Jiang, B. et al. (Eds.) (2025). Proceedings of the 33rd International Conference on Computers in Education.
Asia-Pacific Society for Computers in Education

The Construction of Evaluation Indicators for Intelligent Educational Software and the Measurement of Its Developmental Level

Bo ZHANG[a*], Yuanyuan LI[b], Qianchao XING[c], Yuhan CHEN[d] & Xinying WANG[e]
[a]*XJTLU Entrepreneur College (Taicang), Xi'an Jiaotong-Liverpool University Bo ZHANG, China*
[b]*Shanghai Normal University Yuanyuan LI, China*
[c]*Shanghai Normal University Qianchao XING, China*
[d]*Minhang Yongde School Attached to East China Normal University Yuhan CHEN, China*
[e]*Shanghai Huangpu District Labor and TechnicalEducation Centre, China*
*Bo ZHANG@corresponing.author

Abstract: The evaluation revealed notable disparities across educational scenarios, with strengths in technical quality and user experience but persistent shortcomings in functionality completeness, business alignment, and attractiveness. In 2024, intelligent educational software in China is transitioning from the early phase to more advanced and optimized stages, yet it remains distant from full maturity. By establishing a systematic evaluation framework, this study offers an empirical basis for guiding the high-quality development of intelligent educational software and promoting the broader digital transformation of education.

Keywords: Intelligent educational software, maturity model, evaluation indicator, level measurement

1. Introduction

With the rapid development of generative AI, big data, cloud computing, and the metaverse, educational software has increasingly acquired intelligent features, becoming a critical medium for integrating AI into education. This study defines such tools as intelligent educational software (IES)—applications that leverage technologies such as IoT, big data, AI, and 5G to improve learning performance, optimize teaching, and support diverse educational scenarios (Wang, Li & Yan, 2024). As a key element of "soft" infrastructure, IES plays a central role in advancing educational digital transformation. However, existing assessments of such software remain largely descriptive, lacking systematic and quantitative evaluation frameworks. Current studies typically evaluate educational software along educational and technical dimensions—for example, through usability, functionality, and student engagement—but face persistent challenges: reliance on subjective judgment, poorly structured indicator systems, limited applicability beyond teaching scenarios, insufficient attention to intelligent features, and a lack of empirical validation (Zhang, Goodman & Gu, 2022). To address these gaps, this study reviewed domestic and international evaluation frameworks and identified four key indicators tailored to IES: Functional Integrality, Requirement Matching, Functional Realization, and Charisma Attribute.

Based on these, the Evaluation Indicators System for IES (2024) was developed through collaboration with educators, software developers, and frontline teachers. Using this system, 503 IES products were evaluated and their developmental stages based on the Maturity Model of Digital Transformation for Education (Zhu, Zhang & Dai, 2024). This research aims to advance the transformation and upgrading of IES, provide empirical evidence for optimizing educational practices, and offer insights for implementing China's educational digital transformation strategies.

2. The Construction of Evaluation Indicators

2.1 Design of Indicators

The central role of IES in China's digital transformation lies in embedding AI technologies into diverse educational scenarios. Existing products provide support for personalized learning, smart classrooms, intelligent campus management, and data-driven assessments, thereby empowering three main stakeholders—students, teachers, and administrators (Zhang, et al., 2024). Educational scenarios have gradually shifted from peripheral management contexts to core teaching and learning environments, creating a dynamic system characterized by personalization, intelligence, and ubiquity. In educational research (Yang, et al.,2022), scenarios are typically categorized into learning, teaching, management, and evaluation. Building on this classification, the Evaluation Indicators System for IES (2024) identifies four first-level dimensions: student learning, teacher instruction, educational administration, and teaching evaluation. Under these dimensions, 16 specific educational scenarios are further defined as second-level indicators (Figure 1).

2.2 Design of Indicators

Drawing on Garrett's user experience model, the evaluation framework incorporates four primary indicators—Functional Integrality (FI), Functional Realization (FR), Requirement Matching (RM), and Charisma Attribute (CA)—to assess the effectiveness of Intelligent Education Systems (IES) *(Garrett, 2007)*. FI examines whether IES provides the core functions required across diverse educational scenarios, with Core Functions (CF) serving as a secondary indicator to capture scenario-specific needs. FR evaluates both functional effectiveness and sustained user adoption, operationalized through the Technology Quality Indicator (TQI), which assesses functionality, reliability, usability, maintainability, efficiency, and portability, and the Experience Quality Indicator (EQI), which emphasizes prioritization, consistency, and resilience in user experience. RM measures the degree to which IES functionalities align with scenario-specific requirements, with Educational Scenario Requirements (ESR) ensuring that software functions are systematically developed to match practical educational needs. Finally, CA addresses innovation potential by considering Advanced Function (AF), which evaluates the novelty of functionalities, and Practice Innovation (PI), which examines whether these functionalities generate new teaching or learning practices. Overall, the framework integrates four first-level dimensions (student learning, teacher instruction, educational administration, and teaching evaluation), 16 second-level scenarios, and the four primary indicators (FI, FR, RM, and CA), each supported by corresponding secondary indicators (Figure 2).

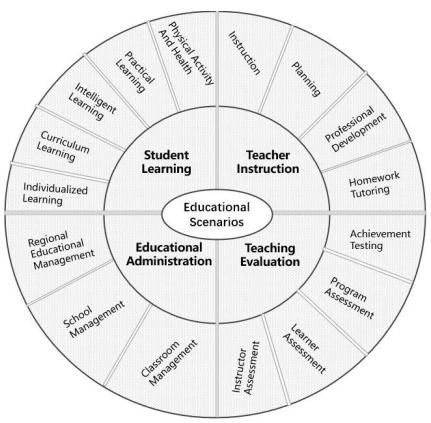

Figure 1. Evaluation Indicators for IES (a)

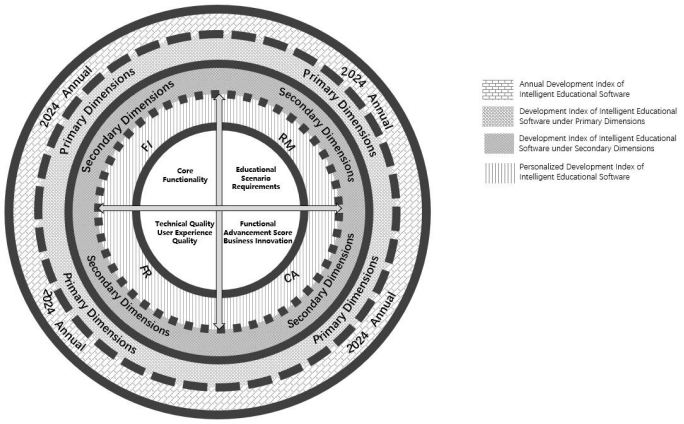

Figure 2. Evaluation Indicators for IES (b)

2.3 Design of Indicator Weights

The Delphi method was employed to determine indicator weights, involving multiple rounds of consultation with frontline educators, software developers, and education experts (Table 1). As this research represents an initial baseline investigation, the Charisma Attribute (CA) was assigned relatively lower weight, while priority was placed on indicators reflecting functional effectiveness and scenario alignment.

Table 1. *Weight Design of Evaluation Indicators for IES*

Primary Dimension	*Secondary Dimension*	*Primary Indicator*	*Secondary Indicator*
Equally Distributed Weights	Equally Distributed Weights	FI, 0.35	CF, 1
		FR, 0.2	TQI, 0.5
			EQI, 0.5
		RM, 0.4	ESR, 1
			AF, 0.5
		CA, 0.05	PI, 0.5

2.4 Quantitative Assessment of Software Scores

The scores of the four primary indicators were first computed and subsequently converted into a percentage scale. Specifically, n denotes the number of core functionalities, m the number of educational scenario requirements, and o the number of charisma attributes. The value ranges for the secondary indicators were defined as follows: CF = 0 or 1; DQI, EQI, ESR = 0 −4; AF, PI = 0−2. Quantitative scoring for each secondary indicator was based on either a three-point or five-point Likert scale.

Step 1. Calculation of primary indicator scores.

$$FI = \frac{\sum_{i=1}^{n} CF_i}{n} \qquad FR = \frac{\sum_{i=1}^{6} TQI_i + \sum_{i=1}^{3} EQI_i}{36} \qquad RM = \frac{\sum_{i=1}^{m} ESR_i}{4m} \qquad CA = \sum_{i=1}^{o} \frac{AF + PI}{2}$$

Step 2. Dimension-level aggregation.

The scores of the four primary indicators under each dimension were calculated using a hierarchical iterative approach.

Step 3. Calculation of the Development Level of IES (DLIES).The final development level scores for each dimension were derived using the following formula:

$$DLIES_i = FI_i * 35\% + FR_i * 20\% + RM_i * 40\% + CA_i * 5\%$$

where i = a, b, c; a represents the year, b denotes the primary dimension, and c corresponds to the secondary dimension.

2.5 Delimitation of Development Stages

Building on expert ratings of development levels, this study further delineates the maturity stages of IES within China's broader educational digital transformation. Drawing on a maturity model of educational digital transformation (Zhu, Zhang & Dai, 2024), a five-level maturity framework was constructed, with score ranges of 0–50, 50–60, 60–70, 70–80, and 80–100, corresponding respectively to the stages of Cognitive, Initial, Intermediate, Optimized, and Mature.

At the Cognitive stage, software lacks core functionalities, exhibits very limited use of intelligent technologies, and shows weak requirement alignment, low realization quality, and no evident innovative attributes. The Initial stage is characterized by the presence of basic functionalities, though integration of intelligent technologies remains limited and software performance across indicators is generally poor. In the Intermediate stage, software demonstrates moderate functional completeness, introduces basic intelligent technologies, and achieves average realization quality and requirement matching, with initial signs of innovative attributes. The Optimized stage reflects substantial functional completeness, widespread application of intelligent technologies, high realization quality, and strong

requirement alignment, with innovation becoming increasingly visible. Finally, at the Mature stage, software exhibits comprehensive core functionalities, effective integration of advanced intelligent technologies, and outstanding performance across all dimensions, with innovation widely and consistently achieved.

3. Investigation into the developmental level of IES

IES should not be viewed merely as a product of intelligent technologies, but rather as a developmental stage in the evolution of educational software toward intelligence. To evaluate the current status, this study identified software registered on the Ministry of Education's Educational Software (APP) Filing Platform (http://app.eduyun.cn) as primary evaluation targets, given its regulatory oversight. Additional widely used and market-recognized applications were also included. Selection criteria required that software: (1) allow registration and trial use; (2) apply to primary education; and (3) integrate intelligent functional features. After screening 3,080 applications, 503 were retained for analysis. These applications were evaluated using the Evaluation Indicators System for IES (2024) and the proposed maturity framework, enabling assessment of both development levels and maturity stages.

3.1 Analysis of Developmental Levels under Secondary-Dimension Scenarios

Evaluation scores across secondary-dimension scenarios ranged from 54.07 to 72.26 (Table 4). Educational administration within school management achieved the highest performance (Optimized stage), while curriculum learning in student learning recorded the lowest (Initial stage), reflecting weak classroom integration. Most scenarios clustered at the Intermediate stage, including homework tutoring, assessment, online instruction, and teacher professional development, showing that intelligent software is beginning to support core teaching and evaluation practices. By contrast, applications concentrated in student learning remain largely at the Initial stage, underscoring substantial gaps but also significant potential for future innovation.

Table 4. *Developmental Levels of IES Across Different Dimensional Educational Scenarios*

Educational Scenario under Secondary Dimensions		***FI***	***FR***	***RM***	***CA***	***Evaluation Score for Development Level***	***Developmental Stage***
Educational Administration	School Management	77.62	75.74	71.89	23.74	72.26	Optimized
	Classroom Management	73.44	77.44	70.11	22.95	70.38	Optimized
Teacher Instruction	Homework Tutoring	65.32	90.73	67.90	14.43	68.89	Intermediate
	Instruction	62.11	89.54	64.05	10.88	65.81	Intermediate
	Professional Development	58.98	88.78	61.06	12.28	63.44	Intermediate
Teaching Evaluation	Planning	56.64	88.06	58.62	10.00	61.38	Intermediate
	Learner Assessment	69.48	83.07	68.13	0.00	68.18	Intermediate
	Achievement Assessment	63.97	86.34	59.56	9.76	63.97	Intermediate
Student Learning	Intelligent Learning	57.35	85.97	48.90	14.90	57.57	Intermediate
	Individualized Learning	54.61	85.01	46.59	12.95	55.40	Initial
	Practical Learning	53.95	84.81	45.90	12.43	54.83	Initial
	Physical Activity and Health	53.72	84.77	45.65	12.38	54.63	Initial
	Curriculum Learning	52.79	84.51	45.10	12.93	54.07	Initial

Educational Scenario under Primary Dimensions	*FI*	*FR*	*RM*	*CA*	*Evaluation Score for Development Level*	*Developmental Stage*
Educational Administration	75.53	76.59	71.00	23.34	71.32	Optimized
Teacher Instruction	60.76	89.28	62.91	11.89	64.88	Intermediate
Teaching Evaluation	66.73	84.71	63.84	4.88	66.08	Intermediate
Student Learning	54.48	85.01	46.43	13.12	55.30	Initial
2024	64.38	83.90	61.05	13.31	64.39	Intermediate

3.2 Analysis of Development Levels of IES in Primary-Level Educational Scenarios

As shown in Table 4, Intelligent Education Systems (IES) varied notably across scenarios. Educational administration achieved the highest score (71.32, Optimized), followed by teaching evaluation (66.08) and teacher instruction (64.88), both at the Intermediate stage. Student learning, however, scored lowest (55.30), remaining at the Initial stage. This reveals a clear disparity: administration and evaluation are relatively advanced, instruction is mid-level, and student learning lags behind.

Administration software recorded high FI and RM scores and the strongest CA performance, but its FR score was the lowest. This suggests bold but uneven experimentation, reflecting both progress and challenges given the complexity of administrative functions. Teacher instruction tools scored moderately in FI and RM but high in FR, indicating efficiency gains in areas such as homework tutoring, though innovation remains limited. Teaching evaluation software showed similar patterns, with strong FR yet modest CA, pointing to partial digital transformation but insufficient integration with evaluative needs.

Student learning software performed relatively well in FR but weakly in FI, RM, and CA. Given the complexity of learning processes, effective design requires both adaptive resources and interactive tools, alongside cautious integration of intelligent technologies. Despite current limitations, this domain holds the greatest potential for future development.

3.3 Annual Analysis of Development Levels of IES

Evaluation of 503 software products indicates that in 2024, the overall development level of IES reached a score of 64.39, placing it within the “Intermediate” stage. FI and RM scored 64.38 and 61.05 respectively, while FR and CA scored 83.90 and 13.31. These findings highlight several trends. First, software design has begun to address the core functional demands of educational practice, though greater precision is needed in identifying user requirements and improving iterative development. Second, while the technical and experiential quality of functionalities generally satisfies educational stakeholders, substantial scope remains for enhancing FI and RM to better align with actual practice. Third, innovation in CA is particularly limited, signaling the need for greater emphasis on charisma attributes to ensure wider acceptance and sustained engagement. Overall, IES has transitioned beyond the initial stage, yet remains far from maturity.

4. Conclusions and Recommendations

This study applied a 2024 evaluation framework comprising four indicators—FI, FR, RM, and CA—to assess 503 IES products across diverse educational scenarios. Three key findings emerged. First, significant disparities exist: educational administration software reached the Optimized stage, teaching evaluation and teacher instruction achieved the Intermediate stage, while student learning software remains at the Initial stage. Second, although FR and user experience scores were relatively strong, FI and RM barely passed, reflecting limited functional coverage and weak stakeholder alignment. Third, overall IES development has

advanced beyond early phases but has not yet reached maturity. Based on these findings, three recommendations are offered.

4.1 Promote High-Quality and Balanced Development of IES

IES supports multiple domains—student learning, teacher instruction, evaluation, and administration. It facilitates personalized learning, instructional efficiency, evidence-based evaluation, and administrative effectiveness. However, development is uneven: administration has advanced rapidly, evaluation and instruction show moderate progress, while student learning lags significantly. Achieving high-quality digital transformation requires prioritizing student learning software, while sustaining innovation in instructional and evaluative tools.

4.2 Strengthen Core Functionalities and User Alignment

Although technical reliability is adequate, insufficient FI and RM scores indicate misalignment between software capabilities and user needs. Governments should establish structured feedback channels between educators and developers to ensure context-relevant functionalities. Researchers should ground design in educational theory and scenario-specific requirements, while developers should enhance usability, simplify operations, and integrate coherent functions. The consistently low CA scores further highlight the need for scenario-adaptive innovation and more intelligent, pedagogy-driven applications.

4.3 Foster Multi-Stakeholder Collaborative Innovation

Advancing IES from intermediate or optimized stages to maturity requires coordinated efforts from government, industry, academia, and schools. Governments should refine regulatory frameworks and provide adoption incentives. Developers should prioritize core functions and iterative improvement in close collaboration with educators. Researchers should generate theoretical insights and practice-oriented feedback, while schools should foreground authentic classroom needs. Building dynamic, multi-stakeholder feedback loops will accelerate innovation, deepen educational transformation, and ensure the sustainable evolution of IES.

Acknowledgements

This work was supported by the XJTLU Research Development Fund Project (Grant No. RDF-24-02-021)

References

Garrett, J., “The Elements of User Experience: User-Centered Web Design, ” Beijing: Mechanical Industry Press, 2007, pp. 34–153.

Wang, C., Li, Y., & Yan, K. (2024). Key Scenarios and Promotion Strategies for Intelligent Education Software Promoting Future Education Development: Excerpt from the 2023 Annual Report on Artificial Intelligence Promoting the Development of Education (Ⅶ). *Chinese Journal of ICT in Education, 30*(07), 73-82.

Zhang, B., Goodman, L., & Gu, X. (2022). Novel 3D contextual interactive games on a gamified virtual environment support cultural learning through collaboration among intercultural students. SAGE Open, 12(2), 21582440221096141.

Zhang, B., Chen, Y., Wang, X., and Li, Y., “Construction of an Evaluation Index System for Intelligent Educational Software Oriented to Student Learning: Excerpt from the 2023 Annual Report on Artificial Intelligence Promoting the Development of Education (Ⅴ),” Chinese Journal of ICT in Education, vol. 30, no. 07, pp. 54-63, 2024.

Zhu, Z., Zhang, B., & Dai, L. (2024). The change and the unchanged in education empowered by digital intelligence technology. *Chinese Journal of ICT in Education*, 30(3), 3-14.

Zhang, B., Goodman, L., & Gu, X. (2023). Telecollaboration tool preferences for online intercultural learning in higher education: Perspectives of Chinese international students. Sage Open, 13(2), 21582440231180087.