an evaluation system of International Chinese Language Smart Education is constructed in this study, analyzes its characteristics, and creates an evaluation index system based on the evaluation content from a case study of an advanced Chinese writing course.

II. Literature Review

A. Technology-Assisted Evaluation Reform

In the past decades, the process of education evaluation has generally included pre-course Diagnostic Evaluation, in-class Formative Evaluation during class, and post-course Summative Evaluation [9]. As research advances, education evaluation has begun to change from result-oriented to process-oriented, while formative evaluation also starts to

Manuscript received September 25, 2022, revised October 18, 2022; accepted November 1, 2022.

The authors are with the School of Information Science, Beijing Language and Culture University, Beijing, 100083, China.

*Correspondence: xujuan@blcu.edu.cn

doi: 10.18178/ijiet.2023.13.2.1805

receive research attention. For example, Stufflebeam *et al*. raised the CIPP evaluation model [10], which is a typical formative evaluation model that consists of four parts, namely Context Evaluation, Input Evaluation, Process Evaluation, and Product Evaluation. These four assessments provide different information for teachers' decision-making [11]. However, all of these assessment methods involve at least one type of technology, and most of them assessed learners' performance by using questions, tests, exams, and questionnaires to collect learning data, so the results are often time-consuming or less scientific. In addition, this evaluation model adopts a unitary approach, which means that there is only one mode of the subject and the object, that is, the teacher assessing the learning outcomes of students.

With the advent of computer-assisted evaluation methods, evaluations have become more valid and sophisticated in terms of measurement, and more flexible in terms of time and place [12]. However, there are still considerable difficulties in rapidly accessing and processing massive educational data in computer-assisted assessment.

In recent years, artificial intelligence (AI) and big data technologies have become more prevalent in education, and new technologies have energized education evaluation reform [13]. It has been pointed out in a paper that through machine learning and natural language processing techniques, the accuracy of evaluation results in students' learning performance has increased [14]. Studies have shown that data-driven education evaluation has been greatly enhanced regarding its functions, objects, content, methods, data sources, data analysis, processes and outcomes Research suggests that data-driven education evaluation has a great improvement in evaluation function, evaluation object, evaluation content, evaluation method, evaluation data source, data analysis, evaluation process, evaluation result, etc. [15]. Liu and Yu *et al*. [16], the study introduced AI technology based on facial expression recognition and voice emotion recognition methods to extract and analyze video streams collected by webcams during online lessons, which helps teachers to shape their evaluation. According to [17], [18], the authors present objectives and practical approaches to evaluation reform in categories of intelligent technology-supported process-based evaluation and comprehensive evaluation, respectively. In addition, there are some evaluation studies specifically oriented to teachers' digital competencies from a technology perspective [19].

B. Technology-Assisted Language Education Evaluation

The first aspect of technology-assisted language education evaluation reform is the change in evaluation techniques. The assessment of learners' language knowledge and skills is among the most fundamental parts of language education and includes both static assessment and dynamic assessment. Computerized tests of language knowledge tests include assessments of phonological, lexical, and grammatical knowledge. Tests of language skills include assessments of listening, speaking, reading, writing and translation. Computer-assisted testing of subjective questions such as speaking and writing is a major area. In the early stage of technology-assisted language learning, computers were able to provide learners with audio of standard pronunciations and score students' pronunciations (with lower accuracy). With the development of AI technologies, especially with the maturation of natural language processing techniques, speech recognition, and speech conversion, speech evaluation technologies are beginning to be applied to the assessment of learners' oral ability [20]. In addition, automatic essay scoring (AES) technology has been available to evaluate the quality of learners' essays [21].

Another major area of research in educational evaluation is the assessment of learning resources, such as textbooks, learning dictionaries, MOOCS, Applications, translation tools, and courseware. Language learners express a strong demand for free online language learning resources. However, the quality of these resources is often mixed and the results should be applied to their development and improvement. Research into teachers' digital competence in language teaching has also received attention. In addition to knowledge of linguistics and pedagogy, teachers should acquire knowledge of educational technology, particularly those of how to integrate technology with pedagogy and language knowledge in the teaching process. For example, teachers should master how to use speech synthesis technology to develop listening audio materials and use flash to present grammar knowledge. Therefore, the evaluation of the digital literacy and competence of language teachers is also a necessary aspect of technology-enhanced reform of the evaluation of language education system.

In contrast, research on technology-assisted teaching evaluation in the field of International Chinese Language Education is rather limited [22], and against this background, an AHP-based evaluation system for distance teaching of Chinese as a second language is constructed in this study. This system includes the evaluation of five components, namely learners, teachers, learning materials, learning supports and services, and teaching platforms. In response to the insufficient interaction in the previous online courses, Liu and Liu [23] put forward a quantitative interaction design evaluation index system for the interaction of the Chinese online course, online teaching platform, and teaching process

So far, research on educational evaluation of International Chinese Language Education with intelligent technology is limited, with few scholars applying data mining and analysis techniques to teaching evaluation. Therefore, this study will fill this gap and apply intelligent technologies to the innovation and reform of the evaluation system of International Chinese Language Education.

III. Chinese Smart Education Evaluation System

A. Chinese Smart Education Evaluation System Structure

Standardization is one of the characteristics of International Chinese Language Smart Education. It is a prerequisite for the organic integration of various technical elements and the efficient empowerment of smart education, which means that the design, development, application, and management of various elements, such as system framework, data resources, and presentation methods, should follow uniform standards. In other words, it is to ensure that all kinds of elements can be docked smoothly [24]. A scientific and appropriate education evaluation system can ensure that International Chinese Language Smart Education develops in a standardized direction.

To promote the construction and application of intelligent technology-assisted evaluation of International Chinese Language Smart Education, this paper proposes and constructs the overall framework of the International Chinese Language Smart Education Evaluation System (ICL-SEES) by Dong [4], as shown in Fig. 1.

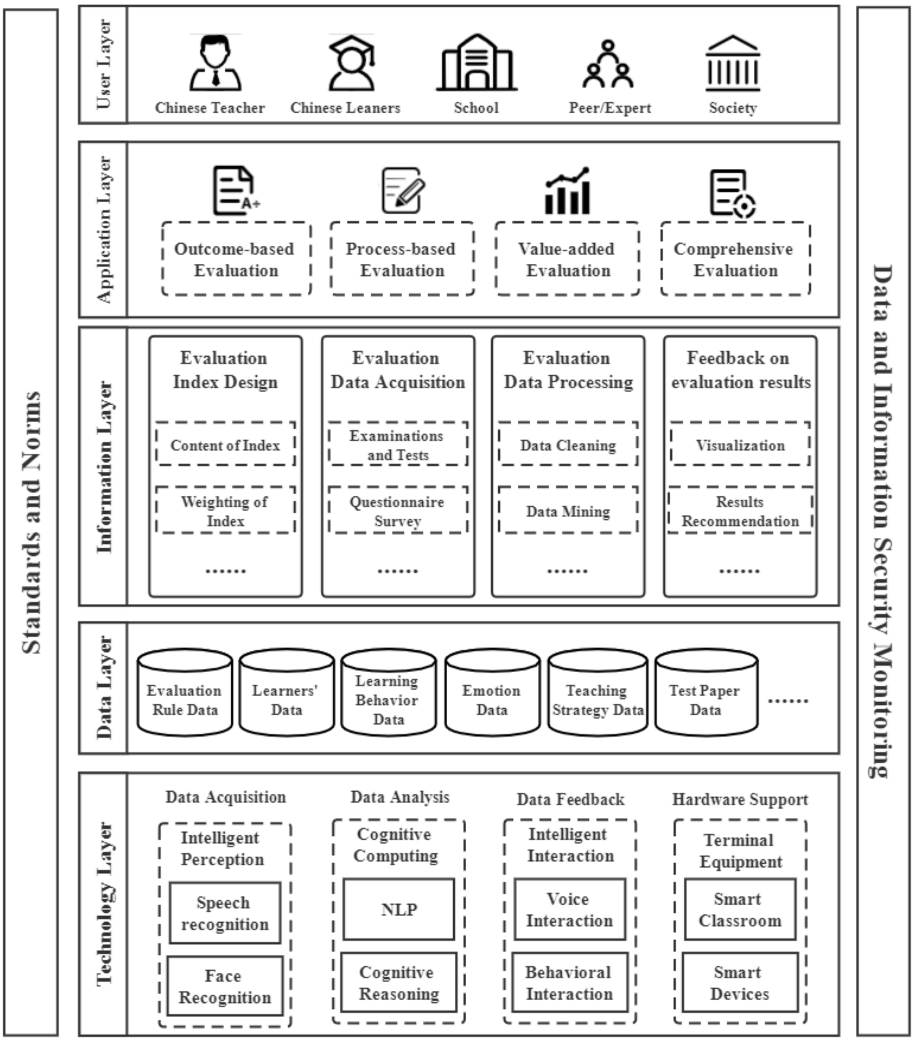

Fig. 1. The ICL-SEES overall framework.

As can be seen in Fig. 1, the ICL-SEES includes a user layer, an application layer, an information layer, a data layer, and a technology layer. This system is diversified and can be demonstrated in the following aspects.

Firstly, the types of evaluation subjects are diversified. The ICL-SEES includes teachers, students, schools, government, and society, as a result, suggestions for the optimization of teaching models can be obtained from different perspectives. Users can log into the ICL-SEES anytime, anywhere, and anyway through the intelligent terminal equipment.

Secondly, the evaluation methods are multi-dimensional. The ICL-SEES includes outcome-based, process-based, value-added, and comprehensive evaluations. These evaluations can serve different purposes in various teaching implementations. Outcome-based evaluation can reflect the students' learning and the quality of teachers' teaching. Process-based evaluation can help teachers keep track of second language (L2) learners' learning in the form of questions or follow-up tests, and help Chinese teachers make real-time adjustments to teaching strategies and arrangements [17]. For example, observing changes in learners' writing scores over the course of a semester. Value-added evaluation measures the extent to which students' language skills have grown over this period and is applied to evaluate school and teacher performance. For example, a comparison can be made between whether the number of vocabulary or grammar points acquired by the learners themselves has changed before and after the lesson. A comprehensive evaluation is a kind of full-view care of the evaluation object, which can be achieved by constructing a scientific comprehensive evaluation index system to provide a systematic, comprehensive, and complete comprehensive evaluation of students [18]. For example, after learning Chinese, L2 learners not only understand the language knowledge and skills but also change their feelings, learning strategies, etc.

Thirdly, the evaluation techniques are diversified. The CSEES adopts various intelligent and big data technologies to achieve intelligent data collection, processing, analysis, and visual output during the whole process of teaching and learning Chinese. Furthermore, the ICL-SEES relies on learning analytics, big data, and visualization technology for quantitative evaluation, and speech intelligent technology can serve to record and recognize the content of teacher-student dialogues.

Fourthly, the evaluation functions are multiple. The ICL-SEES can act to guide, regulate, motivate, diagnose and teach so that the teaching system can be optimized and upgraded and learners can develop comprehensively.

B. Chinese Smart Education Evaluation Content Model

Content evaluation is the key component of an evaluation system. The general idea of the International Chinese Language Smart Education content evaluation model is to evaluate the teacher's teaching quality through the student's learning. Based on the definition of the connotation of smart education [25], the object of smart pedagogy evaluation can be identified as four major components, namely smart environment (or called smart learning environment), smart pedagogy, smart learning, and teaching support and service, which represent the perspectives of developers, teachers, learners, and administrators respectively. These four evaluation contents can be considered as four micro-models, from which the four micro-models, specific evaluation contents, guidelines and indicators, and other building block components can be decomposed, as shown in Fig. 2.

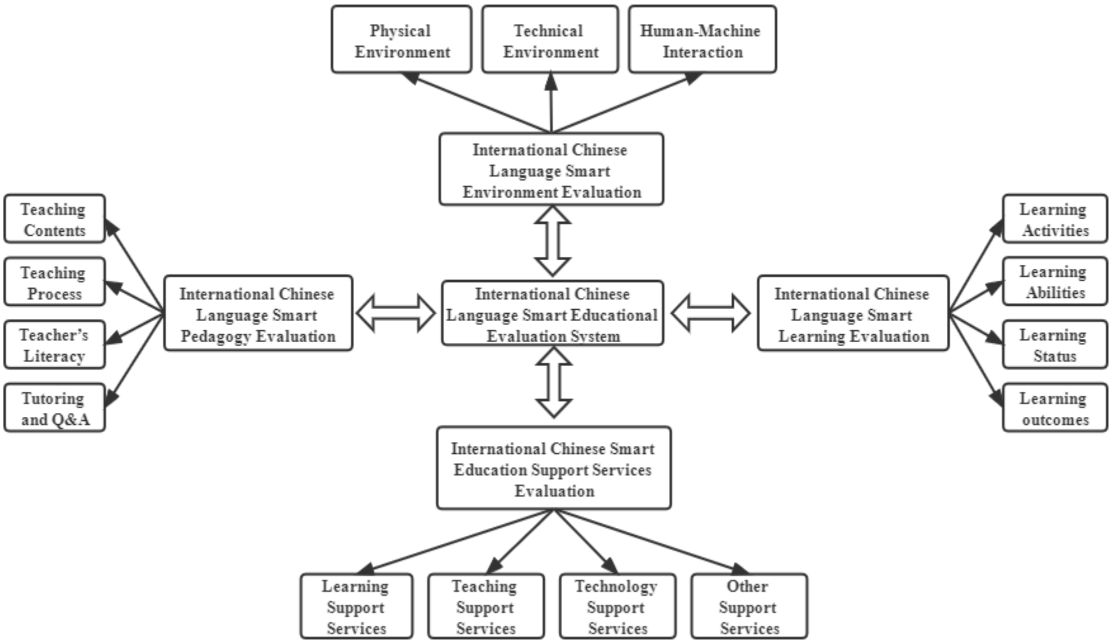

Fig. 2. International Chinese language smart education evaluation model.

It can be seen from the Fig. 2 that the smart environment evaluation model includes the evaluation of the physical, technical, and emotional environments composed of different resources, mainly for the smart classroom, the smart education cloud, and human-computer interaction. The evaluations of smart learning and the evaluation of smart pedagogy are oriented toward two main subjects, the learner and the teacher. The evaluation of smart pedagogy includes the evaluation of teaching contents, teaching process, teacher's literacy, tutoring and Q&A. Evaluation of smart learning includes learning activities, learning competencies, learning status, and learning outcomes. Learning activities refer to the learners' participation in the activities in the pre-class, in-class, and post-class stages. Finally, the evaluation of smart education support services includes four aspects, namely teaching support services (such as teacher training and teaching design resources), learning support services (including enrollment and guidance on resource use), technology support services (such as system operation and maintenance as well as network maintenance) and other support services (including financial support and complaint mechanism). For the Chinese language education system, smart teaching support services have the same importance as the other three evaluation elements, but their impact on teaching itself is peripheral, indirect, and implicit. To further improve the teaching evaluation model, this study includes them as evaluation elements but does not regard them as the key objects of evaluation.

C. Features of Chinese Smart Education Evaluation

First of all, the evaluation methods are scientific. With the help of artificial intelligent technology tools, the International Chinese Language Smart Education evaluation is a data-driven and evidence-based evaluation. The Chinese smart education evaluation is based on data evidence. The teaching and learning process can be quantified by collecting data on language teaching and learning from Chinese Smart Classroom, Smart Education Cloud, and other platforms. Through the collaboration of educational evaluation researchers, information technology professionals, and education and teaching personnel in the field of International Chinese Language Education, a scientific evaluation index system, weight factors, and algorithm model are constructed according to different evaluation objectives, objects, and contents.

Secondly, the evaluation data acquisition is coincident. Assisted by technologies such as brain-computer interface and face recognition, physiological, behavioral, and emotional data of L2 learners can be collected, such as pronunciation, essays, and answers. Data on the whole process of language learning will be mined and analyzed through high-speed computing technology to construct a personalized developmental map of students' language skills, which results in a "whole-person" evaluation. At the same time, data on affective and cognitive states are collected in a seamless and companionable manner based on various interactive technologies [26]. Multi-modal learning analytics technology ensures dynamic tracking of individual development and interpersonal data of learners.

Thirdly, the evaluation results are more accurate. Through the use of techniques such as data mining, and cognitive computing, the diagnosis and analysis of L2 learners' language abilities can be more accurate. To realize the optimization of multidimensional and global data processing and analysis, data based on the language learning process will be feedback based on intelligent technology. Big Data for Education will provide teachers with real-time and accurate feedback information on L2 learners. At the same time, supported by visualization technology, the ICL-SEES will provide users with intuitive and vivid evaluation results through highly personalized interactive data visualization charts, and push evaluation results accurately to users through technologies such as intelligent recommendation engines.

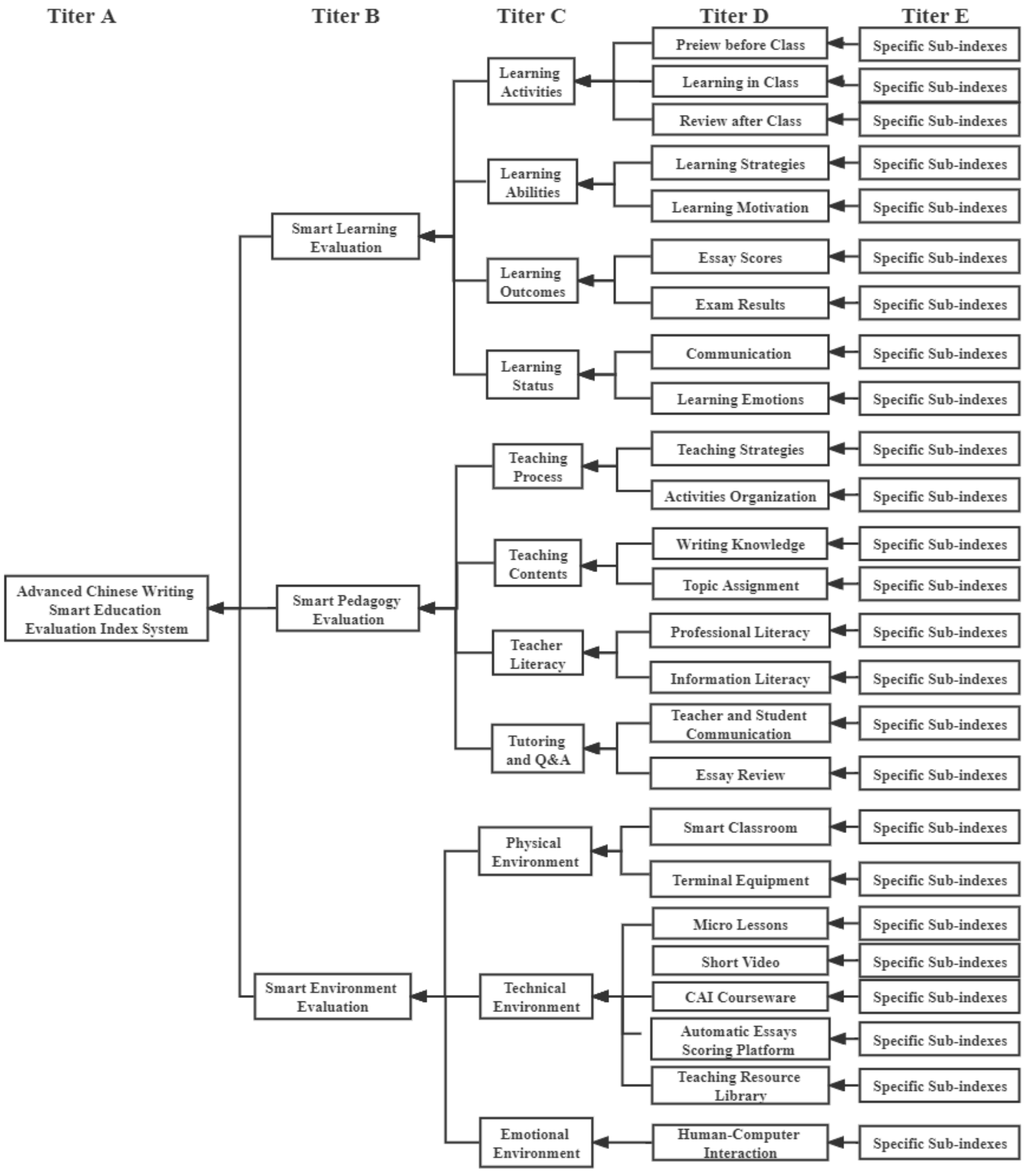

Fig. 3. Smart education evaluation index system.

IV. Research Design

Due to the space limitation of the paper, we cannot conduct an in-depth analysis of all aspects of ICL-SEES. In this paper, a smart education evaluation index system is constructed to make scientific, objective, and accurate evaluations. The establishment of this index system helps to determine what kinds of data need to be collected in the teaching and learning process. It can also help to develop and utilize learning software to achieve multidimensional and continuous collection of educational data for writing courses, with the support of learning analytics, big data, visualization, and other information technologies.

In addition, support services for smart education are not incorporated into this index system as an evaluation object, as it is not significant for this course.

A. Modeling Path

There are three main steps in the construction of the evaluation index system for intelligent education in advanced Chinese writing. Firstly, the basic framework of the evaluation level is determined according to the content model of smart education evaluation. The evaluation objects include smart learning, smart pedagogy, and smart environment. As mentioned above, the evaluation contents of smart learning and teaching support and services are not calculated as they were not very relevant to this study, so their weights are not calculated. Secondly, experts in the field of International Chinese Language Education field are invited to make several rounds of revisions to the framework, and a formal evaluation framework is finally determined. Finally, the Analytic Hierarchy Process (AHP) is adopted to determine the weights of each index and to build a scientific, effective, and standardized evaluation index system.

B. Methodology

Delphi and AHP are the major research methods used in this study.

The Delphi method is a common tool for relatively accurate estimation of a large number of indicators that cannot be quantitatively analyzed. This method applies expert opinion formation - statistical feedback - opinion adjustment to verify expert opinions and to suggest the probability of the proposed evaluation index. After several rounds of consultation, the opinions of experts gradually converged, and the revised smart education evaluation index system was finally obtained [27, 28]. The Sub-criteria were scored by inviting eight experts in the field of International Chinese Language Education. Furthermore, these experts come from two different fields of Chinese writing teaching and Chinese educational technology, separately, with each expert being asked to complete a survey questionnaire containing 6 multiple-choice questions.

AHP as a comparative evaluation theory measures the relative priority scales of intangible objects through a matrix of paired ratios based on human judgment. The basic idea of the Analytic Hierarchy Process is to stratify the problem to be analyzed, decompose it into different constituent factors according to the nature of the problem and the overall objectives to be achieved, and coalesce the factors cohesively at different levels according to their associated influence and their affiliation to form a multi-level analysis structure model. Finally, the problems are compared and ranked in terms of their strengths and weaknesses. The steps in building index weights using APH are: developing the hierarchical structure → putting up the pairwise relative matrixes across the survey → discovering the eigenvalues and eigenvectors → accumulating the relative loads [29, 30].

1) Developing the hierarchical structure of the evaluation system

In this study, the evaluation index system of this advanced Chinese writing course is designed into five tiers. Tier A is evaluation objective, tier B is evaluation element, tier C is evaluation content, tier D is evaluation index, and tier E is evaluation sub-indexes. This system is shown in Fig. 3.

Notably, the interpersonal interaction evaluation of the emotional environment evaluation of the intelligent environment evaluation has been placed in the counseling and answering section of the intelligent teaching evaluation, and the evaluation of the emotional environment evaluation here focuses on the evaluation of "human-computer interaction".

2) Constructing a comparative judgment matrix

In this study, a paired comparison method and a 1-9 scale evaluation method were applied to compare each component element two-by-two. The subjective value explanation is presented in the form of a 9-point scale, in which the empirical basis was proposed by [31]. The ratio of the importance of each element was solved to calculate the weight of the element in the index system.

In the case of elemental tier B, the element scale is derived and the hierarchical analysis model of B1, B2, and B3 is constructed, and the judgment matrix can be found in Table I.

TABLE I: THE CROSS-TABULATION ANALYSIS

	B1	B2	B3
B1	1	2.4/1	6.2/1
B2	1/2.4	1	4.2/1
B3	1/6.2	1/4.2	1

Expressed in a matrix as:

$$B = \begin{bmatrix} 1.000 & 2.400 & 6.200 \\ 0.417 & 1.000 & 4.200 \\ 0.161 & 0.238 & 1.000 \end{bmatrix}$$

3) 3) Determine the weight of each indicator

In this study, the sum-product algorithm is adopted to determine the weight of each indicator, which is given by equation (1).

$$V_n = \sum_{i=1}^{n} b_{nj} \tag{1}$$

According to the third-order judgment, the matrix is constructed for a total of 3 items of index 1, index 2, and index 3 for the AHP study (with the sum-product algorithm being the calculation methods), with eigenvectors of 1.837, 0.907, 0.255 and the corresponding weight values of 61.243%, 30.245%, and 8.513%, respectively for the total three items.

4) Consistency test

For consistency tests, typically, the smaller the consistency ratio (CR) value is, the better the consistency of the judgment matrix. According to [32], the consistency index is tolerable if the CR rate < 0.1, and the judgment matrix satisfies the consistency test. Otherwise, it means that there is no consistency, and the judgment matrix should be adjusted appropriately and adjusted and then analyzed. Accordingly, the maximum eigenvalue can be calculated by combining the eigenvectors first and using equation (2).

$$\lambda_{\max} = 1/n \sum_{i=1}^{n} \frac{(BW)i}{Wi} \tag{2}$$

The calculation result shows that $\lambda_{\max}$ = 3.026. The consistency index (CI) is then calculated in this study by using equation (3).

$$CI = \frac{(\lambda_{max} - n)}{n-1} \quad (3)$$

The calculation shows that CI=0.013. The random index (RI) is directly checked according to the order n of the judgment matrix. The mean of the largest characteristic root (λ*max) is identified and defined by randomly constructing a positive reciprocal inverse matrix by drawing numbers from one to nine and their reciprocals:

$$RI = \frac{\lambda^{*}_{max} - n}{n-1} \quad (4)$$

and finally, the CR value is calculated. The consistency ratio (CR) is estimated for more analysis, whereas ACI is the average index of randomly created weights [33]:

$$CR = \frac{CI}{RI} \quad (5)$$

The consistency test results are shown in Table II.

TABLE II: SUMMARY OF CONSISTENCY TEST RESULTS

Maximum characteristic root	CI	RI	CR	Result
3.026	0.013	0.520	0.025	Pass

The CR value of 0.025<0.1 is calculated for the 3rd-order judgment matrix, indicating that the judgment matrix of this study satisfies the consistency test and the calculated weights are consistent.

In addition, the weights of the indicators of tier E are directly assigned by a domain expert, as these indicators can be flexibly adjusted and set.

According to the above modeling path and method, the weight of advanced Chinese writing smart education evaluation is gradually determined. The calculation of the score for a specific index "i" will use the following equation:

$$W_i = W_B \times W_C \times W_D \times W_E \quad (6)$$

In equation (6), the weight, and the letters (B, C, D, E) indicate the corresponding indicator in respective tiers.

V. RESULTS

A. Smart Learning Evaluation (61.243%)

The evaluation subjects of the Smart Learning elements are Chinese teachers and learners, and the evaluation objects are learners, and a combination of online and offline evaluation methods can be used. The weights of smart learning evaluation are shown in Table III.

TABLE III: SMART LEARNING EVALUATION INDEX MODEL

Tier 1 (C)	Tier 2 (D)	Tier 3 (E)
Learning Activities (33.613%)	Preview before class (23.335%)	(1) Viewing of micro lessons (50%) (2) Discussion of essay ideas (20%) (3) Completion of individual exercises (30%)
	Learning in Class (54.619%)	(1) Writing of essays (80%) (2) Participation in the discussion (20%)
	Review after Class (22.046%)	(1) Use of the platform for free practice (40%) (2) Revision of essays after feedback (40%) (3) Independent improvement exercises after class (20%)
Learning Abilities (11.204%)	Learning Strategies (50.000%)	(1) Use of various writing strategies (60%) (2) Use of other learning strategies (40%)
	Learning Motivation (50.000%)	(1) Use a variety of learning methods such as cooperative, inquiry, and independent (70%) (2) Ability to achieve learning objectives in a relatively short period (50%)
Learning Outcomes (27.819%)	Essay Scores (57.143%)	(1) Machine automatic scoring (40%) (2) Teacher scoring (60%)
	Exam Results (42.857%)	(1) Entrance test scores (10%) (2) Midterm test scores (30%) (3) Final exam results (60%)
Learning Status (27.363%)	Communication (66.667%)	(1) Faculty-student communication (50%) (2) Peer-to-peer communication (50%)
	Emotional state (33.333%)	(1) Students' confidence in learning (30%) (2) Students' attitude towards difficulties (30%) (3) Students' desire for knowledge and curiosity (40%)

B. Smart Pedagogy Evaluation (30.245%)

The evaluation subjects of the Smart Pedagogy elements include teachers, peers/experts, and school administrators, and the evaluation targets are Chinese teachers, mainly using online evaluation. The weights of smart pedagogy evaluation are shown in Table IV.

TABLE IV: SMART PEDAGOGY EVALUATION INDEX MODEL

Tier 1 (C)	Tier 2 (D)	Tier 3 (E)
Teaching Process (31.806%)	Teaching Strategies (50.000%)	(1) Integrated use of multiple teaching methods to motivate students (15%) (2) Selecting an appropriate writing pedagogy according to the content (25%) (3) Use of modern educational technology to assist in teaching (15%) (4) Differentiated instruction based on students' characteristics (45%)
	Activities Organization (50.000%)	(1) Organize various activities with fun (25%) (2) Students have plenty of opportunities to practice (25%) (3) Activities are arranged to meet the needs of the learners (25%) (4) Activities are arranged to match the learners' abilities (25%)

Teaching Contents (16.514%)	Writing Knowledge (42.857%)	(1) Teaching a variety of genres (20%) (2) Teaching genre with practicality (20%) (3) Teaching genre learners is actionable (20%) (4) Teaching genres that meet the needs of learners (40%)
	Topic Assignment (57.143%)	(1) The topic task is novel (20%) (2) Topic tasks are actionable (20%) (3) Topic tasks are close to daily life (20%) (4) Topic tasks meet the needs of learners (40%)
Teacher Literacy (38.855%)	Professional Literacy (75.000%)	(1) Linguistic literacy (30%) (2) Pedagogical literacy (20%) (3) Psychological literacy (15%) (4) Communication literacy (15%) (5) Cultural literacy (20%)
	Information Literacy (25.000%)	Information Awareness (30%) Information literacy (20%) Information capacity (40%) Information Ethics (10%)
Tutoring and Q&A (12.825%)	Teacher and Student Communication (66.667%)	(1) Assignment of homework (20%) (2) Provide targeted coaching (40%) (3) Focus on the overall growth of the learner (20%) (4) Attention to learners' psychology and emotions (20%)
	Essay Review (33.333%)	(1) Essay correction with commentary (50%) (2) Focusing on common problems in essays (25%) (3) Conduct individual question and answer sessions (25%)

C. Smart Environment Evaluation (8.513%)

The evaluation subject of the Smart Environment element is the learner, the evaluation objects are teaching environment and resources, and the evaluation method is mainly online. The weights of smart environment evaluation are shown in Table V.

TABLE V: SMART ENVIRONMENT EVALUATION INDEX MODEL

Tier 1 (C)	Tier 2 (D)	Tier 3 (E)
Physical Environment (17.697%)	Smart Classroom (66.667%)	(1) Classroom type meets the requirements of teaching objectives (20%) (2) Hardware terminals in classrooms to meet learning needs (30%) (3) Classrooms facilitate teachers to organize teaching activities (30%) (4) Classrooms can intelligently capture and analyze teaching data (20%)
	Smart Terminal (33.333%)	(1) Use of supplementary writing tools such as smart pens and styluses (50%) (2) Use of intelligent robots to assist learning (50%)
Technical Environment (26.302%)	Micro Courses (28.301%)	(1) Instructional design (25%) (2) Teaching presentation (25%) (3) Technology realization (25%) (4) Teaching effectiveness (25%)
	Short Video (4.810%)	(1) Meets demand (20%) (2) Clarity of content (20%) (3) Pronunciation norms (20%) (4) With subtitles (20%) (5) Teaching effectiveness (20%)
	CALL courseware (24.281%)	(1) Instructional design (25%) (2) Teaching presentation (25%) (3) Technology realization (25%) (4) Teaching effectiveness (25%)
	Teaching Resource Library (16.171%)	(1) Meets demand (25%) (2) Discourse specification (25%) (3) Informative data (25%) (4) Resource diversity (25%)
	Essay Platform (26.436%)	(1) In line with demand (50%) (2) Diverse functions (50%)
Human -computer Interaction (56.001%)	Applicability (41.005%)	Effective in helping learners to develop independent learning
	Aesthetics (12.466%)	Beautiful interface design
	Operability (33.914%)	Easy for learners to get started
	Durability (12.614%)	Can be used repeatedly and is less likely to fail

VI. DISCUSSION

As shown in Table III, from the perspective of smart learning evaluation, the first finding reveals that learning activities are the dominant evaluation index according to the weight. Among them, classroom learning (54.619%) is substantially more important than the before-class (23.335%) and post-course (22.046%) stages. In the traditional classroom, the main evaluation indicators are the learning results, including the usual essay practice grades and the final exam results. For smart learning, the reduced importance of learning outcomes and the equal attention to the learner's learning state fully highlight the focus on the cognitive subjectivity of students in smart learning and the evolution of result-oriented learning to process-oriented learning. The indicator with the least weight is learning ability (11.204%), while in the opinion of experts, the learning strategies (50.000%) and learning styles (50.000%) are equally important, with learning strategies being the cognitive strategies of learners to control and adjust internal processes, and learning styles being the means to externalize learning.

As shown in Table IV, from the perspective of smart pedagogy evaluation, another finding indicates that teachers' literacy (38.855%) is the paramount evaluation indicator, and teachers' professional literacy has a significant impact on smart pedagogy based on the weight of experts' scores. Teachers' professionalism includes expertise and competence in linguistics, pedagogy, psychology and more. In addition, the teaching process (31.806%) also plays an important role. Teachers' choice of teaching strategies and the organization of teaching activities can effectively influence and mobilize L2 learners' learning status. Finally, although teaching content (16.514%) and interactive feedback (12.825%) occupy less weight, they are still relevant evaluation indicators. The two secondary indicators of teaching content did not differ significantly, but the ratio of teacher-student communication (66.667%) is more important than essay review (33.333%).

As shown in Table V, from the perspective of smart environment evaluation, the third finding indicates that the emotional environment (human-machine interaction) has the greatest weight of 56.001%. Especially, the applicability (41.005%) and operability (33.914%) of the machine reflect the concept of technology for teaching. In addition, software resources exert greater influence (26.302%) than hardware resources (17.697%). Among various types of software resources, micro-courses have the highest importance, accounting for nearly 30%; composition intelligent assessment platform and CALL courseware have a higher proportion, teaching resource library has an average proportion, and short video occupies the lowest proportion. This reveals that short videos are not an essential tool for advanced Chinese writing teaching. Among hardware resources, smart classrooms are much more important than smart learning terminals, as they can meet the requirements of teaching and learning in many aspects, such as context creation, inspiring thinking, information acquisition, resource sharing, and multiple interactions.

VII. Conclusion

Intelligent technology and big data technology empowerment provide new possibilities for education evaluation reform, which can effectively overcome the limitations of traditional education evaluation and make evaluation more diversified, accurate, and optimized. At the same time, intelligent technology can effectively promote reforms such as result-based evaluation, process-based evaluation, value-added evaluation, and comprehensive evaluation.

This study establishes an evaluation index system for advanced Chinese writing smart education with 3 first-level indicators, 11 second-level indicators, 28 third-level indicators, and 120 fourth-level indicators. Based on this education evaluation system, a combination of online and offline methods is adopted at different stages, such as before, during and after class, and teachers, students, experts/peers, schools, and society are invited to participate in the evaluation. Also, teaching evaluation technologies such as learning analytics, intelligent voice, big data, and visualization can be adopted virtually to provide scientific and objective teaching evaluation for writing teaching.

Meanwhile, there are limitations to this system that suggest some future research directions. Firstly, there is a need to establish smart education evaluation standards for different course types (comprehensive, speaking, listening, reading, and translation). Secondly, more experts should be invited to discuss the framework and indicators, so that the framework and indicators can be constantly updated. Third, a user-friendly intelligent platform for smart education evaluation systems, including websites and digital applications, needs to be developed to support users to evaluate Chinese language education.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Ruiling Ma conducted the research, analyzed the data, and wrote the paper; Juan Xu provided writing guidance and funding for the project; all authors approved the final version.

Funding

This work was supported in part by the 2022 Key Project of International Chinese Language Research of the Center for Language Education and Cooperation under Grant 22YH50B.

Acknowledgment

Appreciation is extended to Xin Wang and Ruo Lin for their comments and suggestions on the choice of terms used in this paper.

This study was financially supported by the 2022 Key Project of International Chinese Language Research of the Center for Language Education and Cooperation under Grant 22YH50B.

References

[1] C. J. Lee, "Course improvement through evaluation," *Springer Netherlands*, 1983.

[1] Y. J. Xiao, "Analysis of CIPP education evaluation model," *Education Science*, vol. 3, pp. 42-45, July 2003.

[2] H. X. Xu, R. Wang, and L. T. Ma, "The critical issues in the reform of educational evaluation," *China Exam*, vol. 8, pp. 20-23, Aug. 2020.

[3] B. Q. Liu, T. T. Yuan, Y. C. Ji., B. Y. Liu, and L. Li, "Intelligent technology enabling education evaluation: Connotation, overall framework and practice path," *China Educational Technology*, vol. 8, pp. 16-24, Aug. 2021.

[4] Q. Dong. (Aug. 2020). Education evaluation has entered a new development stage, and the trend of intelligent evaluation has emerged. [Online]. Available: https://www.eol.cn/news/xueshu/201908/t20190807_1677069. shtml

[5] The Xinhua News Agency. *The Overall Plan for Deepening Education Evaluation Reform in the New Era*. [Online]. Available: http://www.gov.cn/zhengce/2020-10/13/content_5551032.htm

[6] Z. K. Yang, "Using information technology to promote the reform and innovation of education and teaching evaluation," *People's Education*, vol. 21, pp. 30-32, Dec. 2020.

[7] R. H. Huang., Y. C. Wang, and Y. L. Jiao, "Education reform in the age of intelligence — On the proposition of two-way empowerment of science & technology and education," *China Educational Technology*, vol. 7, pp. 22-29, Jun. 2021.

[8] K. K. He, J. F. Lin, and W. L. Zhang, *Instructional System Design*, Beijing China, Higher Education Press, 2006, pp. 170-175.

[9] H. C. Jiang and Y. T. Liu, "Construction of teaching quality evaluation system of higher vocational project-based curriculum based on CIPP model," *International Journal of Information and Education Technology*, vol. 11, no. 6, pp. 262-268, 2021.

[10] L. L. Ma, "Construction of evaluation index system of comprehensive practical activity curriculums based on CIPP model," *Teaching and Management*, vol. 3, pp. 115-118, Mar. 2020.

[11] R. H. Huang, H. L. Z. Liu, and X. R. Li, "Trends in computer-assisted assessment," *e-Education Research*, vol. 5, pp. 15-21, May 2002.

[12] X. Tao. (Aug. 2021). Mining the educational value in process data. [Online]. Available: http://www.moe.gov.cn/jyb_xwfb/xw_zt/moe_357/jyzt_2020n/2020_zt21/zhuanjiawenzhang/202012/t20201214_505246.html

[13] J. Wiley, P. Hastings, D. Blaum, A. J. Jaeger, S. Hughes, P. Wallace, T. D. Griffin, and M. A. Britt, "Different approaches to assessing the quality of explanations following a multiple-document inquiry activity in science," *International Journal of Artificial Intelligence in Education*, vol. 27, pp. 758-790, Feb. 2017.

[14] X. M. Yang, J. N. Gu, and B. B. Xie, "The architecture and practice of data-driven education evaluation system in the internet + era," *Journal of Zhejiang Normal University (Social Sciences)*, vol. 44, pp. 16-26, Aug. 2019.
[15] J. J. Liu and X. F. Wu, "Prototype of educational affective arousal evaluation system based on facial and speech emotion recognition," *International Journal of Information and Education Technology*, vol. 9, no. 9, pp. 645-651, 2019.
[16] B. Q. Liu, Y. K. Yu, and T. T. Yuan, "Process evaluation empowered by intelligent technology: goals, paths and typical scenarios," *Modern Educational Technology*, vol. 32, no. 5, pp. 14-23, May 2022.
[17] B. Q. Liu, B.Y. Liu, J. Hu, and T. T. Yuan, "Comprehensive evaluation of intelligent technology empowerment in the new era: Challenges, paths, scenarios and technical applications," *Journal of China Examinations*, no. 6, pp. 6-15, June 2022.
[18] J. E. Lawrence and U. A. Tar, "Factors that influence teachers' adoption and integration of ICT in teaching/learning process," *EMI. Educ. Media Int*., vol. 55, no. 1, pp. 79-105, 2018
[19] J. V. Doremalen, L. Boves, J. Colpaert, C. Cucchiarini, and H. Strik, "Evaluating automatic speech recognition-based language learning system: A case study," *Computer Assisted Language Learning*, vol. 29, no. 4, pp. 833-851, Mar. 2016.
[20] R. L. Ma, G. Cao, M. Wang, and J. Xu, "The application of AES in international Chinese language smart education," in *Proc. 2022 IEEE 4th International Conference on Computer Science and Technology in Education*, 2022, pp. 22-26.
[21] J. Xu, *Evaluation of Distance Learning Chinese as a Second Language*, Beijing China, Chinese Press, 2007.
[22] J. Xu, "Interactive design evaluation of digital teaching of Chinese as a foreign language," in *Proc. the 9th International Society for Chinese Language Teaching*, Beijing, China, Higher Education Press, 2009.
[23] L. Liu and X. H. Liu, "Some thoughts on international Chinese language wisdom education," *Language Teaching and Linguistic Studies*, no. 5, pp. 1-9, Oct. 2022.
[24] Z. T. Zhu, "New developments of smarter education: from flipped classroom to smart classroom and smart learning space," *Open Education Research*, vol. 22, no. 1, pp. 18-25, June, 2016.
[25] F. R. Lin and C. M. Kuo, "Mental effect detection using EEG data in e-learning contexts," *Computers & Education*, vol. 122, pp. 63-79, March 2018.
[26] N. Khamis, "Establishing evaluation criteria for assessing Novices" ability in applying object-oriented concept using Delphi approach," *International Journal of Information and Education Technology*, vol. 6, no. 7, pp. 543-549, 2016.
[27] D. W. S. Tai, R. C. Zhang, Y. T. Wang, and R. Wang, "Determining digital literacy competencies in technical senior high schools using fuzzy Delphi analysis," *International Journal of Information and Education Technology*, vol. 7, no. 8, pp. 612-616, 2017.
[28] D. S. Wahyuni, K. Agustini, and G. Ariadi, "An AHP-based evaluation method for vocational teacher's competency standard," *International Journal of Information and Education Technology*, vol. 12, no. 2, pp. 157-164, 2022.
[29] T. L. Saaty and L. G. Vargas, "Uncertainty and rank order in the analytic hierarchy process," *Eur. J. Oper. Res.*, vol. 32, no. 1, pp. 107-117, 1987.
[30] T. L. Saaty, "A scaling method for priorities in hierarchical structures," *J. Math. Psychol.*, vol. 15, no. 3, pp. 234-281, 1977.
[31] T. Kleickmann *et al*., "Teachers' content knowledge and pedagogical content knowledge: The role of structural differences in teacher education," *J. Teach. Educ.*, vol. 64, no. 1, pp. 90-106, 2013.
[32] M. Heritage and J. Heritage, "Teacher questioning: The epicenter of instruction and assessment," *Appl. Meas. Educ.*, 2013, doi: 10.1080/08957347.2013.793190.

Ruiling Ma is a PhD student of the School of Information Science, Beijing Language and Culture University.

He obtained his master degree in linguistics and applied linguistics. He now majors in language intelligence and technology. His main research interests include language educational technology and teaching second language writing.

Juan Xu is a professor and Ph.D. supervisor of the School of Information Science, Beijing Language and Culture University.

She obtained her Ph.D. in engineering at the China Agricultural University. She has been teaching at Beijing Language and Culture University since 1997. She is now a Ph.D. supervisor in language intelligence and technology. Her main research interests include computer-assisted teaching of Chinese as a second language, online Chinese teaching, and the application of Chinese education technology. She has published over 40 papers in academic journals. As the project leader, she has completed 15 research projects, including the National Social Science Foundation, the Ministry of Education's Science and Technology Research Key Project, and the Ministry of Education's Humanities and Social Sciences Research Project.