International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 2542-2557

https://doi.org/10.70517/ijhsa464214

Exploring the Application of AI Adaptive Learning Models in the Graded Teaching and Evaluation System for Orchestral Instruments

Ge Song[1,*]
[1] Conservatory of Music, Luoyang Normal University, Luoyang, Henan, 471934, China
Corresponding authors: (e-mail: songge@lynu.edu.cn).

Abstract In current teaching practices, teachers assess students' mastery of specific knowledge points through quizzes and in-class questioning, which makes it difficult to provide targeted learning and hinders students' personalized development. In response, this paper proposes an Ability Point Tracking Model (APTM) based on an exploration of Bayesian Knowledge Tracking (BKT) and Deep Knowledge Tracking (DKT). This model uses neural networks to encode students' learning behaviors and predicts student performance through the dot product of ability indicator vectors and student state vectors. Compared to models like BKT and DKT, the APTM model offers greater interpretability. The APTM model was applied in practice using orchestral instrument teaching content. By analyzing students' test responses on the day they completed the knowledge, one week later, and one month later, the model assesses students' knowledge mastery and changes over time, generating diagnostic reports from both class and individual perspectives. The diagnostic structure effectively helps students identify their strengths and weaknesses and assists teachers in implementing differentiated instruction.

Index Terms BKT, DKT, APTM, graded instruction in orchestral and instrumental music

I. Introduction

Orchestral instruments refer to all the instruments in a symphony orchestra, typically performed by a group of fifty to one hundred and twenty people, or even hundreds of musicians. It consists of four main sections: woodwinds, brass, percussion, and strings [1]-[3]. Symphonic music primarily takes the form of five major genres: symphonies, concertos, symphonic overtures, symphonic suites, and symphonic poems. However, the scope of symphonic music can also be expanded to include certain orchestral works with distinctive textural characteristics, such as fantasias, ballads, rhapsodies, variations, and various dance pieces performed by symphony orchestras [4]-[7]. The unique characteristics of symphonic music—its rigorous structure, rich expression, and comprehensive artistic techniques—have established it as a distinctive and widely beloved form of performance. This is an undisputed view, and symphonic music plays an indispensable and pivotal role among the diverse array of musical genres. As a result, orchestral instrument instruction has garnered widespread attention [8]-[11].

However, due to the large number of orchestral instruments and the significant differences in playing difficulty between pieces, graded teaching in orchestral instrumental music facilitates systematic learning and creates a progressive learning process [12]-[14]. Graded teaching in orchestral instrumental music provides learners with an objective and fair evaluation system, which helps improve their musical skills and literacy [15], [16]. However, the effectiveness of graded instruction requires systematic evaluation to implement corresponding improvement measures and strengthen the rigor of talent cultivation [17], [18]. With the development and application of artificial intelligence (AI), its use in instructional evaluation has also gradually emerged. Among these, AI adaptive learning models, as a key manifestation of AI in education, can automatically adjust and optimize performance based on different circumstances [19]-[22]. In the evaluation of graded teaching of orchestral instruments, this model achieves observable evaluation of graded teaching by combining multimodal data collection, personalized learning path design, intelligent feedback, and dynamic evaluation indicators [23]-[25].

This paper first proposes a new knowledge tracking model—the APTM model—based on research into the standard Bayesian knowledge tracking model and the deep knowledge tracking model. Then, the APTM model is compared with benchmark models on the ASSISTments dataset and Jimyi dataset to evaluate its performance. Subsequently, the model is integrated into a graded teaching evaluation system for orchestral instruments, collecting Q-matrix and response data from three tests administered on the same day, one week, and one month after students complete the knowledge content. Finally, cognitive diagnostic reports are generated for both classes and individuals based on the diagnostic results to assist teachers in adjusting teaching plans and help students achieve personalized learning.

II. Adaptive learning model design

II. A. Knowledge Tracking Algorithm

The algorithm model proposed in this paper is mainly applied to knowledge tracking models. It calculates users' mastery of knowledge points based on the predicted accuracy rate of test questions. Therefore, this section will start from the development history of knowledge tracking and conduct research on Bayesian knowledge tracking and deep knowledge tracking.

II. A. 1) BKT

The essence of the BKT model is to view the learner's learning process as a standard hidden Markov process, with the results of historical behavior serving as observed variables and mastery of knowledge points as state variables. Under the current observable variables, the Bayesian formula is used to calculate the probability of the learner's knowledge mastery [26].

The BKT model is based on the following four assumptions: The learning process is a discrete transition from an unmastered state to a mastered state for each knowledge point. The state of a knowledge point is a binary latent variable, either mastered or unmastered. The score, i.e., whether the response is correct or incorrect, is a binary observable variable. Learners do not experience forgetting. Additionally, BKT models are only applied to individual knowledge points, meaning that one BKT model is trained for each knowledge point. If a question involves multiple knowledge points, separate BKT models must be trained for each knowledge point. The Bayesian knowledge tracking model is shown in Figure 1.

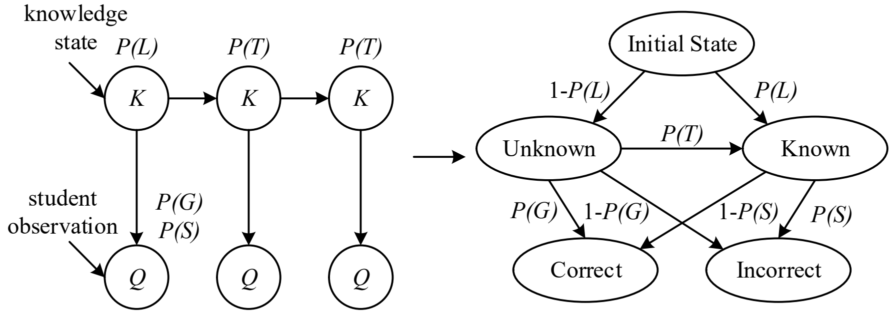


Figure 1: BKT model architecture

The learner's mastery of each knowledge point is influenced by four parameters: $P(L), P(T), P(G)$, and $P(S)$. Among these, $P(L)$ is the prior probability, representing the initial probability of the student mastering the knowledge point before beginning to learn. $P(T)$ is the learning probability, representing the conversion probability of the student mastering the knowledge point after learning. $P(G)$ is the guessing probability, which represents the probability of a student guessing correctly without mastering the knowledge point. $P(S)$ is the error probability, which represents the probability of a student answering incorrectly despite having mastered the knowledge point. The model continuously updates the mastery probability of knowledge points based on the sequence of answers provided by the student. When $P(L)$ reaches 0.95, it can be considered that the learner has mastered the knowledge point.

Students may either answer a question correctly or incorrectly. For correct answers, there are two possible scenarios: one where the student has mastered the knowledge point and answers correctly without error, and another where the student has not mastered the knowledge point but guesses correctly during the answering process. Therefore, the probability of a student answering a question correctly is the sum of these two probabilities, which is:

$$P(L_{n-1}) \times (1 - P(S)) + (1 - P(L_{n-1})) \times P(G) \tag{1}$$

For incorrect answers, there are two possible scenarios: one is that the student has mastered the knowledge point but made a mistake, and the other is that the student has not mastered the knowledge point and guessed incorrectly during the answer process. Therefore, the probability of a student answering incorrectly is the sum of the probabilities of these two scenarios, which is:

$$P(L_{n-1})\times P(S)+(1-P(L_{n-1}))\times(1-P(G)) \tag{2}$$

Based on Equations (1) and (2), and considering the factors of guessing and mistakes, the formulas for calculating the probability of mastering knowledge points when students answer correctly or incorrectly are given by Equations (3) and (4):

$$P(L_{n-1}\mid correct)=\frac{P(L_{n-1})\times(1-P(S))}{P(L_{n-1})\times(1-P(S))+(1-P(L_{n-1}))\times P(G)} \tag{3}$$

$$P(L_{n-1}\mid incorrect)=\frac{P(L_{n-1})\times P(S)}{P(L_{n-1})\times P(S)+(1-P(L_{n-1}))\times(1-P(G))} \tag{4}$$

During the learning process, students' mastery of knowledge points will change, i.e., state transitions in the hidden Markov model. In the classic BKT model, the learning probability $P(T)$ is used to represent the transition probability of students' mastery of knowledge points after learning, thereby obtaining the following knowledge point mastery state estimation function:

$$P(L_n\mid action_n)=P(L_{n-1}\mid action_n)+(1-P(L_{n-1}\mid action_n))\times P(T) \tag{5}$$

II. A. 2) DKT

The BTK model has two key issues. The first is that each BTK model can only model one knowledge point, which causes the model to ignore the relationships between knowledge points. The second issue is that binary state variables are not suitable for identifying the mastery of knowledge points in the learning process. Since deep learning models (DKT) can automatically extract specific features such as difficulty levels from the content of questions, they avoid the high cost of manual feature annotation. Additionally, with the rise of "Internet+Education," it is easy to accumulate massive amounts of response interaction data, making it easier to encode questions as vectors and calculate the relationships between questions [27]. The structure of DKT is shown in Figure 2. The student's historical response performance is treated as a time series and input into a recurrent neural network (RNN). The RNN's output is the probability of the student answering each question. The difference between the model's prediction and the actual performance is calculated using a binary cross-entropy loss function, and the model parameters are updated through backpropagation. The RNN computes the hidden variable h_t at each time step, representing the student's knowledge state at time t. The hidden variable h_t is calculated using the current input response x_t and the knowledge state at time $t-1$ h_{t-1}. h_t is then input into a Sigmoid function to obtain the final prediction. The length of the output sequence is the same as the number of input items, with each item representing the probability that the model predicts the learner will answer the corresponding question correctly, i.e.:

$$h_t=Tanh(w_x x_t+w_h h_{t-1}+b) \tag{6}$$

$$y_t=Sigmoid(h_t+b) \tag{7}$$

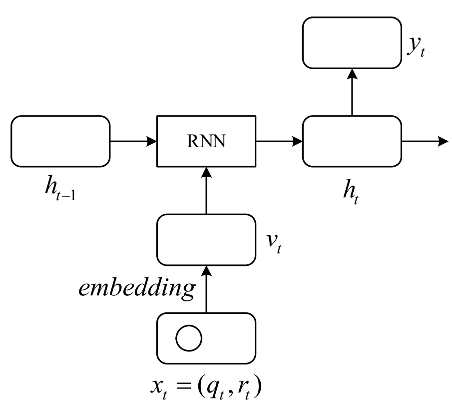


Figure 2: DKT model architecture

To improve model interpretability, this paper proposes the Dynamic Key-Value Memory Network Model (DKVMN) [28], whose model structure is shown in Figure 3. The model adds two matrices: a static key matrix that stores the knowledge concepts contained, and a dynamic value matrix that stores and continuously updates the learner's mastery of knowledge points.

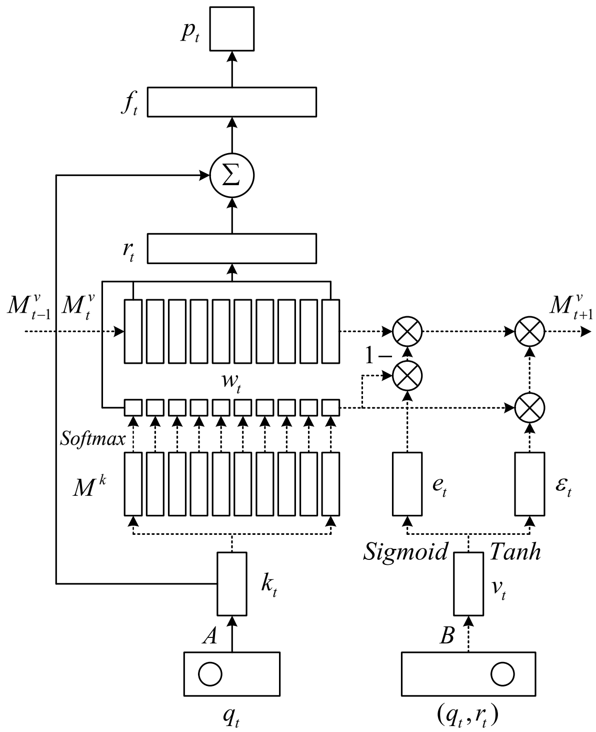


Figure 3: Structure of DKVMN Model

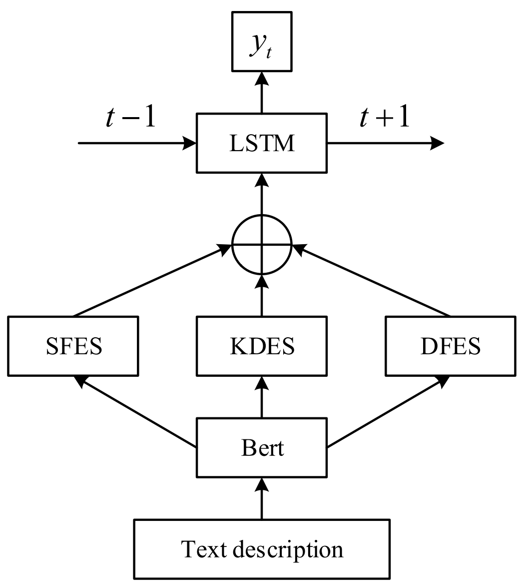


Figure 4: Structure of the EHFKT model

In DKVMN, the model first randomly initializes two N-column matrices, namely the key matrix M^k and the value matrix M^v. In terms of model input, q_t represents the exercise label at time t, which is embedded into a low-dimensional vector k_t. By calculating the cosine similarity between each element of k_t and M^k, the weight w_t is obtained to update the M^v matrix, which represents the student's mastery of knowledge. In the model's prediction output, w_t and M^v are weighted and summed to obtain the assessment vector r_t of the student's knowledge mastery status. Finally, it is concatenated with the vector k_t representing the exercise, and after passing through a fully connected layer, the probability of the student answering correctly at the next time point is obtained. DKVMN uses two matrices to record knowledge point information and the student's knowledge mastery state, enabling it to simultaneously express both the exercise and the student's knowledge point level, thereby enhancing the model's interpretability. Building upon DKVMN, the proposed Deep-IRT[29] model structure is shown in Figure 4. IRT theory is also incorporated into the network, utilizing statistics to enhance the model's interpretability.

II. B.Building an adaptive learning system

Adaptive learning technology is based on domain models and learner models. The adaptive engine creates personalized learning paths for each student, intelligently recommends learning content and tests, and continuously provides feedback and iterations throughout the learning process to enable each student to achieve their learning goals with maximum efficiency.

II. B. 1) Domain Model

The domain model serves as the foundation for the system to adaptively present course resources based on the learner model. To effectively integrate and reuse diverse knowledge data and case repository resources, and better meet the needs of upper-level personalized applications, this paper constructs the domain model based on common metadata standards such as IEEE LOM and CELTS-3. IEEE LOM is a metadata standard developed by the IEEE Learning Technologies Standards Committee, defining the syntax and semantics of learning object metadata. It uses a minimal attribute set to manage, retrieve, and evaluate learning objects.

II. B. 2) Learner Model

Building a student-centered teaching system requires the creation of a learner model for each student, which records individual characteristics and reflects individual differences. This paper designs a learner model based on four important adaptive dimensions: basic information, learning style, learning behavior, and cognitive level.

(1) Basic Information

Basic information refers to static data such as a student's name, gender, class, login username, and password. This information is primarily derived from the details provided by students during system registration and is generally not directly related to the learning process but is essential for system management.

(2) Learning Style

Learning style refers to the different learning tendencies exhibited by each learner during the learning process, including learning motivation, emotions, attitudes, and cognitive tendencies, as well as the degree of acceptance and preference for certain learning strategies and methods.

(3) Cognitive Level

A student's cognitive level is estimated by monitoring their performance within the system to assess their mastery of a particular knowledge area. Understanding a student's cognitive level is a prerequisite for adaptive learning. The adaptive learning model continuously monitors each learner's current knowledge and skill levels, updating in real-time based on student behavior. For knowledge areas and skill weaknesses that students need to address, the system pushes resources to students according to rules defined in the adaptive engine.

(4) Learning Behavior

Learning behavior refers to all learning-related behavioral operations performed by learners during the learning process. In terms of recording learning behavior, computer-based learning systems have a significant advantage over traditional classroom teaching models, as they can accurately and detailedly record learners' learning behavior, facilitating analysis of learners. Students' learning behavior largely reflects their actual learning situation. This paper selects several important learning behaviors as constituent elements of the learning behavior dimension in the learner model, which are recorded in the system in the form of logs.

II. B. 3) Predictive models

The prediction model tracks students' learning progress based on their learning behavior and predicts their future behavior and performance.

The ability point tracking model (APTM) designed in this paper uses neural networks to encode students' learning behavior, represents the abilities required to achieve a case with an ability indicator vector, and represents students' current knowledge and ability status with a student status vector. These two vectors are placed in the same dimensional vector space, and the dot product between them is used to predict student performance. Additionally, this tracking model can explain the relationships between competency points inferred from student behavior and how specific student behaviors influence competency points.

The input for the competency-based tracking model is student behavior information, and the output is the predicted values for competency point achievement.

Let the ability points corresponding to students' learning behaviors at each historical moment t be $k_t \in 1,...,N$, while also recording the positive or negative orientation of the learning behavior, $c_t \in 0,1$. The positive and negative aspects of learning behavior are defined in the domain model for the learning object. For example, answering a question correctly or downloading course materials is positive and is represented by 1, while answering a question incorrectly or failing an assignment is negative and is represented by 0.

To predict the performance of learning behavior at the next time point, the achievement of the ability points corresponding to the learning behavior at time point k_{t+1}, denoted as c_{t+1}, is treated as a Bernoulli variable. The objective of the ability point tracking model is to determine the parameters of the Bernoulli distribution at each time point:

$$p_{t+1} = P(c_{t+1} = 1 \mid k_{1:t+1}, c_{1:t}) \tag{8}$$

$$c_{t+1} \sim Bernoulli(p_{t+1}) \tag{9}$$

The model consists of three parts: a student status encoder, a competency indicator encoder, and a competency status query engine. The student status encoder converts student behavior with time information into a student status vector, the competency indicator encoder converts the competency points required to achieve a case into a competency indicator vector, and the competency status query engine predicts the achievement of student competency points or the performance of a particular learning behavior based on the two encoded vectors. The specific model is shown in Figure 5.

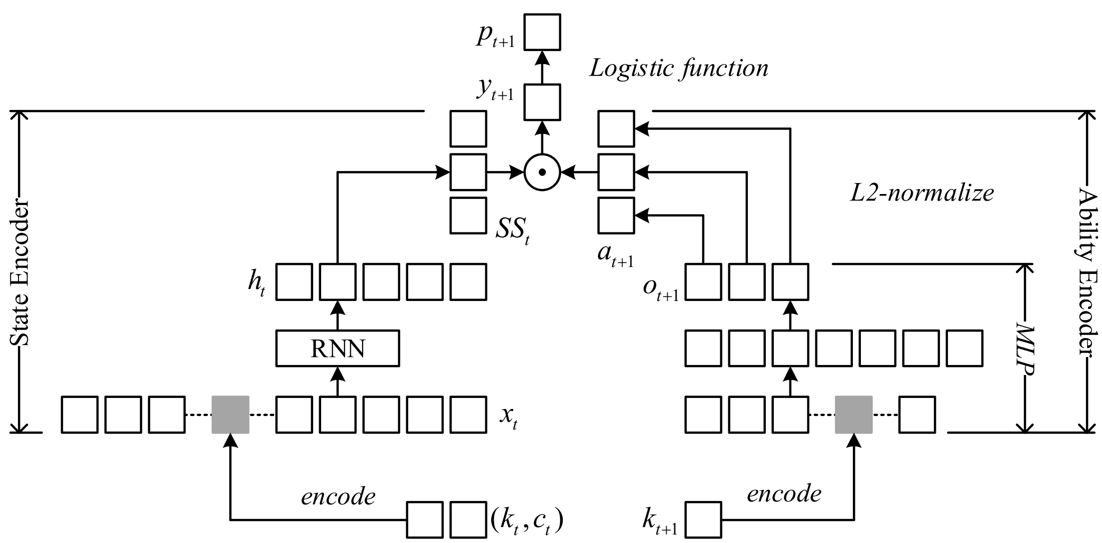


Figure 5: APTM Capability Tracking Model

1) Input

The input requires conversion using the Q matrix to convert learning behavior based on learning objects into a model of increase or decrease based on the achievement of competency points. The model accepts two parts of input: one is the current student behavior information, which is encoded using one-hot encoding according to the following formula:

$$x_t = \{0,1\}^{2N} \tag{10}$$

$$x_t^k = 1 \text{ if } c_t = 0 \tag{11}$$

$$x_t^{k+N} = 1 \text{ if } c_t = 1 \tag{12}$$

where N is the number of ability points, and k is the ability point corresponding to the learning behavior at time t. Another input is the ability point k_{t+1} corresponding to the behavior performance to be predicted, which is also represented by one-hot encoding, a vector of length 2N, where the k_{t+1} th value is 1 and the rest are 0.

2) Student state encoder

Given the input x_t, the recurrent neural network first computes a hidden state vector h_t, which can be viewed as a continuous encoding of information related to the student's historical behavior data. The variables are connected through a simple network:

$$h_t = RNN(x_t) \tag{13}$$

$$SS_t = w_{h,SS} \times h_t + b_{h,SS} \tag{14}$$

3) Ability indicator encoder

Input k_{t+1}, and output the ability state vector a_{t+1} through a multilayer perceptron:

$$o_{t+1} = ReLU(W_1 \cdot ReLU(W_0 \cdot k_{t+1} + b_0) + b_1) \tag{15}$$

$$a_{t+1} = L2_{normalize}(o_{t+1}) \tag{16}$$

Corresponding to the model prerequisites mentioned above, the ReLU activation function and L2-normalize method ensure that the capability metric vector has unit length and non-negative values. Due to these restrictions, the square of the Euclidean distance between two capability vectors has a linear relationship with the cosine distance, which is used to calculate the similarity between capability metric points:

$$d_{Euclidean}(a_1, a_2)^2 = \| a_1 - a_2 \|^2 = 2(1 - a_1 \cdot a_2) = 2d_{cosine}(a_1, a_2) \tag{17}$$

4) Ability Status Querier

Obtain two vectors S_t and a_{t+1} with the same dimensions through the student status encoder and ability indicator encoder. Use the inner product of these two vectors to obtain a logit value y_{t+1}:

$$y_{t+1} = SS_t \cdot a_{t+1} \tag{18}$$

$$p_{t+1} = \sigma(y_{t+1}) \tag{19}$$

The dot in the first equation denotes the inner product, and $\sigma(u) = \frac{1}{1 + \exp 1(-u)}$ in the second equation is an activation function.

The parameters of the Bernoulli distribution used by the final model to predict student performance at each moment are:

$$p_{t+1} = P(c_{t+1} = 1 \mid k_{1:t+1}, c_{1:t}) \approx \sigma(y_{t+1}) = \sigma(SS_t \cdot a_{t+1}) \tag{20}$$

5) Optimization

Calculate the cross-entropy error between the given probability estimate and the actual learning behavior positive/negative tendency at each moment. Sum the error terms for T=1, ⋯ t-1 to obtain the total error. Use the Adam algorithm with backpropagation to optimize the model parameters θ_{model}:

$$E(\theta_{\mathrm{mod}el} \mid c_{t+1}, p_{t+1}) = -[c_{t+1} \log p_{t+1} + (1 - c_{t+1}) \log(1 - p_{t+1})] \tag{21}$$

$$E_{total}(\Theta_{\mathrm{mod}el} \mid c_{2:t+1}, p_{2:t+1}) = \sum_{t=1}^{T-1} E(\Theta_{\mathrm{mod}el} \mid c_{t+1}, p_{t+1}) \tag{22}$$

II. C.Experiments and Evaluation

This section validates the rationality and effectiveness of the APTM model based on real data sets. For each data set, learners are divided into training sets and test sets at a ratio of 8:2. The APTM is trained using the practice data of 90% of the learners in the training set, and hyperparameter search is performed on the remaining 10% of the data using the automatic machine learning tool nni.

II. C. 1) Experimental setup

(1) Datasets

Two real-world datasets, the ASSISTments dataset and the Jimyi dataset, were used. Each dataset underwent certain preprocessing operations, and the statistical information of the two preprocessed datasets is shown in Table 1.

Table 1: Statistics of two data sets

Statistic	ASSISTments	Jimyi
Number of learners	19850	1200
Practice sequence quantity	19850	56365
Practice record	678632	4052631

Number of knowledge structure nodes	100	800
The number of sequential relationships in the knowledge structure	1023	963
The number of similar relationships in the knowledge structure	1536	1030

(2) Benchmark Methods

1) BKT

BKT is a model based on HMMs. Given a practice sequence, BKT uses HMMs to model the learner's implicit cognitive state as a set of discrete binary variables.

2) DKT

DKT applies a recurrent neural network model to the exercise sequence, simultaneously estimating the learner's cognitive state/mastery level for each concept. At each time step, DKT takes the current exercise record as input and outputs a vector representing the learner's mastery level across all concepts, with values ranging between 0 and 1.

3) DKT+

DKT+ is an extended variant of DKT aimed at addressing two main issues in the DKT model. One is that the DKT model cannot reconstruct the input, and the other is that the prediction results of the DKT model are inconsistent across different time steps. By introducing three regularization terms, the authors improved the loss function of the original DKT model, thereby enhancing prediction consistency.

4) DKVMN

DKVMN is another classic model for knowledge tracing. DKVMN has the ability to model and utilize the latent relationships between concepts, outputting the learner's proficiency level for each concept.

5) GKT

GKT is a GNN-based knowledge tracking method that only utilizes the temporal order of knowledge structure learning. At each time step, GKT aggregates the states of neighboring nodes to infer the new state of the current node, updating not only the state of the concept currently being practiced but also the states of its neighboring concepts.

(3) Evaluation metrics

This paper evaluates the model from a classification perspective. During the evaluation process, the learner's responses to practice questions are defined as a discrete binary variable, where 0 represents an incorrect response (negative sample) and 1 represents a correct response (positive sample). Two classic classification metrics—area under the ROC curve (AUC) and F1 score—are used to measure the model's performance. An AUC score of 0.5 indicates that the model's performance is no better than random guessing, while a higher AUC indicates better model performance. The F1 score can be interpreted as the weighted average of precision and recall, with the F1 score reaching its optimal value at 1 and its worst value at 0. A higher F1 score indicates better model performance.

II. C. 2) Predicting Learner Performance

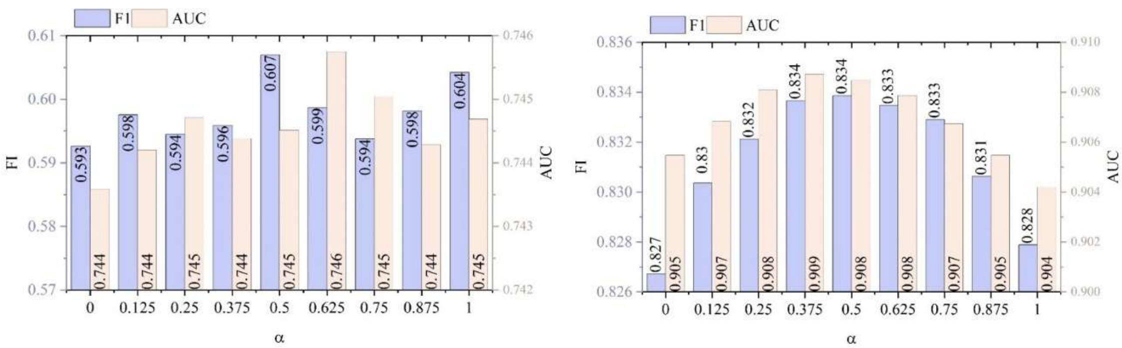
The performance comparison of different knowledge tracking models on the student performance prediction task is shown in Table 2. It can be seen that the APTM proposed in this paper achieves better performance than other benchmark models in terms of AUC and F1 on all datasets.

Table 2: Comparison of different knowledge tracking models

Data set	Eval	BKT	DKT	DKT+	DKVMN	GKT	APTM
ASSISTments	AUC	0.62	0.621	0.657	0.7	0.719	0.749
	F1	0.563	0.565	0.573	0.588	0.599	0.612
Jimyi	AUC	0.819	0.833	0.842	0.875	0.899	0.913
	F1	0.743	0.747	0.747	0.75	0.778	0.822

II. C. 3) Parameter Sensitivity

In APTM, the moderating parameter a plays a crucial role, which can adjust the weights of different learning influences propagated from the conceptual similarity relationship and the learning order relationship in the spatial dimension. When a is small, the learning impact comes more from the learning order relationship. Conversely, when a is large, the model pays more attention to the learning effect of propagating similar relationships from concepts. Experiments were carried out on different a, where a was selected from 0~1. The effect of a on the model effect is shown in Figure 6, and (a) and (b) represent the experimental results of the model on the ASSISTments dataset and the Jimyi dataset, respectively. When a increases, the effect of APTM increases at the very beginning. However, in both datasets, APTM's prediction effect decreases after a is increased to a certain level. These results suggest that it is important to balance the learning influence from the learning order relationship and the conceptual similarity relationship to improve the prediction effect of the model.



(a) ASSISTments Data set (b) Jimyi Data set

Figure 6: The effect of a on the model effect

II. C. 4) Case Study

The evolution of learners' cognitive states in the Junyi dataset is illustrated in Figure 7. It shows the learners' mastery levels of three concepts at the beginning (T=0) and end (T=20) of the learning process (on a scale of 0 to 1).

In Region I, it is evident that the mastery level of Concept 1 saw a significant improvement between Time Step 2 and Time Step 3. Additionally, the mastery levels of Concepts 2 and 3 also showed some improvement. In Region II, at Step 16 of the exercise sequence, when learners were practicing exercises related to Concept 2, they were unable to fully grasp the concept, leading to confusion and a decline in their mastery level. However, their mastery of Concept 1, which is a prerequisite concept in the learning sequence, remained largely unchanged. This observation indicates that APTM models the propagation process of learning influence in the learning sequence as a unidirectional process. From these observations, it can be seen that APTM provides better explainability of state evolution for knowledge tracking tasks by tracking the propagation process of learning influence between concepts in the knowledge structure.

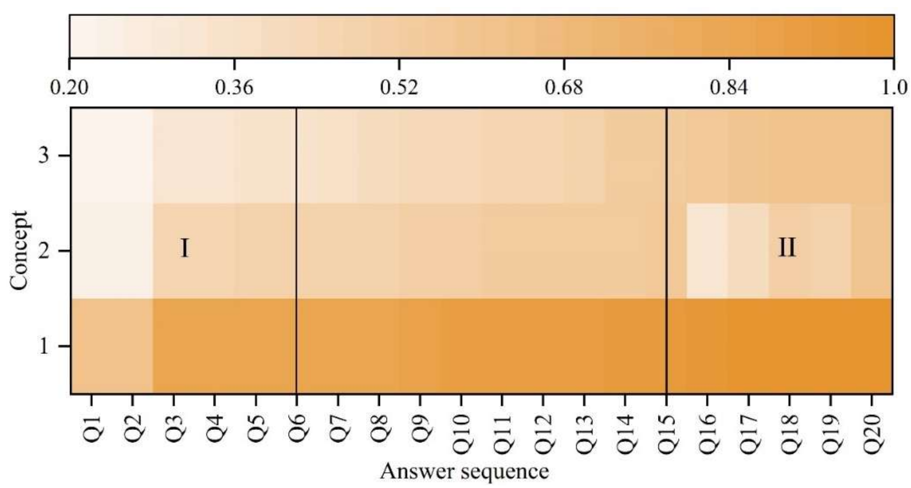


Figure 7: Junyi data focus learners' cognitive status evolution examples

III. Application of the model in the graded teaching and evaluation system for orchestral instruments

III. A. Experimental Design

This study selected the APTM model to conduct a cognitive diagnosis of orchestral instrument recognition for 120 students in the second-year (1) and (2) classes of the Music College at University A. By analyzing the data from three rounds of responses—on the same day, one week, and one month after learning the content—the study identified students' strengths and weaknesses, generated a cognitive diagnostic report, and provided insights to assist teachers in implementing remedial instruction and tailored teaching approaches. Students can use this information to address their weaknesses and engage in personalized learning, while parents can gain a comprehensive understanding of their children's progress to provide targeted support.

III. B. Design of Diagnostic Tests

III. B. 1) Determination of cognitive attributes

Since this study requires data from three rounds of responses from students on the day they complete the course, one week later, and one month later, after consulting with relevant teachers at the school and considering factors such as time and students' class schedules, the content of this study's assessment was determined to be sensory cognitive attributes (K1), technical cognitive attributes (K2), and cultural cognitive attributes (K3) in graded instruction for orchestral instruments.

III. B. 2) Establishment of the Q matrix

This study was conducted in three separate tests, with teachers providing the test papers. Each paper contained 15 questions, all of which were objective-type questions, including multiple-choice and fill-in-the-blank questions. The specific relationship matrix between the questions and cognitive attributes, i.e., the Q matrix, is shown in Tables 3 to 5, which represent the Q matrices for the three sets of data collected on the day of instruction, one week after instruction, and one month after instruction, respectively. A value of 1 indicates that the question assessed a cognitive attribute, while a value of 0 indicates that it did not. As shown in the Q matrix below, in actual tests, it is common for a single question to assess multiple cognitive attributes. In such cases, it is difficult to determine which specific attributes a student has mastered based on that question alone. Therefore, a cognitive diagnostic model is needed to assist in cognitive diagnosis.

Table 3: Q matrix in the same day

Cognitive attribute	K1	K2	K3
Q1	1	0	0
Q2	1	0	0
Q3	1	0	1
Q4	1	1	0
Q5	0	1	0
Q6	1	1	0
Q7	1	1	1
Q8	0	0	0
Q9	0	0	0
Q10	1	0	0
Q11	0	1	1
Q12	1	0	1
Q13	1	0	1
Q14	1	1	1
Q15	1	1	1

Table 4: Q matrix in the weekly measurement

Cognitive attribute	K1	K2	K3
Q1	1	0	0
Q2	0	1	1
Q3	1	0	1
Q4	1	1	1
Q5	1	0	1
Q6	1	0	0
Q7	0	1	1
Q8	0	0	0
Q9	1	0	1
Q10	1	0	1
Q11	1	1	1
Q12	1	1	1

Q13	1	1	0
Q14	0	1	1
Q15	0	1	0

Table 5: Q matrix in lunar measurement

Cognitive attribute	K1	K2	K3
Q1	0	1	1
Q2	0	0	1
Q3	0	1	0
Q4	1	0	1
Q5	1	0	1
Q6	1	1	1
Q7	0	0	0
Q8	1	1	0
Q9	1	0	0
Q10	0	0	0
Q11	1	1	1
Q12	1	1	0
Q13	0	1	0
Q14	0	1	1
Q15	0	1	1

III. C. Analysis of experimental results

After teachers have completed grading the exams, Excel and SPSS software are used to enter and statistically analyze student information such as ID numbers, correctness of answers, and time taken to complete the exam. Descriptive statistics are conducted on student performance, and the APTM model is employed for cognitive diagnosis.

III. C. 1) Descriptive statistical analysis

(1) Overall Situation

The specific distribution of student scores is shown in Figure 8. (a) to (c) represent the histograms of the score distributions for the daily test, weekly test, and monthly test, respectively. Using 60% and 90% of the full score as the cutoff points, this study divides the scores into three intervals: [0, 9), [9, 13.5), and [13.5, 15), representing poor mastery, good mastery, and excellent mastery, respectively. Analysis revealed that scores in all three tests were predominantly concentrated in the [9, 13.5) interval, indicating that students demonstrated good mastery of the orchestral instrument teaching content.

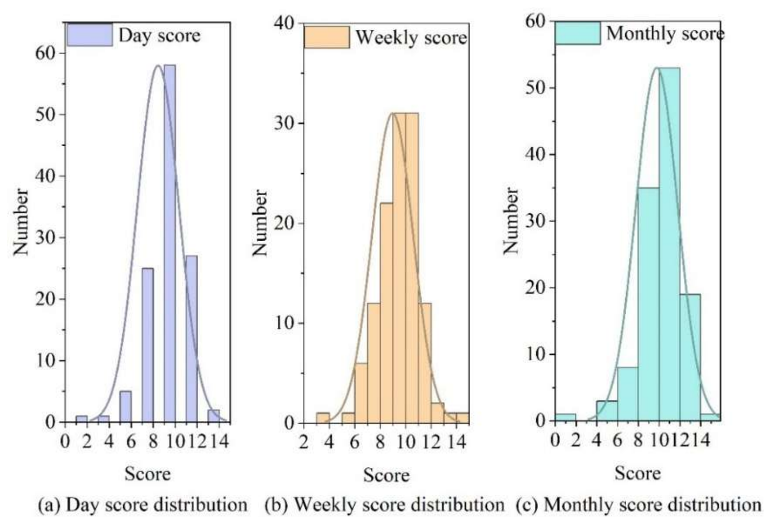


Figure 8: Specific student score distribution

(2) Class Comparison Situation

The experimental subjects came from two different classes. Due to differences among students, the mastery of knowledge may also vary between classes. Additionally, diagnostic reports need to be compiled on a class-by-class basis in the later stages. Therefore, this study conducted a comparative analysis of the situations of students in the two classes.

The descriptive analysis of scores across different classes is shown in Table 6. Class (1) had an average score rate between 8.33 and 8.38 across three different testing periods, indicating average performance. Class (2) had an average score rate between 10.62 and 12.81 across three different testing periods, indicating better performance. The standard deviation of Class (2) was lower than that of Class (1) at all three time points, indicating that Class (2) demonstrated greater stability. The median scores for both classes were above 8 points, indicating that half of the students scored above 8 points and were able to grasp the content of the orchestral instrument instruction section.

Table 6: Different class scores descriptive analysis

	Class	N	Mean	SD	Max	Min	Middle
Day survey	(1)	60	8.33	2.31	14	1	9
	(2)	60	10.62	1.5	14	3	11
Weekly survey	(1)	60	8.33	2.54	15	4	8
	(2)	60	11.9	0.72	15	7	12
Lunar survey	(1)	60	8.38	2.09	14	1	9
	(2)	60	12.81	0.99	15	6	14

(3) Score rates for each knowledge point

The orchestral and instrumental music teaching section consists of three knowledge points. The scores of all students are tallied to obtain the score rate for each knowledge point, which is the ratio of the total number of correct answers to questions related to that knowledge point among all students to the total number of attempts at answering questions related to that knowledge point among all students. The statistics on the accuracy rates of each question are shown in Figure 9. For example, in the monthly test, for knowledge point K1, there were a total of 5 questions. The total number of attempts by 120 students was 600, with 421 correct answers, resulting in a score rate of 70.2%. Students demonstrated the best mastery of knowledge point K3 and the worst mastery of knowledge point K1.

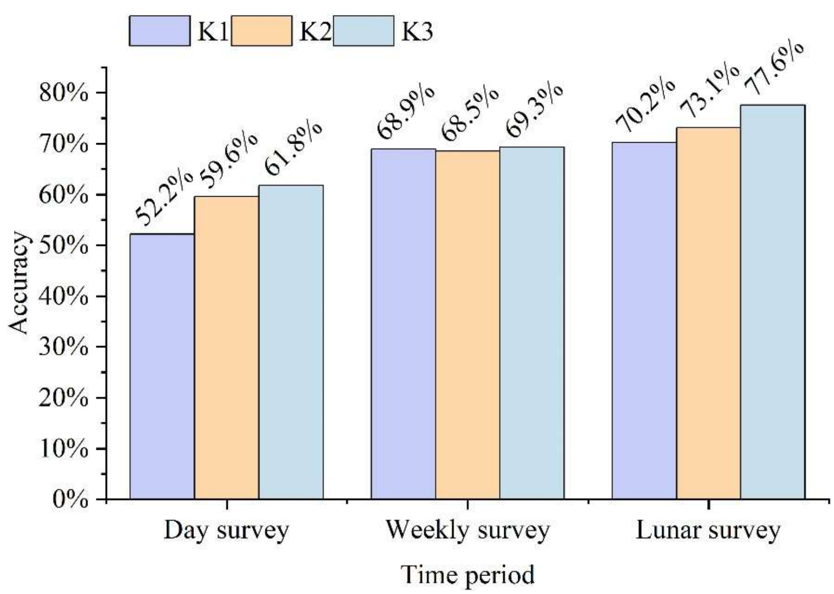


Figure 9: Statistics of the accuracy of each topic

III. C. 2) Analysis of diagnostic results based on APTM

(1) Overall Student Performance

Through diagnostic analysis using the APTM model, the mastery levels of all students participating in the test for each knowledge point are shown in Table 7. At the beginning of the test, the average mastery level for knowledge point K3 was the highest, indicating that students had a solid foundation in the “cultural cognition” knowledge point, while the average mastery level for knowledge point K1 was the lowest, corresponding to relatively weak knowledge reserves in “sensory cognition.” By the end of the test, among all knowledge points, students' overall mastery of knowledge point K3 remained the highest, while their mastery of knowledge point K1 remained the lowest. This indicates that despite the daily learning and test practice activities over this period, students were still unable to master knowledge point K1 effectively, making this area a weak point in their academic performance. By calculating the average increase in mastery levels between the beginning and end of the test, we can track changes in students' mastery of relevant knowledge points and assess the quality of their learning during this period. Among these, the average increase in mastery of knowledge point K2 reached 0.321, indicating significant progress in the learning of “technical cognition” content. Notably, the average increase in mastery of knowledge point K1 was negative, which is a signal of academic regression.

Table 7: Students master the whole knowledge

Knowledge point number	Star		End		Average added value
	M	SD	M	SD	
K1	0.565	0.13	0.511	0.211	-0.054
K2	0.543	0.149	0.864	0.141	0.321
K3	0.903	0.079	0.911	0.092	0.008

To conduct an in-depth analysis of this teaching weakness, this study visualizes students' overall mastery of knowledge point K1. The overall mastery of knowledge point K1 by students is shown in Figure 10. For the “sensory cognition” knowledge point, students' overall mastery level is relatively low, with few students fully mastering it and many not mastering it at all. Additionally, the overall mastery level is highly dispersed, indicating that students' overall learning outcomes for this knowledge point are poor. Teachers should appropriately adjust their teaching strategies to strengthen explanations and practice for this knowledge point.

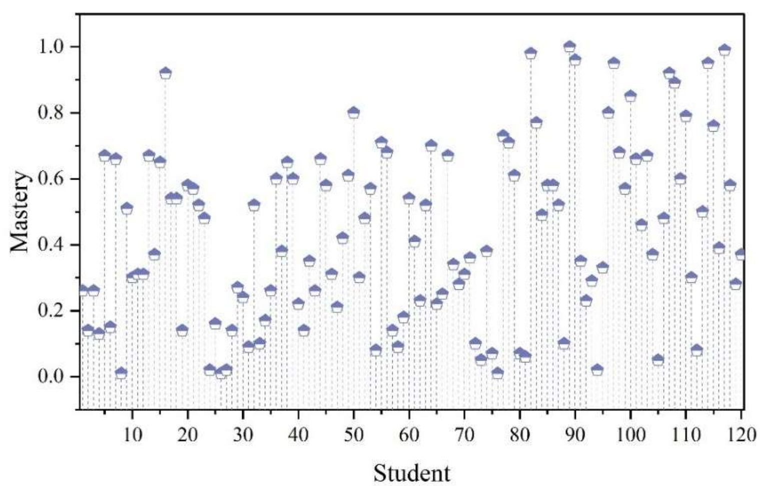


Figure 10: Students' overall knowledge of the knowledge point K1

(2) Different groups

The differences in students' mastery of various knowledge points among different class groups are shown in Figure 11. Observing the performance differences among classes, it can be seen that Class (2) has a significantly higher mastery level of all knowledge points than other classes, demonstrating the highest overall knowledge level of the class. Class (1) has the lowest level of mastery across all knowledge points, indicating that students in Class (1) need to practice more diligently and review and reinforce the relevant learning content. Targeted training should be conducted on these three knowledge points to promptly identify and address gaps, thereby improving the overall knowledge level of the class.

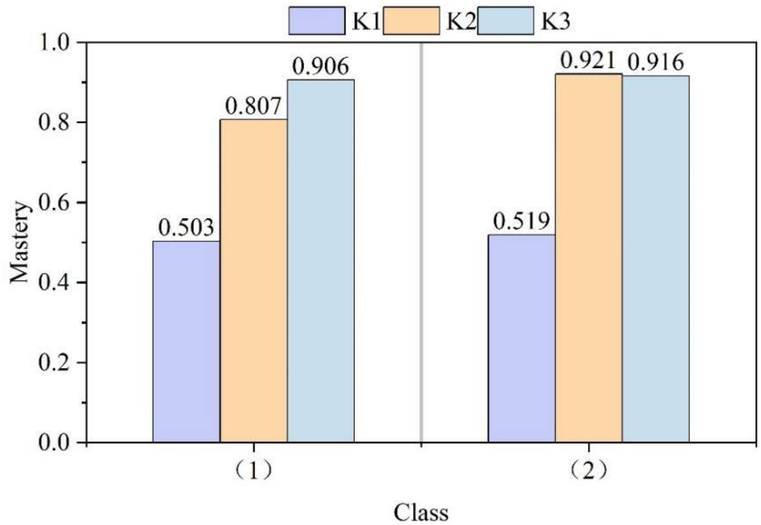


Figure 11: Students finally have a horizontal difference in knowledge points

An in-depth analysis of the knowledge point value-added situation of class groups can explore the academic progress of different classes. The average value-added results of students in each class in terms of knowledge point mastery during this learning period are shown in Figure 12. Overall, the average value-added of all classes on knowledge point K1 was negative, indicating a decline in mastery of this point. However, there were varying degrees of improvement in the mastery of the other two knowledge points, with the greatest progress made on knowledge point K2.

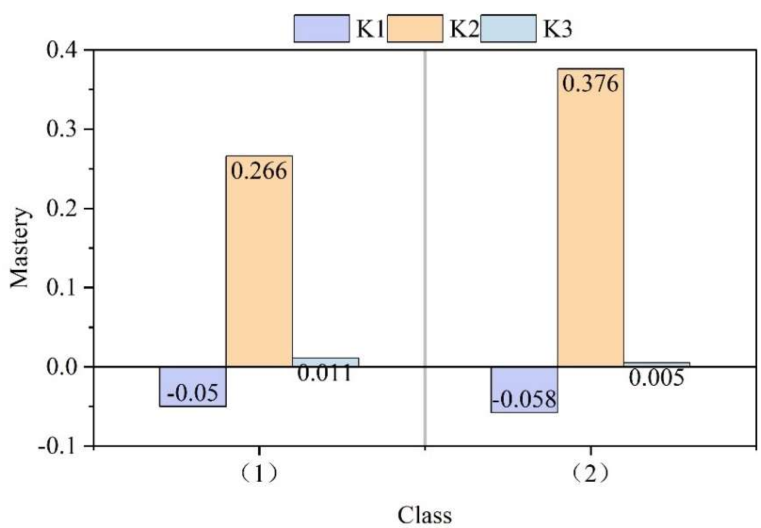


Figure 12: The average appreciation of students in the knowledge point

Specifically, for Class 2, which demonstrated the highest level of mastery of the final knowledge points, the students also achieved the greatest incremental progress across all knowledge points, with their average incremental gains significantly exceeding those of Class 1.

(3) Individual Students

Deeply analyzing and visualizing students' potential knowledge levels rather than scores is a powerful approach to moving away from the “score-centric” mindset. Through in-depth knowledge tracking, one can dynamically analyze individual students' mastery of knowledge points and predict future performance outcomes, which collectively form the basis for personalized diagnostics. Based on each student's personalized diagnostic information and visualized results, both students and teachers can clearly identify weaknesses in academic performance, providing precise and effective intelligent support for academic evaluation. Evaluation results are the core content presented to users. They must not only present diagnostic information in a scientific and comprehensive manner but also be easy for users to read and understand key information.

This study visualizes the evaluation results of individual students. Knowledge tracking was conducted on the students selected in the previous section (Class 1), and the visualization results of the mastery of knowledge points for three of these students are shown in Figure 13. The horizontal axis represents the sequence of answers composed of the last 15 answer records, with each record labeled with the knowledge point number corresponding to the question. The vertical axis represents each student's mastery of the knowledge points in the answer sequence, which is a decimal number between 0 and 1. The higher the value, the higher the student's mastery of the knowledge points. Based on the visualization results of knowledge tracking, we can clearly observe that as the test progresses, students' proficiency in related knowledge points gradually improves. In the later stages of the test, Student A has achieved a high level of knowledge, with mastery of most knowledge points approaching 1, and demonstrates consistent performance. However, Student C's knowledge level remains low, with significant fluctuations in mastery of various knowledge points, resulting in highly unstable learning performance. Student B's performance falls between the two extremes.

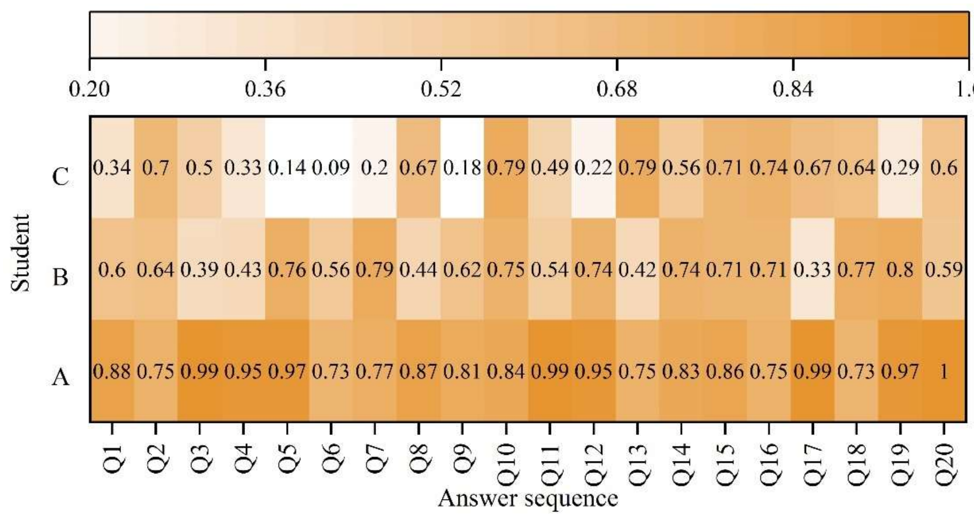


Figure 13: Visual results of knowledge tracking for three students

In summary, based on the above diagnostic analysis, teachers can gain an intuitive understanding of the weak areas in teaching and thus conduct targeted training guidance and remedial learning.

IV. Conclusion

This study is based on the needs of intelligent education. It addresses the shortcomings of traditional evaluation methods, which do not explore students' knowledge mastery from a micro perspective, by improving the knowledge tracking model for cognitive diagnosis. This aims to enhance the promotional role of educational evaluation in teaching and learning and meet students' personalized learning needs. The main findings of this study are divided into the following two parts:

(1) Based on an exploration of the BKT model and DKT model, this paper proposes a case-based ability point tracking model (APTM). Using public datasets, it concludes that APTM achieves the best performance in terms of AUC and F1 scores.

(2) Based on the APTM model, this study conducts cognitive diagnosis of sensory cognitive attributes (K1), technical cognitive attributes (K2), and cultural cognitive attributes (K3) in graded teaching of orchestral instruments. Class diagnostic reports and individual student diagnostic reports are generated based on the model analysis results to assist teachers in tailoring instruction to individual students and enabling students to address their specific weaknesses.

References

[1] Weinzierl, S., Lepa, S., Schultz, F., Detzner, E., von Coler, H., & Behler, G. (2018). Sound power and timbre as cues for the dynamic strength of orchestral instruments. The Journal of the Acoustical Society of America, 144(3), 1347-1355.

[2] Fischer, M., Soden, K., Thoret, E., Montrey, M., & McAdams, S. (2021). Instrument timbre enhances perceptual segregation in orchestral music. Music Perception: An Interdisciplinary Journal, 38(5), 473-498.

[3] Zhu, B. O. (2019). Research on the phase characteristics of orchestra instrumentation development based on artistic value. International Journal of New Developments in Engineering and Society, 3(2), 157-163.

[4] De Souza, J. (2018). Orchestra Machines, Old and New. Organised Sound, 23(2), 156-166.

[5] Amini, K., Biegert, J., Calegari, F., Chacón, A., Ciappina, M. F., Dauphin, A., ... & Lewenstein, M. (2019). Symphony on strong field approximation. Reports on Progress in Physics, 82(11), 116001.
[6] Holman, P. (2017). From Violin Band to Orchestra. From Renaissance to Baroque, 241-256.
[7] Fischer, M., & McAdams, S. (2025). Instrument Timbre Combinations Influence the Relative Prominence of Perceptual Layers in Orchestral Music. Music Perception: An Interdisciplinary Journal, 1-18.
[8] McAdams, S., Gianferrara, P. G., Korsmit, I. R., Goodchild, M., & Soden, K. (2025). Factors Contributing to Instrumental Blends in Orchestral Excerpts. Music & Science, 8, 20592043251326391.
[9] Valle, A. (2015). Residual orchestras: Notes on low profile, automated sound instruments. In The Virtuous Circle. Design Culture and Experimentation. Prooceedings of the Cumulus Conference, Milano 2015 (pp. 717-729). McGraw-Hill Education.
[10] Pellauer, M., Clemons, J., Balaji, V., Crago, N., Jaleel, A., Lee, D., ... & Emer, J. S. (2023). Symphony: Orchestrating sparse and dense tensors with hierarchical heterogeneous processing. ACM Transactions on Computer Systems, 41(1-4), 1-30.
[11] Ruggiero, A., De Simone, M. C., Russo, D., & Guida, D. (2016). Sound pressure measurement of orchestral instruments in the concert hall of a public school. Int. J. Circuits Syst. Signal Process, 10, 75-81.
[12] Feldman, E., Lutch, M., Contzius, A., & Bugaj, K. (2020). Instrumental music education: Teaching with the musical and practical in harmony. Routledge.
[13] Grey, A. N. (2020). Rote instruction in secondary instrumental music classrooms: A review of the literature. Update: Applications of Research in Music Education, 39(1), 59-69.
[14] Stringham, D. A., Thornton, L. C., & Shevock, D. J. (2015). Composition and improvisation in instrumental methods courses: Instrumental music teacher educators' perspectives. Bulletin of the Council for Research in Music Education, (205), 7-25.
[15] Jaatinen, J., Pätynen, J., & Alho, K. (2019). Octave stretching phenomenon with complex tones of orchestral instruments. The Journal of the Acoustical Society of America, 146(5), 3203-3214.
[16] Lofdahl, H. A., & Zabanal, J. R. A. (2019). Incorporating Technology in The Performance-Based Orchestra Classroom. American String Teacher, 69(3), 19-24.
[17] Pope, D. A., & Barnes, G. V. (2015). Influence of education, performance level, and presentation medium on evaluations of string orchestra performances. Bulletin of the Council for Research in Music Education, (203), 7-22.
[18] OKEKE, I. N. (2023). Problems and prospects associated with the teaching and learning of western orchestral musical instruments: The Nnamdi Azikiwe University experience. PROJECT, THESIS AND DISSERTATION.
[19] Rybinski, K., & Kopciuszewska, E. (2021). Will artificial intelligence revolutionise the student evaluation of teaching? A big data study of 1.6 million student reviews. Assessment & Evaluation in Higher Education, 46(7), 1127-1139.
[20] Chen, Z. (2022). Artificial intelligence evaluation for mathematics teaching in colleges under the guidance of wireless network. Mobile information systems, 2022(1), 3201004.
[21] Gligorea, I., Cioca, M., Oancea, R., Gorski, A. T., Gorski, H., & Tudorache, P. (2023). Adaptive learning using artificial intelligence in e-learning: A literature review. Education Sciences, 13(12), 1216.
[22] Minn, S. (2022). AI-assisted knowledge assessment techniques for adaptive learning environments. Computers and Education: Artificial Intelligence, 3, 100050.
[23] Yuldashev, S., Akramov, M., Tursunova, F., Abdullaeva, K., Shakhmurodova, D., & Ubaydullayev, A. (2024, May). A Development of AI Connected System with Adaptive Assessments Method for Evaluation Methods in Education Field. In 2024 4th International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) (pp. 826-830). IEEE.
[24] SH, U. A., & Supriyono, N. M. R. P. (2023). Optimizing Learning Through Artificial Intelligence: Evaluating the Impact of Adaptive Learning Technologies on Student Outcomes. Sinergi International Journal of Education, 1(3), 138-149.
[25] Pfeiffer, A., Bezzina, S., Dingli, A., Wernbacher, T., Denk, N., & Fleischhacker, M. (2021). Adaptive LEARNING and assessment: From the TEACHERS'PERSPECTIVE. In INTED2021 Proceedings (pp. 375-379). IATED.
[26] Meng Lingling,Zhang Mingxin,Zhang Wanxue & Chu Yu. (2019). CS-BKT: introducing item relationship to the Bayesian knowledge tracing model. Interactive Learning Environments,29(8),1-11.
[27] Liang Yu,Peng Tianhao,Pu Yanjun & Wu Wenjun. (2022). HELP-DKT: an interpretable cognitive model of how students learn programming based on deep knowledge tracing. Scientific Reports,12(1),4012-4012.
[28] Feng Xu,Kang Chen,Maosheng Zhong,Lei Liu,Huizhu Liu,Xianzeng Luo & Lang Zheng. (2024). DKVMN&MRI: A new deep knowledge tracing model based on DKVMN incorporating multi-relational information. PloS one,19(10),e0312022.
[29] Chun-Kit Yeung. (2019). Deep-IRT: Make Deep Learning Based Knowledge Tracing Explainable Using Item Response Theory. CoRR, abs/1904. 11738,