International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-21

https://doi.org/10.70917/ijcisim-2026-0374
Article

The Path of a Computer-Supported Collaborative Learning Approach Based on Data Mining Techniques to Promote Students' Intercultural Communication Competence in the English Language Discipline

Chang Liu [1,*]

[1] School of Digital Commerce, Zhejiang Technical Institute of Economics, 310018, Hangzhou, China
* Correspondence author: zjg226@126.com

Abstract: Computer-supported collaborative learning (CSCL) methodology can facilitate communication among learners in online learning platforms. This study proposes a practical strategy for English majors to apply CSCL teaching model and students' cross-cultural communication competence cultivation. By mining the characteristics of students' learning behaviors, combined with an improved density clustering algorithm, collaborative learning cluster construction is realized. Finally, practice teaching is carried out to verify its effectiveness from the improvement of students' intercultural communication ability. Through the paired-sample test of the knowledge dimension of the experimental class and the control class, it is found that the paired-sample t-test results of the pre and post-tests of the knowledge dimension of the experimental class show that $T=-18.605$, $P=0.000$, and the P-value is lower than the significance of 0.05, which can be obtained that there is a significant difference between the performance of the pre and post-tests of the experimental class before and after the experiment, and that the post-tests of the experimental class are superior to the pre-tests, which leads to the conclusion that based on the The computer-supported writing learning method based on data mining technology plays a positive role in the improvement of students' intercultural communication skills.

Keywords: CSCL; data mining; density clustering; intercultural communication skills

1. Introduction

With the advance of globalization, people have more contact with multiple cultures, and they encounter problems that need to be solved in cross-cultural scenarios in their work and life, and cross-cultural communication skills are in the spotlight. Under the new situation, participation in international competition requires a large number of talents who not only know the international economic rules and global thinking. And it also needs talents with intercultural communication ability, which is not only the embodiment of personal comprehensive quality, but also an indispensable ability in international communication and cooperation [1-2]. For English teaching, the cultivation of intercultural communication ability has far-reaching significance. First of all, the cultivation of intercultural communication skills helps to improve the comprehensive quality of students. In cross-cultural communication under the background of multiculturalism, students need to understand the values, ways of thinking and behavioral norms of different cultures under the premise of cultural self-confidence, which will help to broaden their horizons as well as enhance their cultural inclusiveness and international vision [3-5]. Secondly, cultivating cross-cultural communication skills helps to enhance students' competitiveness in employment [6]. In the current globalized employment market, talents with intercultural communication skills are more favored by employers. Of course, intercultural communication ability also helps to broaden the individual's international vision and intercultural cognition. By communicating and interacting with people from different cultural backgrounds, individuals can gain a deeper understanding of the values, ways of thinking and behavioral norms of different cultures, thus broadening their own horizons and cognitive scope [7-8]. Finally, intercultural communication competence helps to promote international communication and cooperation [9]. In the context of the development of globalization, the contact and cooperation between countries are getting closer and closer, only people with intercultural communication skills can better adapt to the international environment and contribute to the development of the country and the nation by participating in international exchange and cooperation.

For English teaching, it is even more crucial to cultivate students with intercultural communication skills. However, there are still some problems in the current situation of cultivating intercultural communication competence in English teaching, such as backward teaching concepts, slightly singular teaching contents and methods, insufficient teachers, and the evaluation system needs to be perfected, etc. [10-13]. Therefore, it is necessary to cultivate students' intercultural communication competence. Ren and Cui [14] suggest that cultivating students' intercultural communication competence needs to pay attention to the teaching of different cultures, recognize the connection between teaching and learning, make students improve intercultural communication competence independently in teaching and strengthen the cultural awareness of their own mother tongue. Pang [15] believes that cultivating students' intercultural communication competence not only requires reforming the curriculum and teaching methods and constructing a multi-faceted evaluation system, but also requires teachers' intercultural literacy in order to support students' learning effects and realize the ultimate intercultural communication. Wang and Wu [16] mention that basic English learning is the foundation for cultivating students' intercultural communication competence, and that the training of speaking will be used to communicate with each other in the form of oral English knowledge to form intercultural communication competence. Li [17] made innovations in teaching strategies and methods to develop students' intercultural communication competence, mainly by introducing cultural background, improving communication experience, curriculum improvement, technology-assisted teaching strategies, interactive, case-based and role-playing teaching forms and critical thinking development to realize students' intercultural communication competence. Aifang et al. [18] analyzed the influence of multicultural literature on intercultural communication competence in English education, mainly through diversified cultural stories, critical thinking enhancement, and open and agreeable attitudes to promote students' intercultural communication competence. Wang [19] designed a program to cultivate intercultural communication competence based on Byram's model of intercultural communication competence from the perspectives of attitude, knowledge, skills, and critical cultural awareness, and suggested that teachers also need to improve their own language proficiency and intercultural communication competence. Zhang [20] mentions that intercultural communication cannot be separated from the learning of cultural knowledge, and in the information age, the application of network information technology to teaching perfection and communication practice is the trend of cultivating intercultural communication competence. Peng [21] uses modern information technology to create and update the model and teaching mode of intercultural communication competence of English in university, using the technology in the teaching environment without the constraints of space and time. Teaching resources can be shared globally, which meets students' individual needs and thus improves intercultural communication competence. Long and Lin [22] introduced to artificial intelligence injected into English teaching in different forms and methods, and established an assessment model of intercultural communication competence for college students, which together make up for the imperfect situation of the teaching methodology and evaluation system in the cultivation of intercultural communication competence, and promote the intercultural communication competence.

Although scholars have put forward many strategies to cultivate students' intercultural communication competence, in the process of practicing, there are still limitations of cultural knowledge learning materials and shallow contents. Most of the communication simulations are role-playing and lack of practical exercises. Students' individualized differences are significant, but teachers do not have a clear perception of this [23-26]. In response to the superficial content, insufficient practice scenarios, and individual differences, computer-supported collaborative learning utilizes computer and network technology to provide learners with a collaborative learning environment, where learners can learn collaboratively at any time and any place, make full use of network resources and communicate across time and space [27]. It is a solution to the time and space limitations and provides a real cross-cultural communication scenario. And data mining technology can transform large amount and multimodal data into useful knowledge and information, which has application advantages in discovering knowledge, visualizing data, correcting data, and integrating resources [28-30]. This provides a scientific and effective method for cultivating students' intercultural communication skills in English language subjects.

The article first analyzes the connotation of CSCL teaching mode, and makes full use of the collaborative and diversified nature of CSCL teaching mode to put forward the practical strategy of applying CSCL teaching mode for English majors. Then it measures the characteristics of students' learning behaviors. Then an improved density clustering algorithm is proposed, on the basis of which the collaborative learning cluster construction algorithm is further designed. According to the clustering algorithm, learners are clustered, and heterogeneous grouping is realized by randomly selecting groups of learners in the basic clusters and regulating the groups with regulated clustering. The article concludes with the practical application of the practical strategies of the CSCL teaching model and the collaborative learning grouping with learners as the experimental objects, and also analyzes the impact of the computer-supported collaborative learning method based on the data mining technology on the four dimensions of students' intercultural communication competence, namely, knowledge, awareness, emotion and behavior, through the independent samples t-test and the paired samples test on the results of the experimental study.

2. Application of the CSCL model of English language teaching and learning

2.1. Connotation of CSCL Teaching Model

The CSCL teaching model is based on constructivism theory, focuses on the establishment and improvement of the application system of collaborative learning environment, and carries out the task-driven practical teaching experience through online live teaching.The CSCL teaching model is originated from multimedia information technology, and integrates the practice of computer-assisted collaborative learning with multiple disciplines, such as linguistics, pedagogy, psychology, etc., so as to provide an innovative example for remote contextualized teaching and learning. It provides an innovative example of distance contextualized teaching.

The CSCL teaching model embodies the technological advantages of collaboration, participation, and sharing in practical application, and obtains dynamic generative data through collaborative model construction, knowledge map visualization operation, and the use of collaborative tools, so as to creatively build the CSCL adaptive learning system, and improve the controllability and operability of the process of English teaching and learning. On the technical level, the use of virtual tutors and real-time interactive machines highlights the intelligent attributes of English language teaching, and the use of CSCL hands-on operational environments can facilitate the development of collaborative scripts within language learners [31]. The configuration of components such as scenarios, roles, and tasks enriches the functional attributes of online teaching and learning, and learners can rely on different ways of knowledge representation to carry out fine-grained collaboration, laying the theoretical and practical foundation for realizing the collaboration paradigm at a high cognitive level. The architecture of the CSCL teaching model is shown in Fig. 1.

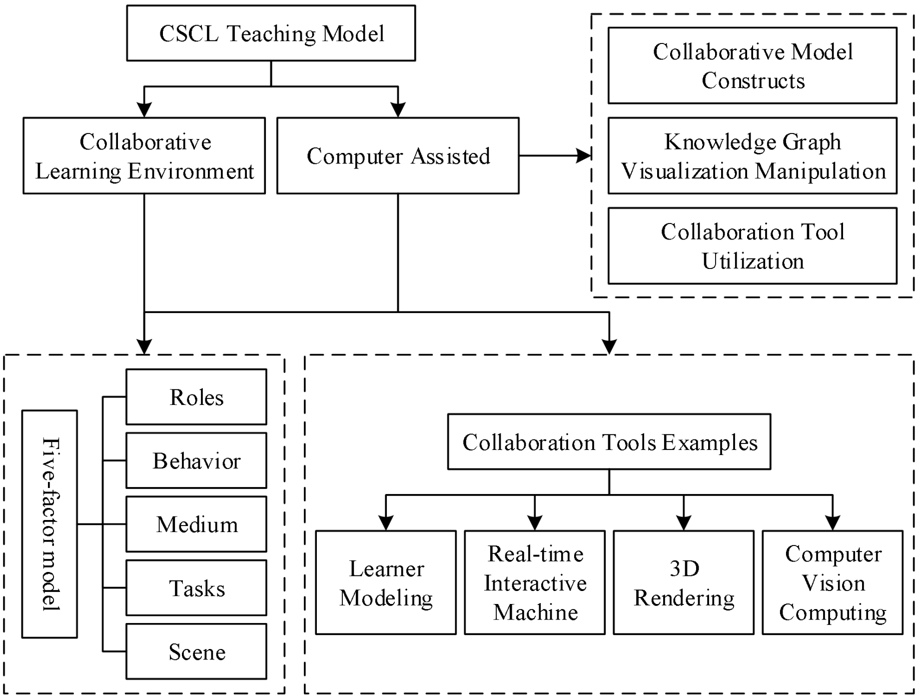


Figure 1. CSCL teaching pattern architecture

2.2. Practical Strategies of Applying CSCL Teaching Model for English Majors

2.2.1. Selection of quality information resources

The selection of English course resources is crucial and has an important impact on the cultivation of students' intercultural communication skills. In view of the current English major course content knowledge points in colleges and universities, teachers can choose some professional educational websites when collecting teaching resources. At the same time, major universities should encourage English major teachers to build their own online courses that meet the characteristics and levels of their own students, cultivate high-quality courses in a gradual manner, form a echelon of college-level-school-level-municipal-provincial-national high-quality online courses, and update and optimize the online courses of English majors continuously.

2.2.2. Scientific division of collaborative groups

The division of computer group collaboration groups will be involved in the later group learning effect of the comprehensive evaluation of the competition, therefore, the use of CSCL mode of teaching teachers in the division of the group must ensure that the fairness of the starting level of each group are on a level. Based on this, teachers need to let students fill out a questionnaire on the CSCL system, which will be accompanied by a mapping exercise. After students fill out the questionnaire, the system can evaluate each student's receptive and communicative abilities based on their grades and time spent on answering the questions, and then divide them into collaborative learning groups of 4-6 students with similar strengths (see Section III for specific methods). Each group should recommend a leader who is responsible for the subsequent division of writing tasks, monitoring the progress of learning tasks, and reporting and summarizing work. In this way, each student will have a corresponding sub-task within the group, and if they encounter any problems in the process of completing the task, they can also seek advice from the teacher through the CSCL platform, so that the whole group will focus on the theme of learning to carry out unremitting exploration. The main teaching mode is shown in the figure below. The English teaching mode based on the CSCL concept is shown in Figure 2.

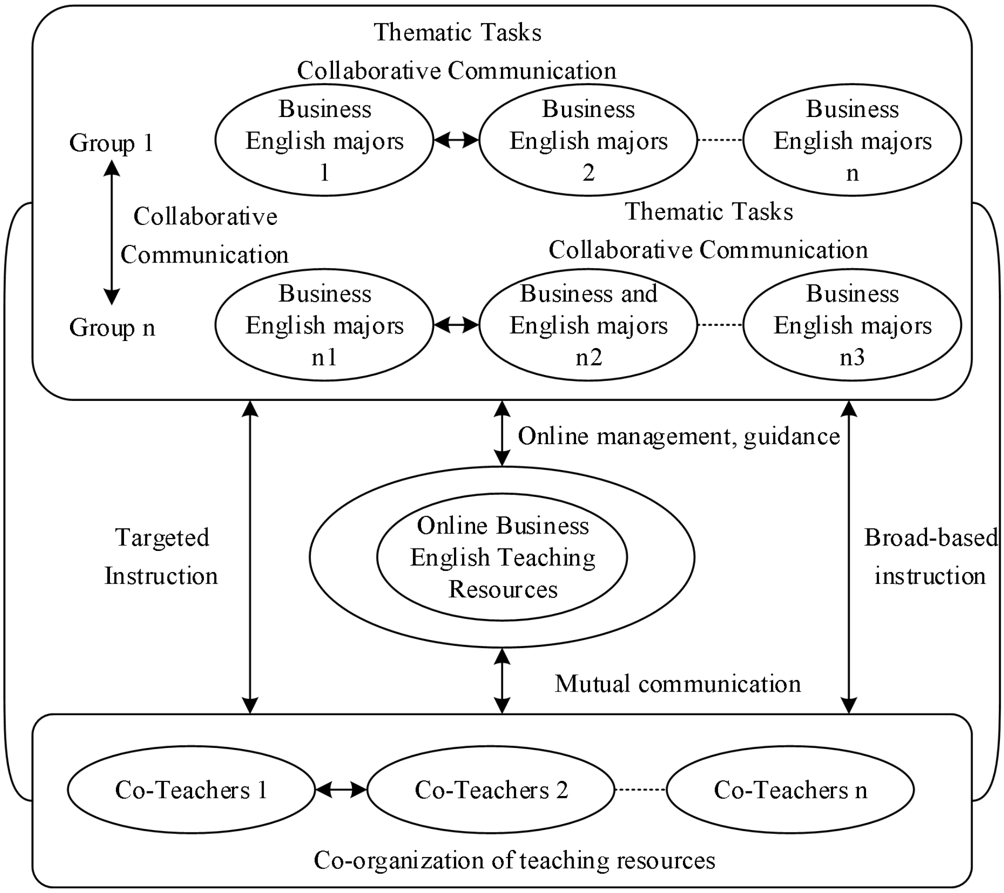


Figure 2. English teaching model based on CSCL concept

(1) Determine learning themes and tasks and create corresponding teaching contexts

After organizing the corresponding teaching resources according to the characteristics of English majors, teachers need to determine the corresponding CSCL learning themes and learning tasks, so as to lay a good foundation for the subsequent creation of teaching situations. Teachers need to solicit students' opinions before determining the business theme, after all, only students who are really interested will be fully engaged in the CSCL theme exploration. Students will actively engage in CSCL theme exploration driven by interest, at this time the teacher needs to prepare for the next learning process of students to build their own knowledge system, one week before the official start of the class issued learning tasks, covering the learning content, learning programs and learning objectives, in the CSCL platform to create a teaching situation of intercultural communication, for the smooth development of CSCL learning to lay a good foundation, while at the same time facilitates the cultivation of students' cross-cultural communication skills.

(2) Comprehensively evaluating the teaching effect and reflecting on the deficiencies of the learning process

Reasonable teaching evaluation can help college students better recognize their own deficiencies in knowledge mastery, and it can also point out the direction for teachers to focus on the next step of teaching, thus further assisting the cultivation of students' intercultural communication skills. The advantages and disadvantages of the teaching evaluation mode will greatly affect the effectiveness of CSCL teaching in English majors. Therefore, in the CSCL mode of teaching, we should not only evaluate the learning results, but also evaluate the whole process of group collaboration and the individual performance of each student, find out the problems existing in the current implementation of the CSCL mode of teaching in the evaluation, and put forward more opinions conducive to the improvement of the collaborative learning mode. Based on this, after the completion of each CSCL business theme exploration, teachers should ask students through online and offline channels to participate in the collaborative learning process of individual evaluation, group evaluation and group summary, and make targeted adjustments in the development of the next CSCL learning arrangements, so as to continuously improve the application of the CSCL learning model in the teaching of English majors [32]. In addition, teachers also need to assess students' English communication ability from the frequency of email sending, the time spent in small discussions, the number of questions asked and answered, and take this as one of the evaluation indexes of CSCL learning effect.

(3) Playing the role of teachers' guidance and creating conditions for independent learning

In the process of applying the CSCL teaching mode in English in colleges and universities, the guiding role of teachers must be emphasized. Although the knowledge covered in this teaching mode focuses on students solving problems in independent inquiry and cooperative communication, teachers still play a crucial role in it. First, teachers are responsible for the collection and organization of learning resources, and are the most important developers and designers of CSCL learning topics, creating a good environment for students to learn independently and work together. Secondly, when students encounter problems that cannot be solved by teamwork in CSCL learning, they need to ask teachers for guidance through e-mail or video. Once again, teachers need to regularly check the progress of students' CSCL learning topic exploration, monitor students' efforts to complete their own tasks in this learning model, and be the leading analyzer and evaluator of evaluation and judgmental questioning. Finally, when a group is slow to make progress because of task exploration, teachers also need to inspire the collaborative group learning atmosphere in a timely manner and help them mediate conflicts among themselves.

3. Methodology for building collaborative learning clusters based on learning behaviours

3.1. Extraction and Measurement of Students' Learning Behavioral Characteristics

3.1.1. Extraction of behavioral features

This section identifies five dimensions of learning behavior. The ELA performance dimension includes quiz score normalization and quiz score ranking. The learning time preference dimension includes eight behavioral characteristics including homework completion in the early morning, homework completion in the morning, coursework completion in class, and video completion in class. The task interaction dimension includes 10 behavioral characteristics including completing homework, completing coursework, completing videos, quiz completion delay index, check-in completion delay index, etc. The interpersonal interaction dimension includes participating in discussions and posting discussions. The system interaction dimension includes the number of times logging into the system and the number of days accessing the system. It is easier to obtain data about students' interaction time in online learning environments, but less interaction time data have been investigated in previous studies [33]. In this paper, the behavioral features extracted from the learning time preference dimension, as well as the completion of homework, completion of courseware, completion of videos, quiz completion delay index, and check-in completion delay index, are behavioral features about interaction time. These behavioral features are measured by the initiation time, deadline time, and submission time of the interaction behaviors, which can better measure the students' behavior of completing learning tasks. By proposing a new method to quantify the behavioral features, constructing metrics with the time data of completion behaviors and then extracting the behavioral features, we can better analyze the online learning characteristics of students, which is helpful to provide reasonable suggestions for students to improve their learning results and help them to get better academic performance.

3.1.2. Measures of Behavioral Characteristics

In this paper, we use Python 3.7 and SPSS 20.0 software to measure the constructed behavioral characteristics of online learning, and the processing methods are summarized as follows:

(1) Behavioral features of interpersonal interaction dimension and system interaction dimension as well are not processed and raw data are used. Missing values are replaced with 0.

(2) Normalization of quiz scores indicates how high or low a student scores on a quiz and is measured according to the max-min normalization method. The calculation process of max-min normalization is shown in Equation (1):

$$X_{normal} = \frac{X - X_{\max}}{X_{\max} - X_{\min}} \tag{1}$$

where X is the original data, $X_{\min}$ is the minimum value in the original data set, and $X_{\max}$ is the maximum value in the data set. The higher the quiz score, the larger the normalized value.

(3) Extraction of Behavioral Features of Learning Time Preference

In order to analyze students' online English learning time preference, this paper extracts the following eight behavioral features based on three metrics, i.e., homework completion time, courseware completion time and video completion time:

Homework completion in the early morning $\left(Homework_{before\_dawn}\right)$, homework completion in the morning $\left(Homework_{morning}\right)$, homework completion in the afternoon $\left(Homework_{afternoon}\right)$, homework completion in the evening $\left(Homework_{evening}\right)$, coursework completion in class $\left(Courseware_{in\_class}\right)$, coursework completion in class $\left(Courseware_{after\_class}\right)$, video completion in class $\left(Video_{in\_class}\right)$, video completion in class $\left(Video_{after\_class}\right)$. The specific metric process is as follows:

Step 1: The values of initial learning time preference behavioral characteristics are all 0.

Step 2: Enter the task completion time $assigement_{time}$, and determine whether the learning task is homework, courseware or video. If it is a homework task, go to Step 3. If it is a courseware or video task, go to Step 4.

Step 3: The input is a homework task, judge $assigement_{time}$ during the early morning period T1 (23:00-7:00), the morning period T2 (7:00-12:30), the afternoon period T3 (12:30-18:00) or the night period T4 (18:00-23:00).

If $assigement_{time} \in T1$, then $Homework_{dawn} += 1$.

If $assigement_{time} \in T2$, then $Homework_{morning} += 1$.

If $assigement_{time} \in T3$, then $Homework_{afternoon} += 1$.

If $assigement_{time} \in T4$, then $Homework_{evening} += 1$.

In all other cases, the assignment behavior feature value is 0.

Step 4: The input is whether the task is a courseware or a video. When the learning task is a courseware

If $assigement_{time}$ is completed in class, then $Courseware_{in\_class} += 1$.

If $assigement_{time}$ is completed during class, then $Courseware_{after\_class} += 1$.

For all other cases, the classroom behavioral feature is 0. Similarly, extract to $Video_{in\_class}$ and $Video_{after\_class}$.

Step 5: Output the extracted study time preference behavioral features.

(4) Metrics for completing homework, completing coursework, and completing video behavioral traits require metrics $time\_n$, $threshold\_n1$, and $threshold\_n2$, where $time\_n$ is the submission time of the student's n th task, $threshold\_n1$ is the 1st threshold value for the n th task, and $threshold\_n2$ is the 2nd threshold value for the n th task. Missing values are replaced by the deadline time of the corresponding task. The specific metric process is as follows:

Step 1: Let the student's n st assignment have a grade of n_level. Enter $time\_n$, $threshold\_n1$, and $threshold\_n2$.

Step 2: Determine how $time\_n$ relates to the critical values $threshold\_n1$ and $threshold\_n2$.

If $time\_n < threshold\_n1$, then $n\_level = 3$.

If $threshold\_n1 < time\_n < threshold\_n2$, then $n\_level = 2$. Otherwise, $n\_level = 1$.

Step 3: Totalize the n_level obtained from all the tasks and finally get the behavioral characteristic values of completing homework, completing courseware and completing video.

(5) The quiz procrastination index and check-in procrastination index are defined using the procrastination index (TS). See formula (2) for details:

$$TS = \sum_{n=1}^{i}\left(1 - \frac{tq_n - te_n}{tf_n - tq_n}\right) \tag{2}$$

Where i is the total number of tasks. te_n is the release time of the n rd task. tf_n is the deadline for the n th task. tq_n is the completion time of the n th task.

After the above processing, the values of all behavioral features except quiz score ranking and learning task completion ranking are in line with the law of increasing function. For example, the larger the value of the behavioral feature of completing courseware, the less time the student used to complete the courseware task and the less procrastination. The smaller the value of the complete video behavioral trait, the more time the student used to complete the video task and the more he/she procrastinated. And the quiz score ranking indicates that the smaller the value, the higher the quiz score and the higher the ranking of the student.

3.2. Density Clustering Based Approach for Collaborative Learning Cluster Construction

3.2.1. Problem description and related definitions

(1) Problem description

In collaborative learning, in order to achieve the goal of matching the comprehensive ability of individual learners with the balanced development of all learners, the principle of "heterogeneity within groups and homogeneity between groups" is generally followed in group construction [34]. The group construction algorithm proposed in this paper first divides the learners through clustering, and then selects a suitable learner from each cluster to form a group, so that the group basically meets the above requirements.

(2) Relevant definitions

Definition 1 A grouping space L is a quaternion $L=(S,C,V,F)$ where S is the set of all learners. C is the set of feature attributes of learners. V is the value domain of the feature attributes. $F: S\times C\rightarrow V$ denotes the feature function.

Definition 2 Assuming that the number of learners in the grouping space is $N(N=|S|)$ and the size of the set of characteristic attributes is $T(T=|C|)$, the values of characteristic attributes can be organized into a $N\times T$ matrix called the characteristic attribute value matrix M :

$$M=\begin{bmatrix} V_1 \\ V_2 \\ \cdots \\ V_N \end{bmatrix}=\begin{bmatrix} V(1,1) & V(1,2) & \cdots & V(1,T) \\ V(2,1) & V(2,2) & \cdots & V(2,T) \\ \cdots & \cdots & \cdots & \cdots \\ V(N,1) & V(N,2) & \cdots & V(N,T) \end{bmatrix} \tag{3}$$

Definition 3 organizes the feature attributes into a feature attribute semantic tree based on their semantic relationships, then the correlation CD of two feature attributes $C(i)$ and $C(j)$ is defined as:

$$CD(i,j)=1-\frac{Dist(C(i),C(j))-1}{Depth(C(i))+Depth(C(j))} \tag{4}$$

Definition 4 Learner distance is defined as the degree of dissimilarity between the characteristic attributes of two learners, which can be defined using the Euclidean distance calculation method with moderating factors:

$$d(x,y)=\sqrt{\sum_{i=1}^{T}(\lambda_i*\sum_{j=1}^{T}(CD(i,j)*(v_{x,i,j}-v_{y,i,j})^2))} \tag{5}$$

Definition 5 The learner distance matrix is defined as the matrix that records the distance between any two learners S_i, S_j in the grouping space and can be expressed as:

$$MD=\begin{bmatrix} d(1,1) & d(1,2) & \cdots & d(1,N) \\ d(2,1) & d(2,2) & \cdots & d(2,N) \\ \cdots & \cdots & \cdots & \cdots \\ d(N,1) & d(N,2) & \cdots & d(N,N) \end{bmatrix} \tag{6}$$

Definition 6 k Nearest-neighbor density is defined as the inverse of the mean of the distances between a learner and the other k nearest learners in the grouping space. Assuming that $y_1,y_2,\cdots,y_k$ is the number of the k closest learner to a learner $S(x)$ in the set of learners S, the k-neighborhood density of a learner $S(x)$ is $D(x,k)=\frac{1}{\frac{1}{k}(\sum_{i=1}^{k}d(x,y_i))}$. In the case where the value of k is determined, $D(x,k)$ it can be abbreviated as D_x.

Definition 7 Let $\{q_i\}_{i=1}^N$ be the set of subscript sequences (satisfying $D_{q_1} \geq D_{q_2} \geq \cdots \geq D_{q_N}$) of the set $\{D_i\}_{i=1}^N$ of learner densities in descending order of size, and the distance δ_i (which can be called the learner's clustering distance) from any learner in a given clustering cluster to the other clusters is defined as $\delta_i = \begin{cases} \min_{j<i}\{d_{q_i q_j}\}, & i \geq 2 \\ \max_{j\geq 2}\{\delta_{q_j}\}, & i = 1 \end{cases}$.

Definition 8 With subgroup g as a separate subgroup space L_g and the learner distance matrix MD_g of subgroup g can be expressed as the adjacency matrix of the connectivity graph G, the group quality of subgroup g is defined as the mean of the weights of the minimum spanning tree of G, which is used as a measure of the subgroup's quality, and is abbreviated as $Q(g)$.

Definition 9 The grouping space L is divided into R subgroups G and each subgroup g has a group mass of $Q(g)$, $R \geq g \geq 1$. In this grouping case, the mean of the group mass is $MQ = \sum_{g=1}^{R} Q(g)$ and the group mass variance is $DG(G) = (\sum_{g=1}^{R}(Q(g) - MQ)^2)/R$.

3.2.2. Description of the methodology

(1) Preprocessing

The main task of the preprocessing module is to generate the dataset-learner distance matrix required by the cluster division and grouping construction module, and the preprocessing process is shown in Figure 3. The preprocessing module first organizes the feature attribute semantic tree, grouping space and grouping target. Then the feature attribute similarity matrix CDM is calculated from the feature attribute semantic tree according to Definition 3, and the learner feature attribute value matrix M is constructed from the grouping space L. At the same time, in order to optimize multi-objective, the influence degree of different feature attributes in the grouping is differentiated, and therefore each feature attribute adjustment factor is formed according to the adjustment of the grouping objective λ_i. Finally, the matrices M, N, and λ_i are used as inputs to generate the learner distance matrix. Finally, the learner distance matrix MD is generated by the learner distance calculation function.

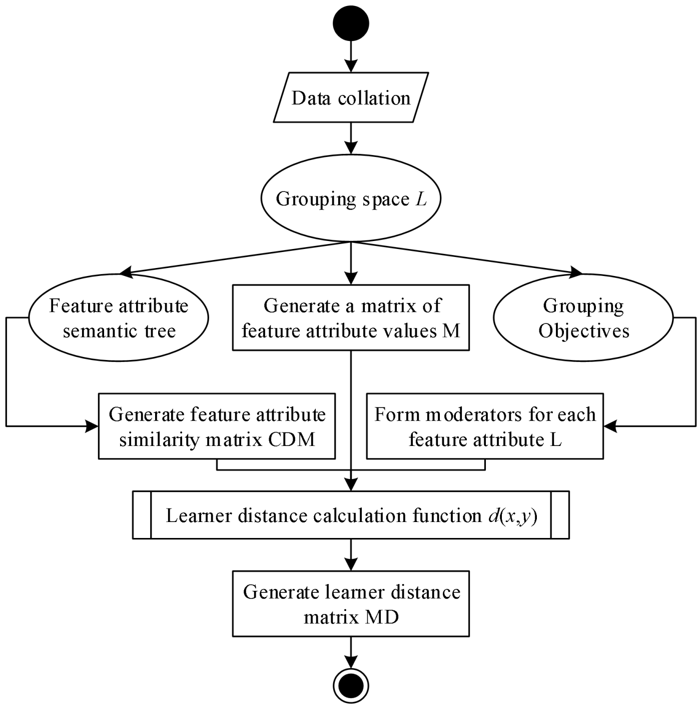


Figure 3. Pretreatment process

(2) Cluster division

The clustering division process is shown in Fig. 4. The learner distance matrix generated by the preprocessing module is used as the data source, and input the planning clustering cluster size cn and group size k to satisfy $cn = \lfloor N/k \rfloor$. Here k represents the group size, and its value is equal to k in the k-nearest-neighbor density computation. The learner density $\{D_i\}_{i=1}^{N}$ is computed sequentially, and the subscript sequence $\{q_i\}_{i=1}^{N}$ is obtained by arranging D_i in descending order. Next, the learner clustering distance is computed to produce set $\{\delta_i\}_{i=1}^{N}$ according to Eq. Then the set of decision parameters $\{D_i \times \delta_i\}_{i=1}^{N}$ is generated, the largest element $\max(D_i \times \delta_i)$ in the set is identified, and the closest cn learners are identified using their corresponding learner $S[i]$ as the clustering center to form clusters. After the new clusters are formed, the rows and columns corresponding to the learners contained in the new clusters are deleted from the learner distance matrix MD to form a new MD. Check the size of MD, if it is greater than cn then go to the next clustering from the beginning of the learner density calculation, otherwise the corresponding learner in the MD is used as a new cluster and the clustering is completed.

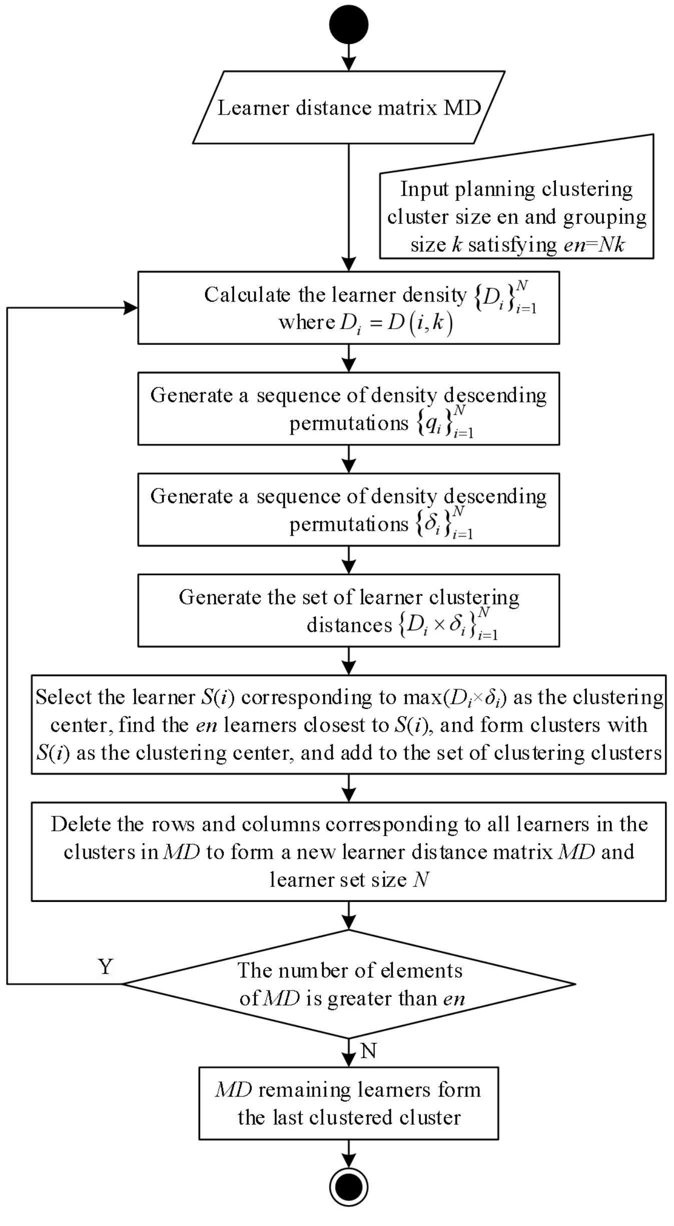


Figure 4. Clustering process

(3) Group construction

After clustering according to the above algorithm, generally speaking, the cluster center density of the last few clusters may have an “abnormal” situation, and this “abnormal” situation can be used to adjust the quality of the final group.

Using the learner distance matrix MD and the generated set of clustering clusters $\{SC_i : 1 \le i \le k\}$ as the data source, input the grouping adjustment factor ξ, calculate the density DC_i and mean DM of each cluster center, and classify $\{SC_i\}$ into two categories according to whether DC_i is greater than $DM \times \xi$, which are the basic cluster set $\{SCI\}$ when $DC_i \ge DM \times \xi$ and $|SCI| = cn$, and the adjustment cluster set $\{SCg\}$ when $DC_g < DM \times \xi$ or $|Cg| < cn$. The grouping construction principles are: ① Each learner can only be assigned to one grouping. ② Each subgroup contains at least k learners. ③ The first $|\{Cl\}|$ members of subgroup g_i consist of one learner member from each of the clusters in the basic cluster set $\{Cl\}$, respectively. ④ The remaining members of grouping g_i are composed of one learner member taken from each cluster of the regulating cluster set $\{Cg\}$ respectively, so that the grouping satisfies the requirements of Principle 5. ⑤ The group quality of all the final subgroups satisfies the minimum variance or less than the set value.

4. Application and analysis of CSCL-based English teaching model

4.1. CSCL Grouping Strategy and Role Design Case Application

4.1.1. Experimental Objects

The subjects selected for this section of the experiment were students in an English teaching class in a school with 56 students, of which 31 were boys and 25 were girls.

4.1.2. Personalized Hybrid Packet Strategy for CSCLs

(1) Pre-processing data

The 56 students in this experiment were numbered according to the last three digits of the school number and organized to generate the learner information of the whole class.

(2) Clustering

In one clustering, there is no need to specify the number of pre-demarcated clusters, using the DBSCAN algorithm after one clustering, the 56 students were divided into four clusters, numbered G1, G2, G3 and G4, with the number of people as G1-12 people, G2-18 people, G3-14 people and G4 -12 people. Then density clustering was performed for G1, G2, G3, and G4 by combining the knowledge level and online interaction degree, respectively. It has been shown that the number of group members should not be too many or too few, according to the needs of subsequent research, it is proposed to divide 56 people into 9 groups, based on the results of the primary clustering to give the K value of the secondary clustering are: KG1=5, KG2=6, KG3=7, KG4=6, respectively.

(3) Cluster division

After two clustering, we can get 4 large clusters after the first clustering and 24 small clusters after the second clustering. According to the principle of “homogeneous between groups, heterogeneous within groups”, group division firstly ensures that the members of each group must belong to the same large cluster, and try not to belong to the same small cluster, and at the same time ensure that the gender of each group is not homogeneous. Taking G1 obtained from primary clustering as an example, after undergoing secondary clustering, the G1 cluster is further clustered into five smaller clusters, G1-1, G1-2, G1-3, G1-4, and G1-5, and the group division is carried out by starting from the node with the largest number of samples in the cluster, one person is selected, and then traversing each cluster in turn, one person is selected from each to reach a total of five people in the group. Due to the uneven number of each cluster obtained from clustering, certain groups were underrepresented, at which point the remaining students from that large cluster needed to be selected and placed into that group until there were no ungrouped students. The final division was Group A to Group I totaling 9 groups.

4.1.3. Implementation of CSCL Collaborative Group Role Scaffolding

In the process of constructing the role scaffolds for group collaboration, it was necessary to combine the individual and the group as a whole to ensure that each group contained the five designed roles and that each member assumed a role. The social network was used to analyze the text and data of group interactions, and the results of the social network analysis showed the degree of access, proximity to the center of access, and mediated centrality of group interactions. Entry shows how much attention a person gets. Out-degree expresses the degree to which a person pays attention to others. Incoming centrality expresses integrative power. Out proximity centrality expresses radiative power. The roles are assigned as follows: "Questioner" and "Problem Analyst" need to have a high level of participation, that is, both out and out of the center are at the high or medium level of the group. The "moderator" is the core figure of group collaboration, and needs to be played by people who are at a high level of all parameters in the group, that is, all parameters are at the middle and high level of the group. The "opinion demonstrator" needs to be a person with a high degree of influence in the group, who has a certain personal opinion on the topic raised, that is, the parameters of the degree of entry and the degree of centrality are at the high level of the group. "Summarizer" is more suitable for people with low levels of various coefficients, usually this type of students are difficult to integrate into close interaction, can listen to the views of others, summarize and summarize knowledge points, and gradually integrate into group interaction.

(1) First round of practice

The duration of this round of collaborative activities is one week, using free combination of grouping, collecting the number of interactions during the group's online and offline collaboration as well as the text of the conversation, and analyzing the degree of collaborative interaction in the free grouping. For example, the total number of effective interaction posts of a group with a better degree of interaction in this round is 33, and the overall network density is 0.4233. It can be seen that the students in the group in this round do not have a high degree of motivation to participate in the interaction, and all of them are inclined to interact with one or two fixed members in the group, and do not form an overall collaborative network of the group.

(2) Second round of practice

In this round, the group was divided according to the practice strategy of CSCL teaching model, and the text dialogue data of the group under the personalized grouping strategy were collected to compare the first round of free grouping and interaction and to verify the effectiveness of personalized grouping in promoting group collaboration. After completing the group division according to the personalized grouping strategy, the group collaborative problem solving is normally carried out in the form of online and offline discussion, and the length of the discussion is one week, and the text data generated by each group online and offline are collected for content analysis to derive the second round of collaborative interaction. Take group A as an example, the second round of collaborative interaction in group A is shown in Table 1, the total number of effective interaction posts in this round is 50, compared with the number of interactions in the first round has been improved, and at the same time, it can be seen that some of the students have produced meaningful in-depth exchanges between them, but there are still “marginal students”. The overall network density in this round is 0.5302, and the interaction density of each group has improved.

Table 1. A group of two rounds of collaborative interaction

NUM	OUTDEGREE	INDEGREE	INCLOSENESS	OUTCLOSENESS	BETWEENNESS
268	19	14	100	100	2.253
271	13	6	83.225	83.225	1.361
264	5	13	83.225	71.254	0.961
269	8	3	71.254	83.225	1.361
266	2	10	83.225	71.254	0.961
270	3	6	71.615	83.225	2.253

(3) Third round of practice

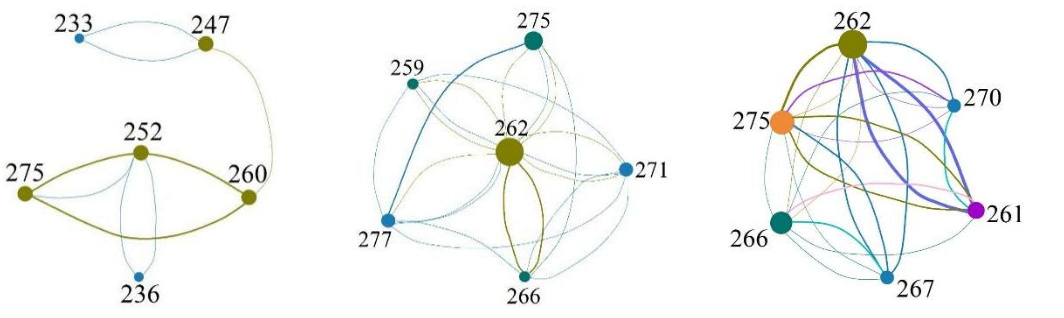
In this round, group role scaffolds are constructed on the basis of personalized grouping, and in the process of carrying out collaborative interactions, the degree of interaction in the first two rounds is compared to verify the effectiveness of role scaffolds in facilitating group interactions. Before conducting this round of collaborative activities, the instructor assigns appropriate roles to group members based on the second round of collaborative interaction table and the principles of role assignment, develops role scaffolds for each group, and the group members fulfill their roles in collaborative activities based on the role scaffolds for their group. Taking Group A as an example, the role scaffolds of Group 5A were designed according to the second round of collaborative interactions and the principles of role allocation, and then the social network analysis was conducted on the number of posts and the content of discussion posts in the collaborative process.The third round of collaborative interactions of Group A is shown in Table 2. During this round of collaboration, the total number of effective interaction posts of Group A is 112, which is a significant increase in the number, and the participation, influence and mediation of all types of roles have been increased to some extent. The overall network density was 0.6241, with more frequent connections and closer interactions among group members.

Table 2. A group of third round of collaborative interaction

NUM	OUTDEGREE	INDEGREE	INCLOSENESS	OUTCLOSENESS	BETWEENNESS
268	34	26	100	100	3.211
271	26	14	100	83.225	2
264	24	26	71.254	83.225	0.336
269	12	4	71.254	100.104	1
266	8	15	71.254	71.271	0.336
270	8	16	83.225	62.548	0

(4) Comparative analysis of practice effects

Comprehensive analysis of group collaboration in the three rounds of practice, personalized grouping strategy compared to the traditional no-intervention grouping, highlighting the subjectivity of the learner in the teaching process, the learner has a stronger motivation to learn, and there is a significant increase in the frequency of effective exchanges between members. On this basis, the construction of group role scaffolding has a positive impact on learner interaction, the degree of interaction significantly increased, and the overall tendency to more intensive social interaction, three rounds of Group A collaboration social network comparison shown in Figure 5, Figure (a) ~ (c) are the first to third round of the Group A collaborative SNA, respectively. from the perspective of the network density, a round of practice the overall network density is less than 0.45, which indicates that the pre-group From the network density point of view, the overall network density of one round of practice is less than 0.45, indicating that in the early stage of the group "non-interventionist" collaboration, there is a lack of leadership members and learners with "prestige" to promote the interaction of the group collaboration, while in the middle and late stages of the network density continues to increase, which indicates that there is an increase in the interaction behaviors of learners in the participating groups, and that the frequency of participation and the quality of the postings are both high. The circulation of knowledge is stronger. Therefore, the personalized hybrid grouping strategy and the construction of group role scaffolding can promote collaboration among learners and achieve better learning results.



(a) The first round of collaborative SNA diagrams (b) The second round a group collaboration SNA diagram (c) Third round a group collaboration SNA diagram

Figure 5. Three round a group of collaborative social network comparison

4.2. Clustering-based grouping application for collaborative learning

This section continues the experiment of collaborative learning grouping with 56 learners in the school, using CLGC to divide the learners into 3 collaborative learning groups. There were 3 groups with 3 group members, 3 groups with 4 group members, and 7 groups with 5 group members. The topics assigned to each group were one of the top three wishes of the group members, which better met the learners' freedom of choice. The average time available for group members to participate in collaborative learning was 65 minutes per day, which accounted for about half of the average daily total learning time, ensuring that learners had sufficient time to collaborate to accomplish the learning objectives. The length of time that some of the groups could participate in collaborative learning is shown in Figure 6. The figure shows the length of time that 100 of the groups could engage in collaborative learning. As can be seen from the figure, the length of time available for collaborative learning varies greatly from group to group, mainly because the learners within the selected period of time have different levels of activity, and the grouping algorithm proposed in the paper tends to classify learners with high levels of activity into the same group, so the length of time available for collaborative learning in these groups is generally higher than that in other groups.

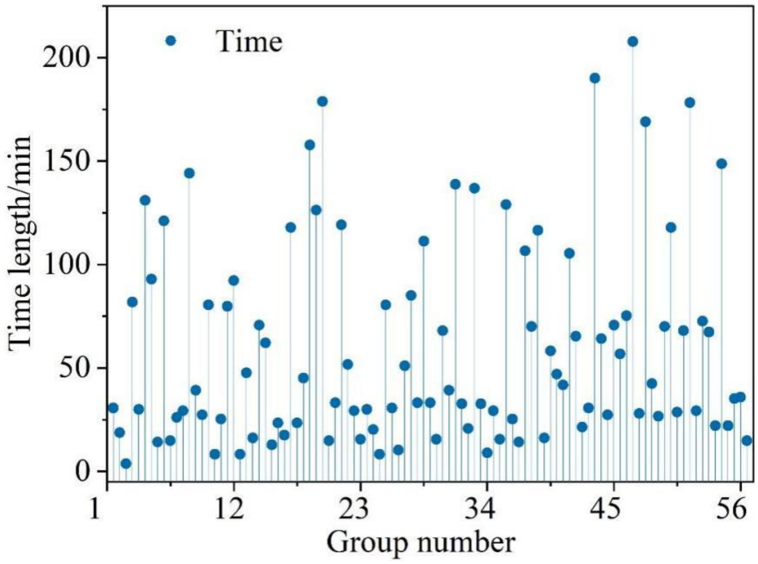


Figure 6. Some groups can participate in collaborative learning hours

4.3. Changes in students' intercultural communication skills

In order to verify whether the computer-supported collaborative learning method based on data mining technology proposed in this paper is beneficial to enhancing students' intercultural communication competence, this section will summarize and analyze the intercultural communication scale data from the pre and post-tests of the experimental and control classes. In this section, another English teaching class chosen as the control class for the pre and post-test study was continued, with a total of 56 students. Among them, the experimental class was taught using the computer-supported collaborative learning method based on data mining technology proposed in this paper, while the control class was taught using the traditional teaching method. It is mainly carried out in the following three aspects:

(1) The intercultural communication competence scales of a total of 112 students from the experimental and control classes were collected and the data were analyzed for mean and standard deviation.

(2) Independent samples t-test was used to compare the pre and post-test data to analyze the differences in intercultural communication competence between the two classes.

(3) Paired samples t-test was used to compare and analyze the pre- and post-test data of the two classes in different dimensions, which is conducive to a fuller and more specific comparison of the impact of teaching computer-supported collaborative learning methodology based on data mining technology in enhancing students' intercultural communication competence.

4.3.1. Analysis of pre-test data for experimental and control classes

The statistical description of the pre-test data of the experimental and control classes is shown in Table 3. As can be seen from the table, the mean values of intercultural communication competence in the knowledge dimension in the experimental and control classes before the experiment were 2.25472 and 2.24898, respectively, and the standard deviations were 0.20376 and 0.20098, respectively.

Table 3. Statistical description of the experimental class and the comparison group

Variable	Group	Number	Mean value	SD	Std.E
Knowledge dimension	Laboratory class	56	2.25472	0.20376	0.02656
	Cross-reference class	56	2.24898	0.20098	0.05499
Consciousness dimension	Laboratory class	56	2.12714	0.31503	0.07667
	Cross-reference class	56	2.18822	0.34845	0.08341

Affective dimension	Laboratory class	56	2.16058	0.27772	0.07342
	Cross-reference class	56	2.24082	0.3823	0.04096
Behavior dimension	Laboratory class	56	2.16094	0.35461	0.10309
	Cross-reference class	56	2.1651	0.35139	0.08346

The independent samples t-test results of the pre-test data of the experimental and control classes are shown in Table 4. The t-test results with equal means show T=0.675, P=0.512, and the P-value is higher than the significance level of 0.05, which shows that there is no significant difference between the two groups of data, so that the two classes' intercultural communication competence level in the dimension of knowledge before the experiment is more or less the same. The mean values of intercultural communication competence in the consciousness dimension of the experimental class and the control class are 2.12714 and 2.18822 respectively, and the standard deviations are 0.31503 and 0.34845, and the results of the t-test of equal means show that T=-1.571, P=0.141, and the P-value is higher than the significance level of 0.05, which indicates that there is no significant difference in the data of the two groups, so it is evident that the experimental class and the control class were about the same level in the knowledge dimension of the two classes before the experiment. The level of intercultural communication competence in the dimension of consciousness is the same. Therefore, by analyzing the data statistics and independent sample t-tests of the knowledge, consciousness, emotion and behavior dimensions of the intercultural communication competence scale of the experimental and control classes before the experiment, it can be seen that the intercultural communication competence levels of the two classes before the experiment are at the same level.

Table 4. Test of t of pre-measured sample

		The same homosexual test		**Average equivalent t test**				
		F	**Sig**	**T**	**df**	**Sig (Double tail)**	**MD**	**Standard error**
Knowledge dimension	Assumed equal variance	2.536	0.097	0.675	102	0.512	0.00574	0.03544
	Unassuming equal variance			0.625	98.431	0.535	0.00574	0.07611
Consciousness dimension	Assumed equal variance	0	1.001	-1.571	102	0.141	-0.06108	0.07008
	Unassuming equal variance			-1.56	107.248	0.115	-0.06108	0.08731
Affective dimension	Assumed equal variance	1.511	0.203	-1.053	102	0.297	-0.08024	0.07268
	Unassuming equal variance			-1.021	105.192	0.304	-0.08024	0.06689
Behavior dimension	Assumed equal variance	0.127	0.734	-0.821	102	0.374	-0.00416	0.05935
	Unassuming equal variance			-0.826	106.685	0.408	-0.00416	0.06031

4.3.2. Post-test data analysis of experimental and control classes

The statistical description of the post-test data of the experimental and control classes is shown in Table 5, where the mean values of intercultural communication competence in the knowledge dimension of the experimental and control classes after the experiment were 3.4542 and 2.2664, with a standard deviation of 0.42385 and 0.45238, respectively.

Table 5. The statistical description of the experimental class and the comparison class

Variable	**Group**	**Number**	**Mean value**	**SD**	**Std.E**
Knowledge dimension	Laboratory class	56	3.4542	0.42385	0.02821
	Cross-reference class	56	2.2664	0.45238	0.08325
Consciousness dimension	Laboratory class	56	3.4100	0.38047	0.02335
	Cross-reference class	56	2.3315	0.32821	0.01516
Affective dimension	Laboratory class	56	3.17297	0.44457	0.11909
	Cross-reference class	56	2.18062	0.40265	0.06008

Behavior dimension	Laboratory class	56	3.24674	0.40073	0.04888
	Cross-reference class	56	2.28043	0.39151	0.07304

The post-test independent samples t-test for the experimental and control classes is shown in Table 6. As can be seen from the table, the t-test with equal means shows T=14.899, P=0.000, and the P-value is lower than the significance level of 0.05, so the difference between the experimental class and the control class in intercultural communication competence in the knowledge dimension after the experiment is relatively significant, and the experimental students' scores are higher than those of the control class. The mean values of intercultural communication competence in the consciousness dimension of the experimental class and the control class after the experiment are 3.4100 and 2.3315 respectively, and the standard deviations are 0.38047 and 0.32821 respectively, and the results of the t-test of equal means show that T=15.798, P=0.000,P-value is lower than the significance level of 0.05, so it can be concluded that the difference between the experimental class and the control class in intercultural communication competence in the consciousness dimension after the experiment is relatively significant, and the scores of students in the experimental class are higher than those in the control class. Awareness dimension is relatively significant, and the scores of the students in the experimental class are higher than those of the control class. Therefore, after the experiment, there is a significant difference between the experimental class and the control class in the four dimensions of intercultural communication competence in terms of knowledge, consciousness, emotion and behavior, and the experimental class students have higher intercultural communication competence than the control class students. This shows that the computer-supported writing learning method based on data mining technology is effective in enhancing students' intercultural communication competence.

Table 6. Test of t of post-measured sample

		The same homosexual test		**Average equivalent t test**				
		F	**Sig**	**T**	**df**	**Sig (Double tail)**	**MD**	**Standard error**
Knowledge dimension	Assumed equal variance	1.582	0.217	14.899	102	0	1.1878	0.11238
	Unassuming equal variance			14.182	108.605	0	1.1878	0.10054
Consciousness dimension	Assumed equal variance	0.159	0.709	15.798	102	0	1.0785	0.07275
	Unassuming equal variance			15.878	104.548	0	1.0785	0.10553
Affective dimension	Assumed equal variance	0.985	0.33	10.693	102	0	0.99235	0.06598
	Unassuming equal variance			10.581	100.856	0	0.99235	0.10939
Behavior dimension	Assumed equal variance	0.144	0.687	12.122	102	0	0.96631	0.0926
	Unassuming equal variance			13.236	105.92	0	0.96631	0.09268

4.3.3. Changes in different dimensions of intercultural communication competence in experimental and control classes

(1) Changes in the knowledge dimension of intercultural communication competence

First, from the independent samples t-test, the experimental class and the control class are at the same level in the knowledge dimension before the experiment, but the difference is obvious after the experiment, with the mean value of 3.4542 and 2.2664, respectively, and the standard deviation of 0.42385 and 0.45238, respectively, and the results of the t-test of the equality of the means show that T=14.899, P=0.000, and the P-value is lower than the significance level of 0.05. So the difference between the experimental class and the control class of intercultural communication competence in the knowledge dimension after the experiment is relatively significant, and the level of intercultural communication competence of the students in the experimental class after the experiment is significantly higher than that before the experiment.

Second, the paired-sample test for the knowledge dimension of the experimental and control classes is shown in Table 7. From the paired samples t-test, the results of the paired samples t-test of the control class in the pre- and post-test of the knowledge dimension show T=-0.05, P=0.937, and the P-value is higher than the significance of 0.05, so that there is no significant difference in the pre- and post-test scores of the control class in the knowledge dimension. The results of paired-sample t-test for the pre and post-test of knowledge dimension in the experimental class show T=-18.605, P=0.000, P-value is lower than significance 0.05, which shows that there is a significant difference in the pre and post-test scores of the experimental class before and after the experiment in the knowledge dimension, and that the post-test scores of the experimental class students are better than the pre-test scores, so that the computer-supported learning method of writing based on the data mining technology has a positive effect on the students' cross-cultural communication competency enhancement is positive.

Table 7. Knowledge dimension matching sample test

		Pair difference						
					The difference is 95% confidence interval			
		Mean value	**SD**	**Standard error mean**	**Lower limit**	**Upper limit**	**T**	**Sig (Double tail)**
Laboratory class	Pre-test of knowledge dimension	-1.1995	0.4728	0.0496	-1.3294	-1.026	-18.605	0.000
	Post-test of knowledge dimension							
Cross-reference class	Pre-test of knowledge dimension	-0.0174	0.4522	0.0874	-0.1	0.1165	-0.05	0.937
	Post-test of knowledge dimension							

(2) Changes in the awareness dimension of intercultural communication competence

First, from the independent samples t-test, there is no significant difference between the experimental class and the control class in the awareness dimension before the experiment, but the difference between the two classes after the experiment is obvious, with the mean values of 3.4100 and 2.3315, and the standard deviations of 0.38047 and 0.32821, respectively, and the results of the t-test of the equality of the means show that the P=0.000,P-value is lower than the significance level of 0.05, so In the dimension of awareness, the level of intercultural communication competence of the students in the experimental class is higher than that of the control class.

Second, the paired samples test for the consciousness dimension of the experimental and control classes is shown in Table 8. From the table paired samples t-test, the paired samples t-test result of the pre and post-test of the consciousness dimension of the control class shows T=-1.231, P=0.238, and the P-value is higher than the significance of 0.05, so there is no significant difference in the pre and post-test scores of the control class in the consciousness dimension. The paired samples t-test results of the pre and post-tests of the consciousness dimension in the experimental class showed T=-17.641, P=0.000, P value lower than significance 0.05. It can be seen that there is a significant difference in the pre and post-test scores of the students in the experimental class before and after the experiment in the consciousness dimension, that is to say, the post-test scores of the experimental class students in the consciousness dimension are significantly better than the pre-test scores, so it shows that the computer-supported writing based on the data mining techniques learning method helps students to improve their intercultural communication skills.

Table 8. Test of the experiment class and the comparison class consciousness dimension

		Pair difference							
					The difference is 95% confidence interval				
		Mean value	**SD**	**Standard error mean**	**Lower limit**	**Upper limit**	**T**	**df**	**Sig (Double tail)**
Laboratory class	Pre-test of Consciousness dimension	-1.2829	0.5122	0.0442	-1.355	-1.1042	-17.641	50	0.000
	Post-test of Consciousness dimension								
Cross-refere	Pre-test of	-0.1433	0.454	0.0721	-0.1861	0.0327	-1.231	52	0.238

nce class	Consciousness dimension Post-test of Consciousness dimension								

(3) Changes in the emotional dimension of intercultural communication competence

First, from the independent samples t-test, the level of intercultural communication competence in the emotional dimension between the experimental class and the control class before the experiment is almost the same, but the difference between the two classes after the experiment is obvious, with the mean values of 3.17297 and 2.18062, and the standard deviations of 0.44457 and 0.40265, respectively, and the results of the t-test of the equality of the mean values show that the T = 10.693, P = 0.000,P-value is lower than the significance level of 0.05, so the students in the experimental class are higher than the control class in the affective dimension.

Secondly, the paired samples test for the affective dimension of the experimental and control classes is shown in Table 9. From the paired samples t-test, the paired samples t-test result of the pre and post-test of the affective dimension of the control class shows T=1.143, P=0.227, and the P-value is higher than the level of significance of 0.05, so there is no significant difference in the pre and post-test scores of the control class in the affective dimension. The paired-sample t-test results of the pre and post-tests of the emotional dimension of the experimental class show T=-13.853, P=0.000, P value is lower than the significance 0.05, which can be obtained that there is a significant difference in the pre and post-test scores of the experimental class before and after the experiments in the emotional dimension, indicating that the post-test scores of the experimental class in the emotional dimension of the students' cross-cultural communication competence are significantly better than the pre-test scores, which indicates that the computerized based on the data mining technology This shows that the computerized writing support learning method based on data mining technology plays a very important role in improving students' intercultural communication ability.

Table 9. Test of emotional dimension matching sample

		Pair difference							
					The difference is 95% confidence interval				
		Mean value	**SD**	**Standard error mean**	**Lower limit**	**Upper limit**	**T**	**df**	**Sig (Double tail)**
Laboratory class	Pre-test of Affective dimension Post-test of Affective dimension	-1.0124	0.4951	0.0551	-1.1131	-0.8497	-13.853	50	0.000
Cross-reference class	Pre-test of Affective dimension Post-test of Affective dimension	0.0602	0.5093	0.0633	-0.0608	0.2026	1.143	52	0.227

(4) Changes in the behavioral dimension of intercultural communication competence

First, from the independent samples t-test, the levels of intercultural communication competence in the behavioral dimensions of the experimental class and the control class before the experiment are almost the same. However, the difference between the two classes after the experiment is obvious, the mean value is 3.24674 and 2.28043 respectively, the standard deviation is 0.40073 and 0.39151 respectively, and the t-test result of equal means shows T=12.122, P=0.000, P value is lower than the significance level of 0.05. It can be seen that the difference between the experimental class and the control class of the intercultural communication competence in the behavioral dimension after the experiment is relatively significant, and the difference of the experimental and control classes of the behavioral dimension is relatively significant. The difference is relatively significant, and the scores of students in the experimental class are higher than those of the control class.

Second, the paired-sample test for the behavioral dimension of the experimental and control classes is shown in Table 10. From the paired samples t-test, the paired samples t-test result of the pre- and post-test of behavioral dimensions in the control class shows T=-1.9567, P=0.053, and the P-value is higher than the significance of 0.05, so it can be seen that the difference between the control class in the pre- and post-test scores of behavioral dimensions is not too significant. The paired sample t-test results of the pre and post-tests of behavioral dimensions in the experimental class showed T=-16.3992, P=0.000, P-value is lower than the significance 0.05. Therefore, there is a more significant difference in the pre and post-test scores of behavioral dimensions of the experimental class before and after the experiment, which also indicates that the post-test scores of the experimental class in behavioral dimensions are significantly higher than the pre-tests, so that the computer-supported writing based on the data mining technology learning method has a significant effect on improving students' intercultural communication skills.

Table 10. Test of behavior dimension of the comparison class

		Pair difference							
					The difference is 95% confidence interval				
		Mean value	**SD**	**Standard error mean**	**Lower limit**	**Upper limit**	**T**	**df**	**Sig (Double tail)**
Laboratory class	Pre-test of behavior dimension	-1.0858	0.49	0.0614	-1.218	-0.9524	-16.3992	50	0.000
	Post-test of behavior dimension								
Cross-reference class	Pre-test of behavior dimension	-0.1153	0.4166	0.0527	-0.2614	0.0337	-1.9567	52	0.053
	Post-test of behavior dimension								

5. Conclusion

The computer-supported collaborative learning model is characterized by strong interactivity, which highlights the cultural attributes of the English subject by means of multivariate interactive learning, so that English teaching adheres to the road of connotative development. In view of this, this study attempts to apply the computer-supported collaborative learning method based on data mining technology to the high school English classroom and explore its role in developing students' intercultural communication skills.

According to the group division based on the practice strategy of CSCL teaching model, in the second round of collaborative interaction situation in Group A, the total number of effective interaction postings was 50, which was improved compared with the number of interactions in the first round, and some students produced meaningful in-depth exchanges with each other. The overall network density was 0.5302, and the interaction density increased in all groups.

A paired-sample test of the four dimensions of the Intercultural Communication Competence Scale found that the posttest scores of the experimental class were significantly higher than the pretest scores in all four dimensions of knowledge, awareness, emotion and behavior, while the pre and posttest scores of the control class did not change significantly. It can be concluded that the computer-supported collaborative learning method based on data mining technology proposed in this paper has a facilitating effect on intercultural communication competence in the dimensions of knowledge, awareness, emotion and behavior.

Funding

1.Major Humanities and Social Sciences Research Projects in Zhejiang higher education institutions, "Research on the Practical Value and Path Choice of Internationalization at Home of Vocational Education", Grant/Award Number: 2020GH070

References

1. Ghatge, A., & Dasgupta, S. (2017). Exploring the Competitive Advantage of Cross-Cultural Communication Training: A Conceptual Semantic Study. International Journal of Management and Development Studies, 6(3), 30-40.
2. Wu, A., Maddula, V., Kieff, M. R., & Kunzel, C. (2021). An online program to improve international collaboration, intercultural skills, and research knowledge. Journal of dental education, 85, 948-951.
3. Sabirova, D. R., Khanipova, R. R. Y., & Sagitova, R. R. (2021). English as a tool for cross-cultural communication. Revista EntreLinguas, e021048-e021048.
4. Sjøen, M. M. (2023). From global competition to intercultural competence: What teacher-training students with cross-cultural teaching experience should be learning. Scandinavian Journal of Educational Research, 67(1), 140-153.
5. Ran, C., & Sarsenbayeva, L. (2021). The importance of cultivation of cross-cultural communication skills of students in the context of multiculturalism. Pedagogy and Psychology, 48(3), 99-10.
6. KA, E., & HORNE, B. (2020). Cultivating Students’ Communication Skills for the Global Workplace: A Cross-Cultural Online Exchange Pro ect with Non-Native Speakers. International ournal of eaching and Education, 8(1), 64-83.
7. Shafazhinskaya, N. E., Shcherbinina, V. M., Ivanova, E. Y., Belyakova, T. E., & Pereverzeva, M. V. (2019). Learning about world art culture as a method of forming a universal cross-cultural communication competence. Humanities & Social Sciences Reviews, 7(6), 1225-1229.
8. Babao, J. N. A., & Adiatma, D. (2023). Intercultural Communication Competence: Unraveling the Role of Cognitive, Affective, and Behavioral Factors. International Journal of Advanced Multidisciplinary, 2(2), 393-397.
9. Wang, G., & Hu, X. (2022). Research on Intercultural Communication in International Business. Open Access Library Journal, 9(7), 1-7.
10. Rofi'i, A., Pambudi, N., Muthawali, D. I., Safar, M., & Kurniadi, P. (2024). INTEGRATING GLOBAL PERSPECTIVES INTO ENGLISH CURRICULUM: A LITERATURE REVIEW ON CROSS-CULTURAL COMPETENCE IN LANGUAGE EDUCATION. Jurnal Review Pendidikan dan Pengajaran (JRPP), 7(2), 3730-3737.
11. Bagui, H., & Adder, F. (2020). Promoting Students' Intercultural Communicative Competence through English Literary Texts: Students' Attitudes and Teachers' Challenges. Arab world english journal, 11(2), 85-93.
12. Qin, X. (2021). Evaluation of English intercultural communication ability based on machine learning and fuzzy mathematics. Journal of Intelligent & Fuzzy Systems, 40(4), 7259-7271.
13. Lee, J. S. (2020). Informal digital learning of English and strategic competence for cross-cultural communication: Perception of varieties of English as a mediator. ReCALL, 32(1), 47-62.
14. Ren, L., & Cui, L. (2017, December). Strategies of Cultivating Students' Intercultural Communication Competence in Senior High School English Teaching. In 2017 International Conference on Economic Development and Education Management (ICEDEM 2017) (pp. 133-137). Atlantis Press.
15. Pang, Z. (2024). Cultivating Intercultural Communication Competence in College English Teaching: Practice and Exploration. International Journal of Educational Teaching and Research, 1(2).
16. Wang, J., & Wu, L. (2020, January). How to cultivate students' Intercultural communicative competence in College English Teaching. In 2019 International Conference on Education Science and Economic Development (ICESED 2019) (pp. 88-91). Atlantis Press.
17. Li, H. (2024). Strategies for cultivating intercultural communicative competence in language education. Frontiers in Educational Research, 7(4).

18. Aifang, X., Ahmad, N. K., & Abd Rahman, S. N. (2024). Impact of multicultural literature on intercultural communicative competence development in English language education: A systematic review. International Journal of Learning, Teaching and Educational Research, 23(11), 308-324.
19. Wang, J. (2024). Analysis of Development of Intercultural Communicative Competence in English Education. The Educational Review, USA, 8(4).
20. Zhang, Y. (2021). Research on the cultivation of College Students' intercultural communicative competence based on the technology of multimedia immersion in 5G Era. In E3S Web of Conferences (Vol. 251, p. 03076). EDP Sciences.
21. Peng, B. (2018). Study on Cultivation of Intercultural Communication Competence of College English Based on Modern Information Technology. International Journal of Emerging Technologies in Learning, 13(9).
22. Long, J., & Lin, J. (2022). An empirical study on cultivating college students' cross-cultural communicative competence based on the artificial-intelligence English-teaching mode. Frontiers in Psychology, 13, 976310.
23. Colomar, M. P. A., & Menn, A. (2024). A virtual exchange to boost students' intercultural communication skills in English. Baltic Journal of English Language, Literature and Culture, 14, 21-34.
24. Banjongjit, B., & Boonmoh, A. (2018). Teachers' perceptions towards promoting intercultural communicative competence in the EFL classroom. REFLections, 25(2), 76-97.
25. Jain, P. (2024). Authentic Materials for Intercultural Communication: Develop Speaking Materials That Encourage Learners to Engage in Authentic Cross-Cultural Communication Effectively in Diverse Global Contexts. Creative Saplings, 3(4), 1-14.
26. Kosareva, L., Evreeva, O., & Zakirova, O. (2019). Formation of language competence: Modern issues and strategies in the area of cross-cultural communication. Space and Culture, India, 7(3), 149-159.
27. Hernández-Sellés, N., Muñoz-Carril, P. C., & González-Sanmamed, M. (2019). Computer-supported collaborative learning: An analysis of the relationship between interaction, emotional support and online collaborative tools. Computers & Education, 138, 1-12.
28. Azevedo, A. (2019). Data mining and knowledge discovery in databases. In Advanced methodologies and technologies in network architecture, mobile computing, and data analytics (pp. 502-514). IGI Global Scientific Publishing.
29. Liang, R., Huang, C., Zhang, C., Li, B., Saydam, S., & Canbulat, I. (2023). Exploring the fusion potentials of data visualization and data analytics in the process of mining digitalization. IEEE Access, 11, 40608-40628.
30. Cai, L., & Zhuang, Z. (2022). A Study on Integrating Multimodal English and American Literature Resources Using Data Mining. Mobile Information Systems, 2022(1), 1567479.
31. Xinglong Zhao & Su Cai. (2011). 'Many to one' mode of CSCL and further discussion. Int. J. of Continuing Engineering Education and Life-Long Learning(1),42-54.
32. Ricardo Valdivia & Miguel Nussbaum. (2009). Using multiple choice questions as a pedagogic model for face - to - face CSCL. Computer Applications in Engineering Education(1),89-99.
33. Jingqian Wang. (2024). Research on Behavior Characteristics of Psychological and Emotional Analysis Based on Big Data Classification Algorithm of Psychological Behavior. Applied & Educational Psychology(7).
34. Jing Wen. (2024). Mining method of students' learning behaviour characteristics in online classroom of colleges and universities based on dense clustering method. International Journal of Continuing Engineering Education and Life-Long Learning(1),77-87.