International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 17 (2025) pp. 1-17

https://doi.org/10.70917/ijcisim-2025-0218
Article

Redefining Teacher-AI Collaboration: a Study of a Collaborative Design Framework for Context-Aware English Lesson Plans

Yitong Dong *

Changchun Institute of Technology, Changchun, Jilin, 130103, China; meredith2081@163.com

Abstract: In the era of Artificial Intelligence, human-computer collaborative teaching has become a new picture of future development in the field of education, and how to utilize AI technology to collaborate on English lesson plan design has not yet been fully studied. Based on this, this paper explores the framework of context-aware English lesson plan collaborative design and improves the Bayesian knowledge tracking to propose the CS-BKT model to obtain students' English knowledge level and facilitate assisting English lesson plan design. The results show that the CS-BKT model possesses a better knowledge state tracking effect with optimal values of AUC, Accuracy, r2 and RMSE metrics, the first three of which are improved by 0.85% to 25.16%, 1.38% to 12.53% and 6.26% to 230.95%, while the latter is reduced by 3.42% to 13.80%. After applying the proposed model and framework, students in the experimental group showed significantly higher results in the latter five tests of their knowledge level than those in the control group (p < 0.05) and obtained higher teacher satisfaction. The context-aware English lesson plan co-design framework integrates context-awareness and artificial intelligence technologies and can promote the overall improvement of English teaching quality.

Keywords: artificial intelligence; Bayesian knowledge tracking; context-awareness; English language teaching

1. Introduction

By the end of 2024, the core industry scale of artificial intelligence (AI) exceeded 600 billion yuan, of which the market scale of AI applications in the field of education reached 89 billion yuan. The integration of education and AI has become an inevitable choice to realize the high-quality development of education and cultivate high-quality talents, and to promote education towards medical student-centered personalization and precision [1-3]. In the field of English teaching, the huge application potential contained in AI is being gradually released, which is reshaping the traditional education ecology in an unprecedented way, bringing an all-round and deep new teaching experience for both teachers and students [4-6]. Among them, AI has a large language model through the learning of a large number of texts, mastering the rich linguistic knowledge and semantic representation, able to understand the vocabulary in the text, grammatical structure, contextual relationships, etc., so as to more accurately capture the textual emotions, which can also be combined with other branches of the AI technology for multimodal learning situation analysis, through the analysis of the classroom learning scene of the student's facial expression or body movements, dynamic adjusting teaching methods, these advantages are due to the powerful language comprehension ability of the Big Language Model [7-11].

A lesson plan is an organizational platform and action guide for teachers to carry out teaching activities [12]. When designing a lesson plan, the teacher arranges the elements of teaching in a scientific, rational and optimized way, forms his or her own teaching ideas and framework, plans the details of the actual classroom teaching actions, and clarifies the expected goals of teaching. Therefore, designing a good lesson plan before class is of great significance in organizing classroom teaching. However, in English education, several studies have refracted that there are three major pain points in lesson plan design: the disconnection between the content of textbooks and real-life language contexts, the backwardness of the feedback system for teaching assessment, and the long time and homogeneity of teachers' design of lesson plans, which makes it difficult for students to engage in cross-cultural learning [13-17]. In contrast, Xiaoju and Sy [18] (2025) found that contextualized teaching (role-playing, situational simulation) is resistant to stress and effective in increasing students' learning engagement and motivation, which contributes to the enhancement of language skills and intercultural communication competence. This shows the importance of contextualization in English language education. In addition, Apriani et al [19] (2020) pointed out that teachers design lesson plans as an obstacle in choosing teaching methods, teaching models, and teaching materials, which may be due to the difficulty in reconciling cultural factors, practicality, and applicability in English teaching as a foreign language. The core capabilities of AI technologies such as big language models, big data analysis, machine learning, and adaptive learning can accurately capture students' learning behavior data and generate personalized learning programs, and also provide teachers with tools such as intelligent lesson planning and automatic evaluation, which can effectively crack the bottleneck of traditional lesson plan design [20-23].

In Williyan et al.'s [24] (2024) review, from lesson planning and content development to instructional presentation and assessment, teachers use AI tools to analyze English language materials, integrate multimedia resources, develop contextualized instructional scenarios and content, and refine instructional assessment to promote student interaction and language acquisition skill development. Murshid and Peter [25] (2025) point out that teachers can use ChatGPT to contextualize global English language textbooks, which makes the English language textbooks localized and more compatible with the application scenarios of non-native English language learners, and the localized textbooks achieve the same pedagogical goals as the original textbooks, and students' classroom participation is also increased. Liu et al [26] (2024) constructed an intelligent dialog-assisted learning system based on a context-aware intelligent learning mechanism, with the help of which students in the English classroom can practice dramatic dialogues, which promotes the enhancement of English speaking and conversation skills, and helps students express themselves in English in real situations. Hwang et al. [27] (2023) designed a program called “Smart RoamLingo” based on AI and recognition technology, which revises and evaluates students' English writing several times to achieve personalized and contextualized feedback in real situations. While Han and Li [28] (2024) analyzed the ChatGPT-based feedback on English writing, which mainly consisted of error correction feedback and overall linguistic expression rhetorical feedback, and the teachers would refine the feedback of ChatGPT to give students more detailed and personalized feedback on their writing. The above study provides a new path for the applicability of teaching contexts and the timeliness of teaching assessment feedback.

In addition, Yue et al [29] (2025) state that AI reshapes teachers' identities in education, transforming them from traditional instructional coaching roles to AI instructional managers, linguistic and cultural moderators, and multitechnology interdisciplinary integrators. Mohamadi et al. [30] (2024) explored the potential use of AI in lesson plan design for English as a foreign language, where the use of the “To Teach AI” tool was effective in increasing the efficiency of planning, personalizing student-centered content, and designing lesson activities aligned with instructional goals. Kerr and Kim [31] (2025) point out that generative AI realizes lesson plan design through the process of instructional topic selection, instructional material generation, lesson flow, and language evaluation under thematic analysis by means of data related to English language teaching and learning, but they also emphasize the limited quality and effectiveness of AI lesson plans. Willner et al [32] (2025) explored AI-based instructional planning for K-12 multilingual learners, which is effective in generating sample instructional plans aligned with standard unit objectives and course goals, but requires attention to protect student privacy and ethical considerations in the use of AI tools. Zhussupova et al [33] (2025) noted that pre-service English teachers' personalization of lesson planning and match with instructional goals were significantly improved with the use of digital multimodal simulation and ChatGPT combined technology, but teachers designed lesson plans that were more engaging and utilized educational resources. Yong [34] (2025) analyzed cases related to pre-service teachers' adoption of AI applications for lesson plan design and instructional reflection; AI possesses significant advantages in design efficiency and innovation, but unstable generation of design solutions and teachers' lack of AI literacy will have an impact on lesson plan design. These studies suggest that collaboration between teachers and AI tools should be facilitated to refine AI-assisted teachers' gaps in empathy, cultural understanding, and critical thinking to promote differentiation and effectiveness in lesson plan design. As Setyaningsih et al [35] (2024) emphasized, teachers need to strictly monitor the practical AI tools to reduce the difficulty and time of lesson plan design, and dynamically adjust the AI lesson plans with their own design and teaching experience to make them fit the teaching objectives and the teaching environment, and to focus on the impact of the scenarios on the teaching and learning while achieving innovation.

In this paper, we investigate a collaborative design framework for English lesson plans to explore an intelligent English education service model based on context-awareness through collaboration between teachers and AI. In order to accurately assess learners' knowledge status and assist in English lesson plan design, the study proposes a cross-skill-Bayesian knowledge tracking model to address the flaw of Bayesian knowledge tracking model that makes knowledge points independent of each other, and conducts comparative experiments with multiple knowledge tracking models on three datasets to explore the knowledge tracking performance of the CS-BKT model. On this basis, teaching experiments are designed to apply the English lesson plan co-design framework and the CS-BKT model to English teaching, and the application effects of the proposed framework and model are examined through questionnaires and students' English knowledge level tests. The CS-BKT model in this paper not only realizes the basic functions of knowledge tracking model, but also takes into account the intrinsic connection between different knowledge or skills, and improves the accuracy of the assessment of learners' knowledge status. The designed practical framework illustrates the collaborative process between teachers and AIs in the design of English lesson plans, and promotes the systematic change of English teaching and learning modes in a "human-computer collaborative" way.

2. A Framework for Collaborative Design of Context-Aware English Lesson Plans

2.1. Teachers collaborate with AI

Teacher-AI collaboration is a new form of teaching characterized by the collaboration between teachers and machine intelligence, and AI technology has become the cognitive outsourcing of teachers by virtue of its powerful computation, recognition, perception, acquisition, tracking, aggregation, analysis, monitoring and other capabilities, so that the collaborative teaching of "AI technology + teacher" has become the norm and the main form of human-machine collaboration. Specifically, data-driven AI technology favors quantitative cultivation, and is able to handle large-scale, fine-grained computation of repetitive tasks in teaching, such as recording learner learning behavior data, diagnosing problems and providing suggestions. Although this kind of precise mining and analysis effectively promotes breakthroughs in education and teaching, and promotes innovation and change in English teaching, digitizing students into quantifiable data to be analyzed and monitored makes education only stay on the level of cold data, lacking its proper humanistic care and moral enhancement, and thus fails to achieve the goal of nurturing human beings. Teachers focusing on qualitative training can make up for the lack of humanity in big data. Teachers analyze the cause and effect of the educational phenomenon behind the information based on the data collected by AI, make teaching decisions, and at the same time take on the tasks of inspirational guidance, lesson plan design, moral cultivation, value shaping, emotional communication, and the cultivation of critical thinking and imagination. Context-awareness refers to the technology that computers "recognize" the user's current environment by means of sensors and other devices or technologies, and make different responses according to the different environments. Therefore, in English teaching, teachers can realize context-awareness through AI technology and intelligent devices, so as to complete the collaborative design of English lesson plans.

2.2. English lesson plan co-design framework

The context-aware English lesson plan co-design framework based on AI collaboration is shown in Figure 1. This paper constructs a practical framework for collaborative design of English lesson plans from three levels, namely, intelligent perception layer, intelligent analysis layer, and intelligent service layer, to explore how context-awareness and the introduction of AI technology can act on the optimization of collaborative design of English lesson plans from a deeper level.

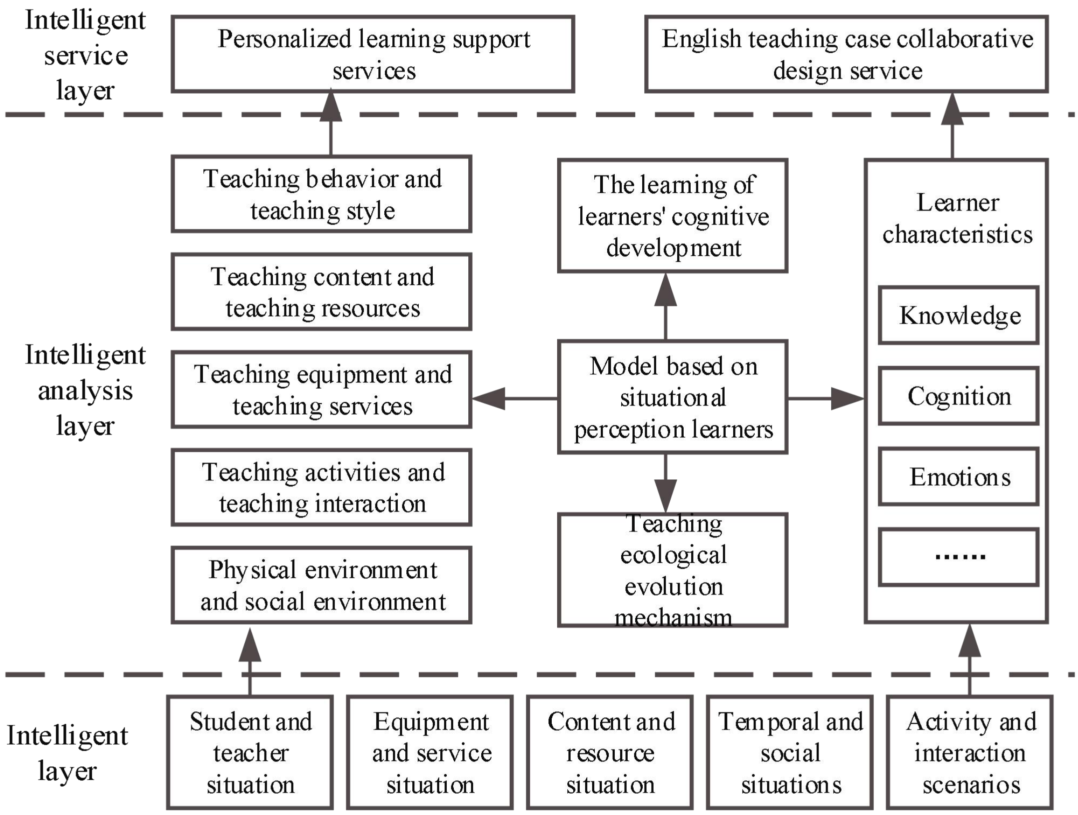

Figure 1. The practical framework of collaborative design of English teaching cases.

2.2.1. Intelligent Sensing Layer

Intelligent perception layer mainly utilizes intelligent perception devices to realize all-round perception and integration analysis of the educational context of the intelligent classroom. Through the full-time and multi-dimensional data collection of "human-machine-object-environment-activity" and other elements of the complete teaching process in the English classroom environment, it realizes the situation of students and teachers, equipment and service context, content and resource context, space-time and social context, activity and interaction context. Such as:

(1) Learner level: collect data on learners' external behavior and internal physiological information using devices such as cameras, eye-trackers, electroencephalography, electrodermography, functional magnetic resonance, and functional near-infrared imaging.

(2) Physical environment level: using intelligent sensing devices such as temperature sensors, light sensors, humidity sensors and other intelligent sensing devices to collect real-time information about the physical environment of the English learning space.

(3) Social environment level: modeling and analyzing the correlation between "human-machine-object" by using the method of social network analysis, so as to realize the organic decomposition of the multiple interaction mechanism in the English classroom.

(4) At the level of teaching content: using the method of knowledge mapping to label the relevant information of English teaching content, and accurately portraying the potential relationship between teaching content and students' prior knowledge.

2.2.2. Intelligent Analytics Layer

The Intelligent Analysis Layer mainly analyzes the correlation between educational context features and learner features in the English classroom environment, constructs a two-way matching mechanism between educational context and learner features, builds a learner model based on context perception, and explores the mechanism of the influence of the elements of the educational context on the learner's knowledge construction, cognitive development, and emotional state.

(1) Learner and Educational Context Characteristics Mining Based on Multimodal Data Fusion

First, the multimodal data fusion method is used to accurately collect and fuse and analyze the academic assessment data, psychological assessment data, external behavior data, and physiological information data of English language learners, so as to accurately portray the knowledge, cognition, emotion and other elements of learners. Second, the multimodal data fusion method is used to accurately analyze the student and teacher context, equipment and service context, content and resource context, spatial and temporal and social context, and activity and interaction context under multiple spatial and temporal conditions, from which key characteristic indicators on English teachers' teaching behavior and teaching style, teaching content and teaching resources, teaching equipment and educational services, teaching activities and learning interactions, and the physical and social environment can be extracted to achieve a precise portrayal of the learners' knowledge, cognition, emotions, and other elements. From these, key characteristic indicators on English teachers' teaching behavior and teaching style, teaching content and teaching resources, teaching equipment and educational services, teaching activities and learning interactions, physical environment and social environment are extracted, and an accurate portrayal of the elements of English education context is realized.

(2) Learner Modeling Based on Context Awareness

focuses on the mechanism of the influence of the creation of educational contexts on learners' knowledge construction, cognitive development and emotional state, explores how the creation of educational contexts affects students' English learning process and learning outcomes, and conducts in-depth excavation and analysis of learners' deep-seated characteristics such as their interest in learning, learning motivation, learning preferences and learning styles.

(3) Analysis of Learners' Cognitive Development Laws

Utilizing context-awareness and AI technologies to achieve accurate monitoring of the whole process of English teaching and exploring the mechanism of the influence of elements of the English educational context on the cognitive development of learners, in order to provide intelligent support services for the design of English lesson plans.

2.2.3. Intelligent Service Layer

(1) Personalized Learning Support Service

Through the collection and analysis of multimodal data to achieve accurate assessment of English learners' behavior, cognition, and emotion, intelligent analysis of learners' learning situation, construction of a two-way matching model between the learner and the educational context, and clarification of how the construction of the educational context plays a role in the cognitive development of the learner, so as to realize the appropriate adjustment of the English educational context.

(2) Collaborative Design Service for English Lesson Plans

Collaborative design service for teachers' English lesson plans is one of the key problems to be solved in the research of English education context-awareness. Through modeling and analysis of the complete educational context, we can achieve accurate monitoring of the whole process of English teaching, help teachers understand the overall cognitive styles and learning preferences of their students, and then choose the most appropriate way of presenting teaching content, optimize the design strategy of lesson plans, and improve the effectiveness of classroom teaching. Through the diagnostic analysis of teachers' teaching process data and teaching effectiveness data, we can help teachers understand their potential problems in professional knowledge, teaching style, teaching strategies, teaching interaction design, teaching activity organization, etc., and provide them with timely and effective teaching improvement strategies.

3. Knowledge tracking model based on CS-BKT

In order to better realize the collaborative English lesson plan design between teachers and AI, based on the English lesson plan collaborative design framework proposed above, we construct a knowledge tracking model based on CS-BKT, which models the students' learning process through educational data mining, assesses the students' English learning status, and assists in the design of English lesson plans.

3.1. Knowledge status tracking process

Bayesian knowledge tracking is a modeling approach that provides real-time feedback in terms of learners' interactions, assesses the learners' knowledge mastery state and predicts their future learning by inputting their question-answering performance. Knowledge state tracking based on Bayesian knowledge tracking mainly consists of three parts: answer sequence input, knowledge state modeling and knowledge level assessment.

3.1.1. Answer Sequence Input

Knowledge status tracking is realized by inputting students' interaction data with the coursework, i.e., students' online answer sequences. After students' independent learning is completed, they participate in the English answer test set by the instructor in the e-learning platform and submit the answer results; after the answer is finished, the e-learning platform generates the learner's answer sequences, which record the results of the learner's answers over a period of time.

3.1.2. Knowledge state modeling

Bayesian Knowledge Tracking (BKT) is used to model the learner's knowledge state by taking the learner's mastery of the knowledge point as a hidden variable and the learner's response as an observed variable to dynamically track the learner's knowledge state. The whole modeling process can be divided into three steps, which are determining the model structure, generating the probability matrix and calculating the mastery probability.

(1) Determine the model structure

First, the model structure of Bayesian Knowledge Tracking needs to be determined. Before determining the model structure, the parameters involved in the model are introduced. Each knowledge node in the BKT model structure contains four parameters: the initial probability parameter $P\left(L_0\right)$, the transfer probability parameter $P(T)$, the guessing probability parameter $P(G)$, and the probability of failure parameter $P(S)$, respectively. Among them, $P\left(L_0\right)$ and $P(T)$ are the learning parameters, which are mainly used to denote the knowledge state of the learner. $P(G)$ and $P(S)$ are the performance parameters, which are mainly used to indicate the learner's question answering.

In the BKT model, the learner's knowledge node is represented by "K", when "$K=1$", it means that the learner has mastered the currently learned knowledge points. When "$K=0$", it means that the learner has not mastered the current knowledge point. The learner's performance node is denoted by "Q". When $Q=1$, it means that the learner has correctly answered the question corresponding to the current knowledge point. When $Q=0$, it means that the learner answered the question corresponding to the current knowledge point incorrectly.

(2) Generating Probability Matrix

Next, the probability matrix is generated based on the model structure. Modeling the influence of the learner's current knowledge node on the performance node and the influence on the next knowledge node can be quantified by the probability matrix. There are three probability matrices of Bayesian knowledge tracking model, which are initial probability matrix, transfer probability matrix and observation probability matrix. Assume that the moment in which the initial knowledge node K_0 is located is t, the moment in which the next knowledge node K_1 is located is $t+1$, and so on, the moment in which the K_n knowledge node is located is $t+n$.

The probability that the learner has mastered the point at the moment t is $P\left(L_0\right)$, then the probability that the learner has not mastered the point at that moment is $1-P\left(L_0\right)$. The probability that the learner transfers from the state of not mastering the point to the state of mastering the point is $P(T)$, then the probability that the learner has not yet mastered the current point after learning is $1-P(T)$. When the learner answers the question, the probability that the learner has not yet mastered the current point and yet still answered the question correctly is $P(G)$, the probability that the learner has mastered the current knowledge point and answered the question incorrectly is $1-P(G)$. The probability that the learner has mastered the current knowledge point but still answered the question incorrectly is $P(S)$, the probability that the learner has mastered the current knowledge point and answered the question correctly is $1-P(S)$.

(3) Calculate the mastery probability

Finally, the Bayesian formula for calculating the learner's mastery probability of the current performance node and the next knowledge node is obtained based on the probability matrix, so as to calculate the learner's mastery probability of the knowledge node in a clearer and more accurate way.

1) When the learner's performance node is "answering the question correctly", the corresponding knowledge node can have two cases: the learner has mastered the current knowledge point or the learner has not mastered the current knowledge point. If the learner has mastered the current knowledge point, the probability that the learner has mastered the current knowledge point and answered the question

correctly is $P\left(L_n\right)(1-P(S))$; if the learner has not mastered the current knowledge point, the probability that the learner has not mastered the current knowledge point and answered the question correctly is $\left(1-P\left(L_n\right)\right)P(G)$. The probability that the learner has answered the question correctly, $P\left(right_n\right)$, is as shown in Equation (1):

$$P\left(right_n\right)=P\left(L_n\right)(1-P(S))+\left(1-P\left(L_n\right)\right)P(G) \tag{1}$$

2) When the learner's performance node is “answering the question incorrectly”, the corresponding knowledge node can have two situations: the learner masters the current knowledge point or the learner does not master the current knowledge point. If the learner has mastered the current knowledge point, the probability that the learner has mastered the current knowledge point but answered the question incorrectly is $P\left(L_n\right)P(S)$. If the learner has mastered the current knowledge point, the probability that the learner has mastered the current knowledge point and answered the question incorrectly is $\left(1-P\left(L_n\right)\right)P(1-P(G))$. The probability that the learner has answered the question incorrectly, $P\left(wrong_n\right)$, is as shown in Equation (2):

$$P\left(wrong_n\right)=P\left(L_n\right)P(S)+\left(1-P\left(L_n\right)\right)P(1-P(G)) \tag{2}$$

3) If the learner's answer is known, the prior probability $P\left(L_n\right)$ of the corresponding knowledge node can be in two cases: the probability of mastery $P\left(L_n \mid right_n\right)$ of the current knowledge node if the learner answers the question correctly is shown in Equation (3):

$$P\left(L_n \mid right_n\right)=\frac{P\left(L_n\right)(1-P(S))}{P\left(right_n\right)}=\frac{P\left(L_n\right)(1-P(S))}{P\left(L_n\right)(1-P(S))+\left(1-P\left(L_n\right)\right)P(G)} \tag{3}$$

The probability of mastery $P\left(L_n \mid wrong_n\right)$ of the current knowledge point in case the learner answers the question incorrectly is shown in Equation (4):

$$P\left(L_n \mid wrong_n\right)=\frac{P\left(L_n\right)P(S)}{P\left(wrong_n\right)}=\frac{P\left(L_n\right)P(S)}{P\left(L_n\right)P(S)+\left(1-P\left(L_n\right)\right)(1-P(G))} \tag{4}$$

Therefore, regardless of whether the learner's answer situation is correct or not, the learning prior probability $P\left(L_n\right)$ can be obtained, and the prior probability $P\left(L_n\right)$ is shown in Equation (5):

$$P\left(L_n\right)=P\left(L_n \mid right_n\right)+P\left(L_n \mid wrong_n\right) \tag{5}$$

4) Update the learner's mastery of the next knowledge point according to the learner's response. it is assumed in the BKT model that forgetting does not occur during the learning process, and the probability of learning the next knowledge point is updated according to the transfer probability $P(T)$. The probability of the learner's mastery of the next knowledge point, $P\left(L_{n+1}\right)$, is shown in Equation (6):

$$P\left(L_{n+1}\right)=P\left(L_n\right)+\left(1-P\left(L_n\right)\right)P(T) \tag{6}$$

3.1.3. Assessment of knowledge levels

After training the knowledge state of the learner, the knowledge mastery level of the learner will be obtained, which mainly includes the student knowledge level assessment table and the student's future answer prediction table, and the knowledge state of the learner at each moment will be visualized, which is more graphic and intuitive to show the change of the knowledge state of the learner with the movement of time.

3.2. CS-BKT model

The BKT model treats each knowledge point as an individual, trains the corresponding parameters and calculates the students' final mastery level. However, in the process of students' English learning, the knowledge points do not exist independently of each other, but are closely related. Therefore, this paper improves the BKT model and proposes the Cross-Skill-Bayesian Knowledge Tracking (CS-BKT) model.

3.2.1. Rationale

The CS-BKT model has a core assumption that students who deepen their understanding of a particular English skill A will likewise deepen their understanding of English skill B to some extent, achieving a touch-and-go effect. Therefore, the CS-BKT model introduces a new parameter matrix of interactions between skills:

$$R_{ij} = \text{Students learn skills } i \text{ and the impact of } j \tag{7}$$

In the skill relationship matrix R, the rows and columns represent the skills corresponding to the questions in the test, and the values at the intersections of the rows and columns are the probabilities of the impact of one skill on another.

The values in the matrix R are only the initial skill influence probabilities, and the probability values in the matrix will continue to change as the student response data are continuously trained.

Therefore, when calculating students' mastery of other skills, we need to take into account the changes in students' mastery of skill k. The structure of the CS-BKT model can be represented in Figure 2, as the answer situation of skill K keeps changing, the corresponding level of students' knowledge of skill K, $p(L)$, also keeps changing. This change also affects the mastery level of other skills, and the parameter of influence is R_{ki}.

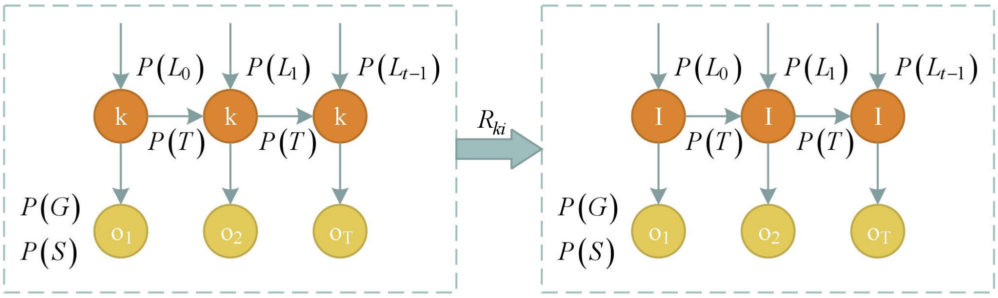

Figure 2. CS-BKT model structure.

Therefore, assuming a total of n skills, comparing to the standard BKT model, an additional n^2 parameters need to be trained. The final level of mastery of a particular skill in the CS-BKT model is the sum of the level of understanding computed in the standard BKT model and the probability of the influence of the other skills on this skill. This can be expressed by the formula:

$$\hat{p}\left(L_{t+1}\right)_u^k = p\left(L_{t+1} \mid obs\right)_u^k + \left(1 - p\left(L_{t+1} \mid obs\right)_u^k\right) \cdot p(T)^k \tag{8}$$

$$\Delta p\left(L_{t+1}\right)_u^k = \hat{p}\left(L_{t+1}\right)_u^k - p\left(L_t\right)_u^k \tag{9}$$

$$p\left(L_{t+1}\right)_u = p\left(L_t\right)_u + R_k \cdot \Delta p\left(L_{t+1}\right)_u^k \tag{10}$$

where $p\left(L_{t+1}\right)_u$ is the level of mastery of student u for all skills, i.e., we are no longer updating a student's level for a single skill, but the full skill mastery status is updated since the skill has an impact on all other skills.

3.2.2. Implementation methodology

According to the above principle, the specific process of CS-BKT model is as follows:

(1) First set the initial parameter values according to the number of knowledge skills.

(2) Input the English responses of different learners.

(3) Input the matrix of influence relationship between initial knowledge skills.

(4) Use gradient descent method to train and update the parameters and knowledge matrix according to the students' answers.

(5) Calculate the students' skill levels using the expectation maximization algorithm.

(6) Recalculate the skill level on the basis of step 5 using the updated values of the relationship matrix between knowledge.

(7) Predict the student's next question answer.

(8) Repeat steps 4 through 7 until the threshold is reached.

3.3. Experimentation and analysis

A series of experiments were conducted to verify the validity of the CS-BKT model. This section will first introduce the datasets used in the experiments, and then introduce the specific settings of the experiments, including data preprocessing, model comparison methods and model evaluation indexes. Finally, the experiments are analyzed.

3.3.1. Experimental data

This paper uses three publicly available datasets, ASSISTment2012, ASSISTmentChall, and NeurIPS2020. The ASSISTment2012 dataset is from an online education platform, ASSISTments, founded in 2004.The data in the dataset was collected from the 2012-2013 school year of A skill enhancement task in which students were required to complete a series of related test questions and achieve a level of mastery.ASSISTmentChall is the dataset used in a data mining competition organized by ASSISTments in 2017.The NeurIPS2020 dataset was released in the educational competition NeurlPS2020.

3.3.2. Experimental setup

(1) Data preprocessing and dataset partitioning

In the data preprocessing stage, the answer times were discretized into data in seconds and the interval durations were discretized into data in minutes. To facilitate the mapping of the data into vectors, the knowledge points, test question numbers, interval durations, and answer durations were converted into the form of ids.

All the datasets were divided into test set and training set in 3:7 ratio. In order to find the optimal parameter settings, all the model training processes were performed using the five-fold cross-validation method and 25% from each training set was randomly selected as the validation set. Finally, the model parameters with the highest AUC on the validation set are selected for testing.

(2) Comparison Methods

To prove the effectiveness of CS-BKT model, DKT, DKT +, DKVMN, SAKT, EKTA and AKT are used to compare with CS-BKT.

(3) Evaluation metrics

In order to observe the effectiveness of the CS-BKT model, the experiments take into account the validation on the regression and classification perspectives, and the four metrics of ROC curve (AUC), prediction accuracy (Accuracy), Pearson coefficient squared (r^2), and root-mean-square error (RMSE) are used respectively:

①Accuracy is a metric commonly used to measure the performance of classification models, and it can indicates the proportion of correctly classified samples to the total number of samples.

② Although Accuracy is simple and intuitive when used as an evaluation metric for classification tasks, it is not applicable to situations where the distribution of samples is unbalanced. In contrast, AUC can be used to deal with unbalanced datasets. AUC is the magnitude of the area under the R0C curve, and the ROC curve plots the relationship between TPR and FPR and can provide a relatively reliable performance metric at any sample proportion. Thus, the AUC can reflect the accuracy of the model's classification even if the data set is very unbalanced. Usually, the closer the AUC value is to 1, the better the model performance is, and conversely, the closer the AUC value is to 0.5, the worse the model performance is, and even when it is equal to 0.5, the model is equivalent to random guessing.

③ r^2 denotes the square of the Pearson coefficient. Pearson's coefficient is a statistic that measures the degree and direction of correlation between two variables in a linear relationship, ranging from -1 to +1. When the two variables are completely positively correlated, the Pearson's coefficient is 1. When the two variables are completely negatively correlated, the Pearson's coefficient is -1. When the two variables are not linearly correlated, the Pearson's coefficient is 0. In this paper, we use the square of the Pearson coefficient to measure the performance of the model. uses the square of the Pearson coefficient to measure the linear correlation between the predicted and actual values of the model to ensure that the values of all evaluation effects remain positive for easy comparison.

④RMSE is a performance metric commonly used in regression models to measure the difference between predicted and true values. It is the value of squaring the prediction error and then averaging it before opening the root sign, indicating the average size of the prediction error.

Where the value domain of AUC, Accuracy, r^2 and RMSE are [0, 1], the larger the value of AUC, Accuracy and r^2, the better the result, and the smaller the value of RMSE, the better the result.

3.3.3. Model Forecast Performance and Comparison

In this experiment, the model parameter with the highest AUC value on the validation set is saved during each training session of the five-fold cross-validation method, and it is used to test the data on the test set. Finally, the average of the five evaluation results on the test set is taken as the final evaluation result of the model.

Figures 3 to 5 show the prediction results of all models on different data sets. It can be seen that the CS-BKT model achieves the highest evaluation results on all data sets and all evaluation indexes, indicating that it has the best prediction effect, i.e., the CS-BKT model constructed in this paper has a more knowledge-tracking effect, which can accurately assess the knowledge status of English learners for the collaborative design of English lesson plans. On the ASSISTment2012 dataset, the AUC, Accuracy, and r^2 results of the CS-BKT knowledge tracking model are improved over the comparison model by 3.82%~15.16%, 2.39%~12.03%, and 26.38%~108.37%, and on the ASSISTmentChall dataset by 6.84%~ 25.16%, 4.86%~12.53%, and 34.49%~230.95% on the NeurIPS2020 dataset, and 0.85%~10.25%, 1.38%~6.57%, and 6.26%~73.14% on the NeurIPS2020 dataset, and the value of the RMSE was reduced on the three datasets by 4.65%~12.48%, 8.42%~ 13.80%, and 3.42%~7.36%, respectively.

In addition, CS-BKT outperforms CS-BKT on the large-scale dataset NeurIPS2020 indicating that the model has the ability to adapt well to large-scale datasets. Meanwhile, CS-BKT has a lower enhancement rate on NeurIPS2020 compared to the other two datasets, which may be due to the short length of most of the learned sequences in NeUrIPS2020, which makes it difficult for the model to effectively learn the features in the sequences. Conversely, the CS-BKT model has the highest lift rate on the ASSISTmentChall dataset, which may be due to the fact that most of the learned sequences in this dataset are long in length, allowing the model to better capture the features on the learned sequences.

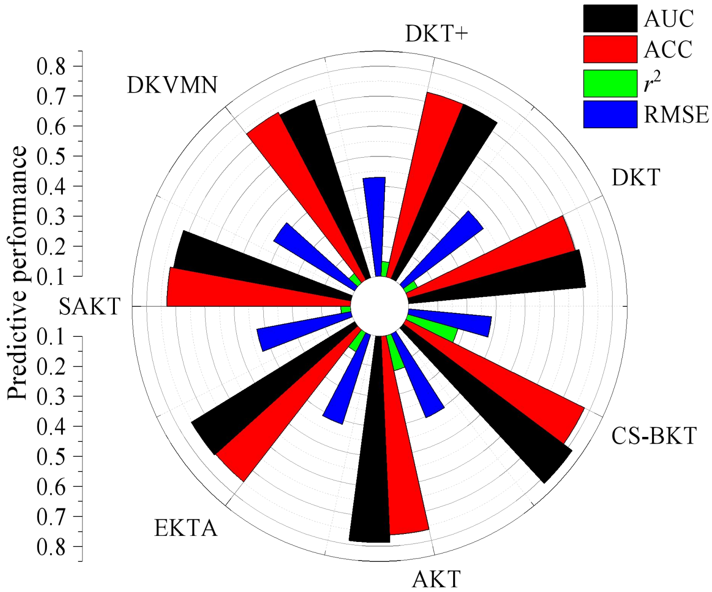

Figure 3. Predictive performance of the knowledge tracking models (ASSISTment2012).

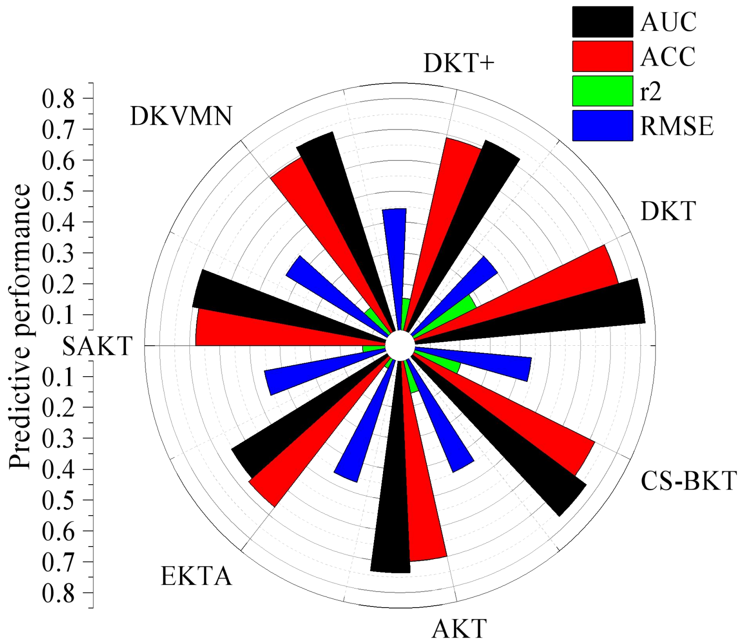

Figure 4. Predictive performance of the knowledge tracking models (ASSISTmentChall).

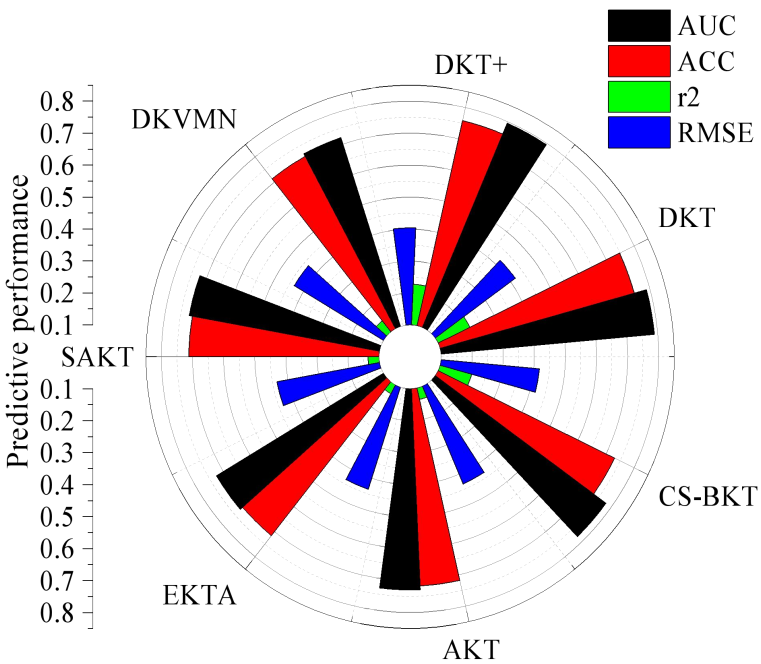

Figure 5. Predictive performance of the knowledge tracking models (NeurIPS 2020).

4. Design framework application and analysis of results

4.1. Framework Application Design

4.1.1. Application purpose

The CS-BKT-based knowledge tracking model is mainly used to continuously diagnose learners' knowledge status by dynamically modeling learners' continuous nonlinear question-answering in the process of English teaching and learning, and to provide accurate data basis for teachers on how to effectively carry out English lesson plan design. Therefore, the purpose of the practical application of the

CS-BKT model and the proposed context-aware English lesson plan co-design framework in teaching and learning is to test its effectiveness and rationality.

4.1.2. Application object

The study integrates learning experience, knowledge background and other considerations, and selects 110 undergraduate students in two classes of the same major of College English in the 2023-2024 academic year of a certain institution as the target of applying the CS-BKT knowledge tracking and context-aware English lesson plan co-design model in the actual teaching and learning process, and randomly identifies the two classes as an experimental group that uses the CS-BKT model and the AI co-design model and a control group that does not use the CS-BKT model and the model, of which there are 56 students in the experimental group and 54 in the control group. The two classes were randomly divided into an experimental group using the CS-BKT model and AI co-designed lesson plans and a control group not using the CS-BKT model and the AI co-designed lesson plans, with a total of 56 students in the experimental group and 54 students in the control group, which are very close in terms of the size of the number of students and the male-to-female ratio, and have a high degree of similarity.

To the maximum extent possible, the experimental group and the control group were identical in terms of course content, teaching faculty, overall and weekly teaching schedule, the main aspects of the teaching process, the use of teaching media and resources, the preparation of test questions and other teaching-related aspects, so as to eliminate as much as possible the impact of various irrelevant factors in the actual teaching and learning process on the final teaching and learning results.

4.1.3. Research tools

(1) Evaluation Questionnaire on the Accuracy of Knowledge Status Tracking Results. It mainly collects learners' evaluation data on the accuracy of knowledge status tracking results, and the evaluation questionnaire consists of five five-level score measurement items. The reliability analysis result of the evaluation questionnaire is Cronbach's α=0.837, and the KMO and Bartlett's sphericity test result is KMO value=0.841 and significant, indicating that the evaluation questionnaire has high reliability and validity.

(2) Teacher Satisfaction Questionnaire. The questionnaire was divided into five survey dimensions: friendliness of tracking method (Q1~Q5), practicality of tracking method (Q6~Q10), suitability of data mining (Q11~Q15), effectiveness of AI synergy (Q16~Q20), and reasonableness of lesson plan design (Q21~Q25), and each dimension consisted of five different scales ranging from “strongly disagree” to “strongly agree” to “strongly agree” to “strongly disagree”. Each dimension consists of five survey items ranging from “strongly disagree” to “strongly agree”, and each dimension has five questions, totaling 25 questions. At the same time, the Cronbach's alpha coefficient and structural validity KMO values of the overall and sub-dimensions of the teacher satisfaction questionnaire are all greater than 0.7, which indicates that the satisfaction questionnaire has high reliability and validity.

4.1.4. Application process

The whole teaching application process lasted for 9 weeks: the learners were pre-tested on their English knowledge level in week 1. From week 2 to 8, the instructor of the course carries out 7 weeks of teaching activities for the learners in the experimental group and the control group respectively. After each week's teaching activities, the two groups of students are tested on their knowledge level, and the students in the experimental group are required to fill in the Questionnaire for Evaluating the Accuracy of the Results of the Tracking of the State of Knowledge. In the 9th week, teachers need to fill in the Teacher Satisfaction Questionnaire. The effective recovery rate of both questionnaires was 100%.

4.2. Analysis of application results

4.2.1. Knowledge status tracking results

The accuracy evaluation of the knowledge state tracking results was conducted seven times during the whole teaching application process of the model, and the accuracy percentage of the knowledge state tracking results is shown in Figure 6. In the first evaluation of the knowledge state tracking results, low accuracy accounted for 23.74%, medium accuracy accounted for 28.22%, and high accuracy accounted for 48.04%, indicating that the knowledge state tracking results had good accuracy at the beginning of the model's English teaching application. In the following five accuracy evaluations, the proportion of low accuracy was 20.74%, 17.79%, 12.80%, 5.75%, 1.79%, showing a decreasing trend, and by the seventh accuracy evaluation the proportion of low accuracy was 0, and the situation of the proportion of medium accuracy also showed a decreasing trend on the whole. In contrast, the proportion of high accuracy in the last six accuracy evaluations was 55.73%, 57.92%, 64.29%, 76.51%, 85.58%, 91.59% in order, showing a rising trend step by step, and these changes indicate that the accuracy of the knowledge state tracking results has steadily risen with the gradual deepening of the pedagogical application of the CS-BKT model.

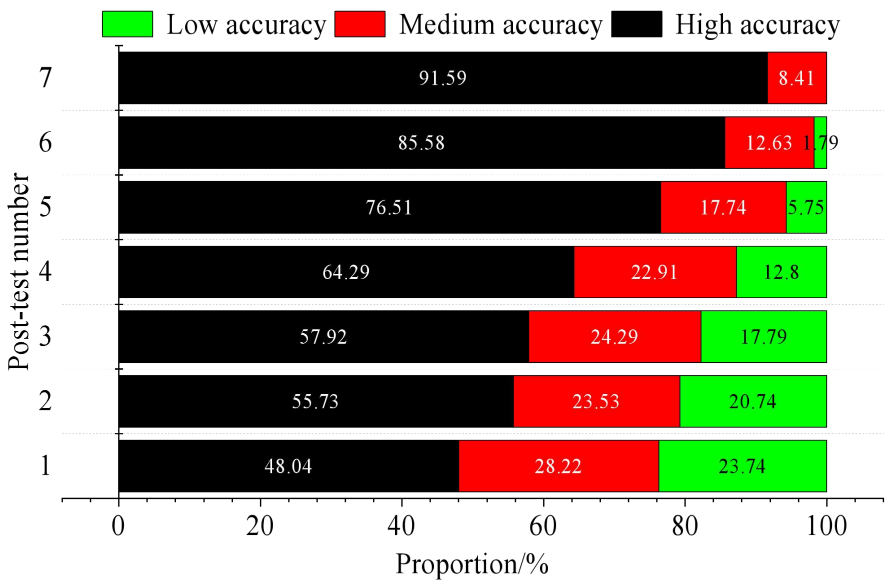

Figure 6. The accuracy of the knowledge status tracking results.

In addition, the answer sequences and corresponding knowledge state tracking results were randomly selected from the answer sequences of the learners in the experimental group for a period of time to be analyzed, and the answer sequences and knowledge state tracking results were visualized with the help of heat maps, and the dynamic changes of the knowledge state tracking results are shown in Figure 7. The horizontal axis represents a question-answer sequence of the learner, the "1" and "0" below the dots represent whether the answer is correct or not on the question, and the numbers represent the corresponding knowledge IDs of the question, and the vertical axis represents all the knowledge points involved in the question-answer sequence of the learner.

At the beginning of the sequence, the learner has only mastered knowledge point K1, and at the end of the sequence, the learner is not proficient in knowledge point K7, which indicates that as the learner continues to learn English and answer the corresponding test questions, the model is able to update the learner's mastery of the corresponding and related knowledge points of the test questions in a continuous manner.

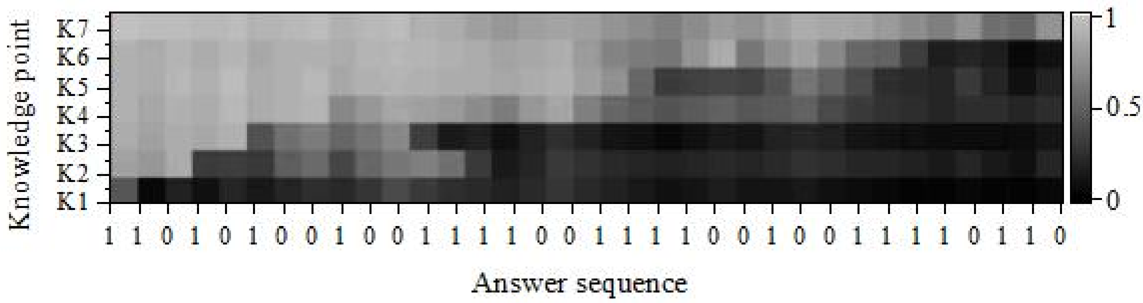

Figure 7. The dynamic change of the knowledge state tracking results.

4.2.2. Analysis of Teaching Effectiveness

The results of the analysis of the seven knowledge level posttest scores are shown in Figure 8. Starting from the 3rd knowledge level posttest, the knowledge level test of the experimental group students was significantly higher than that of the control group students (p-values of 0.035, 0.018, 0.011, 0.005, and 0.002, respectively, all less than 0.05), and these results indicate that the research proposes that the CS-BKT knowledge-tracking model and the AI synergistic lesson plan design can improve teaching effectiveness through the learners' dynamic and accurate diagnosis of knowledge status, assisting teachers in data-supported lesson plan design before class to enhance teaching effectiveness.

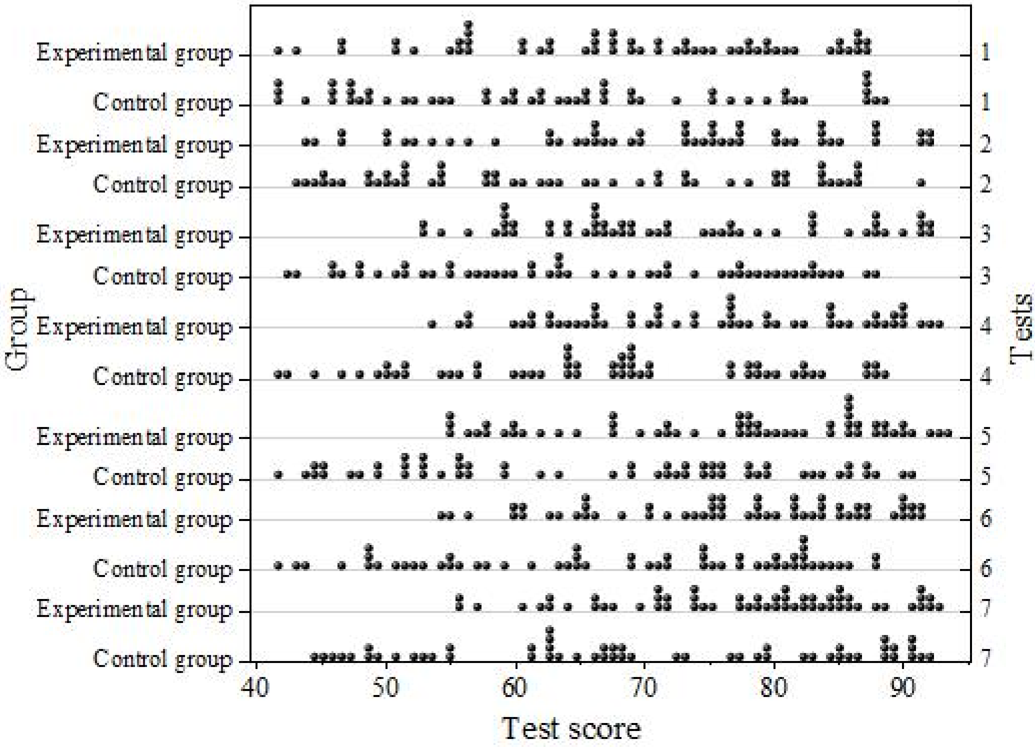

Figure 8. Results of seven knowledge proficiency tests.

4.2.3. Satisfaction analysis

The results of the teacher satisfaction questionnaire are shown in Figure 9, where A~E represent the increasing degree of “Strongly Disagree” to “Strongly Agree”, and ‘1’ and “0” represent choosing the option and not choosing the option, respectively. “1” and “0” represent choosing the option and not choosing the option. Teachers' choices in the five survey dimensions of friendliness of tracking method, practicality of tracking method, suitability of data mining, effectiveness of AI synergy, and reasonableness of lesson plan design are concentrated in general agreement, comparative agreement, and strong agreement, accounting for 16%, 60%, and 24%, respectively, which indicates that the collaborative design of context-aware English lesson plans and the CS-BKT knowledge tracking model in this paper have been recognized by the teachers. knowledge state tracking method and AI synergistic method have good practicality in English teaching lesson plan design.

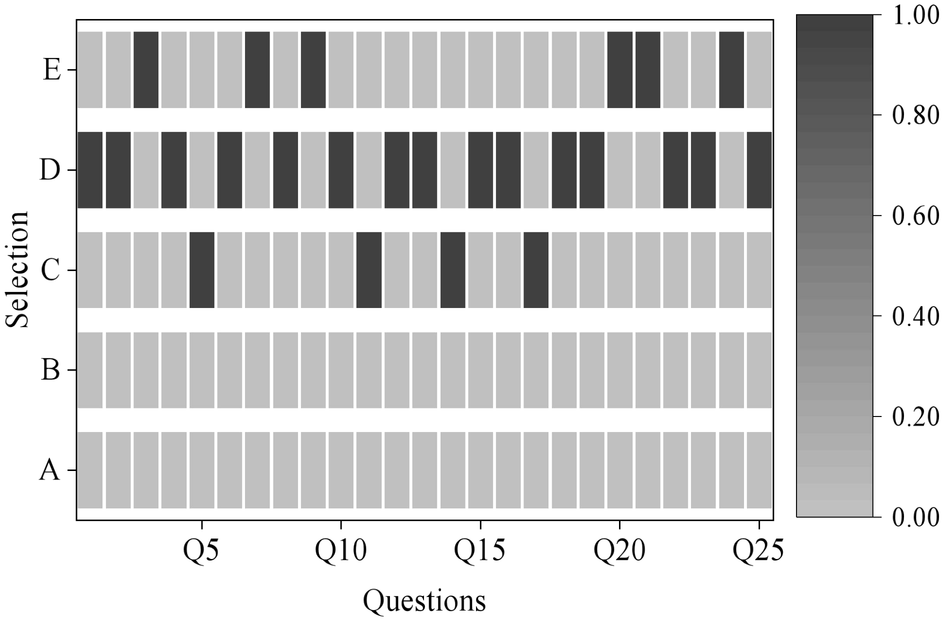

Figure 9. Questionnaire results of teacher satisfaction.

5. Conclusion

The development of intelligent technology has accelerated the deep integration of AI technology with education and teaching in various disciplines. In this paper, we study the collaborative English lesson plan design between teachers and AI, propose a context-aware English lesson plan collaborative design framework, and construct the CS-BKT knowledge tracking model. The experimental analysis finds that the model has good knowledge tracking performance, and the utility of the knowledge tracking model and the context-aware English lesson plan collaborative design framework is examined at the same time.

Compared with the comparison model, the AUC, Accuracy, and r^2 results of the CS-BKT model on the three datasets improved by 0.85% to 25.16%, 1.38% to 12.53%, and 6.26% to 230.95%, and the RMSE decreased by 3.42% to 13.80%. And the percentage of high accuracy is always the largest in the accuracy evaluation of students' knowledge status tracking, indicating the accuracy of CS-BKT model in predicting students' English knowledge status. After applying the CS-BKT model and the proposed framework in English teaching, the results of the latter five knowledge level tests of the students in the experimental group were significantly different from those of the students in the control group at the 5% level. At the same time, the percentage of "Comparatively agree" and "Strongly agree" in the teachers' satisfaction survey reaches 84%, and the teachers very much agree that the proposed model and framework are friendly in tracking method, practical in tracking method, suitable in data mining, effective in AI collaboration, and reasonable in the design of lesson plans. Rationality.

This paper constructs a context-aware English lesson plan collaborative design framework based on the educational context-awareness model of "human-machine-object-environment-activity", which is divided into three levels: intelligent perception layer, intelligent analysis layer, and intelligent service layer. It is divided into three layers: intelligent perception layer, intelligent analysis layer, and intelligent service layer. The practice framework integrates AI technology and context awareness to give the idea of collaborative design of English lesson plans. Its combination with the knowledge tracking model promotes collaboration and mutual assistance between teachers and AI, promotes the efficiency and intelligence of collaborative design of English lesson plans, and is conducive to the development of English teaching quality.

Funding

This research was supported by the Ministry of Education Project for Industry-Academia Cooperation and Joint Education and Training: "Research on the Reform of ESP-based Teaching Model for Civil Engineering English Teaching under the Background of New Liberal Arts" (231107623234314).

References

1. Qian, L., Cao, W., & Chen, L. (2025). Influence of artificial intelligence on higher education reform and talent cultivation in the digital intelligence era. Scientific Reports, 15(1), 6047.
2. Tiwari, R. (2023). The integration of AI and machine learning in education and its potential to personalize and improve student learning experiences. International Journal of Scientific Research in Engineering and Management, 7(02).
3. Mariam, G., Adil, L., & Zakaria, B. (2024). The integration of artificial intelligence (ai) into education systems and its impact on the governance of higher education institutions. International Journal of Professional Business Review: Int. J. Prof. Bus. Rev., 9(12), 13.
4. Loor, M. A. M., Solorzano, D. M. A., Katherine, A., & Moreira, V. (2024). Integration of artificial intelligence in english teaching. Journal of Cleaner Producfion, 289, 125834.
5. Fattah, H. A., Vadivel, B., Shaban, A. A., & Shanmugam, K. (2023). Enhancing english language education: the impact of ai integration in the classroom. Journal of Humanities and Education Development, 5(6), 116-123.
6. Liu, Y., & Ren, L. (2022). The influence of artificial intelligence technology on teaching under the threshold of "Internet+": based on the application example of an English education platform. Wireless Communications and Mobile Computing, 2022(1), 5728569.
7. Kohnke, L., & Zou, D. (2025). The role of ChatGPT in enhancing English teaching: A paradigm shift in lesson planning and instructional practices. Educational Technology & Society, 28(3), 4-20.
8. Sun, L. (2025). Enhancing intercultural competence of Chinese English majors through AI-enabled Collaborative Online International Learning (COIL) in the digital era. Education and Information Technologies, 30(6), 7995-8027.
9. Yang, H., Zhao, Y., Wu, Y., Wang, S., Zheng, T., Zhang, H., ... & Qin, B. (2025). Large language models meet text-centric multimodal sentiment analysis: A survey. Science China Information Sciences, 68(10), 1-29.
10. Yamazaki, T., Yoshikawa, K., Kawamoto, T., Mizumoto, T., Ohagi, M., & Sato, T. (2023). Building a hospitable and reliable dialogue system for android robots: a scenario-based approach with large language models. Advanced Robotics, 37(21), 1364-1381.
11. Bonner, E., Lege, R., & Frazier, E. (2023). Large Language Model-Based Artificial Intelligence in the Language Classroom: Practical Ideas for Teaching. Teaching English with Technology, 23(1), 23-41.
12. Nor, H., Hizriani, N., & Kadariyah, N. L. (2022). Teachers' materials development, lesson planning, and teaching skills readiness in teaching English. PANYONARA: Journal of English Education, 4(1), 32-64.
13. Emiliasari, R. N. (2019). Lesson planning in EFL classroom: A case study in lesson plan preparation and implementation. Wiralodra English Journal (WEJ), 3(2), 367-375.
14. Sudipa, I. N., Aryati, K. F., Susanta, I. P. A. E., & Anggayana, I. W. A. (2020). The development of syllabus and lesson plan based on English for occupational purposes. International Journal of Psychosocial Rehabilitation, 24(4), 290-300.
15. Saputri, M. (2017). English teachers' difficulties in designing lesson plan based on 2013 curriculum. The State Islamic Institue of Surakarta.
16. Darma, S. D., & Suwarno, B. (2018). ENGLISH TEACHERS'DIFFICULTIES IN DESIGNING A LESSON PLAN (RPP) BASED ON KTSP (A STUDY ON ENGLISH TEACHERS AT VOCATIONAL HIGH SCHOOLS IN BENGKULU CITY). JOALL (Journal of Applied Linguistics and Literature), 2, 72-82.
17. Azizova, D. (2024). LESSON PLAN FOR GENERAL ENGLISH AND SUMMARY DESCRIPTION. Nordic_Press, 3(0003).
18. Xiaoju, Z., & Sy, R. A. G. (2025). CONTEXTUALIZING ENGLISH LANGUAGE LEARNING: DESIGNING ENGLISH LESSON FOR STUDENTS MOTIVATION. International Journal of Economics, Business and Social Science Research, 3(2), 48-101.
19. Apriani, E., Supardan, D., & Umami, M. (2020, November). Independent learning: English teachers' problems in designing a good lesson plan in new normal era at man rejang lebong. In International Conference on the Teaching English and Literature (Vol. 1, No. 1, pp. 72-78).
20. Zhao, J. (2025). Advancing english language education: A mixed-methods analysis of AI-driven tools' impact on engagement, personalization, and learning outcomes. Education and Information Technologies, 1-41.
21. Gokturk-Saglam, A. L. (2025). ChatGPT in the classroom: pre-service English language teachers' perspectives on AI integration in language assessment. Learning: Research and Practice, 1-19.
22. Kaw, A., Yalcin, A., Clark, R. M., Gomes, R. B., Serrano, L., Scott, A., & Lou, Y. (2024). On Building and Implementing Adaptive Learning Platform Lessons for Pre-Class Learning in a Flipped Course. ASEE Computers in Education, 14(2), 1-23.
23. Liu, J. (2024). Enhancing English language education through big data analytics and generative AI. Journal of Web Engineering, 23(2), 227-249.
24. Williyan, A., Fitriati, S. W., Pratama, H., & Sakhiyya, Z. (2024). AI as Co-Creator: Exploring Indonesian EFL Teachers' Collaboration with AI in Content Development. Teaching English with Technology, 24(2), 5-21.
25. Murshid, V. H., & Peter, S. (2025). Contextualizing a Global English Textbook with ChatGPT to Enhance ESL Teaching and Learning in an Indian Classroom. AsiaCALL Online Journal, 16(1), 282-294.
26. Liu, Y. F., Hwang, W. Y., & Su, C. H. (2024). Investigating the impact of context-awareness smart learning mechanism on EFL conversation learning. Interactive Learning Environments, 32(8), 4122-4137.
27. Hwang, W. Y., Nurtantyana, R., Purba, S. W. D., Hariyanti, U., Indrihapsari, Y., & Surjono, H. D. (2023). AI and recognition technologies to facilitate English as foreign language writing for supporting personalization and contextualization in authentic contexts. Journal of Educational Computing Research, 61(5), 1008-1035.

28. Han, J., & Li, M. (2024). Exploring ChatGPT-supported teacher feedback in the EFL context. System, 126, 103502.
29. Yue, Y. (2025). The Evolution of Foreign Language Teachers' Roles in the Age of AI. International Journal of Social Science, Education and Humanities, 1(2), 45-53.
30. Mohamadi, L., Asadi, M., & Taheri, R. (2024). Transforming EFL Lesson Planning with ‘To Teach AI’: Insights from Teachers’ Perspectives. Technology, 2(4).
31. Kerr, R. C., & Kim, H. (2025). From Prompts to Plans: A Case Study of Pre-Service EFL Teachers' Use of Generative AI for Lesson Planning. English Teaching, 80(1), 95-118.
32. Willner, L. S. (2025). AI-Powered Instructional Planning for Integrated Content, Literacy, and English Language Development for K-12 Multilingual Learners. GATESOL Journal, 34(1), 17-34.
33. Zhussupova, R., Nurzhanova, A., Beisembayeva, Z., Baibatyrova, A., Shadiev, R., Nurzhanova, Z., ... & Alipbay, B. (2025, July). Lesson Plan Designing in the Digital Age: A Multimodal Approach with ChatGPT. In International Conference on Innovative Technologies and Learning (pp. 74-83). Cham: Springer Nature Switzerland.
34. Yong, D. I. N. G. (2025). New Pathways for Teacher Professional Development: A Case Study of Pre-Service Teachers Using AI for Lesson Planning and Reflection. Artificial Intelligence Education Studies, 1(2), 63-78.
35. Setyaningsih, E., Asrori, M., Zainnuri, H., & Hariyanti, Y. (2024). Exploring high school EFL teachers' experiences with magic school AI in lesson planning: Benefits and insights. Voices of English Language Education Society, 8(3).