J. COMBIN. MATH. COMBIN. COMPUT. 127a (2025) 3161--3176

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Research on Real-time Dynamic Adjustment Strategy of Industry-Teaching Integration Practical Training Process in Higher Vocational Education Based on Reinforcement Learning

Yusong Liu[1], Jianjun Song[2,✉]

[1] *Graduate School, Sehan University, Mokpo, 57447, South Korea*

[2] *School of Electronic Technology and Engineering, Shanghai Technical Institute of Electronics & Information, Shanghai, 201411, China*

ABSTRACT
The era of big data in education has come, data-driven intelligent decision-making has become the development trend in the era of big data, and precise teaching has become the keyword in the era of big data. This paper establishes a real-time dynamic teaching strategy adjustment decision-making model based on the learning characteristics in the process of industry-teaching integration practical training in higher vocational education, and uses Markov decision-making and Q-learning algorithms to solve the optimal teaching strategy in each stage of practical training and learning, which assists the teachers in decision-making and precise intervention. The results of the practical training teaching experiment found that the students in the experimental group, after the dynamic adjustment and intervention strategy implementation of the industry-teaching integration practical teaching, the scores of the practical training theory and application knowledge test were significantly improved ($P<0.05$), and the students' self-efficacy control sense, sense of effort, and sense of competence were all improved to different degrees. In addition, the scores of depth of understanding ($P=0.000$) and strategic approach ($P=0.000$) in practical training learning competencies also increased significantly. The strategy proposed in this study is able to capture the dynamic characteristics of educational data and use the multi-stage dynamic decision-making method to study the development of teaching strategies, which can provide stronger support for accurate teaching decisions and industry-teaching integration of practical training learning.

Keywords: reinforcement learning, Markov decision-making, Q-learning algorithm, teaching strategy, industry-education integration

1. Introduction

With the development of global economic integration, economic and trade exchanges among countries

✉ Corresponding author.
E-mail address: 15000292536@163.com (J. Song).
Received 15 March 2024; Revised 20 April 2024; Accepted 30 December 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-179

have become more and more frequent, and the international status of a country is closely related to its own level of economic development. Entering the new era, China's economy has entered a new period of high-quality development, with the consolidation of the basic position of agriculture, the gradual development of industry in the direction of medium and high end, and the continuous growth of the proportion of service industry in GDP. However, the current level of China's market development has not yet adapted to the needs of high-quality development of the economy, how to realize the transformation of high-speed economic growth into the development of the new normal economy is the key issue of economic transformation and development needs to be resolved [1-3]. Along with the upgrading and change of China's industrial structure and the rise of the fourth industrial revolution, industries with high scientific and technological content play the role of “leader” in the process of promoting the rapid development of the regional economy, and technical and skilled personnel play an important supporting role in the optimization and upgrading of industrial structure and the regional economy to gain competitive advantages, and we try to cultivate more high-level talents through the integration of industry and education, so as to meet the needs of high-quality economic development. The integration of industry and education is aimed at cultivating more high-level technical and skilled talents [4-6]. Influenced by historical inertia, in the past period, the “mainstream thinking” of Chinese colleges and universities was to focus on cultivating academic talents, ignoring the role of technical and skilled talents in the development of society, resulting in the shortage of high-level skilled talents, and the phenomenon of “labor shortage” has appeared one after another in many regions of China. The phenomenon of “labor shortage” has emerged one after another in many areas of China, and there is an imbalance between the number of talents cultivated by higher vocational education and the demand of the society. Therefore, in the context of economic transformation, the state vigorously develops vocational education, advocates the integration of industry and education, cultivates more excellent technical and skilled talents, promotes economic transformation and upgrading, and promotes the high-quality development of China's economy [7-9].

Reinforcement learning is a branch of machine learning. With the explosion of Alpha Go, reinforcement learning has become one of the popular research fields of machine learning. Intelligent bodies learn and interact with the environment to maximize rewards. Intelligent bodies receive feedback in the form of rewards and punishments as they act in the environment and use this feedback to improve their behavior over time. This learning process is driven by repeated trial and error with the goal of maximizing cumulative rewards over time. This acquisition of data through interaction with the environment makes it adept at making sequential decisions, and it learns optimal strategies for sequential decision-making tasks based on reward signals and reinforcement learning algorithms [10-12]. When the reinforcement learning algorithm is applied to the practical training of industry-teaching integration in higher vocational education, the real-time dynamic adjustment strategy can be obtained, which helps to deepen the integration of industry-teaching, school-enterprise cooperation, improve the guarantee system and measures of vocational education, and better promote the high-quality development of vocational education [13-15].

The integration of industry and education is an important way for the high-quality development of higher vocational education. Literature [16] systematically analyzed the research dynamics and trends in the field of industry-education integration through a literature review, and the results pointed out that the development of the field of industry-education integration has gone through three stages, and the findings help scholars, policymakers and practitioners better understand the development trend of this field. Literature [17] emphasized the role of collaborative education and high-quality talent cultivation in higher vocational education by the integration of industry and education human platform, and pointed out that innovating the form of industry-university-research integration and actively building a platform for school-enterprise cooperation can help higher vocational colleges and universities to cultivate high-quality skilled talents needed by the society and enterprises. Literature [18] takes the development of vocational education in Sanmenxia City as an example, analyzes the current situation of vocational education, discusses its future development trend, and puts forward relevant countermeasures to promote the cultivation of high-quality technical talents in vocational education. Literature [19] constructs an industrial integration and talent cultivation system based on the talent supply and demand mechanism of vocational education and new high-quality productivity, and empowers new high-quality productivity through vocational education to provide strong talent support for the integration of industry and education in high-quality development of higher vocational education. Literature [20] specifies the essence of school-enterprise cooperation and tries to apply all kinds of intelligent technologies in vocational education industry-university-research integration in order to effectively utilize the powerful resource integration capability of artificial intelligence to reconstruct the multidimensional capability of vocational education industry-education integration. Literature [21] indicates that the concept of community can promote the integration of industry and education from "integration" to real "integration", and by analyzing the reasons for the deviations in the development of the current situation of the integration of industry and education model, it is proposed that the synergy of community planning and development, the synergy of community governance, the synergy of community multi-party education, and the synergy of the concept of community governance are the main reasons for the development of the integration of industry and education model. By analyzing the reasons for the deviations in the current development of the integration model, we propose four innovative mechanism research strategies based on the community concept: synergy of community planning and development, synergy of community governance, synergy of community multi-nurturing mechanism, and synergy of community integration effect system, in order to enhance the effect of integration of higher vocational education. Literature [22] highlights that the most popular forms of learning in the field of vocational education and training are technology reinforcement learning and blended learning, and uses the Australian Defence Force (ADF) Vocational College as an example to explore the usefulness of two technology integration frameworks, "Replace-Enhance-Modify-Redefinition" (SAMR) and "Technology Integration Matrix" (TIM), in understanding the use of blended technologies in vocational education and training courses to improve the effectiveness of higher vocational education training.

This paper proposes a multi-stage teaching strategy adjustment and decision-making model based on reinforcement learning according to the characteristics of the practical training process of industry-teaching integration in higher vocational education and the teaching needs, based on the characteristics of students' performance and the learning effectiveness of practical training. The Markov decision-making process is used to define the state estimation function and action value function in the model. Then the Q-learning algorithm is used to solve the function to get the optimal teaching intervention strategy under each stage, so as to realise the real-time dynamic adjustment of the teaching strategy in the process of industry-teaching-integration practical training in higher vocational education. Then we design the accurate teaching strategy adjustment and intervention model for the process of industry-teaching integration practical training by combining the learning state, effectiveness analysis and strategy decision-making method. In this study, two groups of students in parallel classes in a higher vocational school were selected as subjects to carry out a comparative experiment of practical training teaching under the traditional model and the teaching intervention model. By analysing the mastery of practical training knowledge as well as the changes in self-efficacy and practical training learning ability of the two groups of students before and after the experiment, the effectiveness of the real-time dynamic teaching adjustment strategy was explored.

2. Method

2.1. *Enhanced learning of grounded theory*

Reinforcement learning [23] is a machine learning method designed to enable an intelligent body to learn what actions to take to achieve a certain goal by interacting with its environment. In reinforcement learning an intelligent body adjusts its behaviour by observing feedback (reward signals) from the environment to maximise long-term cumulative rewards. The environment is the external system in which the intelligence is located, which receives the actions of the intelligence to make corresponding state transitions and generate reward signals. Intelligentsia play the role of learners and decision makers in reinforcement learning, and are responsible for observing the environment and adapting strategies based on feedback from the environment. A state is a variable that describes a particular situation or configuration of the environment, and the intelligence is in a state at each time step, and its decision depends on the current state. Actions are actions or decisions that an intelligent body can take in a particular state, and the intelligent body influences the state transitions of the environment by choosing appropriate actions. Reward is the immediate feedback provided by the environment to the intelligent body, representing the degree of goodness or badness of the intelligent body's actions, and the goal of reinforcement learning is to obtain maximised long-term cumulative rewards. Strategies are the rules or functions by which an intelligent body chooses actions in a particular state, and the goal of reinforcement learning is to find the optimal strategy to maximise long-term rewards to the intelligent body.

Markov Decision Process (MDP) [24] is a commonly used modelling tool in reinforcement learning for describing the interaction process between an intelligent body and its environment. It is based on the Markov property that the future development of the current state depends only on the current state and the action taken currently is independent of the previous state, so MDP can formally describe the decision-making problem of an intelligent body, which includes the elements of state space, action space, transfer probability, reward function and discount factor. The state space represents the set of states that the environment is in. The action space represents the set of actions that an intelligent body can take, and in each state the intelligent body can choose to perform an action in the action space. The state transfer probability represents the probability distribution that after executing an action the environment will take the intelligent body to a new state, i.e., the probability that the next state will occur given the current state and the action executed. Reward function represents the immediate reward that an intelligent body receives for taking a particular action in a particular state, where the reward function can be either deterministic or random. Discount factor represents the discount rate of future rewards is used to measure the importance of current rewards versus future rewards, and the discount factor usually takes values between 0 and 1.

The basic goal of MDP is to find a strategy function, i.e., to choose an optimal action in each state, a process called strategy, which usually uses a value function or an action-value function to evaluate the goodness of the strategy, and to obtain the optimal strategy through the optimisation of the value function or the action-value function.

2.2. Decision-making model for adjusting practical training strategies for industry-teaching integration

2.2.1. Reinforcement learning-based decision-making model construction. The decision of real-time teaching guidance strategy adjustment in the process of industry-teaching integration practical training in higher vocational education is a dynamic decision-making process, with different teaching methods and contents in each stage, and the students' physical state, cognitive ability, and skill level are also changing. The process of industry-teaching integration in real-time training even requires continuous trial and error to achieve the optimal guidance strategy, so reinforcement learning is chosen to solve this multi-stage accurate teaching guidance strategy problem. Reinforcement learning algorithm based on multi-stage not only improves the collaborative performance, but also improves the training efficiency of the samples and the convergence speed of the model.

According to the introduction of the basic theory of reinforcement learning above, based on the optimal function of K stage $d_1^{'}(G_1), d_2^{'}(G_2), \ldots, d_k^{'}(G_k)$, a multi-stage educational strategy decision model can be constructed to further optimize to get the optimal industry-teaching integration of practical training and learning outcomes Y. Firstly, it is necessary to analyse the characteristics and influencing factors of each stage, and then find out the optimal decision function by using mathematical modelling methods. Such a model will help teachers to better guide students in the process of industry-industry integration, so that students can achieve the optimal practical training learning outcomes at each stage.

At this point, the optimal decision function [25] at optimal practical training learning outcome Y is noted as:

$$d' = d_1^{'}(G_1), d_2^{'}(G_2), \ldots, d_k^{'}(G_k) \tag{1}$$

It can be noted as:

$$d' = \max[Y \mid A_1 = d_1^{'}(G_1), A_2 = d_2^{'}(G_2), \ldots, A_K = d_k^{'}(G_k)] \tag{2}$$

Among them:

$$G_k = (X_1, A_1, X_2, A_2, \ldots., X_k)k(k = 1, \ldots k) \tag{3}$$

In order to solve the optimal training learning effect Y, it is necessary to find an optimal decision function $d^{'} = (d_1, d_2, \ldots, d_k)$. In the multiple phases of the production-teaching-integration practical training learning $k(k = 1, \ldots, k)$ the learning characteristics X_k, the production-teaching-integration practical training teaching programme A_k, the production-teaching-integration practical training teaching phases G_k and so on are random variables, it is not possible to directly solve the optimal decision function $d^{'} = (d_1, d_2, \ldots, d_k)$, at this time, you can use the reinforcement learning method to dynamically optimise the solution of this problem.

The reinforcement learning process of an intelligence can be considered as a Markov decision process (MDP) which can be represented by the quintuple $E =< S, A, R, P, \gamma > 1$. Where S denotes the state space, A denotes the action space, the reward function $R(s_t, a_t)$ denotes the reward obtained after taking action a_t in state s_t, and the state transfer function $P(s_t, a_t, s_{t+1})$ denotes the probability of taking action a, state s, and transferring to state s_{t+1}. The larger the value of discount factor $\gamma(0 \le \gamma \le 1)$ the more emphasis is placed on the results accumulated over time.

Reinforcement learning has a Markovian nature, denoted as:

$$V(s_t, a_t, s_{t+1} \mid s_0, a_0, s_1, a_1, \ldots, s_t, a_t) = V(s_t, a_t, s_{t+1} \mid s_t, a_t) \tag{4}$$

An intelligent body obtains a strategy by continuously executing actions in the environment and getting feedback from the environment, thus obtaining a strategy π. According to the theory of Markovian decision-making process, the state estimation function can be defined as:

$$V_\pi(s) = \sum_{\alpha} \pi(s, a) \sum_{\alpha} P(s, a, s^{'})[R(s, a, s^{'}) + \gamma V_\pi(s^{'})] \tag{5}$$

Setting $Q_\pi(s, a)$ represents the expected reward value of state s after strategy π performs action a. From Markov decision process theory, the action value function [26] can be defined as:

$$Q_\pi(s, a) = \sum_{x'} P(s, a, s')[R(s, a, s') + \gamma V_\pi(s')] \tag{6}$$

2.2.2. Decision-making model solution methods:

1) Q-learning algorithm. The classical learning algorithm for reinforcement learning is the Q-learning algorithm [27], which can be applied to complex domains because it is model-free, does not need to learn or apply a transformation model, considers each action in the entire action space at each iteration, and ensures the final convergence of the learning process. The learning state characteristics of students tend to change dynamically, and the state of each stage is related to the selection of action variables, so the Q-learning algorithm in reinforcement learning is chosen as the basic algorithm for solving the problem.

The Q-learning algorithm uses the state action value function $Q(s,a)$ as the valuation function, and the behavioural assessment of the intelligent body is achieved through the function $Q(s,a)$, which is capable of evaluating the effect of a certain action taken by the intelligent body in a specific state s, and considering the sum of discounts for obtaining a return in the future, at which point the optimal policy can be recorded as:

$$Q(s_t,a_t)=R+\gamma\max_{a_{t+1}}Q(s_{t+1},a_{t+1}) \tag{7}$$

Since there is a bias during each reinforcement learning process, it can be recorded as:

$$\sigma(s_t,a_t)=R+\gamma\max_{a+1}Q_{t-1}(s_{t+1},a_t)-Q_{t-1}(s_t,a_t) \tag{8}$$

Therefore, the Q-learning algorithm can be updated as:

$$\begin{aligned}Q_t(s_t,a_t)&=Q_t(s_t,a_t)+\alpha\sigma(s_t,a_t)\\&=Q_t(s_t,a_t)+\alpha[R+\gamma\max_{a+1}Q_{t-1}(s_{t+1},a_t)-Q_{t-1}(s_t,a_t)]\end{aligned} \tag{9}$$

Where, in the above equation $Q_t(s_t,a_t)$ is the value of state one action stored in the Q-table, α is the learning rate where $a\in(0,1)$, the Q-value of (s_t,a_t) can be updated based on the current error, and R is the reward value brought about by transferring from state s_t to s_t+1. The simplified basic iterative formula of the Q-learning algorithm is as follows:

$$Q(s_t,a_t)=R+\gamma\max_{a_{t+1}}Q(s_{t+1},a_{t+1}) \tag{10}$$

After finishing it is noted as:

$$Q(s,a)\leftarrow Q(s,a)+\alpha[R+\gamma\max_a Q(s',a')-Q(s,a)] \tag{11}$$

In the last stage K, Definition:

$$Q_k(s_k,a_k)=\max[Y\mid G_k=s_k,A_k=a_k] \tag{12}$$

$$V_k(s_k)=\max Q_k(s_k,a_k) \tag{13}$$

$$d_k^{'}(s_k)=\arg\max Q_k(s_k,a_k) \tag{14}$$

Where Q_k is the expected value of learning effectiveness of industrial-teaching integration practical training in higher vocational education, V_k is the maximum practical training learning effectiveness, and $d_k^{'}$ is the optimal decision function.

According to the formula analogy, in the K-1 stage, defined:

$$Q_{k-1}(s_{k-1},a_{k-1})=\max[V_k(G_k)\mid G_{k-1}=s_{k-1},A_{k-1}=a_{k-1}] \tag{15}$$

$$V_{k-1}(s_{k-1})=\max Q_{k-1}(s_{k-1},a_{k-1}) \tag{16}$$

$$d_{k-1}^{'}(s_{k-1})=\arg\max Q_{k-1}(s_{k-1},a_{k-1}) \tag{17}$$

Among them, the expected value of the effectiveness index of the learning stage $G_{k-1}(X_1,A_1,X_2,A_2,\ldots,X_{k-1})$ of the practical training for industry-teaching integration is Q_{k-1}, and the optimal practical training teaching effect is V_{k-1}.

By analogy, the optimal decision function can be obtained as:

$$d^{'}=(d_1^{'},d_2^{'},\ldots,d_k^{'}) \tag{18}$$

At this point, the multi-stage optimal decision function $d^{'}$ based on the Q-function is defined as follows:

$$d'=d_1^{'}(G_1),d_2^{'}(G_2),\ldots,d_k^{'}(G_k) \tag{19}$$

2) Convergence analysis. Convergence conditions for Q-learning algorithms [28] often include the following.

The state space s is traversable. In each state, the intelligent can take different actions and eventually return to some state that has been visited. Traversability of the state space is a necessary condition for the convergence of the Q-learning algorithm.

The reward function R is cumulative. In each state, the reward states of the intelligent body are cumulative, which again implies that the update of the Q-function can converge to some value.

The reward discount factor γ is in the interval [0, 1]. γ is used to balance the relationship between current and future rewards, and γ in the interval [0, 1] ensures that the Q-learning algorithm is able to effectively balance exploration and utilisation.

The learning rate α is appropriate. The learning rate is used to adjust the update magnitude of the Q-function. An appropriate learning rate can ensure that the Q-function can converge gradually, while avoiding the problem of excessive oscillation or slow convergence.

Because the Q-learning algorithm guarantees infinite traversal of the state-one action space, the Q -value will remain constant as the intelligent body continues to adopt the optimal policy, thus achieving convergence. When the Q -value converges, the optimal policy can be calculated based on the converged Q -value. Assume a Q-learning intelligent body, in a Markov decision process $\hat{Q}(s,a) \leftarrow r + \gamma \max \hat{Q}(s^{'},a^{'})$ and assume that $\overline{Q}(s,a)$ is initialised to an arbitrary value and the reward discount factor γ is in the interval [0, 1]. If there is no limit to the number of times each state one action can be revisited, then $\hat{Q}(s,a)$ converges to $Q(s,a)$ when $a \rightarrow \infty$ for all s and a.

2.3. *Design of dynamic adjustment strategies for the practical training process of industry-teaching integration*

This study distills the core problem of dynamic adjustment strategies and precise interventions in the process of industry-teaching integration practical training in higher vocational education, i.e., with the goal of promoting students' learning in practical training, we provide each student with suitable and targeted instructional strategy adjustments and interventions based on the students' learning status and characteristics in the process of industry-teaching integration practical training. The core problem of precision teaching intervention can be transformed into an optimization problem, i.e., to find the optimal precision teaching intervention based on students' learning status and characteristics in order to maximize students' learning effects. The essence is to find a one-to-one mapping of learning states and characteristics to instructional interventions to maximize learning outcomes. The precise practical training teaching and intervention strategy designed based on the dynamic adjustment decision-making model of practical training is shown in Figure 1, which consists of three parts: the learning state and characteristics, the learning effectiveness analysis and the teaching intervention decision-making in the process of production-teaching-integration practical training for higher vocational students. The three parts form a unified whole, and the characterization of learning state and features is the premise of precise adjustment strategy and practical training teaching intervention, the learning effectiveness analysis is the guarantee of dynamic strategy adjustment and precise teaching intervention, and the dynamic adjustment decision model of practical training teaching based on reinforcement learning is the key to real-time dynamic adjustment and precise teaching intervention. It should be noted that the precise teaching strategy adjustment and intervention in the process of industry-teaching integration practical training in higher vocational education should focus on dynamics. Chronology and contextualization are the basic features of education, so educational data are dynamic data, which determines the dynamic nature of teaching intervention. With the passage of time and contextualization, students' cognitive level, learning behaviors, and other learning states and characteristics are changing. The teaching effects and teaching objectives of concern also vary at different stages of industry-education integration teaching. The learning state and characteristics of students and teaching objectives in the process of practical training and teaching have changed over time. In this case, only dynamic and precise teaching interventions can capture these changes in the teaching process in time, make timely adjustments, and provide suitable teaching interventions for each student at each stage of the teaching process, meet the needs of each student at different stages, and better promote student learning. At the same time, teaching interventions are post-effective. The process of practical training and teaching in higher vocational education is not a simple combination of different teaching stages, but a whole, in which the various stages are interconnected and affect each other. Within a certain teaching time limit, the impact of a certain stage of teaching adjustment and intervention on the learning effect of practical training may be reflected in the subsequent teaching stage, which has a certain degree of after-effectiveness.

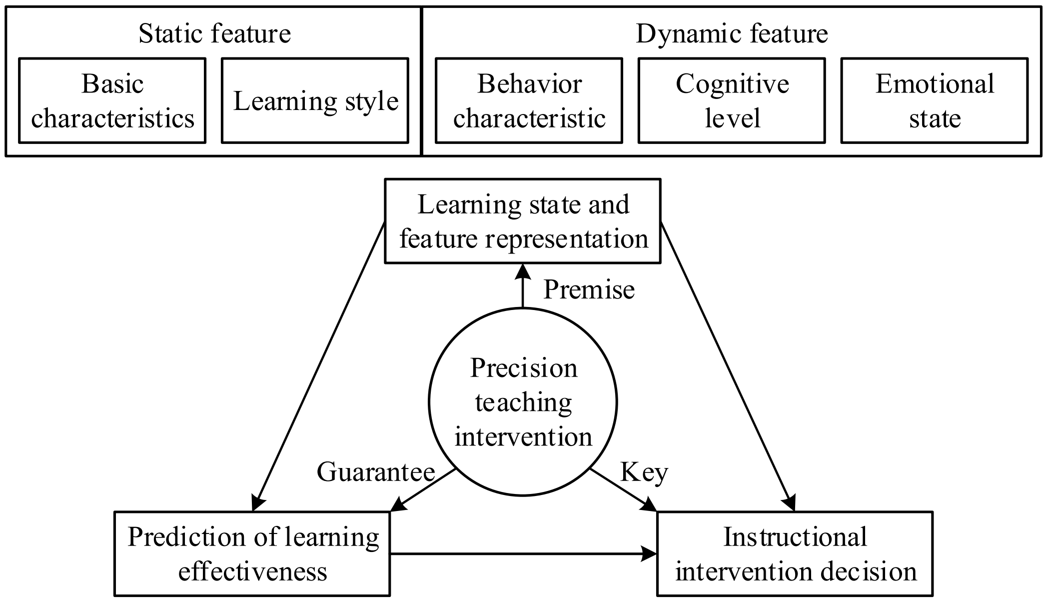


Fig. 1. The process of integration of the training process was dynamically adjusted

3. Results and discussion

3.1. Educational experiment setup

3.1.1. Student subjects. The experimental class (Group E) and the control class (Group C) were composed of two parallel classes totaling 104 students in the second year of a senior vocational school in Bang County, Henan Province. The selection of subjects was based on the following considerations: first, the quality of teaching and learning integration in this higher vocational school used to occupy the upper class level in the whole county, but in recent years, due to a variety of reasons, it has led to a decline in the quality of teaching and learning integration, and now it belongs to the middle-level higher vocational school, which is quite representative of the higher vocational schools. Secondly, the sophomore students are the key period of practical training for industry-teaching integration, which is more typical of higher vocational students as the experimental object. Thirdly, the current situation of the practical training process of the integration of production and education of parallel class students can represent the teaching level of the integration of production and education of the majority of students in this school.

3.1.2. Experimental model. This experiment adopts a balanced group stratified experimental design mode, the independent variable is the dynamic adjustment strategy of the integration of industry-teaching practical training process in higher vocational education based on reinforcement learning, and the students in Groups E and C carry out the integration of industry-teaching practical training process under the application of the dynamic adjustment and intervention strategy of practical training teaching proposed in this paper and under the traditional teaching mode, respectively. The dependent variables are the levels of students' practical training for integration of production and education at different levels, which include self-efficacy, theoretical knowledge mastery and learning ability of practical training for integration of production and education. The students are divided into three levels of upper-middle, middle and lower-middle according to their total grades before the practical training of production-teaching integration, and through the experimental treatment, the interference of non-experimental factors is excluded, to find out the actual effect of the experimental factors on the influence of students with different levels of achievement on the mastery of practical training knowledge. At the same time, we analyze the actual effects of the self-efficacy and learning ability of students in different modes in the process of industry-teaching integration practical training.

3.1.3. Experimental tools. This experiment involves 2 scales, the Student Learning Self-Efficacy Scale and the Deep Learning Competency Scale. All three scales use a five-level scoring scale. The Learning, Production and Teaching Integration Practical Training Learning Self-Efficacy Scale is based on the Learning Self-Efficacy Scale for Higher Teacher Students by the Department of Psychology of Zhejiang Normal University et al. The Practical Training Learning Ability is analyzed using the Questionnaire on the Status of Deep Learning. The process of scale development mainly followed this order, according to the integration of practical training in higher vocational education and the specific characteristics of higher vocational students, the existing scale was analyzed in depth, and the question items were added, deleted and adjusted. Further corrections were made by discussing the scale topics and wording expressions with peers and grassroots teaching staff. The experimental scale topics were finalized through item analysis of the predicted data, and in order to reduce systematic errors and avoid subjects' oriented responses, some of the topics were taken from each scale and scored in reverse, and then recoded uniformly at the end.

3.2. Analysis of students' mastery of practical training knowledge

3.2.1. Comparative analysis of theoretical knowledge scores. This paper analyzes the results of higher vocational students' understanding and mastery of practical training content based on the test answer results, and the results of the comparative analysis of the mastery of theoretical knowledge of practical training are shown in Table 1. The overall results show that the students in Group E showed a significant improvement in their theoretical knowledge achievement based on the real-time dynamic teaching strategy adjustment and intervention method after the application of the industry-teaching integration practical training teaching. Specifically, before the experiment, the theoretical scores of students in Group E and Group C at the upper middle, middle and lower middle achievement levels were 32.56 and 32.67, 29.64 and 29.12, and 22.34 and 23.47, respectively, with the significance values ranging from 0.889-0.963, and none of them differed significantly. The students in Group E, through the implementation of the Real-time Dynamic Teaching Strategies Adjustment and Intervention Practices , which resulted in a significant increase in the correct rate of the theoretical knowledge test for all three levels of students ($P=0.036$, 0.023, $0.047<0.05$), demonstrating the effectiveness of teaching strategy decision-making and adjustment based on reinforcement learning in enhancing the students' mastery of theoretical content in the process of practical training.

Table 1. Knowledge of practical training is analyzed (Theoretical knowledge)

Grade level	Group	N	Before	After	T	P
Medium bias	Group E	50	32.56	42.61	4.715	0.036
	Group C	54	32.67	39.64	1.101	0.078
	T	——	1.526	2.632	——	——
	P	——	0.924	0.047	——	——

Medium	Group E	50	29.64	36.42	3.163	0.023
	Group C	54	29.12	32.64	1.269	0.062
	T	——	1.423	2.154	——	——
	P	——	0.889	0.036	——	——
Moderate deflection	Group E	50	22.34	29.65	4.446	0.047
	Group C	54	23.47	24.12	0.938	0.086
	T	——	1.594	3.260	——	——
	P	——	0.963	0.025	——	——

3.2.2. Results of the Applied Topics Test. The scores of practical training application knowledge before and after the practice for students of different achievement levels in the two groups of subjects are shown in Table 2. Before the experiment, there was no significant difference between Group E and Group C in terms of horizontal and vertical comparisons of students at all achievement levels ($P<0.05$). After the experiment, the pre- and post-tests of the upper-middle achievement group and the lower-middle achievement group of Group E reached significant differences ($T=2.212$, 2.765, $P=0.023$, $0.028<0.05$), which reflected that under the adjustment of the real-time dynamic teaching strategy of the students of Group E, the scores of the assessment scores of the application questions in the process of practical training of the integration of industry and education were significantly improved, indicating that the students' understanding and mastery of the teaching and learning of the practical training were improved. In contrast, the students in group C performed generally in the understanding and mastery of this practical training course, and the average score of the post-test application knowledge achievement of the students at the middle and upper achievement levels was 25.63 points. This also clarifies that traditional teaching without dynamic adjustment and intervention of teaching strategies may lead to students' lack of in-depth and thorough understanding of the theoretical knowledge of practical training. It is inferred that effective teaching methods or strategies for industry-education integration need to be based on digital and intelligent means and centered on the organic integration of industry and education.

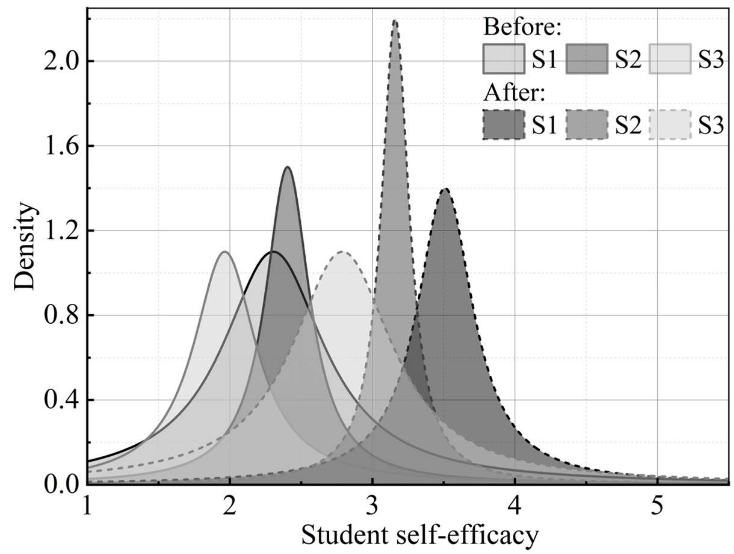
Table 2. Knowledge of practical training is analyzed (Application knowledge)

Grade level	Group	N	Before	After	T	P
Medium bias	Group E	50	25.64	31.24	2.212	0.023
	Group C	54	24.87	25.63	0.629	0.073
	T	——	0.644	2.857	——	——
	P	——	0.098	0.038	——	——
Medium	Group E	50	20.36	29.86	2.663	0.052
	Group C	54	21.48	20.34	0.977	0.062
	T	——	0.846	2.648	——	——
	P	——	0.067	0.05	——	——
Moderate deflection	Group E	50	15.64	22.34	2.765	0.028
	Group C	54	16.98	17.52	0.841	0.091
	T	——	1.003	2.485	——	——
	P	——	0.078	0.047	——	——

3.3. Analysis of changes in students' learning self-efficacy

The results of self-evaluation of students' self-efficacy before and after the experiment in each group are shown in Fig. 2, (a) and (b) represent the results of self-efficacy analysis in Group E and Group C, respectively. S1-S3 in the figure represent self-efficacy control, self-efficacy effort and self-efficacy ability, respectively. According to the self-efficacy analysis results of Group E students, Group E students' self-efficacy control, effort and competence have been improved to different degrees after the practical teaching of industry-teaching integration under dynamic adjustment and intervention strategies. Among them, the self-efficacy effort score increased more, with the mean score increasing from 2.25 to 3.93. This indicates that the teachers of Group E, through the real-time dynamic teaching strategy adjustment and the implementation of the intervention model in the process of practical training, firstly, have a new understanding of the students' practical training learning activities and their process, and are able to liberate themselves from the confinement of the traditional teaching activities, make bold attempts, and reform the classroom teaching sharply. Secondly, it breaks the traditional dull and dogmatic way of practical training teaching in the past, and makes the classroom teaching lively and interesting, with teachers happy to teach and students happy to learn. In addition, teachers in the experimental process noticed the important role of emotional expectations on students, greatly stimulate the learning enthusiasm and confidence of students, especially backward students. Specifically, through the implementation of precise teaching strategies such as establishing learning objectives, role transformation, attribution training, stimulating interest, appropriate evaluation, encouraging competition, etc., most of the students in Group E were able to have a new understanding of their own learning process in practical training. In practice, due to the ingenious introduction of learning attribution training, flexible and diverse rewards and punishments, diversified evaluation and competition mechanism adopted by the teachers, the self-control ability of most of the students, especially the middle-aged and poor students, on their own practical training and learning behaviors is greatly improved, and the students have full affirmation of their own learning ability in the enthusiastic expectations of the teachers and their own successes, and they are more willing to and eager to achieve better practical training results through their own efforts. They are more willing and eager to achieve better practical training results through their own efforts.

(a) Group E

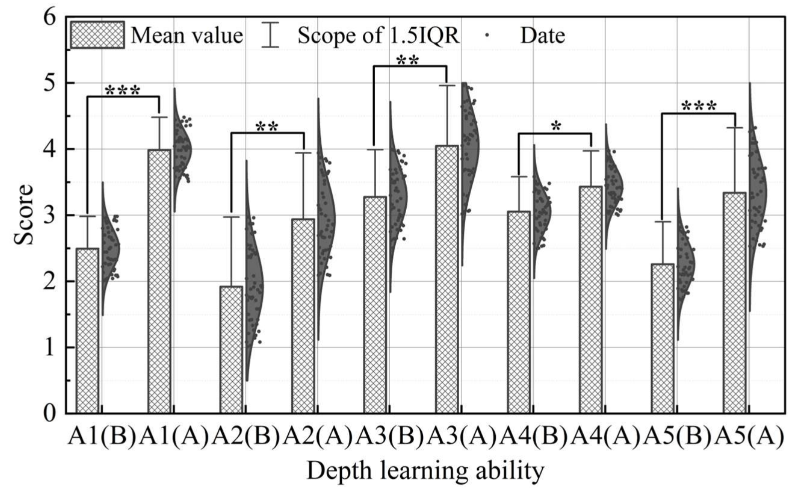


(b) Group C

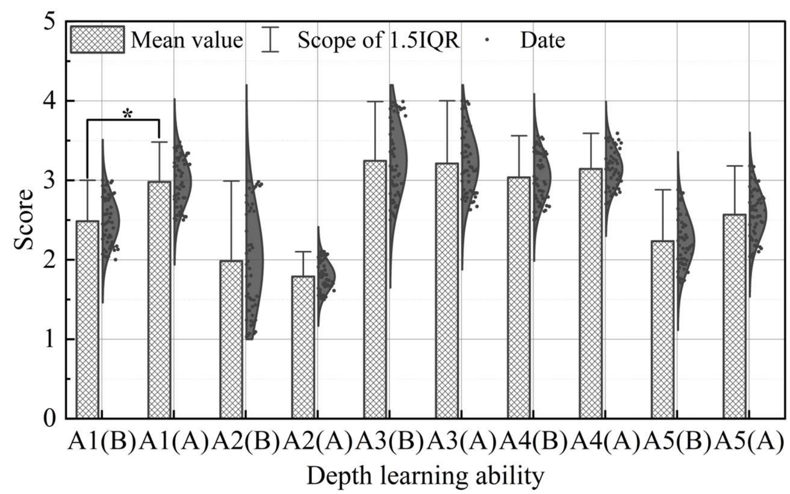
Fig. 2. Student self-efficacy analysis results

3.4. Results of the analysis of practical training and learning competencies

3.4.1. Comparative analysis of pre- and post-test data. In order to analyze the development of the depth learning ability of the subject higher vocational students in the process of industry-teaching integration of practical training before and after the experiment, a comparative analysis of the pre-test and post-test data was carried out firstly, and the results of the comparison of the students' practical training learning ability before and after the experiment are shown in Fig. 3, with (a) and (b) representing the results of the comparative analysis of the students of Group E and Group C respectively. A1-A5 in the figure represent depth understanding, critical thinking, practical innovation, career planning and strategic approach in practical training learning ability respectively. The "(A)" and "(B)" after the serial number represent post-experiment and preexperiment, respectively. The distribution of the bars in the figure can visualize the average score of each dimension of ability, and the distribution of data points can reflect the distribution of students' scores. From the pre- and post-test comparison data of the students in Group E, there is a greater degree of development in the dimensions of depth understanding, critical thinking, practical innovation, career planning and strategic approach in the process of production-teaching integration practical training for higher vocational students, which shows a significant difference. Significant differences in the dimensions of deep understanding (P=0.000) and strategic approach (P=0.000) passed the test at the P<0.001 level, while the dimensions of critical thinking (P=0.004) and practical innovation (P=0.008) were found to pass the test at the P<0.01 level. Combining the pre- and post-test data of Group C students, excluding the changes in students' practical training learning ability brought about by daily teaching and students' natural growth, it indicates that the adoption of the real-time dynamic teaching strategy adjustment and intervention model under the enhanced learning decision-making helps to enhance the effect of various dimensions of the in-depth learning ability of the higher vocational students in the process of practical training of the integration of industry and education. From the pre- and post-test comparison data of the students in Group C, the students had a certain degree of development in the three dimensions of deep understanding, career planning and strategy approach, with the average scores of 2.49 and 2.98, 3.04 and 3.14, 2.24 and 2.57 in the pre- and post-tests, respectively, but they did not show any significant difference (P>0.05). Side by side, it reflects that with the daily practical training teaching and the natural growth of students, some of the depth learning ability has some development, but not enough to reach the level of significant change, so it is necessary to intervene with measures in practical training activities to promote the significant improvement of students' practical training learning ability.



(a) Group E



(b) Group C

Fig. 3. The results of the analysis of the study ability

3.4.2. Comparative analysis of data from different groups of students. The results of the comparison of the posttest data of the practical training and learning ability of the students in Group E and Group C are shown in Table 3. From the posttest data, the students in Group E showed significant differences in the dimensions of deep understanding, critical thinking, practical innovation, strategic approach, and career planning, as well as in the total scores, which were significantly higher than the data of the students in Group C. The mean scores of the students in Group E and Group C on the posttest were 4.05 and 3.05, respectively. In the practical innovation and strategic approach dimensions, the mean posttest scores of students in Groups E and C were 4.05 and 3.21, 3.34 and 2.57, respectively, and passed the test at the $P<0.01$ level ($P=0.001$). In addition, the other dimensions as well as the total scores also showed significant differences ($P<0.05$). Overall, the practical training teaching practice of industry-teaching integration through the dynamic teaching strategy adjustment and intervention model effectively promotes the development of students' practical training learning ability in all dimensions.

Table 3. Analysis of data difference of students' learning ability

Dimension	Group E		Group C		T	P
	Mean	SD	Mean	SD		
Deep understanding	3.98	0.63	2.98	0.7	1.721	0.045
Critical thinking	2.94	0.54	1.79	0.76	1.780	0.008
Practical innovation	4.05	0.67	3.21	0.97	2.005	0.001
Career planning	3.43	0.84	3.14	0.61	2.065	0.007
Strategy	3.34	0.6	2.57	0.62	2.115	0.001
Total score	17.74	2.65	13.69	4.25	1.721	0.045

4. Conclusion

This study constructs a multi-stage real-time dynamic teaching strategy adjustment decision model based on reinforcement learning to realize the timely adjustment and precise intervention of teaching strategies in the process of industry-teaching integration practical training in higher vocational education. The results of the practical training teaching experiment show that the students in the experimental group applying the dynamic adjustment strategy have significantly improved their scores on the practical training theory and application knowledge test before and after the experiment ($P<0.05$), indicating that the students' understanding and mastery of practical training teaching have been improved. It was also found that the students in the experimental group, after the practical teaching of industry-teaching integration under the application of the dynamic adjustment and intervention strategy, showed different degrees of improvement in their sense of self-efficacy control, sense of effort, and sense of competence in the process of practical training. In the dimensions of in-depth understanding ($P=0.000$) and strategic approach ($P=0.000$) in practical training learning ability, the significant difference in the score comparison between before and after the experiment passed the test of $P<0.001$ level.

The above results prove that the dynamic teaching adjustment strategy based on reinforcement learning has a positive effect on the improvement of the learning effect of industry-teaching integration practical training in higher vocational education, but the scope of content and experimental objects covered in this study is relatively small, thus the actual results obtained are only preliminary. In the future, the research object will be further expanded and more extensive experimental research will be conducted to further test its feasibility and effectiveness.

References

[1] Bei, J. (2018). Study on the “high-quality development” economics. China Political Economy, 1(2), 163-180.

[2] Wei, L., & Yanbin, C. (2021). China's Economic Growth and High-Quality Development: 2020-2035. China Economist, 16(1), 2-17.

[3] Pan, W., Wang, J., Lu, Z., Liu, Y., & Li, Y. (2021). High-quality development in China: Measurement system, spatial pattern, and improvement paths. Habitat international, 118, 102458.

[4] Chen, G. (2024, September). Research on Multi-coordination Mechanism of Highly Skilled Personnel Training in Vocational Colleges. In 2024 10th International Conference on Humanities and Social Science Research (ICHSSR 2024) (pp. 1468-1474). Atlantis Press.

[5] Yangyang, L., & Lisha, M. (2019, October). Study on the Cultivation of Applied Talents and the Reform of Curriculum System in Universities Based on the Integration of Production, Teaching and Research. In 2019 International Conference on Advanced Education, Service and Management (Vol. 3, pp. 155-160). The Academy of Engineering and Education.

[6] Wang, Y., & Ma, L. (2023, November). Research on training mechanism of innovative talents in

combination with production, study and research in universities. In 2023 8th International Conference on Modern Management and Education Technology (MMET 2023) (pp. 134-140). Atlantis Press.

[7] Lan, L., & Muda, W. H. N. B. W. (2024). Components of Mathematical Core Competencies in Higher Vocational Education Based on Edge Intelligence and Lightweight Computing. Pakistan Journal of Life and Social Science, 22(2).

[8] Karminskaya, T. D., & Islamutdinov, V. F. (2021). Influence of higher and vocational education on the economic development of the khmao-yugra region. Ekonomika Regiona= Economy of Regions, (2), 445.

[9] Gao, F., & Zhang, P. (2020). Performance evaluation of industry-education integration in Higher Vocational Colleges: An evidence from China. International Journal of Emerging Technologies in Learning (iJET), 15(23), 208-219.

[10] Ray, S. (2019, February). A quick review of machine learning algorithms. In 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon) (pp. 35-39). IEEE.

[11] AbuKhousa, E., El-Tahawy, M. S., & Atif, Y. (2023). Envisioning architecture of metaverse intensive learning experience (MiLEx): Career readiness in the 21st century and collective intelligence development scenario. Future Internet, 15(2), 53.

[12] Vithayathil Varghese, N., & Mahmoud, Q. H. (2020). A survey of multi-task deep reinforcement learning. Electronics, 9(9), 1363.

[13] Xiaonuan, W. (2024, September). Optimization and design of industry-education integration service platform based on artificial intelligence. In AIP Conference Proceedings (Vol. 3131, No. 1). AIP Publishing.

[14] Wang, P. (2023). Under the Background of Industry-Education Integration--Real-Time Environmental Inspection--Exploration and Practice of Teaching Mode. Journal of Education and Educational Research, 5(1), 174-177.

[15] Gong, X. (2024). Performance evaluation of industry-education integration in higher education from the perspective of coupling coordination-an empirical study based on Chongqing. PloS one, 19(9), e0308572.

[16] Zuo, J., Huang, F., & Ye, Y. (2024). A Bibliometric Analysis of Industry-Education Integration Research Based on Web of Science. In Artificial Intelligence, Medical Engineering and Education (pp. 311-320). IOS Press.

[17] Chen, J. (2022). Research on Collaborative Education Model from the Perspective of Industry-Education Integration and School-Enterprise Cooperation. Advances in Educational Technology and Psychology, 6(11), 106-110.

[18] Wang, J. (2024). Developing Vocational Education and Deepening the Integration of Industry and Education Providing Human Resources Support for Building a Strong Provincial Centre-Taking the Development of Vocational Education in Sanmenxia as an Example. Journal of Social Science Humanities and Literature, 7(5), 49-54.

[19] Hu, K., & Li, X. (2024). Research on Industry-education Integration and Talents Cultivation of Vocational Education Empowering New Quality Productive Forces. Journal of Higher Vocational Education (ISSN: 3005-5784), 1(2), 211.

[20] Aiwen, W. (2022, August). Research on University-Industry Cooperation of Vocational Colleges under the Background of Artificial Intelligence. In 2022 2nd International Conference on Big Data Engineering and Education (BDEE) (pp. 228-232). IEEE.

[21] Wang, Y., Jiang, Y., Tian, S., Zheng, Y., Zhou, S., & Huang, H. (2023). Research on the four collaborative innovation mechanisms of industry-education integration based on the concept of community. Journal of Educational Technology and Innovation, 5(4).

[22] Reich, A., Rooney, D., & Lizier, A. L. (2021). Using technology integration frameworks in vocational education and training. International Journal of Training Research, 19(2), 93-106.

[23] Yuhong Hu,Deguang Wang,Ming Yang & Jiahan He. (2025). Integrating reinforcement learning and supervisory control theory for optimal directed control of discrete-event systems. Neurocomputing128720-128720.

[24] Cyrille Vessaire,Pierre Carpentier,Jean Philippe Chancelier,Michel De Lara & Alejandro Rodríguez Martínez. (2024). Complexity Bounds for Deterministic Partially Observed Markov Decision Processes. Annals of Operations Research(prepublish),1-38.

[25] Hyeong Soo Chang. (2013). On functional equations for K th best policies in Markov decision processes. Automatica(1),297-300.

[26] Asl Hamed Jabbari & Uchibe Eiji. (2023). Online Reinforcement Learning Control of Nonlinear Dynamic Systems: A State-action Value Function Based Solution. Neurocomputing.

[27] Negar Baharvand & Alireza Shameli Sendi. (2024). Optimizing service function chaining and scheduling with deep Q-learning for enhanced quality of experience in edge networks. Cluster Computing(1),27-27.

[28] ZhichengKou & JinggaoSun. (2024). Model-free inversion-based iterative learning control algorithm with adaptive gain: Achieving superior robustness and convergence. Optimal Control Applications and Methods(5),2312-2324.