Research on English Teaching Reading Quality Evaluation Method Based on Cognitive Diagnostic Evaluation

Yana Yi

Henan Institute of Technology, Xinxiang, 453000, China

Correspondence should be addressed to Yana Yi; 13673523391@hait.edu.cn

Received 21 February 2022; Revised 15 March 2022; Accepted 28 March 2022; Published 4 May 2022

Academic Editor: Chin-Ling Chen

Copyright © 2022 Yana Yi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to improve the evaluation effect of English teaching reading quality, this paper applies the cognitive diagnosis evaluation algorithm to the teaching reading quality evaluation process. Moreover, this paper compares the verification effects of Q-matrix from the perspective of nonparameterization and parameterization and considers their performance through simulation research. In addition, this paper applies the method to multilevel scoring empirical data to verify the effectiveness of the cognitive diagnostic evaluation algorithm. Finally, this paper combines cognitive diagnosis and evaluation algorithms to construct an English teaching reading quality evaluation system. The experimental research results show that the English teaching reading quality evaluation method based on cognitive diagnostic evaluation proposed in this paper can effectively improve the reading quality of English teaching.

1. Introduction

In the English adaptive reading system, we use the length of the reading material, sentence structure, article type, vocabulary phrase situation within the concept, and other language ability characteristics to characterize the difficulty of reading. If we use people to obtain feature variables manually, this will be a very ambitious task. However, the current level of information technology is still unable to automatically obtain these characteristics. Therefore, the function of this link is to obtain the language feature quantity of the reading material as automatically as possible and manually label the remaining feature quantity [1].

In the field of natural language processing, machine reading comprehension tasks play an important role. Humans develop their understanding of texts by doing reading comprehension as they grow up. Machine reading comprehension is to let the machine imitate the behavior of human reading comprehension and improve the machine’s comprehension ability by reading a given article to the machine and answering relevant questions. Early machine reading comprehension mainly extracted features manually, designed task rules, and then used conventional machine learning classification algorithms to make predictions. This early machine reading comprehension system can be thought of as an expert system that can answer questions in a particular domain according to a set of logical rules derived from professional knowledge. Because it only focuses on a certain field, the design is simple and easy to implement, but the application is narrow, lacks common sense, and has poor versatility, which cannot meet the needs of current practical applications. With the continuous improvement of the quality of data sets, the rapid expansion of scale, and the emergence of deep neural networks, machine reading comprehension has developed rapidly. At this stage, machine reading comprehension automatically learns task-related features and rules through deep learning networks and uses various attention mechanisms to obtain interactive information, which greatly improves the performance of the model.

Computer-assisted instruction is abbreviated as CAI, which is an automated computer-aided instruction technology that has developed rapidly in the past 30 years. CAI integrates many disciplines, such as computer science, pedagogy, psychology, and electronic education. Moreover, it has many advantages, such as the scientific nature of
knowledge combination, the intuitiveness of expression, the efficiency of classroom teaching, the timeliness of feedback correction, the initiative of learning activities, the simulation of process demonstration, the scalability and reproducibility, and so on. Therefore, it shows great vitality [2]. Carrying out computer-assisted teaching is not only in keeping with the trend of world education modernization, but also the needs of China’s education reform and training of talents. At present, there are many forms of computer-assisted English teaching, such as helping teachers prepare lessons, select questions, write papers, read, test, and help students fill in exercises, memorize words, and correct mistakes. However, most of them are simple, fixed, and unchanging models, which cannot be different from person to person; that is, there is no intelligence [3].

This article combines cognitive diagnostic evaluation methods to evaluate and analyze the quality of English teaching reading, build an intelligent system, and verify the performance of the system to improve the level of English reading teaching.

Organizational structure of this paper is as follows: The first part studies the current situation of reading in English teaching, puts forward relevant problems, and lays the foundation for the creation of this paper. The second part analyzes the current situation of mobile reading, summarizes related technologies, and leads to the research content of this paper. The third part analyzes the main algorithmic cognitive diagnostic evaluation methods in this paper, which provides the algorithm basis for intelligent reading quality evaluation. The fourth part constructs an English teaching reading quality evaluation system based on cognitive diagnostic evaluation and analyzes the effect of the system. Finally, the conclusions of this paper are summarized in the conclusion section.

The main contribution of this paper is to compare the validation effects of Q-matrices from nonparametric and parametric perspectives, consider their performance through simulation studies, and apply the method to empirical data on multilevel scoring to verify that the method works in empirical studies. The performance in the data provides an effective method for intelligent system data processing.

2. Related Work

When studying the development status of mobile reading, the author also combed the existing literature. Literature [4] elaborated on the understanding of the concept of mobile reading behavior and carried out a quantitative analysis of mobile reading behavior, mainly through questionnaire surveys. It includes the purpose of the user’s mobile reading, reading context, reading frequency, and main reading content. Literature [5] provides some inspiration for related research on the understanding of mobile reading, such as personalized and social reading. In the research that affects the efficiency of mobile reading, literature [6] studies the information presentation problem of mobile terminals from the aspect of reading satisfaction with mobile reading. Literature [7] focuses on the usability and ease of use of reading products in the bookstore. The point of view in the article is that information presentation is an important part of the construction of a good reading experience because information acquisition is the essence of reading. In mobile reading, it is very important to present information in a way suitable for mobile reading. Literature [7] took context analysis and related theories as the starting point, conducted an in-depth analysis of the group’s mobile reading behavior, and finally extracted ten main elements that affect the user experience of mobile reading devices.

The task of machine reading comprehension is based on a given article and article-related questions, modeling the article and article-related questions by constructing a suitable model, and selecting or generating answers to the questions. It uses triples (C, Q, and A) to describe the problem, where C represents the article, Q represents the article-related question, and A represents the answer to the question [8]. Machine reading comprehension can be divided into four types: closed type, multiple-choice type, text extraction type, and answer generation type. Close machine reading comprehension is based on contextual information to predict the words that are dug out in the sentence of the article, and most of the words that are dug out are entity words. The characteristic of multiple-choice machine reading comprehension is that after the machine understands a given article, it starts from related questions and selects a correct answer from the set of candidate answers [9]. Text extraction machine reading comprehension requires the machine to extract a certain piece of continuous text from a given article as the answer to the question. It should be noted that the answer may be a word or a sequence of multiple consecutive words [10]. Compared with a single entity of the close class, the task of text extraction is more demanding and more challenging. Answer generation machine reading comprehension is the most flexible reading comprehension task. It has a variety of answers. It may require a model to extract a certain fragment from the original text as an answer like text extraction. It may also require a model to understand the article. After summary, generate the answer yourself (the original text does not) [11]. The answer generation category is closer to human answering questions, involving multiple rounds of question and answer, multithop reasoning, and other technologies. The BiDAF model published in [12] is mainly for experiments on the CNN/DM data set, which belongs to the close-filled machine reading comprehension, and has been innovated in the coding layer and the information interaction layer. The coding layer of BiDAF is formed by stacking multiple subcoding layers, using vector representations of various granularities. First, it maps each character of the input word into a vector and uses the CNN network and Max-pooling method to combine all the character vectors of each word into a character-level vector representation. Then, use the pretrained word vector GloVe to query the word-level vector representation of each word. Finally, the character-level vector representation and word-level vector representation of each word are spliced before and after to form a vector representation with a larger dimension, and then input to the two-way LSTM network to obtain the same
dimension of each word, which contains certain context information. The vector representation. Literature [13] proposed a new two-way attention flow mechanism. This two-way attention mechanism calculates query-to-context (Q2C) attention and context-2-query (C2Q) attention and obtains mutual information such as problem-perceived article representation and article-perceived problem representation. The R-Net introduced in [14] belongs to the text extraction machine reading comprehension. It uses a pretrained language model FusionNet’s ideas for multiple fusions in the information interaction layer. Specifically, SDNet combines the GloVe word-level vector representation and the fixed weight and weighted recombined BERT vector representation at the coding layer. Residual squaring between the observation response and the ideal response on the item is correctly specified. The calculation formula of this method under multilevel scoring can be expressed as [18]

\[
R_j^p = \sum_{i=1}^{N} (u_{ij} - E(U_{mj}|q_j, \tilde{\alpha}_i))^2.
\]

Among them, \(E(U_{mj}|q_j, \tilde{\alpha}_i)\) represents when the attribute vector of the \(j\)th question is \(q_j\), the expected value of the \(i\)th subject’s score on this question. The superscript \(P\) in \(R_j^p\) represents the multilevel scoring, which is to distinguish it from the statistics in the second-level scoring, and \(\tilde{\alpha}_i\) is the estimated value of the participant’s attribute mastery mode. Further, if the statistics are classified according to the knowledge attribute vector \(\alpha\) of the subjects, then formula 1 can be transformed into the following form:

\[
R_j^p = \sum_{m=1}^{\infty} \sum_{C_m \in C_m} \left[ u_{ij} - E(U_{jm}|q_j, \tilde{\alpha}_i) \right]^2.
\]

Among them, \(C_m\) is the first potential mastering category of \(m\), where \(E(U_{jm}|q_j, \tilde{\alpha}_i) = \Sigma_{h=1}^{H} h \cdot P(Y_t = h|\theta)\). If considering the different distributions of each attribute mastering mode, formula 2 can be further transformed into the following formula [19]:

\[
R_j^p = \sum_{m=1}^{\infty} \sum_{C_m \in C_m} \left[ u_{ij} - E(U_{jm}|q_j, \tilde{\alpha}_i) \right]^2.
\]

\(\pi_{ma}\) is the posterior probability distribution of \(C_m\), and formula 3 can be understood as the sum of the expected residual squares between the observation score and the expected score.

For the subject’s classification problem, if the \(\eta_i(\hat{\alpha}_i, q_j)\) corresponding to the knowledge attribute mastering mode \(\hat{\alpha}_i\) can minimize \(d_{wh}(u_i, \eta_i(\hat{\alpha}_i, q_j))\), then \(\hat{\alpha}_i\) is the estimated value of the subject’s knowledge attribute mastering mode. The specific formula is as follows [20]:

\[
d_{wh}(u_i, \eta_i(\hat{\alpha}_i, q_j)) = \sum_{j=1}^{N} \frac{1}{\bar{P}_j(1 - \bar{P}_j)} |u_{ij} - \eta_i(\hat{\alpha}_i, q_j)|.
\]

In the above formula, \(h\) represents the \(h\)th ideal attribute mastering model, \(1/\bar{P}_j(1 - \bar{P}_j)\) is the weighted part of the question, and \(\bar{P}_j\) represents the proportion of subjects who answered the question \(j\) correctly or the probability of answering the question correctly. \(\bar{P}_j(1 - \bar{P}_j)\) is the variance of the participant’s observation and answer on question \(j\). Therefore, this indicator is more inclined to choose topics with small variance. Through the previous analysis, a two-step iterative algorithm can be used to estimate the attribute vector of the question based on the \(R_j^p\) statistic. The first step is to use the weighted Hamming distance \(d_{wh}(u_i, \eta_i(\hat{\alpha}_i, q_j))\) to classify the subjects, and the second step is to use \(R_j^p\) to estimate the attribute vector of the subject based on the classification of the subjects.

3. Cognitive Diagnosis and Evaluation Methods

In this paper, the cognitive diagnostic evaluation algorithm is used as the English teaching reading quality evaluation algorithm to provide a reference for the construction of the subsequent English teaching reading quality evaluation system, and then the algorithm is studied.

3.1. Matrix Verification Metrics. In this study, the non-parametric Q-matrix verification method is expanded, and the corresponding Q-matrix verification index is extended to the multilevel scoring problem, and the corresponding method is denoted as \(R^q\). Based on the \(q\) vector of a certain item in the test, if the sum of the residual squares RSS between the observation response and the ideal response on this item can be minimized, it indicates that the \(q\) vector of the item is correctly specified. The calculation formula of this method under multilevel scoring can be expressed as [18]

\[
R_j^q = \sum_{i=1}^{N} (u_{ij} - E(U_{jm}|q_j, \tilde{\alpha}_i))^2.
\]

3.2. Algorithmic Simulation Assumptions. This study simulates the data under the conditions of different number of subjects (400, 600, 800, and 1000) and different error \(q\)
vectors (5%, 10%, and 15%). There are a total of $4 \times 3 = 12$ experimental conditions, each of which is repeated 100 times. In this study, R software was used to generate simulation data.

(1) **Mock Test Q-Matrix**

For the Q-matrix, the knowledge attributes required for an answer category refer to the attributes required by the candidate to correctly reach the category after completing all the previous steps.

(2) **Simulation of Q-matrix Containing Errors**

The initial Q-matrix $Q_0$ is constructed on the basis of the real Q-matrix, and the questions with incorrect calibration in $Q_0$ are randomly selected according to three proportions (5%, 10%, and 15%, resp.). The selected wrong item is randomly selected among $2^k - 2$ possibilities (cannot be a vector of all 0s and a correct vector).

(3) **Simulate the Knowledge State of the Subjects**

It is assumed that the knowledge state of the subjects obeys a uniform distribution; that is, the number of subjects under each knowledge attribute mastery mode is similar.

(4) **Simulation Question Parameters**

The simulation of the question parameters is completed according to the following rules; that is, the highest category is $S_j(h|\alpha_{j|h} = 1) = 0.9$, and the lowest category is $S_j(h|\alpha_{j|h} = 0) = 0.1$. When there are more than two answer categories, the middle category probability is randomly selected from the uniform distribution $U(S_j(h|\alpha_{j|h} = 0), S_j(h|\alpha_{j|h} = 1))$, and it is ensured that the more attributes are mastered, the greater the category probability is.

(5) **Simulate the Subject’s Answer**

After simulating the student’s knowledge state, Q-matrix and item parameter values in the above steps, the response probability of the participant on the item is calculated. Participants’ response scores were simulated based on the item response function of the sequential GDINA model.

(6) In order to evaluate the performance of the $R^P$ method in the Q-matrix estimation, the indicators at the Q-matrix level are, respectively, used to evaluate the estimation accuracy, which includes the number of successful estimations, the evaluation indicators of the item level, the number of criteria for the item mode, the evaluation index of the attribute level, and the average criterion number of the item attributes. In terms of estimating efficiency, the average number of iterations and the average running time are used to evaluate.

Among them, the number of successful estimates $N_{\text{success}}$ is as follows: it represents the number of times the Q-matrix is completely correctly estimated in 100 batches of randomly generated data.

\[
N_{\text{success}} = \sum_{r=1}^{100} I(Q^{PT} = Q^P). 
\]

Among them, the indicator function $I(Q^{PT} = Q^P)$ indicates whether the $r$th estimated Q-matrix is exactly the same as the real Q-matrix, that is, whether the Q matrix containing errors is successfully corrected. $I(Q^{PT} = Q^P) = 1$ means successful estimation; otherwise, it is 0.

The criterion number of question patterns (PMN) is as follows: the algorithm calculates the consistency (Pattern Match Number, PMN) between the measurement pattern of all questions in the Q-matrix estimated each time and the measurement pattern of the real Q-matrix question (Pattern Match Number, PMN) and calculates the average of 100 experiments.

\[
PMN = \frac{\sum_{r=1}^{100} \sum_{j=1}^{J} I(q_j = \tilde{q}_j)}{100},
\]

where $I(q_j = \tilde{q}_j)$ indicates whether the attribute vector of item $j$ in the $r$ batch of data is correctly estimated. 1 indicates the correct estimation; otherwise, it is 0.

Item attribute average criterion number (ARN) is as follows: the algorithm calculates the average correct recovery amount of item attributes in 100 repeated experiments, which reflects the probability of knowledge attribute recovery (Attribute Recovery Number, ARN).

\[
ARN = \frac{\sum_{r=1}^{100} \sum_{j=1}^{J} I(q_{jk} = \tilde{q}_{jk})}{100 \times J},
\]

When the Q-matrix is not successfully restored, the item model criterion number PMN and the item attribute average criterion number ARN describe the extent to which the estimation method restores the item attribute vector. The higher the PMN and ARN, the more accurate the estimation of the method.

The average number of iterations (AIN) is as follows: the algorithm calculates the average number of iterations in 100 repeated experiments (Average Iterative Number).

\[
AIN = \frac{\sum_{r=1}^{100} \text{ItNum}_r}{100},
\]

where ItNum $r$ refers to the number of iterations used in the $r$th estimation.

Average running time (ART) is as follows: the algorithm calculates the average running time ART (Average Running Time) of the Q-matrix verification process in 100 repeated experiments.

\[
ART = \frac{\sum_{r=1}^{100} t_r}{100}
\]

$t_r$ refers to the running time used in the $r$th experiment. Analyzing the same batch of data, the smaller ART indicates the higher efficiency of the algorithm.

Figure 1 and Figure 2 show the changes in the number of successful estimates of the Q-matrix under different types of grouping. Figure 1 is grouped according to different numbers of subjects, and Figure 2 is grouped according to different Q-matrix error ratios.
As shown in Figure 1, it describes the trend of estimating the Q-matrix with the number of people using the $R^p$ method. It is easy to see the influence of the increase in the number of people and the decrease in the error rate on the Q-matrix estimation effect under the nonparametric method. Obviously, the larger the number of people and the lower the error rate, the better the Q-matrix estimation effect.

In order to further verify the performance of the RP method in estimating the Q-matrix, this study examined the estimation success rate of the method step by step.

Figures 3 and 4, respectively, show the estimated success numbers of the attribute vectors corresponding to the correct answer categories under the conditions of grouping by the number of people and grouping by the error ratio when using the $R^p$ method to estimate the Q-matrix. The number of answer categories in the entire Q-matrix is 39. Therefore, the closer this indicator is to 39, the more accurate the estimate is.

The basic idea of the Q-matrix estimation algorithm is as follows: when the number of subjects is large enough, we analyze the test data based on the correctly defined Q-matrix to obtain model parameters and subject classification parameters. The expected response distribution and the observed response distribution obtained by calculation should be equal, which can be represented by the following pseudocode:

$$Q_{true} = Q_{obs}$$

$$P(R|Q_{true}) = \hat{P}(R).$$

Among them, $Q_{true}$ is the Q-matrix (also called the estimated value of the matrix under the alternative multilevel scoring diagnostic test), $Q_{true}$ is its true value, $P(R|Q_{true}, P)$ represents the distribution of the answer vector $R$ determined by the model parameters and the overall distribution, and $\hat{P}(R)$ represents the observed distribution of the answer vector $R$. At the same time, there are
\[ P(R|Q^p, P) = \sum_a \sum_{j=1}^{l} P(R_j|Q^p, a). \]  
(12)

\[ \bar{P}(R) = \frac{1}{N} \sum_{i=1}^{N} I(R_i = R). \]

Among them, \( p_a \) represents the distribution of the attribute mastering model \( a \) in the population, and \( R_i \) represents the response vector of the subject \( i \). A key concept of the \( S^p \) method is the \( S^p \) matrix (\( T \)-matrix); its function is to describe the expected response distribution.

1. For a single question (we assume that the highest score of the question is \( H \); that is, the score interval is \([0, H]\), and there is a total of \( H + 1 \) answer categories), then there are

\[ P(R^j = 1|Q^p, P) = \sum_a p_a P(R^j = 1|Q^p, a) = B_{Q^p}(j)P, \]

\[ P(R^j = H|Q^p, P) = \sum_a p_a P(R^j = H|Q^p, a) = B_{Q^p}(j)P. \]

(13)

2. For item pairs (here for convenience, we assume that each item has the same answer category. In practical applications, the answer category of each item may be different), then there are

\[ P(R^{i1} = 1, R^{i2} = 1|Q^p, P) = \sum_a p_a P(R^{i1} = x_{i1}|Q^p, a) P(R^{i2} = x_{i2}|Q^p, a) = B_{Q^p}(j_1, j_2)P, \]

\[ P(R^{i1} = H, R^{i2} = H|Q^p, P) = \sum_a p_a P(R^{i1} = H|Q^p, a) P(R^{i2} = H|Q^p, a) = B_{Q^p}(j_1, j_2)P. \]

(14)

Therefore, there are a total of \( C_J^2 \) question pairs. In the same way, there are three combinations of topics, up to the combination of \( J \) topics. In this way, the \( T \)-matrix of size \((2^{JH} - 1) \times 2^J\) can be constructed:

\[
T = \begin{bmatrix}
\begin{array}{ccc}
\mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_K \\
1 & \begin{bmatrix} p_{a_{1,1}} & P_{a_{1,1}} & \cdots & P_{a_{1,K,1}} \\
2 & \begin{bmatrix} \vdots & \vdots & \cdots & \vdots \\
1\nu2 & \begin{bmatrix} P_{a_{1,1,1}} & P_{a_{1,1,2}} & \cdots & P_{a_{1,1,K,1}} \\
& \vdots & \cdots & \vdots \\
\end{array}
\end{array}
\end{array}
\end{bmatrix}
\end{bmatrix}
\]

(15)

We assume that the row vectors of the \( T \) matrix are \( B_{Q^p(1)}, \ldots, B_{Q^p(1,2)} \), respectively; then formula 9 can be expressed as

\[
T(\alpha^p) = \begin{bmatrix}
B_{Q^p(1)} \\
\cdots \\
B_{Q^p(1,2)}
\end{bmatrix}.
\]

(16)

Further, according to formula 5 to formula 8, the \( T \)-matrix can be expressed as

\[
T_{QP} = \begin{bmatrix}
P(R^1 = 1|Q^p, P) \\
\cdots \\
P(R^1 = 1, R^2 = 1|Q^p, P)
\end{bmatrix}.
\]

(17)

The \( \beta \) vector is another important concept of this method, which is the column vector corresponding to equation 10. The score vector is the ratio of the number of people answering the question combination, and it represents the distribution of observation scores. When \( N \to \infty \), if all parameters are calibrated correctly, according to the law of large numbers, there is

\[
\beta = T(\alpha^p)P.
\]

(18)

Therefore, the objective function of the matrix verification under the multilevel scoring diagnostic test can also be expressed as [21]

\[
S_P(\alpha^p) = |T(\alpha^p)P - \beta|.
\]

(19)

Among them, \([..] \) is the Euclidean distance. Since the parameters are unknown, it is necessary to estimate each parameter based on the model and use maximum likelihood estimation to estimate each parameter.

This research is based on the following assumptions. We assume that a \( Q \)-matrix has been defined by experts in the relevant knowledge domain, that is, the initial \( Q \)-matrix. Moreover, this initial \( Q \)-matrix contains only partial errors (i.e., some of the elements are incorrectly defined, and the remaining elements are correctly defined).

This paper simulates the data under different number of subjects (800, 1000, 2000, and 4000) and different error attribute ratios (5%, 10%, and 15%). There are a total of \( 4 \times 3 = 12 \) experimental conditions, and simulations are performed 100 times under each condition. In this study, R software was used to generate simulation data.

1. Test the true value of the \( Q \)-matrix.

The algorithm uses restricted \( Q \)-matrix, which is denoted as \( Q^1 \).

2. The algorithm simulates the initial \( Q \)-matrix.
Based on the real Q-matrix, the algorithm simulates the initial Q-matrix containing a certain proportion of errors, denoted as $Q^0_P$, and $Q^0_P$ contains the incorrectly calibrated $q$ vector. The wrong questions are randomly selected according to a certain percentage (5%, 10%, and 15%, resp.), and the selected wrong items are randomly selected among the possibilities in $2^K - 2$.

(3) Simulate the participant’s knowledge state.

The knowledge state of the subjects is simulated in a uniform distribution; that is, the number of subjects under each knowledge attribute mastery mode is similar.

(4) Simulate question parameters.

The simulation of the topic parameters is completed according to the following rules; that is, the highest category is $S_j(h|\alpha_{jik} = 1) = 0.9$, and the lowest category is $S_j(h|\alpha_{jik} = 0) = 0.1$. When there are more than two answer categories, the middle category probability is randomly selected from the uniform distribution $U(S_j(h|\alpha_{jik} = 0), S_j(h|\alpha_{jik} = 1))$, and it is ensured that the more attributes are mastered, the greater the category probability is.

(5) Simulate the subject’s answer.

After simulating the student’s knowledge state, Q-matrix, and item parameter values in the above steps, the response probability of the participant on the item is calculated. Participants’ response scores were simulated based on the item response function of the sequential GDINA model.

$$P(X_j = h|\alpha_c) = \prod_{x=0}^{h} S_j(x|\alpha_c). \quad (20)$$

Among them, $S_j(x|\alpha_c)$ is the probability that the participant scores at step $x$ on question $j$. It can use common cognitive diagnosis model functions, such as DINA or GDINA.

(6) Evaluation index

The evaluation indicators used in this study are the same as those in study one. The algorithm uses the $S^P$ method to estimate the results of the Q-matrix of the multilevel scoring diagnostic test. This research mainly focuses on the effect of Q-matrix estimation when the sample is large. On the whole, the performance of the parameterized $S^P$ method is when the test sample is large. In most cases, the estimated success rate of the entire Q-matrix exceeds 50% [22].

It can be further seen from the results that as the sample increases, the success rate of the $S^P$ method in estimating the Q-matrix will increase. As the correct elements in the Q-matrix increase, that is, the error rate is reduced, the success rate of the $S^P$ method in estimating the Q-matrix will also increase.

Figures 5 and 6 show the changes in the number of successful estimates of the Q-matrix in the case of grouping by different types. Figure 5 is grouped according to the number of participants, and Figure 6 is grouped according to the Q-matrix containing different error ratios.

Figure 5 describes the change trend of the Q-matrix estimated by the $S^P$ method with the number of people. It is easy to see the influence of the increase in the number of people on the parameterization method. Under different error ratio conditions, as the sample size increases, the effect of Q-matrix estimation gradually improves. When the error ratio of Q-matrix is 10% and 15%, and the number of subjects is 800, the correct estimation rate of Q-matrix is slightly higher than 50%. However, when the error rate is 5% and the number of subjects is different, the number of successful estimates of the Q-matrix is significantly higher than the error rate of 10% and 15%. Moreover, under the condition of a sample size of 4000, the success rate of Q-
matrix estimation reaches 80%. From Figure 6, it is easy to see the influence of the increase in the number of people on the estimation effect of the parameterization method. Under the condition of different error ratios, the increase in the number of people is still more obvious for the improvement of the estimated success rate, which is different from the $R^p$ method.

In order to further evaluate the performance of the $S^p$ method in estimating the $Q$-matrix, we examined the estimated success rate of the method step by step.

For the change trend of the successful estimation of the average category attribute vector in the entire $Q$-matrix, we refer to Figures 6 and 7. The average $Q$-matrix element estimates the correct number, referred to in Figure 8.

It can be seen that on the one hand, as the proportion of error elements in the $Q$-matrix increases, whether it is the estimated success rate of the entire $Q$-matrix, the average estimated success rate of the answer category attribute vector in the $Q$-matrix, or the average estimated success rate of the elements in the $Q$-matrix will have a corresponding decline. On the other hand, as the number of subjects increases, the estimation accuracy indicators will increase to varying degrees.

It can be seen from Figure 6 that the estimated success times of the average correct answer category attribute vector increases as the proportion of errors decreases. When the sample is 4000, the average number of categorical attribute vectors of estimated errors is less than one under different error ratio conditions. However, when the number of subjects is 800, 1000, and 2000, the average number of categorical attribute vectors estimated to be incorrect reaches 2 or more under the conditions of 10% and 15% error ratios. It can be seen from Figure 7 that as the sample size increases, the estimated success times of the average answer category attribute vector increase significantly.

It can be seen from Figures 8 and 9 that when the error ratio contained in the $Q$-matrix is only 5%, 1000 test data can achieve a higher element estimation success rate. However, when the error rate reaches a higher 10% or 15%, the estimated success rate of the elements in the $Q$-matrix is not high when the number of subjects is 800, 1000, and 2000. Furthermore, when the sample size reaches 4000, a higher $Q$-matrix element estimation success rate can be obtained. Moreover, when the sample size is 4000, the success estimates of the $Q$-matrix maintain a high and similar level under different sample size conditions. In general, the $S^p$ method will obtain a relatively stable $Q$-matrix element estimation success rate when the sample reaches 4000 subjects.

### 4. Research on the English Teaching Reading Quality Evaluation Method Based on Cognitive Diagnostic Evaluation

This article regards students’ learning English as a structured activity to solve problems. In this activity, the problem solver starts from the starting state, goes through a series of intermediate states, and finally reaches the goal state. Artificial intelligence research shows that there are two ways to reach a goal. One is to give all possible states and test whether the final state meets the desired goal, and the other is to use the available additional information to find the correct solution path. In the process of learning or reading English, the text given is the initial state of the problem. The goal is to connect the content of the text with the reader’s existing knowledge structure. Moreover, problem-solving activities form a problem space. In this problem space, according to the additional information that can be obtained, seek the correct solution path. This process is shown in Figure 10(a).

Figure 10(b) describes a very simple network used to analyze English sentences. This kind of network is called "S" network. In the network shown in Figure 10(b), from one state to another, the input signal must be confirmed as a verb. That is, one state must be connected to another state through necessary actions. Also, only the lead from the initial state is allowed, and there will always be a line starting from the initial state until the send line. The other is the NP network. The connection between the states is divided into a search line and a classification line, as shown in Figure 10(c).
Figure 9: The number of attribute elements of the question successfully estimated by the $S^p$ method, range [0, 195].

Figure 10: (a) Description of the problem space. (b) Simple transition network. (c) Simple NP network.
The design of the inference engine of the English Reading Teaching Expert System is based on fuzzy inference, reasoning combined with artificial neural network models, nonmonotonic reasoning, and mixed control strategies. The inference engine repeatedly matches the rules in the knowledge base against the conditions or known information in the field of English reading teaching. The reasoning network drawn in Figure 11 helps teachers and students understand the reasoning mechanism and interpretation mechanism of the expert system.

On the basis of the above research, this paper uses experimental teaching methods to study the English teaching reading quality evaluation system proposed in this paper and calculates the evaluation effect and teaching effect, as shown in Figure 12 and Figure 13.

From the above research, it can be seen that the English teaching reading quality evaluation method based on cognitive diagnostic assessment proposed in this paper can effectively improve the reading quality of English teaching.

5. Conclusion

The network platform can enable students to share resources. Moreover, it provides an appropriate amount of input every day to keep students in a dynamic language learning environment. Moreover, some students with higher
learning ability and learning level can also share the high-quality resources they have obtained in extracurricular reading. In the open era, students have a wealth of life information, personalized life experience, and innovative learning methods. Moreover, in the process of cooperative learning, inquiry learning, and autonomous learning, they have also formed a rich and colorful curriculum resources between each other, which become the activity carrier of reading resources. In addition, the process of developing and using curriculum resources is the process of student learning. This article combines cognitive diagnostic evaluation methods to evaluate and analyze the English teaching reading quality, build an intelligent system, verify the performance of the system, and improve the level of English reading teaching. The experimental research results show that the English teaching reading quality evaluation method based on cognitive diagnostic evaluation proposed in this paper can effectively improve the reading quality of English teaching.

When estimating the Q-matrix in steps, under each condition, the two methods have higher success estimation rates in the first step, but with the increase of the answering steps, the two methods are more sensitive to the attributes corresponding to the later steps. The success rate of vector estimation will gradually decrease, which is also the main reason for the low success rate of estimation of the entire Q-matrix. The next step requires in-depth research on methods to improve the success rate of attribute vector estimation corresponding to the following steps.

Data Availability
The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest
The author declares no competing interests.

Acknowledgments
This study was sponsored by Henan Institute of Technology, China.

References
[1] A. M. Songbatumis, "Challenges in teaching English faced by English teachers at MTsN Taliwang, Indonesia," Journal of foreign language teaching and learning, vol. 2, no. 2, pp. 54–67, 2017.
[2] J. C. Richards, "Teaching English through English: proficiency, pedagogy and performance," RELC Journal, vol. 48, no. 1, pp. 7–30, 2017.
[3] L. Susanty, Z. Hartati, R. Sholihin, A. Syahid, and F. Y. Liriwati, "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, vol. 5, no. S1, pp. 303–316, 2021.
[4] A. S. Fatimah, S. Santiana, and Y. Saputra, "Digital comic: an innovation of using toondo media as teaching media for teaching English short story," English Review: Journal of English Education, vol. 7, no. 2, pp. 101–108, 2019.
[5] B. Ayçiçek and T. Yanpar Yelken, "The effect of flipped classroom model on students’ classroom engagement in teaching English," International Journal of Instruction, vol. 11, no. 2, pp. 385–398, 2018.
[6] N. Guzachchova, "Zoom technology as an effective tool for distance learning in teaching English to medical students," Bulletin of Science and Practice, vol. 6, no. 5, pp. 457–460, 2020.
[7] M. S. Hadi, "The use of song in teaching English for junior high school student," English Language in Focus (ELIF), vol. 1, no. 2, pp. 107–112, 2019.
[8] A. Mahboob, "Beyond global Englishes: teaching English as a dynamic language," RELC Journal, vol. 49, no. 1, pp. 36–57, 2018.
[9] H. Sundari, “Classroom interaction in teaching English as foreign language at lower secondary schools in Indonesia,” Advances in Language and Literary Studies, vol. 8, no. 6, pp. 147–154, 2017.
[10] K. Kim and Y. Park, “A development and application of the teaching and learning model of artificial intelligence education for elementary students,” Journal of The Korean Association of Information Education, vol. 21, no. 1, pp. 139–149, 2017.
[11] O. Zawacki-Richter, V. I. Marin, M. Bond, and F. Gouneur, “Systematic review of research on artificial intelligence applications in higher education–where are the educators?” International Journal of Educational Technology in Higher Education, vol. 16, no. 1, pp. 1–27, 2019.
[12] S. C. Yang, W. K. Vong, R. B. Sojitra, T. Folke, and P. Shaffo, “Mitigating belief projection in explainable artificial intelligence via Bayesian teaching,” Scientific Reports, vol. 11, no. 1, p. 34, 2021.
[13] Y. Lee, "An analysis of the influence of block-type programming language-based artificial intelligence education on the learner’s attitude in artificial intelligence," Journal of The Korean Association of Information Education, vol. 23, no. 2, pp. 189–196, 2019.
[14] J. M. Alonso, "Teaching explainable artificial intelligence to high school students," International Journal of Computational Intelligence Systems, vol. 13, no. 1, pp. 974–987, 2020.
[15] R. Yang, "Artificial intelligence-based strategies for improving the teaching effect of art major courses in colleges," International Journal of Emerging Technologies in Learning (iJET), vol. 15, no. 22, pp. 146–160, 2020.
[16] P. Ribeiro, H. Simões, and M. Ferreira, "Teaching artificial intelligence and logic programming in a competitive environment," Informatics in Education, vol. 8, no. 1, pp. 85–100, 2009.
[17] V. Rampton, M. Mittelman, and J. Goldhahn, "Implications of artificial intelligence for medical education," The Lancet Digital Health, vol. 2, no. 3, pp. e111–e112, 2020.
[18] Q. Zhou, "Reforms in teaching the course of introduction to artificial intelligence," Solid State Technology, vol. 64, no. 2, pp. 6331–6335, 2021.
[19] K. I. M. Haejoong and L. E. E. Sangmin, "Do less teaching, do more coaching: toward critical thinking for ethical applications of artificial intelligence," Journal of Learning and Teaching in Digital Age, vol. 6, no. 2, pp. 97–100, 2021.
[20] P.-H. Lin, A. Wooders, J. T.-Y. Wang, and W. M. Yuan, “Artificial intelligence, the missing piece of online education?” *IEEE Engineering Management Review*, vol. 46, no. 3, pp. 25–28, 2018.

[21] C. Guan, J. Mou, and Z. Jiang, “Artificial intelligence innovation in education: a twenty-year data-driven historical analysis,” *International Journal of Innovation Studies*, vol. 4, no. 4, pp. 134–147, 2020.

[22] V. Sugumaran, Z. Xu, P. Shankar, and H. Zhou, *Application of Intelligent Systems in Multi-Modal Information Analytics*, vol. 929, Springer, New York, NY, USA, 2019.