International Journal of
Humanities, Education, and Social Sciences

e-ISSN : 3026-0892
p-ISSN : 3026-1422
Index: Harvard, Boston, Sydney, Dimensions, Lens, Scilit, Semantic, Google, etc

https://doi.org/10.58578/IJHESS.v4i2.8980

Toward a Digital-Enabled Evaluation Model for Graduate Education Quality: A Study of Chinese Universities

Yue Liao, Jing Quana, Shiyan Chen, Can Zeng
Chongqing University of Technology, Chongqing, China
jquan@cqut.edu.cn

Article Info:

Submitted:	Revised:	Accepted:	Published:
Jan 27, 2026	Mar 15, 2026	Apr 8, 2026	Apr 13, 2026

Abstract

In the context of high-quality development in higher education, evaluating graduate education quality has become increasingly important for assessing the cultivation of high-level talent. This study aimed to construct a graduate education quality evaluation system grounded in comprehensive competency development, encompassing five key dimensions: digital awareness, digital technology knowledge and skills, digital application, digital social responsibility, and professional development. The study employed the Analytic Hierarchy Process to determine indicator weights, Data Envelopment Analysis to measure institutional efficiency across China, panel regression to identify factors influencing educational value-added, and K-means clustering to reveal patterns of resource allocation. The findings indicate that digital technology knowledge and skills, together with digital application, carry the highest weights in the evaluation system. The results further show substantial variation in efficiency across institution types and regions, while also demonstrating that high investment does not necessarily translate into high educational value-added, thereby underscoring the importance of resource utilization efficiency and appropriate training models. The study concludes that a more differentiated evaluation mechanism and optimized resource allocation are necessary to enhance graduate education quality. These findings provide a scientific basis for improving graduate education evaluation and offer practical implications for policy and institutional quality enhancement.

Keywords: Graduate Education Quality; Analytic Hierarchy Process; Data Envelopment Analysis; Panel Regression; K-Means Clustering

Introduction

In the context of intensifying global competition in higher education, the quality of graduate education has emerged as a critical strategic asset and a key indicator of a nation's innovative capacity and long-term core competitiveness. As the highest level of talent cultivation within the national education system, graduate students not only shoulder the fundamental responsibilities of advancing scientific research, driving technological innovation, and cultivating high-level professional reserves, but also constitute a pivotal intellectual force propelling sustained socio-economic development and transformative technological progress. However, the prevailing graduate education quality evaluation system predominantly relies on static, output-oriented indicators—such as publication counts, graduation rates, and short-term employment outcomes—while largely overlooking dynamic, process-based measurements of cultivation effectiveness and longitudinal capability enhancement. This structural limitation results in an evaluation framework that is unable to comprehensively capture students' developmental trajectories, adaptive learning processes, and incremental competency growth throughout their academic and research endeavors. At the national policy level, China's "Overall Plan for Deepening the Reform of Educational Evaluation in the New Era" explicitly advocates for the systematic improvement of the graduate education quality evaluation system. It emphasizes shifting toward a more holistic model that integrates comprehensive assessment of academic progress, research outputs, and overall professional competence, thereby aligning evaluation practices with the broader objectives of national talent development and innovation-driven growth. In this context, how to construct a scientific, operational, and dynamic evaluation system that can reflect the growth of graduate students has become an urgent issue to be addressed.

Given the diverse types of higher education institutions in China and the uneven regional development, a unified evaluation standard struggles to cover the differences in institutional positioning, discipline development, and student cultivation objectives across various institutions. Therefore, constructing a scientific evaluation system that integrates capability development orientation and classification evaluation principles, with a focus on educational digitization, to dynamically reflect graduate students' growth processes and digital competency enhancement levels, holds significant importance for promoting high-quality development in graduate education and achieving evaluation-driven development. In terms of practical value, the research results of this study can provide important decision-making references for the reform and development of graduate education in China. On the one hand, for universities, a scientific and reasonable graduate education quality evaluation system can help them accurately grasp the advantages and disadvantages of their own talent training work, identify the key links and weak links affecting education quality, and then optimize the talent training program, adjust the curriculum system, scientifically allocate educational resources, and improve the precision and adaptability of talent training, so as to continuously enhance the core competitiveness of their graduate education. On the other hand, for educational administrative departments, this study can provide a theoretical basis and practical reference for formulating differentiated graduate education evaluation policies and resource allocation plans. By considering the differences in school types, regional characteristics, and development stages of various institutions, it can promote the rational flow and balanced allocation of educational resources, narrow the development gap between different regions and types of institutions, and help build a more fair, efficient, and sustainable graduate education quality ecosystem.

Research on the evaluation of graduate education quality in China exhibits a trend toward interdisciplinary integration and multidimensional development. Grounded in the Plan–Do–Check–Act (PDCA) cycle, recent studies argue that iterative feedback loops are essential for sustainable improvement of graduate-education quality. Shi and Niu (2023) demonstrate that embedding PDCA within university-level quality-assurance procedures continuously refines teaching processes and, by extension, the competencies of master's and doctoral graduates. This finding aligns with broader higher-education literature in which PDCA is framed as a dynamic alternative to static compliance models (Li et al., 2023).

Complementing the process perspective, Zhang and Wang (2020) combine combination-weighting techniques with a comprehensive fuzzy-evaluation model to operationalise graduate-programme quality. Their indicator system spans inputs (faculty, finance), process (curriculum delivery, supervision) and outputs (publications, employment outcomes), echoing the classic "input–process–output" architecture later formalised by Wang et al. (2022) for STEM graduate training. The inclusion of entropy-weighted indices and machine-learning classifiers has been shown to increase the discriminative power of such systems across heterogeneous disciplines (Wu, 2022).

A second stream of research emphasises multi-stakeholder participation. Chauveron et al. (2021) contend that incorporating student, industry and community voices mitigates evaluator bias and strengthens the legitimacy of programme assessments. Their youth-character evaluation framework—adapted by Ferreira et al. (2023) for peer-feedback loops in clinical nursing education—offers a transferable template for graduate programmes seeking to balance academic and societal expectations. Albers et al.'s theory-of-change approach for learning networks provides a promising but under-utilised remedy (Albers, et al., 2024). Institutional adoption barriers—particularly ambiguous criteria and insufficient staff capacity—are frequently documented in case studies of regional universities (Zhang et al., 2020). Zhang et al. (2025) integrated data-mining tools to automate indicator extraction from learning-management systems.

This study takes educational digitization as its entry point, focusing on the transformation and upgrading of the Chinese graduate education quality evaluation system, aiming to expand the application boundaries of competency-based evaluation theory in higher education within a digital environment. By introducing an evaluation perspective that combines "static level" and "dynamic value-added," this study constructs an integrated evaluation framework that incorporates process-oriented, developmental, and personalized characteristics. This contributes to the refinement of existing educational evaluation theoretical frameworks and deepens theoretical understanding of core graduate competencies such as digital literacy, research capabilities, and interdisciplinary integration skills. Additionally, the research helps address the evolution of the essence of education and talent cultivation objectives in the digital age, providing theoretical support for achieving a deep integration of evaluation and cultivation.

The findings of this study hold significant practical value for driving transformative changes in graduate education quality practices. On one hand, establishing a scientifically sound and reasonable graduate education quality evaluation system can provide decision-making basis for universities to optimize talent cultivation programs, scientifically allocate educational resources, and enhance the quality of educational provision, thereby enhancing the precision and adaptability of talent cultivation. On the other hand, this research can also provide theoretical references for education authorities to formulate differentiated evaluation policies tailored to specific institutions and local conditions, promoting the rational flow and balanced development of regional higher education resources, and contributing to the construction of a more equitable and efficient graduate education quality ecosystem.

Methods

Data Sources

To protect the privacy of universities and graduate students, this study did not employ actual institutional data. Instead, a simulated dataset was developed based on the operational framework of graduate education. The simulation was informed by official statistics from the Chinese Ministry of Education and indicator characteristics drawn from publicly available annual reports of selected universities. This approach ensures that the dataset reflects real-world patterns in terms of overall distribution, inter-indicator relationships, and logical consistency, thereby supporting the validity and reliability of the subsequent analysis and findings.

Construction of Graduate Education Quality Evaluation Indicators

The graduate education quality system is based on the "Teacher Digital Literacy" standards **Error! Reference source not found.**, combined with the digital practices of university teachers in teaching and research, to form a two-level evaluation indicator system. The criteria-level indicators consist of five items: digital awareness, digital technology knowledge and skills, digital application, digital social responsibility, and professional development. The scheme-level indicators consist of 15 items, which further specify and quantify the criteria-level indicators to form a two-tiered indicator system tailored to the actual conditions of universities. The specific evaluation indicator system is shown in Table 1.

Table 1 Indicators and Explanations of Indicators

Principle-level indicators	Program-level indicators	Explanation
digital awareness	Understanding the trend toward digitization in education	Level of awareness of trends in the digital transformation of education

	Teaching Impact Sensitivity	Sensitivity to the impact of digital technology on teaching
	Policy Learning Frequency	Frequency of actively learning about digital education-related policies Ability to use professional software such as programming/ modeling/simulation (e.g., Python, SPSS, etc.)
Digital Technology Knowledge and Skills	Professional software usage skills	Frequency of actively learning about digital education-related policies Ability to use professional software such as programming/ modeling/simulation (e.g., Python, SPSS, etc.)
	Familiarity with office/teaching software	Teachers' proficiency in commonly used office/teaching software (such as PowerPoint, WPS, Word)
	Online teaching platform functions	Master the various functions of online teaching platforms (such as Yuka Classroom, SuperStar, MOOC, etc.).
Digital Applications	Proportion of multimedia content	The proportion of multimedia/micro-lessons/ MOOC courses introduced into teaching
	Frequency of use of teaching platforms	Frequency of use of online teaching platforms
	Use of teaching data analysis	Conducting teaching analysis activities using data analysis platforms (number of times per semester/whether teaching analysis reports are produced, etc.)
Digital Social Responsibility	Participation in network governance	Teachers' participation in the governance of online spaces (reporting inappropriate content, opposing AI cheating, etc.)
	Level of mastery of online ethics	Teachers' understanding of online ethics and copyright awareness
	Student Data Privacy Protection	The standardization of student information and data privacy protection practices (whether course data permissions are set, whether data is anonymized)
Professional Development	Online training frequency	Frequency of online participation in teacher training, professional development, and lectures
	Participation in resource platform construction	Whether teachers participate in the development of teaching platforms and digital resources
	Digital scientific research and teaching innovation	Whether teachers use digital technology for research/teaching innovation (e.g., creating MOOCs, writing teaching blogs)

Confirmation of indicator weights

The Analytic Hierarchy Process (AHP) is a method that breaks down complex problems into several levels and factors, and then uses pairwise comparisons to determine the relative importance of each factor, thereby comprehensively deriving the priority order of decision-making options. It converts qualitative problems into quantitative analysis and can effectively handle multi-objective, multi-criteria decision-making problems.

Constructing a judgment matrix. The judgment matrix $A = (a_{ij})_{n\times n}$ is based on comprehensive considerations, comparing factors pairwise, and quantitatively representing the importance of one factor relative to another. It has the following properties:

$$①a_{ij} > 0$$

$$②a_{ii} = 1, (i = 1,2,\dots,n)$$

$$③a_{ij} \cdot a_{ji} = 1, (i,j = 1,2,\dots,n)$$

Criterion-level judgment matrix: The relative importance of each factor in the criterion layer relative to the objective layer is determined through expert scoring or pairwise comparison methods. Solution layer judgment matrix: For each criterion, the relative importance of each solution in the solution layer is compared, and a judgment matrix is constructed. This study adopts the 1-9 scale method to quantify the relative importance of factors in the judgment matrix, with specific scale definitions and explanations as shown in Table 2.

Table 2 Explanation of quantity scales

Scale	Definition	Explanation
1	Equally important	When comparing two factors, they are equally important.
3	Slightly important	When comparing two factors, the row factor is slightly more important than the column factor.
5	Significantly important	When comparing two factors, the row factor is significantly more important than the column factor.
7	Much more important	When comparing two factors, the row factor is much more important than the column factor.
9	Extremely important	When comparing two factors, the row factor is extremely important compared to the column factor.
2，4，6，8	Equally important	The intermediate value representing the above adjacent judgments.
The reciprocals of 1 to 9 1~9	Slightly important	If factor a_i is compared with factor a_j to obtain judgment r_{ij}, then factor a_j is compared with factor a_i to obtain judgment r_{ji}=1/r_{ij}

In this task, the opinions of 10 experts were collected using the Delphi method, and pairwise comparisons and scoring were performed for the five criteria in the criterion layer, as well as for the fifteen criteria in the solution layer, resulting in the judgment matrix.

Calculate the corresponding weights. If $A = (a_{ij})_{n\times n}$ is the judgment matrix, and $W = (W_1, W_2, ..., W_n)$ is the weight vector of A. Using the eigenvector method, first calculate the maximum eigenvalue λ_{max} of the judgment matrix using formula 2.1; then calculate the geometric mean of each row element using formula 2.2, and finally normalize the result to obtain the weights using formula 2.3.

$$A \cdot W = \lambda_{max} \cdot W \quad (2.1)$$

$$\bar{W}_i = \sqrt[n]{\prod_{j=1}^{n} a_{ij}} \quad (2.2)$$

$$W_i = \frac{\bar{W}_i}{\sum_{i=1}^{n} \bar{W}_i} \quad (2.3)$$

where W_i is the weight of factor a_i, and $\sum_{i=1}^{n} W_i = 1, W_i \geq 0 (i = 1,2, ..., n)$. The weights obtained from the scores given by ten experts are averaged to obtain the criterion layer weights and scheme layer weights in Tables 3 and 4.

Table 3 Weighting of criteria layer indicators

Guideline-level indicators	Weighting	Guideline-level indicators	Weighting
Digital awareness	0.1116	Digital Social Responsibility	0.0702
Digital knowledge and skills	0.3775	Professional Development	0.1809
Digital applications	0.2597		

Table 4: Weights of indicators at the program level

Scheme-level indicators	weighting	Scheme-level indicators	weighting
Awareness of trends in education digitization	0.5736	Use of teaching data analysis	0.1245
Sensitivity to teaching impact	0.2864	Participation in network governance	0.1721
Frequency of policy learning	0.1399	Level of understanding of network ethics	0.5237
Professional software usage skills	0.1824	Protection of student data privacy	0.3042

Scheme-level indicators	weighting	Scheme-level indicators	weighting
Familiarity with office/teaching software	0.4657	Frequency of online training	0.3659
Mastery of online teaching platform functions	0.2471	Participation in resource platform construction	0.1567
Proportion of multimedia content	0.2931	Innovation in digital scientific research and teaching	0.3651
Frequency of teaching platform use	0.4676		

The final integrated weights are obtained from formula (2.4) as shown in Table 5.

Integrated weight = criterion layer weight * indicator layer relative weight （2.4）

Table 5 Integrated weights

Scheme-level indicators	Weight	Scheme-level indicators	Weight
Awareness of digitalization trends in education	0.064014	Use of teaching data analysis	0.032333
Sensitivity to teaching impact	0.031962	Participation in network governance	0.012081
Frequency of policy learning	0.015613	Level of understanding of network ethics	0.036764
Professional software usage skills	0.068856	Protection of student data privacy	0.021355
Familiarity with office/teaching software	0.175802	Frequency of online training	0.066191
Mastery of online teaching platform functions	0.09328	Participation in resource platform construction	0.028347
Proportion of multimedia content	0.076118	Innovation in digital scientific research and teaching	0.066047
Frequency of teaching platform usage	0.121436		

Consistency test. The consistency of the comparison and judgment process must be tested to determine whether the calculated weights are reasonable.

$$CI = \frac{\lambda_{max} - n}{n-1} \qquad (2.5)$$

$$CR = \frac{CI}{RI} \quad (2.6)$$

The consistency index CI is calculated using the above formula. It is an indicator that measures the consistency of the judgment matrix, where n is the order of the judgment matrix. CR is the consistency ratio, which is an indicator of whether the judgment matrix has satisfactory consistency. If CR < 0.1, the consistency of the judgment matrix is considered acceptable. The results of the consistency test of the criterion layer indicators (top three experts) calculated in this study are shown in Table 6.

Table 6 Consistency test of criteria-level indicators (top three experts)

expert	Consistency Index CI	Consistency ratio CR	Passed inspection
expert 1	0.0583	0.0520	Passed
expert2	0.0480	0.0429	Passed
expert3	0.0306	0.0273	Passed

Value-Added Evaluation Model for Graduate Education Quality

In order to comprehensively assess the relative efficiency of different graduate education institutions in terms of resource allocation and utilization, this study selected indicators such as research funding investment, faculty mentoring time, investment in experimental platform construction, scholarship and grant amounts, and academic exchange funding to reflect the input of human, material, and financial resources during the educational process; output indicators include thesis quality scores, the quantity and quality of research papers, participation in research projects, innovation capability assessment scores, and the number of academic exchanges, to reflect the level of educational outcomes and students' comprehensive capabilities.

Data Envelopment Analysis Model and Efficiency Measurement. Data Envelopment Analysis (DEA) is a non-parametric efficiency evaluation method based on linear programming, suitable for comparing the relative effectiveness of decision-making units (DMUs) under conditions of multiple inputs and multiple outputs. Unlike traditional regression analysis, DEA does not assume a functional relationship between inputs and outputs. Instead, it measures the relative efficiency of each DMU by constructing an efficiency frontier, thereby objectively evaluating the resource utilization levels of various graduate education units without relying on specific production function assumptions. This study adopts the CCR model, which assumes constant returns to scale (CRS) and reflects overall technical efficiency (TE). The technical efficiency values for the development of graduate education quality at selected universities from 2023 to 2025 are shown in **Table 7**.

Table 7 Technical Efficiency Values for the Development of Graduate Education Quality at Selected Universities

School Name	2023 Efficiency Value	2024 Efficiency Value	2025 Efficiency Value
Double First-Class Universities 1	0.026	0.026	0.025
Double First-Class Universities 2	0.031	0.031	0.030
Double First-Class Universities 3	0.028	0.029	0.027
Double First-Class Universities 4	0.027	0.027	0.026
Double First-Class Universities 5	0.028	0.028	0.027

Panel regression model. To explore the impact of university classification attributes (double first-class universities/ordinary universities) and regional location (eastern/central/western) on the added value of graduate education, this study constructed a panel data regression model, controlling for other potential confounding factors, to quantify the differentiated effects of different dimensions. The four panel regression models constructed are as follows:

Model 1: The impact of school type on the added value index

$$GI_{jt} = \beta_0 + \beta_1 Type_{jt} + \epsilon_{jt}$$

Among these, GI_{jt} denotes the value-added index of university j in period t, $Type_{jt}$ is the dummy variable for the university (1 for double-first-class universities and 0 for ordinary universities), and ϵ_{jt} represents the random error term at the individual time point.

Model 2: The impact of region on the value-added index

$$GI_{jt} = \beta_0 + \beta_1 East_{jt} + \beta_2 Central_{jt} + \epsilon_{jt}$$

Among these, $East_{jt}$ denotes the region dummy variable (1 for eastern universities, 0 otherwise), $Central_{jt}$ denotes the region dummy variable (1 for central universities, 0 otherwise), and the reference group is $East_{jt} = 0$ and $Central_{jt} = 0$ (western universities).

Model 3: Simultaneously considers the impact of school type, region, and university investment on the value-added index

$$GI_{jt} = \beta_0 + \beta_1 Type_{jt} + \beta_2 East_{jt} + \beta_3 Central_{jt} + \beta_4 infrastructure_{jt} + \beta_5 Training_{jt} + \beta_6 Resources_{jt} + \beta_7 Size_{jt} + \epsilon_{jt}$$

Where $infrastructure_{jt}$t represents the logarithmic form of the digital infrastructure investment of university j in period t, $Training_{jt}$ represents the logarithmic form of the teacher training funding of university j in period t, $Resources_{jt}$ represents the logarithmic form of the digital resource investment of university j in period t, and $Size_{jt}$ represents the size of university j in period t.

Model 4: The impact of interaction terms (school type and region) on the value-added index

$$\begin{aligned} GI_{jt} = \beta_0 + \beta_1 Type_{jt} + \beta_2 East_{jt} + \beta_3 Central_{jt} + \beta_4 (Type_{jt} \times East_{jt}) \\ + \beta_5 (Type_{jt} \times Central_{jt}) + \beta_6 infrastructure_{jt} + \beta_7 Training_{jt} \\ + \beta_8 Resources_{jt} + \beta_9 Size_{jt} + \epsilon_{jt} \end{aligned}$$

Among these, $Type_{jt} \times East_{jt}$ represents the synergistic effect of double first-class universities in the eastern region, and $Type_{jt} \times Central_{jt}$ represents the synergistic effect of double first-class universities in the central region.

By estimating regression coefficients, we can quantify the extent to which university classification and region influence the value added of digital competence. Positive coefficients indicate that the factor promotes the value added of digital competence, while negative coefficients indicate that the factor inhibits the value added of digital competence. The magnitude of the absolute value of the regression coefficient reflects the strength of the factor's influence.

K-means Cluster Analysis. To identify the resource allocation characteristics of different universities in terms of support for the value-added development of teachers' digital competencies, this study constructed a clustering model based on key input dimensions. The core idea of this model is to classify universities based on their actual investments in teacher support resources, thereby uncovering their intrinsic differences and potential strategic pathways in promoting the development of teachers' digital competencies. Three indicators were selected: investment in teacher digital technology training (training), investment in digital infrastructure construction (infrastructure), and investment in the development of digital educational resources (resources). These serve as the core clustering dimensions reflecting the “hardware and software” resource supply of universities' graduate education quality support systems. To ensure the stability of the clustering results, this study took the average values of the variables for each university during the observation period, eliminating the impact of annual fluctuations to reflect long-term investment strategies. Additionally, the data was standardized to eliminate differences in units of measurement and enhance the reliability of the clustering.

K-means clustering is a classic unsupervised learning algorithm aimed at dividing a dataset into K mutually exclusive clusters, ensuring that data points within the same cluster are as similar as possible and those across different clusters are as dissimilar as possible. Its core principle involves iteratively optimizing to minimize the sum of squared distances between data points and their respective cluster centers. The algorithm steps are as follows:

(1) Initialization. The number of clusters K must be predefined, and K data points are randomly selected as initial cluster centers.

(2) Assign data points. Calculate the distance between each data point and all cluster centers, typically using Euclidean distance, then assign each data point to the cluster corresponding to the nearest cluster center.

(3) Update cluster centers. For each cluster, recalculate its center position as the mean of all data points in that cluster.

(4) Iteration and convergence. Repeat steps (2) and (3) until either the cluster centers no longer change significantly (i.e., convergence) or the predefined maximum number of iterations is reached, satisfying both termination conditions.

The optimization objective of K-means is to minimize the squared error:

$$J = \sum_{i=1}^{K} \sum_{x \in C_i} ||x - \mu_i||^2 \qquad (2.7)$$

Among them, C_i represents the $i.$ th cluster, μ_i represents the center of the $i.$th cluster, and x represents the data point.

Empirical Analysis and Results

AHP Weight Analysis

The results of the indicator weight calculation based on the AHP show that the importance of each dimension in the graduate education evaluation system varies significantly (see Table 3). At the criterion level, "digital technology knowledge and skills" has the highest weight (0.3775), indicating that in the process of graduate education, the ability of mentors and students to master professional software and online research platforms is a core support; "Digital Application" follows closely behind (0.2597), highlighting the critical role of teaching platform usage frequency and research data analysis capabilities in graduate students' innovative practices.

From the perspective of the scheme layer indicators (see T**able 4**), "Familiarity with Office/Teaching Software" (0.4657), "Teaching Platform Usage Frequency" (0.4676), and "Mastery of Online Ethics" (0.5237) lead in weight within their respective criterion layers. This reflects the emphasis placed by graduate education on the normative use of basic tools and research data ethics.

Consistency test results show that the consistency ratio (CR) of the judgment matrices of the 10 experts is all less than 0.1 (see **Table 6**), with the CR values of the top three experts being 0.0520, 0.0429, and 0.0273, respectively, indicating that the weight allocation has a high degree of logical consistency, and the indicator system design is scientifically reliable.

DEA efficiency calculation results

The calculation of the efficiency of graduate education development in universities from 2023 to 2025 using the DEA model shows (see **Table 7)** that the average technical efficiency value of double first-class universities (0.028) is higher than that of non-double first-class universities (0.025), but the overall trend is a slight decrease year by year. This indicates that high-quality universities are experiencing diminishing marginal returns in the matching of resource input and output amid digital transformation. The phenomenon is closely related to the tendency in graduate education to prioritize scale over quality: the expansion of enrollment at some Double First-Class universities has led to a substantial increase in the student-supervisor ratio, the dispersion of key research resources (such as digital platform access and project funding), and inadequate adaptation of training models to digital needs, ultimately resulting in a decline in overall resource utilization efficiency and educational output quality.

Regional difference analysis shows (**Figure 1**) that the median efficiency value of universities in the central region is the highest (0.035), but the dispersion is significant as indicated by a larger interquartile range. This reflects that some universities in the central region (such as regional key normal universities and specialized characteristic institutions) have achieved efficient graduate training through precise resource allocation—focusing on advantageous disciplines and aligning digital resources with training objectives—while others suffer from resource waste due to unreasonable investment structures and inadequate integration of digital technology with teaching. The efficiency values in the eastern region are stable with a narrow distribution but contain low-efficiency outliers, possibly due to abundant research funding yet redundant management, overlapping digital infrastructure investments, and insufficient synergy between resource input and academic output in some institutions. The efficiency value in the western region is relatively balanced (0.0352), benefiting from the precise implementation of national policies supporting graduate education in the western region, such as targeted funding for digital infrastructure and paired mentor training programs that effectively mitigate regional resource gaps.

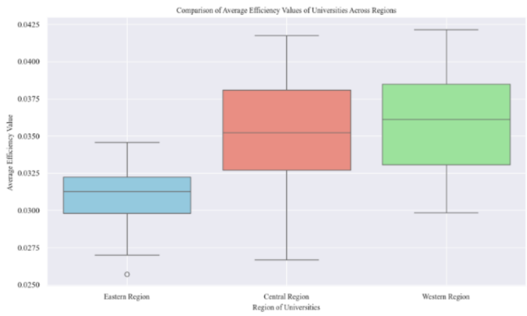

Figure 1 Comparison of average efficiency values among universities in different regions

From an annual trend perspective (**Figure 2**), the technical efficiency value of non-Double First-Class universities has consistently exceeded that of Double First-Class universities over the 2023–2025 period, with the gap widening to 0.003 by 2025. This trend indicates that non-Double First-Class universities, constrained by relatively limited resource endowments, place greater emphasis on refined resource allocation and efficient utilization in graduate education.

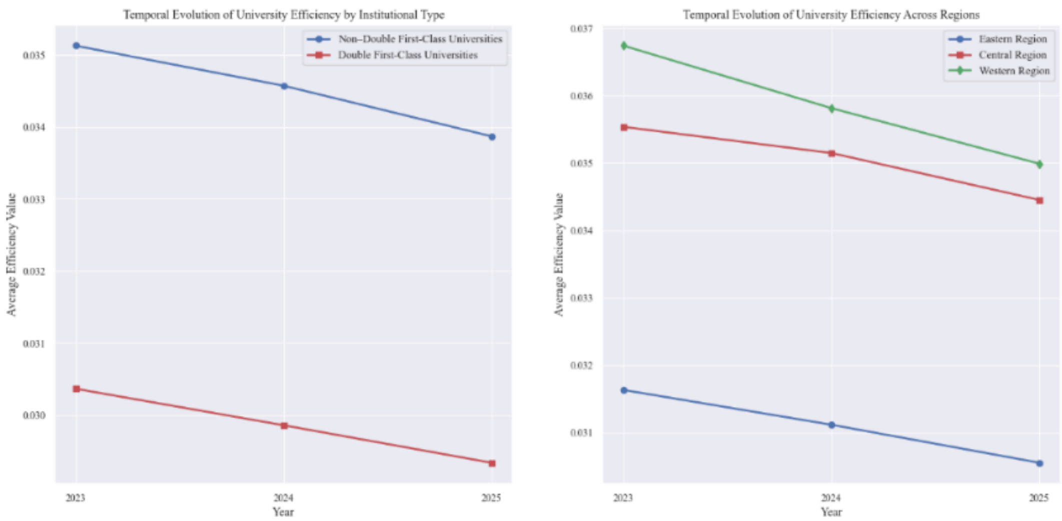

Figure 2: Annual Trends in Efficiency Values of Universities by Type and Region

Panel Regression Results Analysis

The Impact of School Type and Regional Factors. Panel regression model 1 (Table 8) shows that school type has a significant positive impact on the value-added index of graduate education, indicating that double-first-class universities, with their superior research platforms, high-level mentor resources, and intensive academic exchange networks, can more effectively promote the improvement of graduate students' digital research capabilities, interdisciplinary integration skills, and innovative practice abilities. These institutional advantages provide graduate students with access to cutting-edge research projects and collaborative opportunities that accelerate competency development. Model 2 (Table 9) further reveals regional differences, with the value-added index in the eastern region being significantly higher than that in the western region, benefiting from the region's abundant digital resources and vibrant academic ecosystem.

Table 8 Regression results for Model 1

Variable	Coefficient	p-value
School type	0.9	0.0327

Table 9 Regression results for Model 2

Variable	Coefficient	p-value
Eastern Region	2.61	0.028
Central Region	-0.29	0.019

Effects of resource investment. Model 3 (Table 10) introduces resource investment variables, showing that investment in digital infrastructure has no significant effect, indicating that simply increasing hardware equipment has limited effect on improving the quality of graduate education. Investment in teacher training has a negative effect, suggesting that current training content is out of step with the needs of graduate supervisors. Investment in digital resources has a significant positive effect, indicating that the effective provision of resources such as open access databases and academic collaboration platforms can directly support graduate research output.

Table 10 Regression results for Model 3

Variable	Coefficient	p-value
School type	0.427	0.0424
Eastern region	0.607	0.044
Central region	0.261	0.026
Investment in digital infrastructure	0.039	0.688

Variable	Coefficient	p-value
Investment in teacher training	-0.0109	0.097
Investment in digital resources	0.074	0.075
Size of higher education institutions	0.0009	0.459

Interaction Effect Analysis. In Model 4 (Table 11), the interaction term coefficient for “Double First-Class × Eastern Region” is -0.116 ($p = 0.032$), indicating that the collaborative advantage of Double First-Class universities in the eastern region exhibits diminishing marginal returns. This may stem from redundant resource allocation and inefficient competition among top institutions in the region, where overlapping investments in digital infrastructure and research directions fail to yield proportional value-added. Meanwhile, the interaction term for “Double First-Class × Central Region” (-0.117, $p = 0.017$) suggests that Double First-Class universities in the central region have failed to fully leverage their regional leadership role. These institutions need to strengthen deep collaboration with local research institutions and enterprises, align their digital talent training with regional industrial development needs, and mitigate the disconnect between academic research and local practical demands.

Table 11 Regression results for Model 4

Variables	Coefficient	p-value
School type	0.245	0.0260
Eastern region	0.289	0.029
Central region	0.245	0.012
Interaction between school type and eastern region	-0.116	0.032
Interaction between school type and central region	-0.117	0.017
Investment in digital infrastructure	-0.657	0.068
Investment in teacher training	0.33	0.092
Investment in digital resources	0.378	0.085
Size of higher education institutions	0.0011	0.040

Cluster Analysis Results

Based on K-means clustering (K=4), the resource allocation models for graduate education in higher education institutions are categorized into four types (Figure 3):

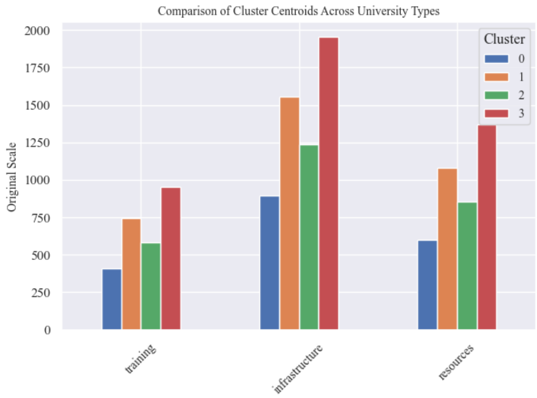

Figure 3 Comparison of cluster center characteristics

Cluster 3 (High-Investment Leadership Type): Includes five double-first-class universities that lead in training, infrastructure, and resource investment. Their per-student research funding for graduate students is 2.3 times that of other categories, and their paper publication volume accounts for 31% of the total sample. However, their efficiency score ranks second, indicating room for improvement in resource utilization.

Cluster 1 (Resource Advantage Type): Nine universities stand out in digital resource investment (weight 0.32), but training investment is insufficient, leading to low scores in graduate students' data analysis capabilities (mean 68 points), necessitating optimization of resource-capability alignment.

Cluster 2 (Moderate Development Type): Twenty-five universities have balanced but limited investment, with significant fluctuations in graduate student training quality (CV=0.18), making them suitable for resource sharing through regional alliances.

Cluster 0 (Low Investment Type): 11 western regional universities have insufficient investment, with graduate student participation in innovation projects at only 12%. Policy support is needed to address these shortcomings.

Conclusions and Recommendations

This study established an evaluation system for graduate education based on the perspective of educational digitization and revealed the key characteristics of current graduate education development through empirical analysis. The main conclusions are as follows: First, in terms of evaluation indicator weights, “digital technology knowledge and skills” (0.3775) and “digital application” (0.2597) occupy a core position at the criterion level, highlighting the foundational role of mastering digital tools and applying them in research practice in graduate education. At the scheme level, the high weights of indicators such as "familiarity with office/teaching software" and "frequency of use of teaching platforms" further confirm the emphasis placed on standardized technical application and routine digital practice in graduate education. Additionally, the inclusion of indicators such as "understanding of online ethics" and "protection of student data privacy" reflects the basic requirements of graduate education in the digital age regarding academic ethics and data security. Second, from the efficiency measurement results, there are significant differences in graduate education development across university types and regions. Although double-first-class universities have overall high efficiency, they exhibit a trend of diminishing marginal returns, with a slight decrease of 1.48% in technical efficiency values from 2023 to 2025, indicating that top-tier universities still have room for optimization in resource allocation and technology transfer. At the regional level, universities in central China have the highest median efficiency but with a dispersed distribution, universities in eastern China have stable efficiency but with low-efficiency outliers, and universities in western China have balanced efficiency but with room for improvement in overall levels, indicating that regional coordinated development remains a key issue for the high-quality development of graduate education. Third, panel regression analysis reveals the key factors influencing the value-added of graduate education. School type and eastern regional characteristics have a significant positive impact on the value-added index, while investments in digital infrastructure show no significant benefits, and teacher training funds even exhibit a negative impact, indicating that the current resource allocation structure suffers from a "heavy on hardware, light on effectiveness" issue. Interaction term results show that the synergistic effects of double-first-class universities in the eastern and central regions are diminishing, suggesting the need to be vigilant against internal friction caused by the excessive concentration of high-quality resources. Fourth, cluster analysis categorizes universities into four types: "high-investment leading," "resource-advantaged," "moderately developed," and "low-investment." Different types of universities exhibit significant differences in resource allocation strategies and efficiency performance. Among these, high-investment universities need to improve resource utilization efficiency, low-investment universities urgently require policy support to narrow the gap, and medium-development universities can achieve resource complementarity through regional collaboration.

Based on the research findings and in conjunction with the "Overall Plan for Deepening the Reform of Educational Evaluation in the New Era," the following policy recommendations are proposed:

First, establish a classification evaluation mechanism to promote differentiated development. For double-first-class universities, establish an "efficiency-first" evaluation orientation, incorporating the alignment of resource investment and research output into performance assessments. This should focus on measuring the utilization efficiency of digital infrastructure, the conversion rate of digital resources into innovative achievements, and the effectiveness of interdisciplinary digital collaboration teams to avoid blind scale expansion. For local universities, emphasize "characteristic development" evaluations, encouraging them to focus on their disciplinary strengths and regional development needs to build digital education brands—such as developing industry-adaptive digital skill training modules for applied disciplines or digital humanities research programs for liberal arts majors. At the regional level, eastern regions should optimize resource allocation mechanisms through data-driven analysis to identify and eliminate redundant investments in overlapping digital projects; central regions should strengthen inter-university collaboration via regional digital education consortia, facilitating the sharing of high-quality digital courses and joint research supervision to balance development gaps; western regions can leverage specialized policies and paired assistance programs with eastern universities to address shortcomings in digital resources and faculty digital literacy training.

Second, optimize the structure of resource allocation to enhance the efficiency of investments. Reduce redundant investments in digital infrastructure without clear application scenarios—such as excessive hardware duplication—and redirect funds toward "targeted training." Develop modular courses combining "research tool practical training" and "academic ethics" for graduate student advisors, covering cutting-edge digital research methods (e.g., big data analytics for academic research) and emerging ethical issues (e.g., responsible AI use in thesis writing), adopting a blended model of online modules and offline workshops to address the mismatch between training and actual needs. Simultaneously, expand the scope of sharing high-quality digital resources, promote the interconnection of academic databases and research platforms across eastern, central, and western universities, and establish a standardized national shared digital resource repository for graduate education to reduce regional resource barriers and ensure equitable access to core digital materials.

Third, improve the digital ethics and capability development system. Incorporate "cyber ethics" and "data privacy protection" into required graduate courses, integrating international norms (e.g., data privacy principles) and national regulations on digital research conduct. Strengthen awareness of academic norms through case-based teaching, analyzing real-world instances of digital research misconduct—such as unauthorized use of sensitive data or AI-assisted plagiarism—to enhance graduate students' ethical decision-making. Establish graduate digital capability development portfolios to dynamically track performance across "technology application—research innovation—ethical practice," including quantitative metrics (e.g., digital tool proficiency scores) and qualitative assessments (e.g., supervisor evaluations of ethical practice). Link evaluation results to scholarship awards, academic qualification certifications, and research funding allocations to form a closed-loop "development—evaluation—incentive" system, supplemented by periodic assessments and customized improvement plans to support continuous competency growth.

Fourth, establish a regional collaboration mechanism to promote balanced development. Leverage the superior digital resources, faculty expertise, and research platforms of eastern universities to form a cross-regional graduate education alliance, adopting models such as "online joint training"—where central and western graduate students enroll in high-quality digital courses from eastern universities—and "remote research collaboration" via cloud-based tools. Implement a unified credit recognition system and joint supervision guidelines to ensure the quality and sustainability of cross-regional training. Establish a "Digital Special Fund for Graduate Education in Central and Western Regions" to prioritize support for basic digital infrastructure construction, core mentor training, and region-specific digital education innovation pilot projects (e.g., integrating local cultural heritage with digital technologies) at low-investment universities, fostering both capacity building and distinctive development.

Competing Interests: The author declares no potential conflicts of interest with respect to the research and authorship pertaining to this article.

Acknowledgements: This work was supported by the action plan for the high-quality development of graduate education of Chongqing University of Technology (No. gzljg2023204; gzlkc202204; gzlkc202406).

Ethical approval: Ethical approval was not required as the study did not involve human participants.

Author contributions: All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Shiyan Chen and Can Zeng. The first draft of the manuscript was written by Yue Liao and Jing Quan. All authors read and approved the final manuscript.

References

Albers, M., Reitsma, M., Benning, K., Gobbens, R. J. J., Timmermans, O. A. A. M. J., & Nies, H. L. G. R. (2024). Developing a theory of change model for a learning and innovation network: A qualitative study. *Nurse Education in Practice, 77*, 103954. https://doi.org/10.1016/j.nepr.2024.103954

Chauveron, L. M., Samtani, S., Groner, M. G., Urban, J. B., & Linver, M. R. (2021). Including diverse stakeholder voices in youth character program evaluation. *American Journal of Evaluation, 42*(2), 221–236. https://doi.org/10.1177/1098214020917218

Ferreira, A., Araújo, B., Alves, J., Principe, F., Mota, L., & Novais, S. (2023). Peer feedback: Model for the assessment and development of metacognitive competences in nursing students in clinical training. *Education Sciences, 13*(12), 1219. https://doi.org/10.3390/educsci13121219

Li, X., Zhang, S., Yu, Q., et al. (2024). Research on the evaluation system of the cultivation quality of master's degree candidates in schools. *Bulletin of Chinese Psychological Sciences, 2*(4), 35–38.

Shi, Y., & Niu, Y. (2023). Research on the quality assurance system of postgraduate education based on PDCA cycle—A case study of a university in Beijing. *Journal of Higher Education Research, 4*(2), 18–23.

Wang, Z., Chen, L., & Liu, H. (2022). Construction of a three-dimensional evaluation model for postgraduate education quality under the background of "double first-class". *Journal of Graduate Education, 3*(1), 56–62.

Wu, M. (2022). Study on Hopf branch of stability of time-delay unified system and performance evaluation based on deep learning. *Mathematical Problems in Engineering, 2022*, 1–11. https://doi.org/10.1155/2022/2173900

Zhang, M., Li, N., Wang, J., & Yang, W. (2025). Research on optimization of teaching quality monitoring system in normal universities based on PDCA cycle theory. *Information and Communication Technology for Education, 13*(3).

Zhang, Z., & Wang, M. (2020). Research on graduate education quality evaluation based on combination empowerment and comprehensive fuzzy model. *IOP Conference Series: Materials Science and Engineering, 768*(5), 052003. https://doi.org/10.1088/1757-899X/768/5/052003