Evaluation of Listening Efficiency in Vocational Colleges Based on AHP-FCE: Empirical Study on AI Agent and Digital Twin Training Practice

Yanyan Zhao
Foundation Department, Engineering University of PAP, Xi'an, China

Abstract: The efficiency of classroom learning is a key index for evaluating the quality of teaching in vocational colleges, directly influencing the effectiveness of technical talent cultivation. This paper selects evaluation index through literature review and questionnaire surveys, builds a student listening efficiency evaluation model based on AHP-FCE that conforms to the characteristics of vocational education, and conducts empirical analysis to diagnose the key issues affecting listening efficiency. Then, based on AI agents and digital twin training technology, improvement measures are proposed, and the effectiveness of the improvements is verified using multivariate statistical methods from three dimensions: statistical significance, practical significance, and overall effect. The research results validate the effectiveness of the evaluation model and improvement measures, providing scientific tools and practical references for the diagnosis and improvement of classroom teaching quality in vocational colleges.

Keywords: Vocational Colleges; Listening efficiency; Analytic Hierarchy Process; Fuzzy Comprehensive Evaluation; Multivariate Statistical Analysis

1. Introduction

As the type of education most closely linked to economic and social development, the quality of classroom teaching in vocational education directly affects the effectiveness of technical and skills training. As an intuitive reflection of classroom teaching quality, the efficiency of classroom listening is the core indicator for measuring students' learning status and teaching effectiveness. However, in current vocational colleges, there is a common teaching dilemma of "high head-down rate and low head-up rate" in the classroom [1]. Problems such as students' lack of concentration and insufficient participation seriously restrict the quality of talent cultivation. How to scientifically evaluate students' listening efficiency, accurately diagnose classroom problems, and then take effective improvement measures to enhance teaching quality has become an important topic of concern for vocational education researchers.

This study aims to construct an AHP-FCE evaluation index model for students' listening efficiency in vocational colleges. Through empirical analysis to diagnose problems, propose targeted improvement measures, and use multivariate statistical methods to verify the improvement effects, it provides theoretical basis and practical tools for improving the teaching quality of vocational college classrooms.

2. Research Foundation

There are some problems in the current research on evaluating the listening efficiency of vocational college students: Firstly, the evaluation criteria are vague, mostly relying on teachers' subjective experience judgments, lacking a systematic index system; Secondly, the evaluation methods are simple, unable to handle the multi-factor and multi-level ambiguous concept of listening efficiency; Thirdly, the effect verification methods are simple, mainly using the comparison of means before and after, which is difficult to fully reveal the multi-dimensional effects of the improvement measures [2]. The Analytic Hierarchy Process (AHP) can decompose complex problems into an ordered hierarchical structure and determine the index weights through pairwise comparisons [3]; Fuzzy Comprehensive Evaluation (FCE) can effectively handle the fuzziness and uncertainty in the evaluation process [4]. The AHP-FCE model formed by combining these two methods provides a scientific methodological support for the evaluation of listening efficiency. Multivariate statistical methods can analyze the significance, robustness and practicality of the improvement effect from multiple perspectives in a more scientific and comprehensive way.

3. Build the Evaluation Index Model

3.1 Establishment of Evaluation Index

In light of the teaching characteristics and actual situation of vocational colleges, this paper conducts a literature study by consulting databases such as CNKI and Wan fang. It also conducts semi-structured interviews with supervisory experts, front-line teachers, and students of different grades [5]. Based on the AHP, it constructs an evaluation index system for classroom efficiency from the perspectives of both teachers and students, consisting of three levels: the target level (students' listening efficiency), the criterion level (5 items), and the solution level (20 items), as shown in Figure 1.

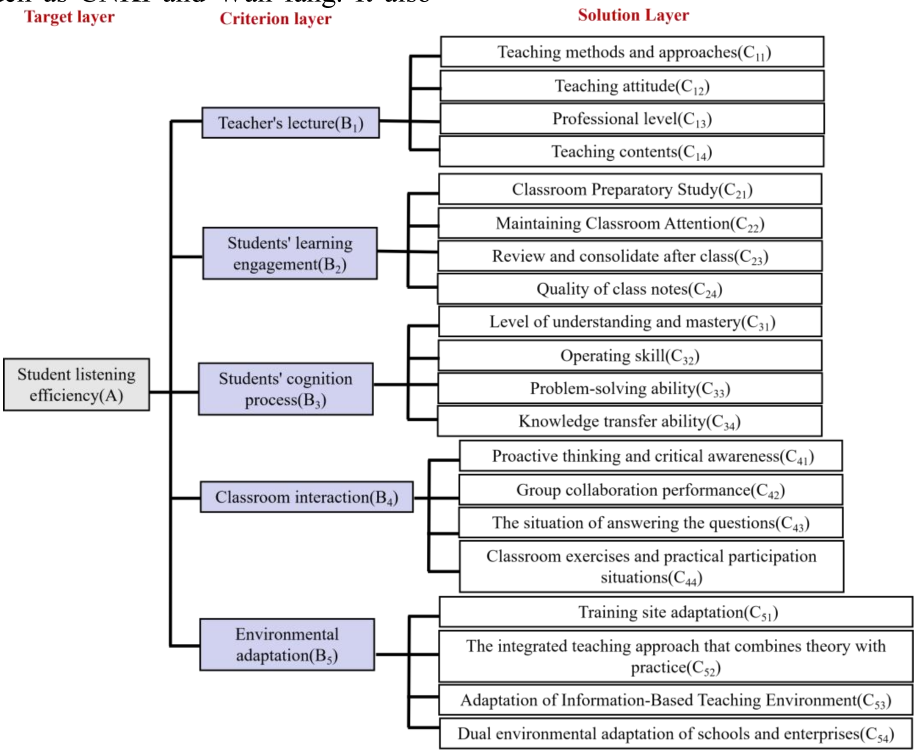

Figure 1. Evaluation Index Model for Classroom Learning Efficiency

3.2 Establishment of the Weight of Evaluation Index

This paper combines qualitative and quantitative analysis methods and uses the AHP to determine the weights of each factor in the evaluation index system. Firstly, based on the index system, 10 experts including professional leaders, key teachers, and enterprise training managers from vocational colleges were invited to form an expert assessment group. From top to bottom, the 1-9 scale method (see Table 1) was adopted layer by layer to compare and score the relative importance of multiple indicators within the same level pairwise to determine the ratios between each factor, and write them in matrix form. Then, a judgment matrix was constructed.

Table 1. Scale Table for Proportions of Relative Importance (Nine-Point Scale)

Comparison between index A and index B	Extremely important	Very important	Important	Somewhat important	Equally important	Slightly important	Unimportant	Very important	Extremely important
Evaluation scale value	9	7	5	3	1	1/3	1/5	1/7	1/9

Note: Take 8, 6, 4, 2, 1/2, 1/4, 1/6, 1/8 as the median values of the above evaluation values.

In practical situations, due to the complexity of the research subjects and the diversity of people's perceptions, the judgment matrices obtained often contradict each other. Therefore, it is necessary to conduct a consistency test on the judgment matrices [6]. Only when the judgment matrices satisfy consistency can objective and correct conclusions be drawn. This paper uses the root mean square method to calculate the eigenvector of the maximum eigenvalue of the judgment matrix for each index, and normalizes it to obtain the weight of each index. Then, the maximum eigenvalue $\lambda_{\max}$, CI and CR of the judgment matrix are calculated.

$$CI = \frac{\lambda_{\max} - n}{n-1}, n \text{ is order of the judgment matrix} \quad (1)$$

$$CR = CI / RI \text{ (The value of } RI \text{ is related to n)} \quad (2)$$

When $n = 4$, RI is set to 0.90; when n=5, RI is set to 1.12. When $n > 2$, if the consistency ratios CR of all index calculated are less than 0.1, it indicates that the consistency test of the matrix has passed [7]. According to this method, the final consistency test results and weight distribution of the evaluation index for the listening efficiency of vocational college students are shown in Table 2 and Table 3.

3.2 The Comprehensive Evaluation Model Based on AHP-FCE

Based on the listening efficiency index system established by AHP, the FCE method is used to quantitatively evaluate the listening efficiency of students, the comprehensive evaluation flowchart based on AHP-FCE as shown in Figure 2. FCE is a comprehensive evaluation method based on fuzzy mathematics, which can convert qualitative evaluations into quantitative results. It is mainly used to solve problems that are unclear in concept, difficult to quantify, and have ambiguous boundaries.

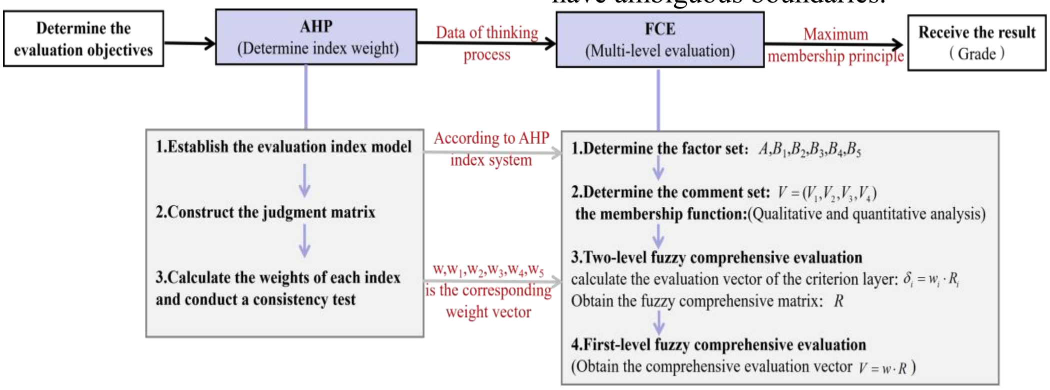

Figure 2. Comprehensive Evaluation Flowchart Based on AHP-FCE

Specific steps are as follows:

(1) Determination of the factor set, the comment set and the membership function

According to the evaluation index system established in the AHP, the factor set of the evaluated object is set as $A, B_1, B_2, B_3, B_4, B_5$. Due to the different importance levels of the factors in the set, the weights assigned to each factor are also different. Here, the weights of each evaluation index determined in the AHP are used to obtain the corresponding weight vector, denoted as $w, w_1, w_2, w_3, w_4, w_5$. The comment set is denoted as $V = (V_1, V_2, V_3, V_4)$ = (excellent, good, pass, fail), including four specific evaluation grades. For quantitative indicators (such as question response rate, in-class practice accuracy rate, practical assessment, etc.), the trapezoidal membership function is used to calculate the membership degree; for qualitative indicators (such as teaching attitude, classroom discipline), students self-assess and the teachers who observe the class independently rate based on the classroom observation scale, and the average value is taken to calculate the membership degree [8].

(2) Two-level fuzzy comprehensive evaluation

This article invites experts to grade the qualitative indicators in the index system based on the four-level evaluation criteria. For the quantitative indicators, calculate them using the trapezoidal membership function. Regarding the instructors, the scores are given by the students of this major. The evaluation matrices $R_i, i = 1,2,3,4,5$ for each index at the scheme level have been obtained. Then, the evaluation vectors for each criterion level are calculated according to the formula $\delta_i = w_i \cdot R_i$.

(3) First-level fuzzy comprehensive evaluation

The evaluation vectors of each criterion layer are combined to form the total evaluation matrix R. This is then combined with the weight vector of the criterion layers to obtain the comprehensive evaluation vector $V = w \cdot R$ of the teaching efficiency. The evaluation result of a certain student's teaching efficiency is determined by the principle of maximum membership degree.

4. Application and Verification

4.1 Determine the Evaluation Index System and Weights for Class Observation Efficiency

This paper conducts an empirical analysis of the listening efficiency of 30 students from a certain class in a vocational college using the AHP-FCE evaluation model. Firstly, based on the evaluation index system shown in Figure 1, 10 experts including professional leaders, key teachers, and enterprise training managers from vocational education institutions were invited to form an expert assessment group. They used the 1-9 scale method to compare and score the relative importance of multiple indicators within the same level in pairs, constructed a judgment matrix, and conducted a consistency test on the judgment matrix, as shown in Table 2. Based on the judgment matrix, the weights of each index affecting the listening efficiency of these 30 students can be calculated, as shown in Table 3.

Table 2. Consistency Test of Each Index

Index	A	B_1	B_2	B_3	B_4	B_5
λ_{max}	5.0686	4.011	4.0316	4.0109	4.0316	4.0109
CI	0.01715	0.0037	0.0105	0.0036	0.0105	0.0036
RI	1.12	0.9	0.9	0.9	0.9	0.9
CR	0.0153	0.0041	0.0117	0.004	0.0117	0.004

Table 3. Weight Distribution of Evaluation Index for Classroom Efficiency

Target layer	Criterion layer	index weight	Solution Layer	index weight	comprehensive weight
Student listening efficiency (A)	Teacher's lecture (B_1)	0.1610	Teaching methods and approaches(C_{11})	0.2272	0.0366
			Teaching attitude(C_{12})	0.1225	0.0197
			Professional level(C_{13})	0.2272	0.0366
			Teaching contents(C_{14})	0.4231	0.0681
	Students' learning engagement (B_2)	0.0986	Classroom Preparatory Study(C_{21})	0.096	0.0095
			Maintaining Classroom Attention(C_{22})	0.4658	0.0459
			Review and consolidate after class(C_{23})	0.2772	0.0273
			Quality of class notes(C_{24})	0.1611	0.0159
	Students' cognition process (B_3)	0.4162	Level of understanding and mastery(C_{31})	0.1225	0.0510
			Operating skill(C_{32})	0.4231	0.1761
			Problem-solving ability(C_{33})	0.2272	0.0946
			Knowledge transfer ability(C_{34})	0.2272	0.0946
	Classroom interaction (B_4)	0.2618	Proactive thinking and critical awareness(C_{41})	0.2772	0.0726
			Group collaboration performance(C_{42})	0.1611	0.0422
			The situation of answering the questions(C_{43})	0.0960	0.0251
			Classroom exercises and practical participation situations(C_{44})	0.4658	0.1220
	Environmental adaptation (B_5)	0.0624	Training site adaptation(C_{51})	0.4231	0.0264
			The integrated teaching approach that combines theory with practice(C_{52})	0.2272	0.0142
			Adaptation of Information-Based Teaching Environment(C_{53})	0.1225	0.0076
			Dual environmental adaptation of schools and enterprises(C_{54})	0.2272	0.0142

As can be seen from Table 3, at the criterion level, the cognitive process and classroom interaction are the core factors influencing the listening efficiency of vocational college students; the teacher's teaching ranks third and is a key factor affecting the listening efficiency; although environmental adaptation has a relatively low weight, the adaptation of the training site ranks 14th in the overall ranking and still has a certain impact on the listening efficiency. At the scheme level, operational skills is the indicator with the highest single weight, reflecting the "skill-oriented" feature of vocational education. Among the top 5 index of the comprehensive weight, there are four index related to hands-on practice, which fully demonstrates the characteristics of "integration of theory and practice, and industry-academia integration" in vocational education.

4.2 Application of Fuzzy Comprehensive Evaluation

Based on the above index system, the FCE was used to evaluate the listening efficiency of the 30 students in this class. Data were collected through multiple sources such as supervising classes, practical operation assessments, classroom observations, and student self-evaluations. Each index was rated on a four-level scale, and the evaluation set was denoted as $V=(V_1,V_2,V_3,V_4)=$ (excellent, good, pass, fail).

(1) Determination of the factor set, the comment set and the membership function

According to the evaluation index system established in AHP, the factor set of the object to be evaluated is determined as

$A=(B_1,B_2,B_3,B_4,B_5)$, $B_1=(C_{11},C_{12},C_{13},C_{14})$

$B_2=(C_{21},C_{22},C_{23},C_{24})$, $B_3=(C_{31},C_{32},C_{33},C_{34})$

$B_4=(C_{41},C_{42},C_{43},C_{44})$, $B_5=(C_{51},C_{52},C_{53},C_{54})$

The corresponding weight vector:

$$w=\left(0.1610\ 0.0986\ 0.4162\ 0.2618\ 0.0624\right)$$
$$w_1=\left(0.2272\quad 0.1225\quad 0.2272\quad 0.4231\right)$$
$$w_2=\left(0.096\quad 0.4658\quad 0.2772\quad 0.1611\right)$$
$$w_3=\left(0.1225\quad 0.4231\quad 0.2272\quad 0.2272\right)$$
$$w_4=\left(0.2772\quad 0.1611\quad 0.096\quad 0.4658\right)$$
$$w_5=\left(0.4231\quad 0.2272\quad 0.1225\quad 0.2272\right)$$

(2) Two-level fuzzy comprehensive evaluation

For the 30 students mentioned above, 10 teachers from the relevant field were invited to grade the qualitative indicators in the index system based on the evaluation criteria. The quantitative indicators were calculated using the trapezoidal membership function. Taking student 1 as an example, the evaluation matrices for each index are as follows:

$$R_1=\begin{pmatrix}0.2&0.5&0.3&0\\0.3&0.5&0.2&0\\0.2&0.6&0.2&0\\0.3&0.5&0.2&0\end{pmatrix}\quad R_2=\begin{pmatrix}0.1&0.4&0.4&0.1\\0.2&0.5&0.3&0\\0.1&0.4&0.4&0.1\\0.1&0.5&0.3&0.1\end{pmatrix}$$

$$R_3=\begin{pmatrix}0.3&0.5&0.2&0\\0.2&0.6&0.2&0\\0.2&0.5&0.2&0.1\\0.2&0.5&0.2&0.1\end{pmatrix}\quad R_4=\begin{pmatrix}0.2&0.5&0.3&0\\0.3&0.4&0.2&0.1\\0.2&0.5&0.2&0.1\\0.2&0.5&0.2&0.1\end{pmatrix}$$

$$R_5=\begin{pmatrix}0.4&0.4&0.2&0\\0.3&0.5&0.2&0\\0.3&0.5&0.2&0\\0.3&0.4&0.3&0\end{pmatrix}$$

The evaluation vectors of each criterion layer are calculated as follows:

$$\delta_1=w_1\cdot R_1=\left(0.26\quad 0.52\quad 0.22\quad 0\right)$$
$$\delta_2=w_2\cdot R_2=\left(0.15\quad 0.47\quad 0.33\quad 0.05\right)$$
$$\delta_3=w_3\cdot R_3=\left(0.22\quad 0.55\quad 0.2\quad 0.03\right)$$
$$\delta_4=w_4\cdot R_4=\left(0.22\quad 0.48\quad 0.24\quad 0.06\right)$$
$$\delta_5=w_5\cdot R_5=\left(0.33\quad 0.44\quad 0.22\quad 0.01\right)$$

$$R=\begin{pmatrix}\delta_1\\\delta_2\\\delta_3\\\delta_4\\\delta_5\end{pmatrix}=\begin{pmatrix}0.26&0.52&0.22&0\\0.15&0.47&0.33&0.05\\0.22&0.55&0.2&0.03\\0.22&0.48&0.24&0.06\\0.33&0.44&0.22&0.01\end{pmatrix}$$

(3) First-level fuzzy comprehensive evaluation

Calculate the comprehensive evaluation vector $V=w\cdot R=\left(0.22\quad 0.51\quad 0.24\quad 0.03\right)$. According to the principle of maximum membership degree, the evaluation of student 1's listening efficiency is good. Using the same method, the fuzzy comprehensive evaluation results of the other 29 students can be obtained. Finally, the fuzzy comprehensive evaluation Results as shown in Table 4, and the parallel coordinates plot—individual student membership profiles as shown in Figure 3.

From Table 4, it can be seen that the overall listening efficiency of this class is mainly "good", accounting for 50%. This is highly consistent with students' regular performance and final exam results, indicating that the constructed AHP-FCE evaluation model has good effectiveness. Further analysis of the relationship between the scores of each criterion layer and their weights reveals two prominent issues:

(1) High-weight index do not match the empirical performance: For example, the weights of the cognitive process and classroom interaction are the highest, and they are the core factors affecting the listening efficiency of vocational college students. However, some students scored low in indicators such as "problem-solving ability" and "active thinking and questioning awareness", indicating that these high-weight indicators perform poorly in actual classrooms and are the main bottlenecks restricting the listening efficiency.

(2) The "can do" dimension of operational skills is insufficient and practical training is weak: The comprehensive weight of operational skills is the highest, but the empirical data scores are generally lower than the "understanding and mastery degree", indicating that students have deficiencies in the "can do" aspect and the practical training teaching needs to be strengthened.

Table 4. Fuzzy Comprehensive Evaluation Results

Student	Overall Evaluation Vector	Attendance Efficiency Evaluation Result
1	(0.22 0.51 0.24 0.03)	Good
2	(0.18 0.47 0.3 0.05)	Good
3	(0.1 0.32 0.42 0.16)	Pass
4	(0.28 0.53 0.16 0.03)	Good
5	(0.08 0.29 0.45 0.18)	Pass
6	(0.32 0.5 0.15 0.03)	Excellent
7	(0.2 0.48 0.28 0.04)	Good
8	(0.12 0.35 0.4 0.13)	Pass
9	(0.25 0.52 02 0.03)	Good
10	(0.15 0.45 0.35 0.05)	Good
11	(0.06 0.24 0.48 0.22)	Pass
12	(0.3 0.55 0.12 0.03)	Excellent
13	(0.22 0.49 0.26 0.03)	Good
14	(0.14 0.42 0.38 0.06)	Good
15	(0.09 0.3 0.44 0.17)	Pass
16	(0.26 0.54 0.17 0.03)	Good

17	(0.19 0.46 0.31 0.04)	Good	24	(0.13 0.4 0.41 0.06)	Good
18	(0.11 0.33 0.43 0.13)	Pass	25	(0.24 0.53 0.2 0.03)	Pass
19	(0.27 0.52 0.18 0.03)	Good	26	(0.17 0.46 0.33 0.04)	Good
20	(0.16 0.44 0.36 0.04)	Good	27	(0.1 0.31 0.44 0.15)	Good
21	(0.07 0.28 0.46 0.19)	Pass	28	(0.29 0.54 0.14 0.03)	Pass
22	(0.31 0.51 0.15 0.03)	Excellent	29	(0.23 0.48 0.26 0.03)	Excellent
23	(0.21 0.5 0.26 0.03)	Good	30	(0.12 0.38 0.38 0.12)	Good

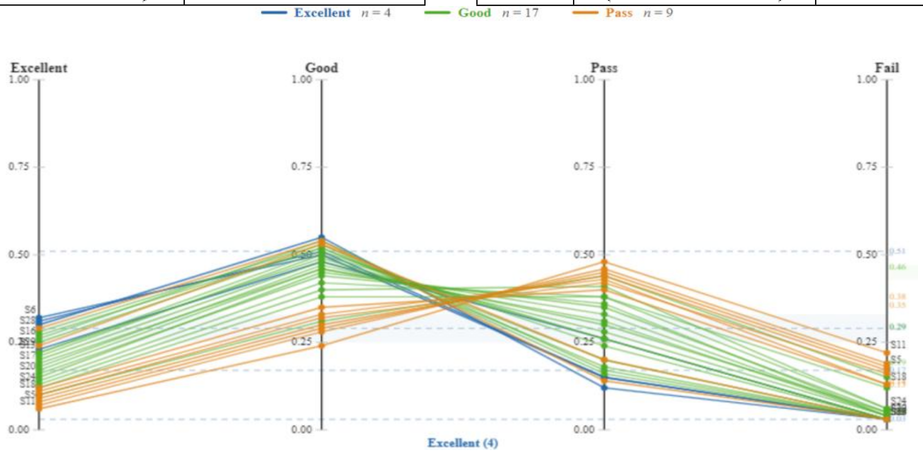

Figure 3. Parallel Coordinates Plot—Individual Student Membership Profiles

4.3 Improvement Measures for Information-Based Teaching Based on AI Agent and Digital Twin Training Practice

Based on the problems identified in the above empirical analysis, targeted improvement measures are proposed by leveraging information technologies such as AI agent and Digital Twin Training. The teaching philosophy of "centering on learning" is adhered to. AI agents are used to solve the problem of "passive learning and insufficient thinking", while Digital Twin Training is employed to address the issue of "insufficient practice and weak training". These two technologies are integrated into the teaching chain of "pre-class preview-in-class collaborative learning- post-class extension", achieving full-process intelligent support from "problem-driven" to "task practice" and finally to "skill mastery". The collaborative mechanism of the AI Agent and Digital Twin Training Practice as shown in Figure 4.

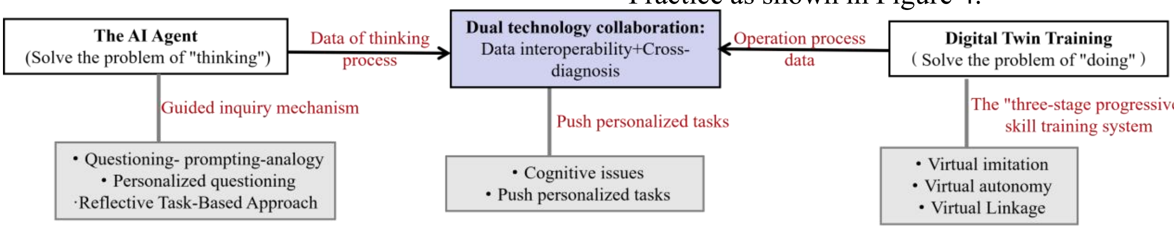

Figure 4. Collaborative Mechanism of the AI Agent and Digital Twin Training Practice

(1) The "guided exploration" mechanism driven by AI Agent

This article uses AI agent as exclusive study companions for students, guiding them to autonomously construct solutions through questioning, prompting and analogy. Before class: Students preview the task list and watch micro-lessons through the online platform. The AI agent generates personalized inquiries based on the students' self-assessment of mistakes during the preview, prompting them to reflect on their knowledge blind spots and changing from "passive reception" to "active examination". During class: In the exploration and analysis section, students first have multiple rounds of guided conversations with the AI, complete the preliminary problem decomposition, and then submit the preliminary solutions on the online platform. The teacher retrieves typical solutions for a class discussion, and students enter the digital twin training system with the exploration results. The AI agent simultaneously records the students' decision sequences during the training and analyzes the data characteristics of their problem-solving abilities.After class: The AI agent generates personalized reflection questions and extension tasks based on classroom performance and training data. For students who lack active thinking, open-ended questions are pushed, requiring them to submit their thinking process records; for students with weak problem-solving abilities, structured case analysis is pushed, requiring them to complete step-by-step reasoning fill-ins; for students with insufficient operational skills, digital twin specialized practice tasks are pushed. Students submit reflection notes through the AI agent, and the AI automatically extracts key growth points and generates a semester ability growth report.

(2) The "three-stage progressive" skill training system supported by Digital Twin Training Practice

This paper constructs a virtual simulation environment that is 1:1 corresponding to the real equipment, and designs a three-stage progressive training method to achieve real-time feedback, automatic scoring, and path recording during the operation process. The first stage is virtual imitation: students follow the step-by-step prompts of the digital twin to complete basic operations, aiming to establish a sense of norms. The second stage is virtual autonomy: teachers turn off the prompts and students complete tasks independently in the virtual environment, with the aim of strengthening proficiency and problem-solving skills. The third stage is virtual-real linkage: when students have achieved proficiency in virtual practice, they can enter the operation of the real equipment to achieve skill transfer.

Before class, students expose the common blind spots in their learning through virtual pre-operation, providing data basis for classroom teaching; during class, the "virtual first, virtual-real linkage" mode is adopted, where students lay the foundation through virtual training and complete skill solidification through real machine operation; after class, students conduct unlimited practice in the virtual environment to achieve autonomous reinforcement of weak links.

(3) Dual Technology Synergy Mechanism

The AI agent and Digital Twin Training form a collaborative closed loop through data exchange. The AI agent records students' thinking processes (question quality, reasoning path), while the Digital Twin records students' operation processes (decision sequence, error type). After the two data streams converge, the AI agent generates personalized diagnoses, tracing the operational mistakes back to cognitive blind spots, or mapping the thinking block points to specific skill deficiencies. Based on the diagnosis results, the AI automatically pushes targeted training tasks (cognitive problems or specialized virtual operation exercises), achieving the integration of "thinking - doing - evaluating - correcting".

4.4 Verification of the Improved Effect Based on Multivariate Statistical Methods

After implementing the aforementioned improvement measures for the students of the same class for two months, the listening efficiency of the students was evaluated again using the AHP-FCE model. During this period, the professional teaching team uniformly executed the measures to ensure that the teaching conditions and teaching staff remained consistent. To fully verify the improvement effect, multivariate statistical methods were used to conduct a comparative analysis of the data before and after.

(1) Descriptive Statistical Analysis

Descriptive statistics describe the central tendency and dispersion of the data by calculating the mean (M) and standard deviation (SD). The results are shown in Figure 5.

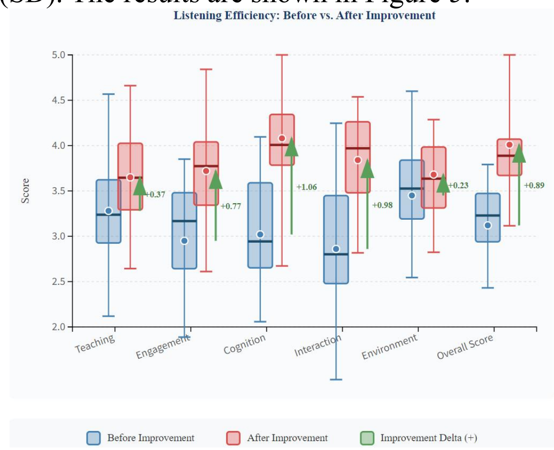

Figure 5. Improvement in Each Dimension of Listening Efficiency before and after (n = 30)

The scores of each dimension have all increased, and the distribution of the improvement differences is highly consistent with the distribution of the AHP weights. The higher the weight of a dimension, the more significant the improvement effect. This verifies the scientificity of the evaluation system and the effectiveness of the improvement measures.

(2) Paired Samples t-Test and Effect Size Analysis

A Paired-Sample t-Test was conducted on the comprehensive scores of 30 students [9]. The results showed that the comprehensive score of students' listening efficiency improved from

3.12±0.54 to 4.01±0.43, with t=12.36 and p<0.001. This indicates that the difference in comprehensive scores before and after the improvement has a highly significant statistical meaning.

To evaluate the practical significance of the improvement measures, the effect size was calculated using Cohen's d. The effect size classification standards are: d>0.8 is a large effect, 0.5<d<0.8 is a medium effect, and d<0.5 is a small effect. The results are shown in Table 5:

Table 5. Effect Size Analysis of Each Dimension

Evaluation Dimension	Teacher Instruction	Student Engagement	Cognitive Process	Classroom Interaction	Environmental Adaptation	Summary Score
Cohen's d	0.74	1.35	1.92	1.86	0.48	1.76

Both the cognitive process and the classroom interaction dimensions are considered to have a significant effect. This indicates that the improvement measures have achieved remarkable practical results in these two core dimensions. The learning engagement dimension comes next, also reaching the level of a significant effect. The comprehensive score is d=1.76, and the overall effect is significant.

(3) Wilcoxon Signed-Rank Test

To verify the robustness of the results, the non-parametric Wilcoxon Signed-Rank Test was used to validate the comprehensive scores of 30 students before and after the improvement. This test does not require the data to follow a normal distribution and is applicable to small samples or situations where the data distribution is unknown. The calculated Z value was 4.78, and p<0.001, which was consistent with the results of the paired sample t-Test. This excluded the influence of data distribution on the results, indicating that the test results are robust.

(4)Multivariate Analysis of Variance

To simultaneously examine the overall effect of the improvement measures across five dimensions and avoid the problem of Type I error inflation caused by multiple single-variable tests, a multivariate analysis of variance method was adopted for analysis [10]. The scores before and after the improvement were taken as the within-group variables, and the scores of the five dimensions were taken as the dependent variables. By comparing the within-group dispersion matrix W and the between-group dispersion matrix B, the between-group differences of multiple dependent variables were tested.

The selected test statistic was $\Lambda = \frac{|W|}{|B+W|}$, which approximately follows an F distribution, and the smaller Λ is, the more significant the between-group difference is. Through calculation, it was obtained that Λ=0.312, F(5,24)=10.58, p<0.001 and η^2=0.688. This indicates that the improvement measures have an extremely significant overall effect across multiple dimensions, and the effect size reaches the large effect level ($\eta^2 > 0.14$). This further validates the comprehensive effectiveness of the improvement measures.

5. Conclusion

The AHP-FCE evaluation model constructed in this paper highlights vocational education-specific index such as operational skills, adaptability of training venues, and adaptation to both school and enterprise environments. The weight distribution is highly consistent with the "skill-oriented" characteristic of vocational education, and it can scientifically diagnose the key issues affecting students' listening efficiency. The improvement measures proposed based on AI agent and Digital Twin Training effectively solved the problems of high index weights, low empirical scores, and weak training. Based on empirical data, the significant effectiveness of the improvement was verified using multivariate statistical methods, providing a theoretical basis and practical tools for the diagnosis and improvement of classroom teaching quality in vocational schools, and having certain reference value for promoting the high-quality development of vocational education. However, there are also some shortcomings. For example, during the empirical analysis, students from different majors and different grades were not covered, and the sample size was small. The universality of the research conclusion needs to be further verified. The improvement measures were implemented for only 2 months, although the short-term effect was significant, the sustainability and long-term stability of the improvement effect have not been verified yet, and further tracking research is needed. In the future, based on this research, the evaluation system will be further improved, the research scope will be expanded, and the mechanism exploration will be deepened to enhance the quality of talent cultivation in vocational schools.

References

[1] Zhang Di, Xue Mingke, Shang Xiaofeng, et al. Research on the Evaluation of Educational Effects of Vocational Farmers' Academic Education Based on AHP-FCE. Journal of Smart Agriculture, 2024(7).
[2] Zhang Bolao, Zhang Yuan. Activating the Classroom: Exploring and Practicing the Paths for Improving the Efficiency of Vocational School Public Subject Classes. Education Research and Innovation, 2026, 2(12).
[3] Saty T L. Analytic Hierarchy Process: Principles, Applications and Issues. Translated by Xu Shubo. Beijing: Coal Industry Press, 1988: 10-35.
[4] Zhang Chonghui, Su Weihua: Expansion and Application Research of Fuzzy Comprehensive Evaluation Method. Beijing: Science Press, 2023
[5] Chen Wenfang, Jiang Jian, Cui Wei. Research on the Construction of Precise Assessment System for Classroom Teaching in Higher Vocational Colleges from the Perspective of "Double High Plan" - Based on the Analysis of the Observation Evaluation Forms of 18 "Double High Plan" Higher Vocational Colleges. Vocational Education, 2024(2): 75-79.
[6] Chen Yan Tai, Chen Guo Hong, Li Mei Juan. Classification of Comprehensive Evaluation Methods and Research Progress. Journal of Management Science, 2004, 7(2): 69-79.
[7] Hong Zhiguo, Li Yan, Fan Zhihua, et al. Research on the Critical Value of Consistency Index of Analytic Hierarchy Process. Mathematical Statistics and Management, 2011(3), 414-423.
[8] Jiang Qi Yuan, Xie Jin Xing, Ye Jun. Mathematical Models. 5th Edition. Beijing: Higher Education Press, 2018: 278-302.
[9] Guo Li. Construction and Research of an Accurate Learning Model Based on Multi-modal Sample t-test. Journal of Shaoguan University, 2020(6).
[10] Yao Yiqian. Research on Improving Online Learning Outcomes of Vocational College Students Based on Multivariate Statistical Analysis. Modern Commerce and Industry, 2025(17): 213-215.