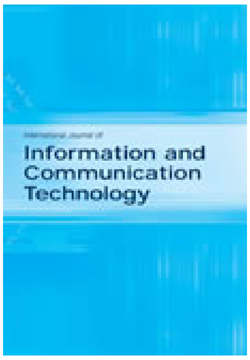


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

The evaluation method of English teaching quality incorporating students' cognitive transfer

Yanting Xu

DOI: 10.1504/IJICT.2025.10072281

Article History:

Received:	20 May 2025
Last revised:	07 June 2025
Accepted:	08 June 2025
Published online:	25 July 2025

The evaluation method of English teaching quality incorporating students' cognitive transfer

Yanting Xu

College of General Education,
Zhangzhou Health Vocational College,
Zhangzhou 363000, China
Email: 17859778852@163.com

Abstract: Current English teaching quality evaluation methods ignore the influence of students' cognitive level on the evaluation results. To this end, this paper first establishes a diagnostic model of students' English cognition, dynamically assigns students to subgroups with similar abilities through K-means clustering, and computes the mean capability attribute value for all groups. Subsequently, the hidden memory information is extracted from the evaluation text, and the cognitive ability vector is extracted from the English practice records. Students' cognitive abilities are transferred to the evaluation text memory information to obtain the cognitive transfer matrix. The attention scheme is employed to integrate the text memory information and cognitive ability representations to categorise the evaluation texts. Simulation results indicate that the suggested approach has an AUC of 0.9725, which is better than the comparison approach and can achieve more accurate English quality evaluation.

Keywords: English teaching evaluation; K-means clustering; cognitive diagnosis; cognitive transfer; attention mechanism.

Reference to this paper should be made as follows: Xu, Y. (2025) 'The evaluation method of English teaching quality incorporating students' cognitive transfer', *Int. J. Information and Communication Technology*, Vol. 26, No. 28, pp.17–32.

Biographical notes: Yanting Xu received her Master's degree at the Guangxi Normal University in 2012. She is currently a Lecturer in the College of General Education at Zhangzhou Health Vocational College. Her research interests include English literature, English teaching research and cluster analysis.

1 Introduction

The traditional assessment of the standard of English instruction is mostly based on quantitative indicators such as test scores and classroom performance, which ignores students' ability to transfer English knowledge and skills from the classroom to practical application scenarios (Sun et al., 2017). Cognitive transfer, as a core concept in educational psychology, emphasises students' ability to flexibly apply knowledge and learn by example in the learning process, and is an important dimension in measuring teaching effectiveness (Donaldson and Mavrogordato, 2018). In English teaching, whether students can transfer the grammar rules and vocabulary expressions learned in the classroom to real contexts such as cross-cultural communication and academic writing directly reflects the degree of achievement of teaching goals (Susanty et al., 2021). However, the current evaluation method for the quality of English intelligent instruction has not yet fully integrated the consideration of students' cognitive transfer ability, and cannot effectively play the role of evaluation as a guide and feedback for teaching (Qian, 2022). Therefore, exploring the quality evaluation method of English intelligent teaching that incorporates students' cognitive transfer can not only fill the gaps in the existing evaluation system and provide a scientific assessment basis for the practice of English intelligent teaching, but also help students to improve their English learning ability in a comprehensive way.

Early methods of assessing English instruction were to diagnose students' learning and thus obtain evaluation results. Zhang et al. (2022) used the online analytical processing (OLAP) model to study the domain of English learning, and interpreted subjects' English learning paths through quiz diagnosis to give feedback on the current mastery of students' knowledge attributes, thus evaluating the quality of teaching. Zhang (2021) used linear regression modelling to diagnose the attributes of liberal arts students in the English language curriculum to test and diagnose their mastery patterns, and based on the diagnostic results, help students to clarify their learning deficiencies, so that teachers can make targeted adjustments to the teaching programme. Ma et al. (2023a) used hierarchical regression analysis to classify the hierarchical attributes, and based on the feedback of students' answers to the test papers, a comprehensive analysis was carried out to evaluate the results. Kou et al. (2023) used the vertical equivalence method of cognitive diagnosis to analyse the teaching of English listening to provide a reference for the development of overall remedial teaching programmes.

Traditional English teaching quality evaluation is mostly led by teachers, and students are in a passive position to receive evaluation. In the evaluation of classroom performance, teachers score students according to their classroom speeches, homework completion, etc., and students seldom have the opportunity to participate in the evaluation process. This single evaluation subject makes the evaluation results may be influenced by teachers' personal subjective factors, and lacks comprehensiveness and objectivity. Due to the uniformity of evaluation standards, it is difficult for teachers to formulate individualised teaching plans and evaluation schemes according to the different situations of students, which is not conducive to teaching students according to their aptitude and meeting the learning needs of different students. The teaching quality evaluation method based on text analysis embeds the unstructured text data into structured vectors through a language model, and then inputs the embedded vectors into a machine learning model to capture the features, so as to obtain the classification results of the evaluation, which improves the objectivity of the evaluation. Li (2024) used different machine learning models such as simple Bayes (NB) (Qi et al., 2022) and logistic regression (LR) (Simonetti et al., 2017) to experimentally compare on student learning assessment data, and obtained a more satisfactory evaluation classification effect. Chen et al. (2021) offered a teaching evaluation approach for smart education and used it to mine textual data for teaching evaluation, and research's outcome implied that students' individual cognitive ability affects the results of teaching evaluation. Hou (2021) used a hybrid method integrating dictionary and decision tree (DT) and support vector machine (SVM) approaches to predict results of teaching assessment and improve the accuracy of teaching evaluation.

However, teaching evaluation methods based on traditional machine learning models mostly rely on a priori knowledge of feature engineering. For complex teaching evaluation texts, the modelling ability is weak, and the accuracy of classification and evaluation is low. Research based on deep learning can learn the complex features and laws in text data and accurately recognise the classification results of text. Ge et al. (2021) proposed a teaching quality evaluation method using an improved bidirectional long and short-term memory network (BiLSTM) combined with Word2Vec, with an evaluation accuracy of 82.04%. Li (2023) combined convolutional neural network (CNN) to extract features and BiLSTM to extract sequences as well as attention mechanism to perceive the advantages of contextual information, and obtained better evaluation results. Most of the above studies relied on text mining to obtain instructional evaluation results; however, students' cognitive abilities can also have a significant impact on evaluation results. Zhai et al. (2023) diagnosed students' cognitive abilities based on the self-coder structure and integrated the cognitive state in the teaching evaluation model for modelling, which improved the evaluation accuracy. Wu et al. (2023) used gated recurrent units (GRUs) to achieve in-depth modelling of students' cognitive migration state and constructed a teaching quality evaluation model based on the transformer model, with an evaluation accuracy rate of 87.96%.

Accurate evaluation of the quality of English teaching can help teachers improve their own teaching ability, but the existing research has neglected the students' implicit knowledge status and cognitive level affecting their evaluation of English courses, and the assessment's analysis of English instruction incorporating students' cognitive status provides ideas for coping with the above issues. For this reason, this study constructs an approach for assessing the quality of English intelligent instruction that incorporates cognitive transfer, taking into account the characteristics of students' cognitive abilities in English courses. Firstly, the model gains information about students' long-term studying interactions, breaks down this interaction sequence into parts, and designs student competency attributes in light of customised characteristics to obtain individual competency changes during the learning process. By utilising the K-means clustering approach, students are adaptively allocated to subgroups comprising peers with comparable abilities, thereby establishing multiple student subgroups that align closely with the proficiency levels of individual student. Meanwhile, working out the group-specific average for the ability attribute among students, so as to complete the modelling of students' cognitive ability in English. Hidden memory information was then extracted from students' English course evaluation texts and cognitive ability vectors were extracted from historical practice records. The cognitive transfer matrix (CTM) is obtained by transferring students' cognitive abilities to the evaluated text memory information. The attention mechanism is introduced to enhance the original text memory information representation and cognitive ability representation in the CTM. Finally, the two enhanced representations are fused as text features to classify the evaluation text. The experimental outcome implies that the evaluation accuracy and AUC of the proposed method are 91.02% and 0.9725, respectively, which are significantly improved compared with the baseline method, and can provide an objective basis for teachers to improve their English teaching methods.

2 Relevant technologies

2.1 Cognitive diagnosis theory

Cognitive diagnosis, as the core of the new generation of test theories, is supported by cognitive psychology, measurement and information technology through the design of tests to assess students' internal cognitive structures and mental processes that are not directly observable through quantifiable tests (Heller et al., 2015). Unlike standardised test theories, cognitive diagnosis is no longer limited to the evaluation of the individual at the macro level, but goes deeper into the micro responses of the subjects (Paulsen and Valdivia, 2022). Cognitive diagnosis can be broadly or narrowly defined as the analysis of the correlation between an individual's intrinsic cognitive characteristics and test scores. Cognitive diagnosis in the narrower sense involves the use of tests to determine students' knowledge or skills to categorise them according to their level of mastery of the knowledge or skills being tested. Determining student knowledge and skills follows a closed loop of decomposition, measurement, analysis and intervention. Deconstruct subject matter content into observable knowledge/skill attributes. Collect multidimensional data through standardised tests and qualitative assessments. Use statistical modelling and dynamic tracking to identify strengths and weaknesses. Based on the diagnostic results, design personalised teaching plans, and form an iterative assessment-teaching optimisation.

Individuals' internal mental processing cannot be directly observed, but can only be indirectly assessed through responses to test items (Ma et al., 2023b). For the goal of realising a more accurate measurement and assessment of the internal mental processing of the subjects, experts in the fields of cognitive psychology, psychometrics, modern statistical mathematics and computer science have fully absorbed the results of psychological research and incorporated them into statistical models to develop measurement models with cognitive diagnostic functions. It is essentially a mathematical-statistical model used to provide diagnostic information. The process of cognitive diagnosis is complex. In the whole cognitive diagnostic process, the design of diagnostic tests determines the accuracy of cognitive diagnostic results. The development of a cognitive diagnostic test can be summarised as the determination of diagnostic goals and the construction of a cognitive model:

1 Defining diagnostic objectives. The objectives of a cognitive diagnostic test include identifying the test subject and a unit or point of knowledge in a specific discipline.

2 Cognitive modelling analyses the mental processes of individuals in the process of solving specific problems, and builds a cognitive processing model to determine the knowledge structure and cognitive processing skills designed for diagnostic goals.

2.2 Attention mechanism

People get the information they need quickly from an image or text because when they look at an image or text that is rich in information, they focus on certain key areas rather than on every detail. In this way, by allocating attention, unimportant information can be ignored and effective information can be found in a short period of time (Brauwers and Frasincar, 2021). This feature has led to the creation of the attention mechanism, which has been widely used in many fields.

In the field of deep learning, as the amount of data continues to grow, effectively extracting data that is helpful in training the model will result in an increase in the performance of our model. Researchers have introduced the attention mechanism into deep learning, which allows a model to process an input by focusing only on the information that is most similar to that input and ignoring other information (Shi et al., 2020). This not only improves the predictive performance of the model, but also provides a degree of interpretability.

The attention mechanism can generally be computed in two steps; the first step computes the attentional weights of all inputs, and the second step weights the obtained attentional weights with the input information for weighted summation (Niu et al., 2021). The key is adopted to compute the attention distribution and the value is adopted for the final weighted summation. First, the similarity or other similarity metrics between the query vector Q and the key vector K are calculated, and then the similarity is normalised to obtain an attention weight value belonging to the range between 0 and 1. Finally, this normalised weight is weighted and summed with the value V to obtain the final information representation.

$$a_n = \frac{\exp\left(s\left(k_n, q\right)\right)}{\sum_{n=1}^{N} \exp\left(s\left(k_n, q\right)\right)} \tag{1}$$

$$attention(Q, K, V) = \sum_{n=1}^{N} a_n v_n \tag{2}$$

where s denotes the scoring operation, k_n and k_v denote the values of key K and value V, respectively.

3 Diagnostic modelling of English cognition based on individual student differences

3.1 Designing a framework for modelling English cognitive ability

Although most of the current research has been effective in modelling students' English learning processes and predicting their future performance, many of these studies do not take into account the impact of individual differences in students' learning processes, nor do they take into account individual differences in ability. However, in the actual process of English language learning, students' English language learning abilities are constantly changing due to previous learning experiences and other factors. To cope with this issue, this paper proposes the individual difference-based cognitive ability diagnostic (IDS-CAD) model based on the attributes of students' abilities. Firstly, IDS-CAD gains the long-term studying interactions of students at earlier time, and then divides this interaction sequence into segments. In light of the customised characteristics, for example students' performance outcomes and answering durations, IDS-CAD constructs students' competence attributes to obtain the changes of individual competence during the learning process. Subsequently, by adopting the K-means clustering approach (El Khattabi et al., 2024), Students are adaptively allocated into subgroups comprising peers of comparable proficiency levels, tailored to their unique attributes, creating multiple student subgroups with comparable proficiency levels aligned to unique capabilities. The mean ability attribute value of each group is calculated to complete the modelling of students' English cognitive ability.

Constructing models to represent students' cognitive aptitude in English language acquisition, the student ensemble is defined as *S*, the set of English exercises as *E*, and the set of knowledge points (KPs) as *K* for each topic. Assuming that each student is working independently in an English learning scenario, the student's response history is documented as $X_t = [(e_1, r_1), (e_2, r_2), \ldots, (e_t, r_t)]$ in the assessment records, in which question $e_t \in E$ corresponds to the item answered by the student at timestamp *T*, r_t stands for the related answer outcome, $r_t = 1$ stands for the answer is correct, $r_t = 0$ stands for the answer is incorrect. Model the knowledge space representation of *K* topics as a matrix $M_K(d_k \times |K|$, encoding interdependencies among KPs. The d_k-dimensional column vector is defined as the latent representation of an individual KP in the embedding space.

3.2 Modelling student cognitive ability attributes

During routine English language instruction, the value of students' proficiency attributes can reflect to some extent the students' proficiency in the knowledge they have internalised. Given that the student-generated data throughout the learning journey constitutes a temporal sequence, the trajectory of each learner's personalised capability development can be derived from the diagnostic information present in their digital interaction history. This study leverages this foundation to formulate the dimensional constructs of students' capabilities. For ease of computation, the relative efficiency score, computed as the ratio of a student's correct-response latency to the population mean latency, is operationalised as the student's ability indicator, as indicated in equation (3) and equation (4).

$$C\left(e_j\right)_{1:g} = \begin{cases} \dfrac{t_j}{A_{ij}}, e_j == 1 \\ 0, \quad e_j == 0 \end{cases} \tag{3}$$

where $C(e_j)_{1:g}$ stands for the variation in a student's proficiency related to question *j* across a temporal segment of duration *g*, t_j denotes the response time of a student correctly answering question *j*, A_{ij} stands for the mean latency of all students who accurately responded to question *j*, which incorporates knowledge component *i*, $e_j == 1$ stands for a learner provided an accurate response, and $e_j == 0$ stands for a learner provided wrong response, and that knowledge component *i* includes *M* correctly answered marks, with the j^{th} response time t_{ij}^{skill}.

The ability attribute scores of students at the i^{th} knowledge component are vectorised for temporal clustering across interval *g*, yielding $r_{1:g}^{i} = (C(x_1)_{1:g}, C(x_2)_{1:g}, \ldots, C(x_n)_{1:g})$, where $r_{1:g}^{i}$ stands for a capability vector characterising performance within temporal window 1:*g*, *i* is the amount of knowledge component. Correctly answering question *j* results in an ability enhancement of $C(x_j)_{1:g}$ for the student at KP *i*.

Following the acquisition of learners' capability attributes, the K-means approach is utilised to partition learners into clusters in terms of their capability metrics, i.e., the whole student cohort is adaptively allocated to subgroups exhibiting reduced inter-individual variability, while considering the diverse learning aptitudes of individual learners. Before initiating the clustering process, the optimal number of clusters was first determined, upon completion of an experimental trial with a range of K parameters.

In the clustering training process, the core position of every student cluster is first identified, independent of the time-based partitioning criteria, and once the centroid is established, it remains fixed throughout the entire clustering procedure. Subsequently, learners are adaptively allocated to distinct groups at every occurrence of time interval T_g. The clustering methodology is formally defined in equation (4).

$$Clu\left(S_a, T_g\right) = \arg\min \sum_{k=1}^{|K|} \sum_{r_{1:g-1}^{i} \in K} \left\| r_{1:g-1}^{i} - \mu_k \right\|^2 \tag{4}$$

where S_a denotes a student a, T_g stands for the time interval, μ_k denotes the focal point of student subgroup, and $r_{1:g-1}^{i}$ stands for a student's capability attributes over a temporal sequence $1{:}g - 1$.

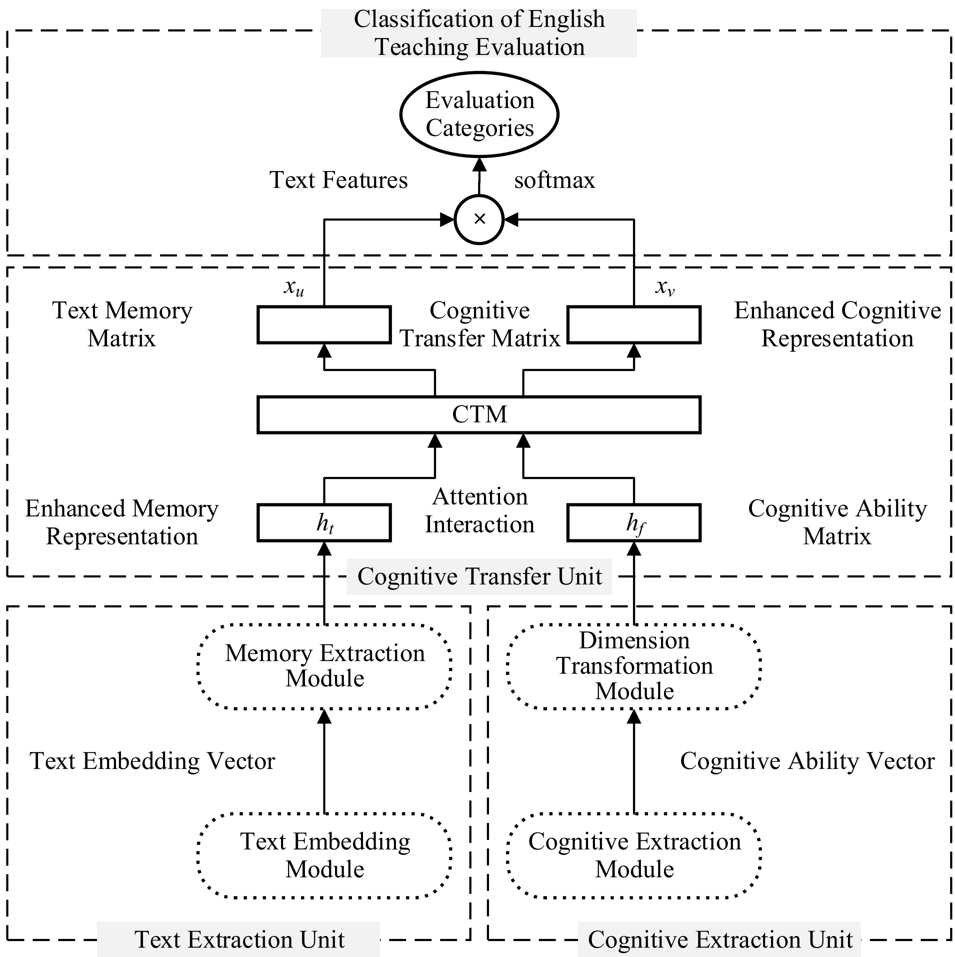
4 A method for assessing the quality of English instruction intelligently by incorporating students' cognitive transfer

4.1 Text embedding and feature extraction for English teaching evaluation

The current English intelligent teaching quality evaluation method ignores the influence of students' implicit knowledge state and cognitive level on the evaluation results, for the goal of coping with the above issues, this paper designs an English intelligent teaching quality evaluation method that incorporates students' cognitive transfer, as indicated in Figure 1. First of all, input the text data of English instruction assessment and historical practice records. The text data of English instruction assessment is converted into text embedding vectors through the text embedding module. The text embedding vector is sent to the memory retrieval module to output the memory matrix of the text. The text embedding vector is converted to a text embedding vector. The text embedding vectors are fed into the memory extraction module to output the text memory matrix. The historical practice records are extracted from the cognitive ability vector by the cognitive extraction module, and then fed into the dimension transformation module to obtain a cognitive ability matrix with the same dimensions as the text memory matrix. The text memory matrix and the cognitive ability matrix were fully connected and interacted to output the CTM. For the purpose of enhancing the accuracy of instructional evaluation, an attention mechanism was designed in the CTM to enhance the representation of textual memory information and cognitive ability. The two enhanced representations of the cognitive transfer unit are fused as the final text features, which are fed into the softmax classifier to classify the English teaching evaluation, and the ultimate outcome comprises the categorised results of teaching assessment.

The embedding of English teaching evaluation texts mainly vectorises unstructured English teaching evaluation texts. Here, a pre-trained BERT language model (Lee et al., 2023) is used for word embedding, which integrates the structure and ideas of common language models such as Word2Vec, Elmo, GPT, etc., and performs well in various natural language processing tasks. Since BERT uses a bi-directional transformer model with a self-attention mechanism for pre-training, it can better capture the linguistic features in the text of learning assessment. Given that the text's length is denoted by l, the word vector obtained by embedding is denoted as $v = \{v_1, v_2, \ldots, v_l\}$, where $v_i \in R^d$ represents the word vector mapped by each word in the English teaching evaluation text and d represents the dimension of the word embedding.

Figure 1 The English teaching quality evaluation model integrating students' cognitive transfer



After the embedding representation of the English teaching evaluation text, the bidirectional gated recurrent unit (BiGRU) is adopted to capture the semantic long-term dependency features in the text. BiGRU allows the model to take into account the contextual information of each word in order to better understand the meaning of the text. Through the left-to-right iterative GRU-R and right-to-left iterative GRU-L architectures, the memory matrix in the output structural chemistry teaching evaluation text vector can be expressed as $h_t = (h_t^1, h_t^2, \ldots, h_t^l)$, where $h_t \in R^{l\times 2d}$ and $h_t^i = [\vec{h}_t^i, \overleftarrow{h}_t^i]$ represent the combination of hidden states in two directions at position i.

The GRU structure is a gated RNN designed to effectively capture long-term dependencies in text sequences. It consists of a reset gate r_t and an update gate z_t, which are used to control how much information h_{t-1}^i needs to be retained in the hidden state memory of the previous word, and how much information needs to be forgotten and how much new information $\tilde{h}_t^i$ needs to be added to the hidden state information of the current word, respectively, where v_t represents the word embedding vector at position t. In addition, σ and tanh represent two types of activation functions to better adapt to and fit complex data patterns and relationships.

4.2 Student cognitive extraction module

For the goal of integrating students' cognitive states into the English teaching evaluation model, students' historical practice records are used to extract students' cognitive ability vectors, which reflect students' proficiency in course knowledge areas. The research adopts the IDS-CAD model to extract the cognitive vector c' in students' practice of history English exercises. The model takes as input the students' historical practice data $o_e \in R^E$ and the students' personal information $o_s \in R^n$, and obtains the historical practice factor matrix $p_e \in R^K$, the difficulty of the KPs $k_h \in R^{E\times K}$, the differentiation degree of the practice $d_e \in R^E$, and the student factor matrix $f_s \in R^K$ from the trainable matrix. The matrix of history practice factors, the difficulty of the KPs, the differentiation of the exercises, and the matrix of student factors are effectively modelled by an interaction function to infer students' cognitive profiles in history practice c'. The interaction function can be expressed as equation (5).

$$c' = p_e \circ (f_s - k_h) \times d_e \tag{5}$$

In the interaction function, n represents the total number of students, E denotes the aggregate count of exercises, and K represents the total amount of points included in the exercises. The output c' of the interaction function is iterated through the fully linked level to achieve the final output of the cognitive ability vector cs = $\{\varphi_1, \varphi_2, \varphi_3, \ldots, \varphi_k\}$ of the student group, where φ_k represents the learners' mastery of the course knowledge. At the end of the module, the cognitive ability vector c_s is updated.

When transferring the student cognitive ability vector c_s to the text memory matrix h_t, the cognitive ability matrix h_f is obtained by nonlinearly transforming c_s to obtain a cognitive ability matrix with the same dimensions as the text memory matrix. Here, the nonlinear transformation function tanh is chosen to dimensionally transform the progress of c_s, and finally the cognitive ability matrix $h_f \in R^{n\times 2d}$ is obtained.

4.3 Student cognitive transfer module

Students' cognitive ability of the knowledge of English teaching courses affects their evaluation of the courses. In order to better capture the characteristics of students' teaching evaluation texts, a cognitive transfer unit is designed to transfer students' cognitive vectors of historical exercises to the evaluation of English teaching texts. The cognitive transfer unit helps the model to better understand the relationship between students' cognitive abilities and text memorisation, and thus to capture the high-level semantics of the relationship. The cognitive transfer unit is mainly divided into two steps.

Firstly, the text memory matrix h_t and the cognitive ability matrix h_f interact through the cognitive transfer unit and output the *CTM*. Second, for the purpose of gaining more discriminative features, the text memory matrix and the cognitive ability matrix were enhanced using the attention mechanism. Specifically, the text memory matrix and cognitive ability matrix are fed into a neural network for full connectivity computation to realise the fusion of features, and then activated by an activation function to generate a comprehensive *CTM*. The construction equation is as follows, where σ is the activation operation, b_{fm} is the bias, and W_{CTM} is the weight matrix.

$$CTM = \sigma\left(h_t \times W_{CTM} \times \left(h_f\right) + b_{fm}\right) \tag{6}$$

For the purpose of enhancing the accuracy and validity of the model, the attention mechanism was introduced into the fused CTM to enhance the features of the text memory matrix and the cognitive ability matrix, as shown in equation (7) and equation (8).

$$\alpha_u^i = \frac{\exp\left(\sum_{i=1}^{n} \alpha_{ij}\right)}{\sum_{n=1}^{l} \exp\left(\sum_{j=1}^{n} \alpha_{nj}\right)} \tag{7}$$

$$\alpha_v^j = \frac{\exp\left(\sum_{i=1}^{i} \alpha_{ij}\right)}{\sum_{n=1}^{n} \exp\left(\sum_{i=1}^{i} \alpha_{in}\right)} \tag{8}$$

where α_{ij} represents an element at a certain position in the CTM, the attention scores for text memory and cognitive ability were calculated by totalling each row and column in the CTM. α_u^i reflects the importance score of text memory in the i^{th} row of the fused *CTM*, while α_v^j represents the importance score of cognitive ability in the j^{th} column. By assigning different attentional weights to each element of the CTM, the model makes it easier to capture important emotional features. After assigning attentional weights to the text memory matrix h_t and cognitive ability matrix h_f, the enhanced memory representation x_u and enhanced cognitive representation x_v were computed using dot product and summation.

4.4 Classification of English smart instruction quality evaluation results

The purpose of classifying the outcomes of English language teaching quality assessments is to leverage memory and cognitive representations, which are optimised by the CTM, as textual descriptors, and to classify these features using classifiers to obtain the evaluation results. First, the augmented memory representation x_u and the augmented cognitive representation x_v were spliced to form a global affective feature. Then, this feature is inputted into the fully linked level for processing, and the probability distribution of different feature categories is calculated by softmax function. Ultimately, the feature class with the highest probability is selected as the classification result y_i of the model for the input English teaching evaluation text, as shown in equation (9), where tanh represents the activation function, W_{fc} is the weight of the augmented cognitive representation, b_{fc} is the bias of the augmented cognitive representation, z is the final English teaching quality evaluation result, z in this paper is taken to be 4, and the teaching quality evaluation is divided into four categories, i.e., excellent, good, pass and fail.

$$y_i = \frac{\exp\left(\tanh\left(W_{fc} \times [x_u, x_v] + b_{fc}\right)\right)}{\sum_{i=1}^{z} \exp\left(\tanh\left(W_{fc} \times [x_u, x_v] + b_{fc}\right)\right)} \tag{9}$$

The training objective of the English intelligent teaching quality evaluation classification is to minimise the cross-entropy loss (loss) between output y_i and the true label y_i', which is also the loss function of the model, as shown in equation (10).

$$loss = -\sum_{i} \left(y_i \log y_i' + (1 - y_i) \log(1 - y_i') \right) \tag{10}$$

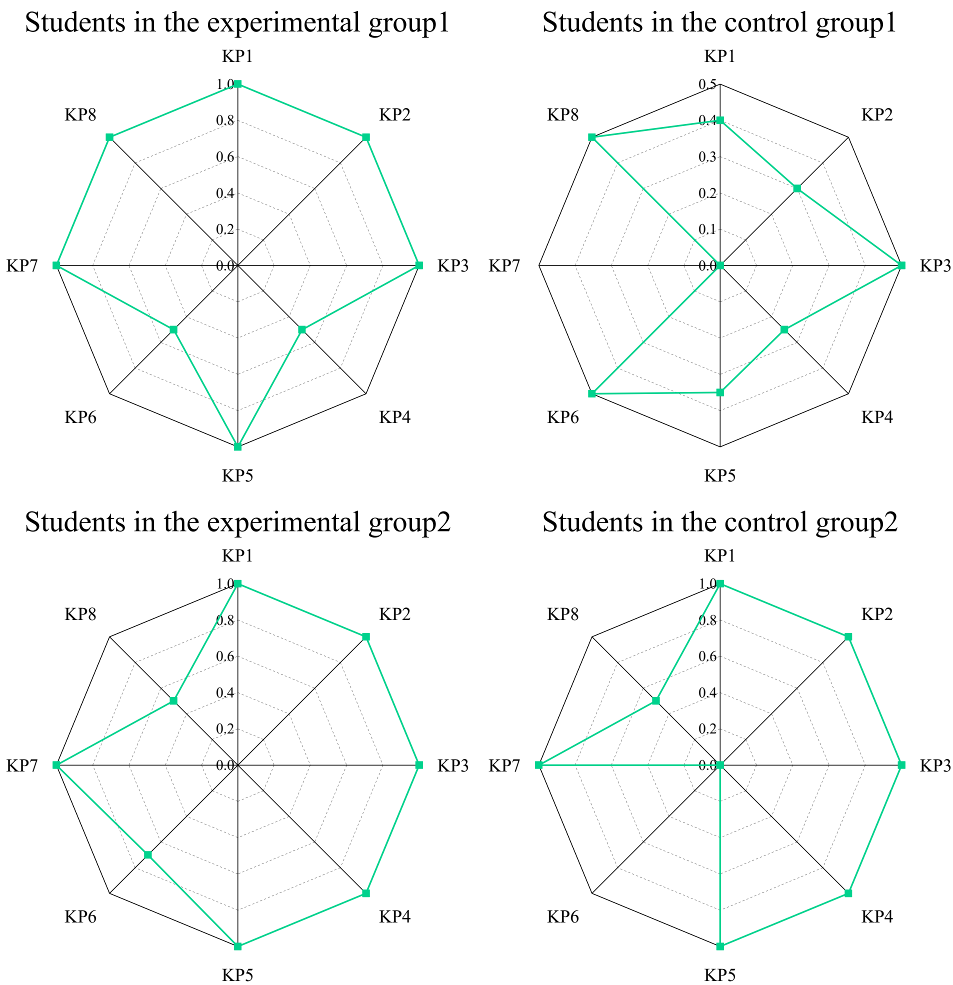
where i represents the sample data index. In addition, in order to prevent overfitting of the model, the early stop and exit rates are set during training.

5 Experimental results and analyses

Due to the lack of a public dataset of high-quality English teaching evaluation, the data used in the experimental part was crawled from a certain teaching evaluation website in the USA. It mainly includes two parts: the evaluation text data of students on English courses and the historical practice data of students. By manually labelling the evaluation text data, the evaluation results are labelled as excellent, good, pass and fail; the content of the historical exercise data mainly includes the number of questions, the labels of the conceptual areas encompassed within the questions, and the results of the students' answers. The total number of data was 7,185, leaving 4,855 higher-quality data after invalid records were removed. 80% of the sample data are used to train the model, and the remaining sample data are used to test the categorisation effect of the model. Among them, the amount of excellent samples is 1,986, the amount of good samples is 1,623, the amount of qualified samples is 817, and the number of unqualified samples is 429. The simulation tool is pyTorch based on python version 3.6, the dimension value of the word embedding part is set to 512, the learning rate and the exit rate of the model are set to 1e–3 and 0.1, respectively, and the batch size and the epoch value at iteration are set to 64 and 30. For the goal of preventing overfitting, the regularisation penalty coefficient is set to 1e–2, and the initialisation parameter of the network conforms to the normal distribution $U(-0.05, 0.05)$.

Before analysing the categorical performance of the proposed evaluation method, this paper evaluates the cognitive diagnostic abilities of students in both experimental and control groups. Radar charts serve to analyse the knowledge mastery of learners, as shown in Figure 2. The left side of Figure 2 shows the attribute probability radar chart of the experimental group students, and the right side shows the control group students.

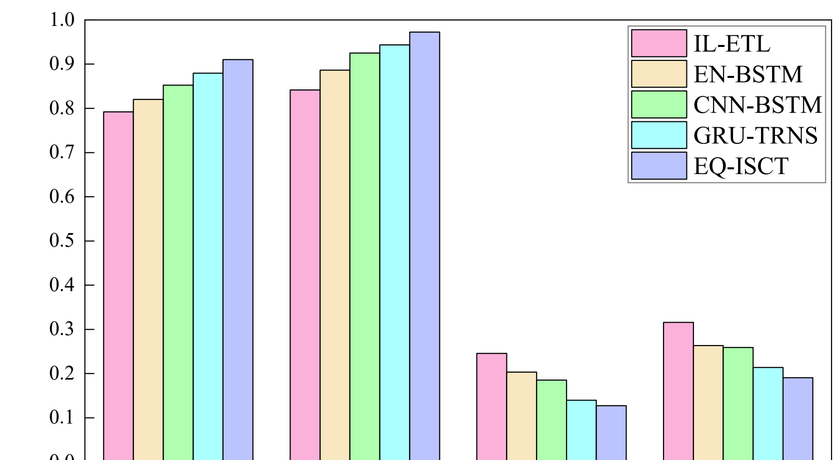
Figure 2 The students' mastery of English knowledge (see online version for colours)



In the first batch of graphs, the data shows that students in the experimental group grasp the various KPs of the English language programme more effectively, and there is a significant change in the level of learning. In the reference group, with a mastery rate of less than 0.5 for all KPs, there was no significant enhancement in learning outcomes. In the second group of figures, the probability distribution of KP mastery across the treatment and reference groups is comparatively uniform. Whereas, for all seven KPs, the probabilities associated with the students in the experimental group exceed 0.8. The probability of learners in the reference group mastering English KPs was also considerably less than that of learners in the treatment group. To summarise, the two learners in the treatment group grasp knowledge better than those in the reference group, i.e., the experimental group demonstrates a superior degree of cognitive growth, and IDS-CAD model can accurately diagnose the students' cognitive ability.

For the goal of assessing the accuracy of the evaluation results of the proposed evaluation method EQ-ISCT, this paper selects the accuracy rate (ACC), AUC, mean absolute error (MAE), and root mean square error (RMSE) as the assessment metrics, and the comparison methods are selected from the current advanced research methods IL-ETL, EN-BSTM, CNN-BSTM, and GRU-TRNS, and the different evaluation methods' performance metrics are compared as shown in Figure 3. The ACC of EQ-ISCT is 91.02%, which is improved by 11.84%, 8.98%, 5.75%, and 3.06% compared to L-ETL, EN-BSTM, CNN-BSTM and GRU-TRNS, respectively. The output range of the AUC metric is within the interval (0, 1), and the closer the AUC value is to 1, it means that the model has better performance and its classification is better (Li, 2024). The AUC of EQ-ISCT is 0.9725, which is closest to 1. Therefore, its classification is better than the other four methods. Comparing the classification accuracy again, the MAE and RMSE of EQ-ISCT are 0.1274 and 0.1907, respectively, which are reduced by 8.93%–48.15% compared to IL-ETL, EN-BSTM, CNN-BSTM, and GRU-TRNS, and show high classification accuracy. The IL-ETL method mainly utilises the idea of integrated learning to mix multiple machine learning models to evaluate the teaching quality, but the traditional machine learning models are highly dependent on manual feature engineering and require domain knowledge to extract effective features, so the evaluation results are not as good as the other four methods. EN-BSTM, CNN-BSTM and GRU-TRNS are all based on the variants of RNN for teaching quality evaluation. The difference is that GRU-TRNS takes into account the influence of students' cognitive ability on the evaluation results. Therefore, the evaluation results of GRU-TRNS are superior to those of EN-BSTM and CNN-BSTM. Based on the above analysis, EQ-ISCT can accurately achieve the quality evaluation of intelligent English teaching.

Figure 3 Comparison of evaluation results of different methods (see online version for colours)



Indicator	ACC	AUC	MAE	RMSE
IL-ETL	0.7918	0.8417	0.2457	0.3158
EN-BSTM	0.8204	0.8863	0.2031	0.2633
CNN-BSTM	0.8527	0.9251	0.1852	0.2591
GRU-TRNS	0.8796	0.9437	0.1399	0.2138
EQ-ISCT	0.9102	0.9725	0.1274	0.1907

This paper also conducts an experimental study on the ablation of the components in the EQ-ISCT. The removal of the student cognitive ability diagnostic unit is denoted as EQ-ISCT/CM, the removal of the cognitive extraction unit is denoted as EQ-ISCT/CFE, and the removal of the student cognitive transfer unit is denoted as EQ-ISCT/CD. The results of the ablation experiments with different components are shown in Table 1. The EQ-ISCT/CFE has the worst performance in all the indexes, which not only has a low classification accuracy, but also has a large classification error, indicating that the EQ-ISCT/CFE has a decisive influence on the classification performance of the EQ-ISCT. The classification of Q-ISCT/CFE and EQ-ISCT/CD did not differ significantly across indicators, suggesting that modelling students' cognitive abilities as well as cognitive state transfer are equally important and both have a significant impact on the model. Therefore, based on the above analysis, it can be seen that EQ-ISCT, which integrates all the components, can realise a more accurate evaluation of English teaching quality.

Table 1 Classification accuracy of each class of beat features under different methods

Method	*ACC*	*AUC*	*MAE*	*RMSE*
EQ-ISCT/CM	0.8687	0.9341	0.1605	0.2358
EQ-ISCT/CFE	0.8519	0.8803	0.1983	0.2617
EQ-ISCT/CD	0.8631	0.9067	0.1872	0.2494
EQ-ISCT	0.9102	0.9725	0.1274	0.1907

6 Conclusions

Data-driven textual analysis can provide direct assessment of English teaching quality. However, existing studies have neglected the impact of students' tacit knowledge status and cognitive level on the English instruction assessment. For this reason, this paper proposes an approach for assessing the quality of English instruction that incorporates cognitive transfer, taking into account the cognitive ability characteristics of students in English courses. Firstly, starting from the process of learners interacting with English practice, this paper establishes the attributes of students' abilities and suggest a cognitive diagnostic model of English in light of students' individual differences. Constructing students' ability attributes based on personalised characteristics, and acquiring variations in individual abilities during the studying process. By employing the K-means clustering approach, students are adaptively allocated to subgroups comprising peers with comparable abilities, creating numerous sub-groups of students who exhibit similar degrees of individuality, and computing the mean capability attribute value of every group of learners, so as to complete the modelling of students' cognitive ability in English. Then, relying on the students' English course evaluation text information and historical practice records, we extracted the textual hidden memory information and practiced cognitive ability vectors, and used the attentional interaction mechanism to transfer the cognitive ability vectors to the evaluation text analysis task. Through this cognitive transfer, the model will not only pay attention to the semantic information contained in the text when extracting text features, but will also be influenced by cognitive factors, which will lead the model to categorise the text of the English teaching evaluation more accurately. The experimental outcome demonstrates that the suggested approach has the highest evaluation classification accuracy and AUC index performance and the highest index performance, which provides a scientific basis for optimising English teaching strategies and enhancing teaching quality.

There is room for further improvement in this paper, for example, due to the constraints of data collection, the characteristics of students' cognitive ability are relatively single, and the influence of multiple factors such as the learning process, teacher-student relationship, and personal value judgment on students' evaluation of English courses is not taken into account. The next step in the study will be to expand the collection of data on student characteristics to obtain a more comprehensive profile of students' cognitive abilities. Meanwhile, in the future, while strengthening the construction of the data corpus for English course evaluation, the multimodal data of the English teaching process can be combined to study the representation of 'behaviour-cognition-emotion' of the teaching and learning subjects. Focus on the research of the relationship between multimodal data of teaching feedback from teachers and students, as well as the influence of data in the English teaching process on text analysis, in order to promote the exploration and practice of new-generation information technology in the reform of English instruction assessment.

Declarations

All authors declare that they have no conflicts of interest.

References

Brauwers, G. and Frasincar, F. (2021) 'A general survey on attention mechanisms in deep learning', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 35, No. 4, pp.3279–3298.

Chen, Y., Wang, X. and Du, X. (2021) 'Diagnostic evaluation model of English learning based on machine learning', *Journal of Intelligent & Fuzzy Systems*, Vol. 40, No. 2, pp.2169–2179.

Donaldson, M. and Mavrogordato, M. (2018) 'Principals and teacher evaluation: the cognitive, relational, and organizational dimensions of working with low-performing teachers', *Journal of Educational Administration*, Vol. 56, No. 6, pp.586–601.

El Khattabi, M-Z., El Jai, M., Lahmadi, Y., Oughdir, L. and Rahhali, M. (2024) 'Understanding the interplay between metrics, normalization forms, and data distribution in K-means clustering: a comparative simulation study', *Arabian Journal for Science and Engineering*, Vol. 49, No. 3, pp.2987–3007.

Ge, D., Wang, X. and Liu, J. (2021) 'A teaching quality evaluation model for preschool teachers based on deep learning', *International Journal of Emerging Technologies in Learning (iJET)*, Vol. 16, No. 3, pp.127–143.

Heller, J., Stefanutti, L., Anselmi, P. and Robusto, E. (2015) 'On the link between cognitive diagnostic models and knowledge space theory', *Psychometrika*, Vol. 80, No. 4, pp.995–1019.

Hou, J. (2021) 'Online teaching quality evaluation model based on support vector machine and decision tree', *Journal of Intelligent & Fuzzy Systems*, Vol. 40, No. 2, pp.2193–2203.

Kou, X., Yang, Z. and Wang, Y. (2023) 'Research on English teaching reading quality assessment based on cognitive diagnostic assessment', *International Journal of Continuing Engineering Education and Life Long Learning*, Vol. 33, Nos. 4–5, pp.388–402.

Lee, Y., Kim, J. and Kang, P. (2023) 'Lanobert: system log anomaly detection based on Bert masked language model', *Applied Soft Computing*, Vol. 146, pp.12–27.

Li, D. (2024) 'An interactive teaching evaluation system for preschool education in universities based on machine learning algorithm', *Computers in Human Behavior*, Vol. 157, pp.50–63.

Li, N. (2023) 'A deep learning algorithm for evaluating the quality of English teaching', *International Journal of Computational Intelligence and Applications*, Vol. 22, No. 3, pp.23–38.

Ma, C., Ouyang, J. and Xu, G. (2023a) 'Learning latent and hierarchical structures in cognitive diagnosis models', *Psychometrika*, Vol. 88, No. 1, pp.175–207.

Ma, L., Yang, H. and Yang, J. (2023b) 'A multimodal teaching quality evaluation for hybrid education based on stepwise regression analysis', *Mobile Networks and Applications*, Vol. 28, No. 3, pp.960–970.

Niu, Z., Zhong, G. and Yu, H. (2021) 'A review on the attention mechanism of deep learning', *Neurocomputing*, Vol. 452, pp.48–62.

Paulsen, J. and Valdivia, D.S. (2022) 'Examining cognitive diagnostic modeling in classroom assessment conditions', *The Journal of Experimental Education*, Vol. 90, No. 4, pp.916–933.

Qi, S., Liu, L., Kumar, B.S. and Prathik, A. (2022) 'An English teaching quality evaluation model based on Gaussian process machine learning', *Expert Systems*, Vol. 39, No. 6, p.e12861.

Qian, L. (2022) 'Research on college English teaching and quality evaluation based on data mining technology', *Journal of Applied Science and Engineering*, Vol. 26, No. 4, pp.547–556.

Shi, L., Wang, Y., Cheng, Y. and Wei, R. (2020) 'Review of attention mechanism in natural language processing', *Data Analysis and Knowledge Discovery*, Vol. 4, No. 5, pp.1–14.

Simonetti, B., Sarnacchiaro, P. and González Rodríguez, M.R. (2017) 'Goodness of fit measures for logistic regression model: an application for students' evaluations of university teaching', *Quality & Quantity*, Vol. 51, No. 6, pp.2545–2554.

Sun, M-h., Li, Y-g. and He, B. (2017) 'Study on a quality evaluation method for college English classroom teaching', *Future Internet*, Vol. 9, No. 3, pp.1–17.

Susanty, L., Hartati, Z., Sholihin, R., Syahid, A. and Liriwati, F.Y. (2021) 'Why English teaching truth on digital trends as an effort for effective learning and evaluation: opportunities and challenges: analysis of teaching English', *Linguistics and Culture Review*, pp.303–316.

Wu, X., Xu, T. and Zhang, Y. (2023) 'Research on the data analysis knowledge assessment of pre-service teachers from China based on cognitive diagnostic assessment', *Current Psychology*, Vol. 42, No. 6, pp.4885–4899.

Zhai, X., Haudek, K.C. and Ma, W. (2023) 'Assessing argumentation using machine learning and cognitive diagnostic modeling', *Research in Science Education*, Vol. 53, No. 2, pp.405–424.

Zhang, X. (2021) 'Evaluating the quality of internet-based education in colleges using the regression algorithm', *Mobile Information Systems*, Vol. 2021, No. 1, p.7055114.

Zhang, Z., Gao, Q. and Chen, F. (2022) 'Evaluating English language teaching quality in classrooms using OLAP and SVM algorithms', *Mobile Information Systems*, Vol. 2022, No. 1, p.9327669.