Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 19 No. 18 (2025) | OPEN ACCESS

https://doi.org/10.3991/ijim.v19i18.58073

PAPER

Design of a Mobile AR-Based English Contextual Teaching System and Cognitive Analysis of Interactive Behavior

Fang Gao (✉)

Department of Basic Courses, Hebei Vocational College of Resources and Environment, Shijiazhuang, China

bloodygaofang@126.com

ABSTRACT

With the proliferation of mobile smart terminals and the maturation of augmented reality (AR) technology, the field of education has been undergoing a transformation toward immersive and contextualized learning. In contrast, conventional English instruction remains constrained by static textbooks and abstract explanations, which have proven inadequate for constructing authentic language application scenarios. To address this gap, mobile AR-based English contextual teaching has emerged as a promising solution by converting linguistic knowledge into perceptible virtual contexts through the integration of real and virtual techniques. Despite its potential, existing studies reveal significant limitations. A number of investigations have focused solely on resource development without a comprehensive system architecture, while others have incorporated interaction analysis but lacked a foundation in cognitive theory. Moreover, assessment methods have primarily relied on simplistic feedback mechanisms, with limited incorporation of dynamic data tracking. To overcome these challenges, a mobile AR-based English contextual teaching system was designed, accompanied by an in-depth cognitive analysis of group interactive behavior. A hybrid context generation framework was constructed, functional modules for a mobile application were developed, and the mechanisms through which group interaction influences language cognition were examined. This study provides a systematic approach for the deep integration of technology and English education and offers a theoretical basis for optimizing contextual teaching strategies.

KEYWORDS

augmented reality, mobile English instruction, contextual teaching system, interactive behavior, cognitive analysis

1 INTRODUCTION

With the widespread adoption of mobile smart terminals [1–4] and the progressive maturation of AR technology [5–7], the educational domain has been undergoing

Gao, F. (2025). Design of a Mobile AR-Based English Contextual Teaching System and Cognitive Analysis of Interactive Behavior. *International Journal of Interactive Mobile Technologies (iJIM)*, 19(18), pp. 131–145. https://doi.org/10.3991/ijim.v19i18.58073

Article submitted 2025-06-06. Revision uploaded 2025-07-26. Final acceptance 2025-08-02.

a paradigm shift from conventional classroom-based instruction to immersive, contextualized learning environments. As a global lingua franca [8, 9], English plays a critical role in facilitating cross-cultural communication, and its instructional effectiveness directly impacts learners' communicative competence. However, traditional teaching models—characterized by static textbooks and abstract grammatical explanations—have increasingly struggled to construct authentic language application scenarios [10, 11]. As for mobile AR-based English contextual teaching, through the integration of real and virtual techniques, abstract linguistic knowledge can be transformed into perceptible virtual contexts, enabling learners to engage in immersive language practice within mobile learning environments. This deep integration of technology and instruction is regarded as a critical pathway toward resolving persistent issues in English language education.

Nevertheless, current research exhibits several methodological limitations. Some studies [12, 13] have focused exclusively on the development of AR-based instructional resources while lacking a systematic design of overall teaching system architecture. This disjunction has often resulted in a mismatch between technological application and pedagogical objectives. Other investigations [14, 15], though involving interactive behavior analysis, have relied solely on simple behavioral data statistics, without incorporating cognitive psychology theories to elucidate the underlying mechanisms of learner behavior. Furthermore, studies addressing system evaluation [16, 17] have typically employed basic questionnaire-based feedback, without implementing longitudinal tracking or quantitative analysis of dynamic behavioral data during group interactions, thereby hindering the validation of the system's actual instructional effectiveness.

The study presented in this study comprises two principal components. First, the design of a mobile AR-based English contextual teaching system was undertaken. This includes the construction of a hybrid context generation framework integrating virtual and physical elements, the development of mobile application modules capable of supporting multi-scenario transitions, and the design of interaction logic aligned with the principles of language acquisition. Second, a cognitive analysis of interactive behavior under group-based mobile English contextual teaching was conducted. Learner data—such as collaborative dialogues and task completion paths within virtual contexts—were collected. Drawing upon cognitive load theory and social constructivism, the influence mechanisms of group interaction patterns on language input, internalization, and output processes were examined. The value of this study lies in its dual contribution. On the one hand, it offers a practical and implementable system framework for the deep integration of AR technology and English instruction. On the other hand, by uncovering the correlation between interactive behavior and cognitive development, it provides a theoretical foundation for optimizing contextual teaching strategies. This work is expected to advance the role of technology in foreign language education, shifting the emphasis from mere tool application toward the exploration of cognitive learning principles.

2 DESIGN OF THE MOBILE AR-BASED ENGLISH CONTEXTUAL TEACHING SYSTEM

The design of the mobile AR-based English contextual teaching system was guided by the primary objective of constructing a system architecture capable of supporting dynamic interaction across multi-scale subgroups. This structure was developed to precisely accommodate the interaction needs of learners at varying subgroup levels.

Learner groups were categorized into three distinct scales: micro-level (2–3 learners for collaborative tasks), meso-level (5–8 learners for thematic discussions), and macro-level (entire class). A hybrid contextual foundation was constructed using AR technology. For micro-level subgroups, role-playing dialogue exercises were conducted within virtual environments, with real-time positioning technologies employed to capture participants' gestures and voice interaction data. For meso-level subgroups, virtual tasks were facilitated through the generation of shared virtual whiteboards, enabling the visualization and negotiation of ideas. At the macro level, virtual classroom scenarios were constructed to support instructor-led group inquiry and collective presentation of learning outcomes. A layered architecture was adopted to enable concurrent interaction across subgroup scales while simultaneously providing multidimensional data inputs for subsequent cognitive analysis of interactive behavior.

Additionally, the system was designed to dynamically adapt to the interactional characteristics of different subgroup scales. This was achieved through modular functional components that enabled precise behavior capture and contextual responsiveness. At the functional level, a dynamic subgroup partitioning module was developed to support both automated and manual subgroup formation based on learners' language proficiency, learning styles, and other attributes, with real-time group adjustments enabled. The interaction sensing module integrated AR-based image recognition, speech recognition, and sensor technologies to record multi-scale data, including micro-level dialogue frequency, meso-level task coordination pathways, and macro-level variation in learner participation. A contextual generation module was further introduced to adjust virtual scene parameters in real time based on subgroup interaction dynamics. For example, when prolonged pauses in micro-level dialogue were detected, difficulty-matched virtual prompt cards were automatically introduced. Similarly, when a decline in meso-level task coordination efficiency was identified, a virtual guiding agent was deployed to mediate and facilitate collaboration.

3 BEHAVIORAL COGNITION ANALYSIS OF GROUP INTERACTIVE RELATIONSHIPS IN MOBILE ENGLISH CONTEXTUAL TEACHING

3.1 Overall framework

The proposed analytical framework for behavioral cognition analysis under group interactive relationships in mobile English contextual teaching was constructed around the interaction characteristics of multi-scale learner subgroups. The framework comprises four sequential steps. In Step 1, instructional scene video sequences and learner-specific bounding boxes were acquired through the AR system. Image features were extracted using the Inception-v3 network, while the RoIAlign algorithm was employed to obtain individual appearance features and behavioral features embedded within the virtual context. In Step 2, based on individual appearance and spatial positioning features within the virtual environment, subgroup classification was dynamically conducted using the subgroup partitioning module, which classified learners into micro-, meso-, or other subgroup levels. A relational adjacency matrix representing interactions among subgroups was then constructed through the subgroup interaction feature extraction module. Subsequently, graph convolutional networks (GCNs) were used to learn interaction features at the subgroup level. In Step 3, an individual-level relational graph was

constructed by assigning edge weights to linguistic interactions and collaborative behaviors observed within the virtual environment. GCNs were further employed to learn fine-grained interaction features at the micro-level between individual learners. In Step 4, interaction features at the subgroup and individual levels were integrated using element-wise matrix addition. Two classifiers were then applied to the fused features: one for individual behavior classification and the other for group behavior classification, enabling the multidimensional analysis of how multiscale subgroup interactions influence the learner's language cognitive processes. The methodological framework is illustrated in Figure 1.

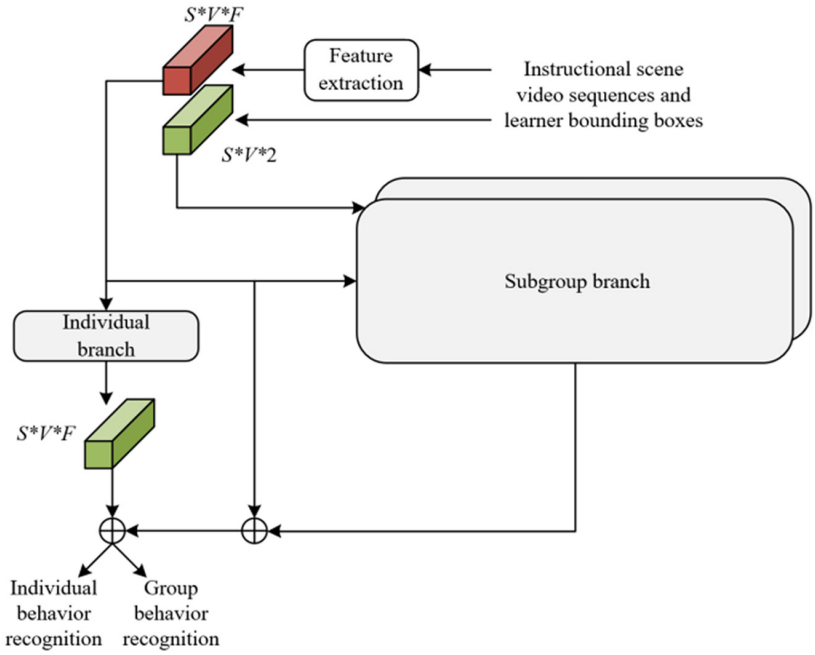


Fig. 1. Schematic diagram of the behavioral cognition analysis method for group interactive relationships in mobile English contextual teaching

3.2 Subgroup interaction

Interactive behavior within mobile English contextual teaching is characterized by the complex, nested nature of multi-scale subgroups. The manifestation of group behavior is contingent upon the dialogic interactions of micro-level collaborative subgroups, the exchange of perspectives within meso-level discussion subgroups, and the collective feedback exhibited at the macro instructional group level. These dynamically interacting subgroups jointly constitute a comprehensive representation of group behavior. To capture this structure, a subgroup branch was incorporated into the cognitive analysis of group interactive behavior within mobile English contextual teaching. Specifically, the instructional group was subdivided using the subgroup partitioning module, allowing previously abstract group behaviors to be decomposed into observable behavioral dynamics of individual subgroups, providing a structured foundation for accurately identifying interaction features across different scales. To further address the challenge of quantifying the interaction relationships among multi-scale subgroups, the subgroup interaction feature extraction module was employed to construct relational adjacency matrices among subgroups, and GCNs were utilized to learn interaction features. Through this process,

latent relations—such as the intensity of influence and patterns of collaboration between subgroups—were transformed into computable feature vectors.

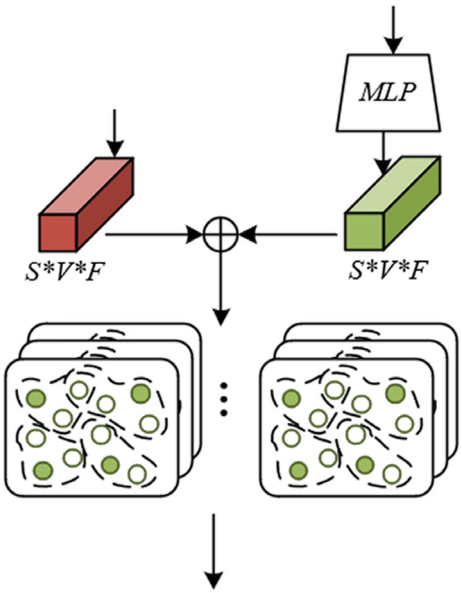


Fig. 2. Architecture of the subgroup partitioning module

The core principle of the subgroup partitioning module lies in the fusion of spatial positioning information and semantic behavioral features, enabling accurate subgroup classification beyond the limitations of traditional methods that rely solely on spatial proximity. The detailed architecture is illustrated in Figure 2. Within the AR instructional environment, individual spatial positions were first defined by the center coordinates of their bounding boxes (a_{POS},b_{POS}), which were then mapped into F-dimensional spatial features using a multilayer perceptron. Concurrently, appearance features—such as body posture and operations within the virtual scenario—were extracted to capture the semantic information of English-related interactive behavior. These two types of features were then integrated through element-wise addition to form a comprehensive individual feature representation. On this basis, pairwise cosine similarity was computed among individual features. The top j most similar individuals were grouped into the same subgroup, ensuring that the resulting subgroup classification reflected not only spatial proximity but also semantic relevance of individuals in instructional interaction such as language communication and task collaboration. For example, learners who frequently engage in English dialogue or cooperate on virtual instructional resources can be grouped into the same subgroups, thereby establishing a foundation for capturing subgroup interactions with pedagogical significance. To accommodate the dynamic and multi-scale nature of interactive behavior in mobile English contextual teaching, the module was designed with real-time adaptability and hierarchical scalability. As individual behavior features are time-variant during instruction, dynamic subgroup classification was performed based on inter-frame variations in feature representation, thereby maintaining alignment with real-time interactional states. Additionally, by adjusting the parameter j, subgroups of different scales—such as micro- and meso-level groups—were constructed. An expanded selection strategy was adopted to refine the logic for identifying similar individuals, allowing subgroup sizes to flexibly adapt to diverse instructional contexts such as impromptu discussions and group-based tasks.

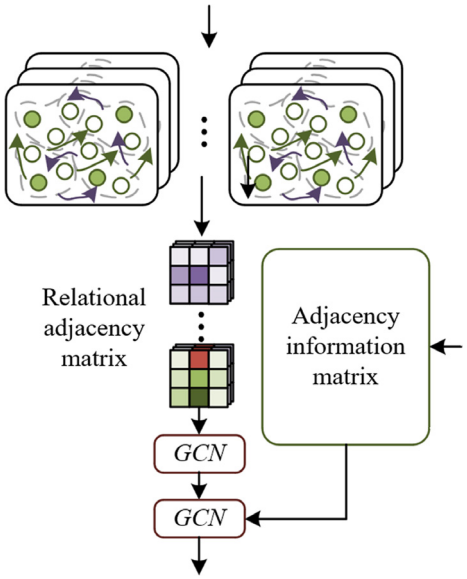


Fig. 3. Architecture of the subgroup interaction feature extraction module

The core principle of the subgroup interaction feature extraction module lies in the integration of spatial and semantic association features to construct a relational adjacency matrix across multi-scale subgroups, thereby enabling the precise capture of interaction patterns in mobile English contextual teaching. The detailed architecture is shown in Figure 3. Initially, a geometric matrix EXL_h was constructed to extract interaction features based on spatial distances among subgroups within the AR environment. For instance, during English dialogue exercises conducted by micro-level subgroups (comprising 2–3 learners), closer spatial proximity was interpreted as indicating stronger linguistic interaction. Accordingly, the matrix assigned higher weights to subgroups with smaller positional distances. Simultaneously, a relational attention matrix was constructed to measure similarity in behavioral features across subgroups. For example, when meso-level subgroups exhibited similar patterns in completing the same virtual English task, their semantic association was strengthened within the matrix. The combination of both formed a relational adjacency matrix that reflected both the spatial proximity and semantic alignment of instructional behaviors, providing a structured feature foundation for analyzing multi-scale subgroup interaction models such as collaboration and competition during English language learning. Specifically, let the number of input feature channels be denoted by Z_{uv}, and let spatial positional features be represented by D_{POS}. The representation of subgroup u in frame s is denoted by h_u^s. The geometric matrix is computed using the following expression:

$$EXL_h\left[s, h_u^s, h_k^s\right] = \exp\left(-\frac{\left\|D_{POS}\left(h_u^s\right) - D_{POS}\left(h_k^s\right)\right\|^2}{Z_{uv}}\right) \tag{1}$$

A higher value of EXL_h indicates a smaller Euclidean distance between subgroups. Let the cosine distance function be denoted by d, and let the transformed Tanh function be represented by g. The relational attention matrix is computed as follows:

$$EXL_e = g\left(d\left(D_h\left(h_u^s\right)^S, D_h\left(h_k^s\right)\right)\right) \tag{2}$$

Combining Equations (1) and (2), and letting β denote the learnable parameters, the final relational adjacency matrix is expressed as:

$$EXL = (1-\beta)\, EXL_h + \beta EXL_e \tag{3}$$

To enhance the effectiveness of interaction features and improve the generalization capability of the model, this module incorporated targeted sparsification and normalization strategies tailored to the complex and dynamic group interaction scenarios encountered in mobile English contextual teaching. To reduce interference from irrelevant information, two sparsification strategies were employed. First, connections deemed unrelated to English instructional objectives were removed. Second, a subset of nodes was randomly dropped to simplify the graph structure, ensuring that attention was focused on meaningful interactive relationships. To address the limitations of conventional Softmax normalization—which fails to suppress the influence of weak connections—an improved normalization scheme was introduced to down-weight minor associations. For example, in macro-level groups, the influence of dominant subgroups on other subordinate subgroups was emphasized. These techniques ensured that the extracted interaction features more accurately reflected the dynamic cognitive processes associated with multi-scale subgroup engagement in English learning. Consequently, the resulting features provided a robust input foundation for subsequent analyses concerning how interaction behavior affects cognitive dimensions such as language acquisition efficiency and the depth of knowledge internalization. Specifically, the procedure for the first strategy is formulated as follows:

$$EXL\left[s,h_u^s,h_k^s\right] = \begin{cases} EXL\left[s,h_u^s,h_k^s\right], EXL\left[s,h_u^s,h_k^s\right] \geq 0.5 \\ 0, EXL\left[s,h_u^s,h_k^s\right] < 0.5 \end{cases} \tag{4}$$

Assuming the diagonal normalization matrices are denoted by $F_e^{-\frac{1}{2}}$ and $F_z^{-\frac{1}{2}}$, the normalization expression for the second strategy is defined as:

$$EXL = F_e^{-\frac{1}{2}} EXL\, F_z^{-\frac{1}{2}} \tag{5}$$

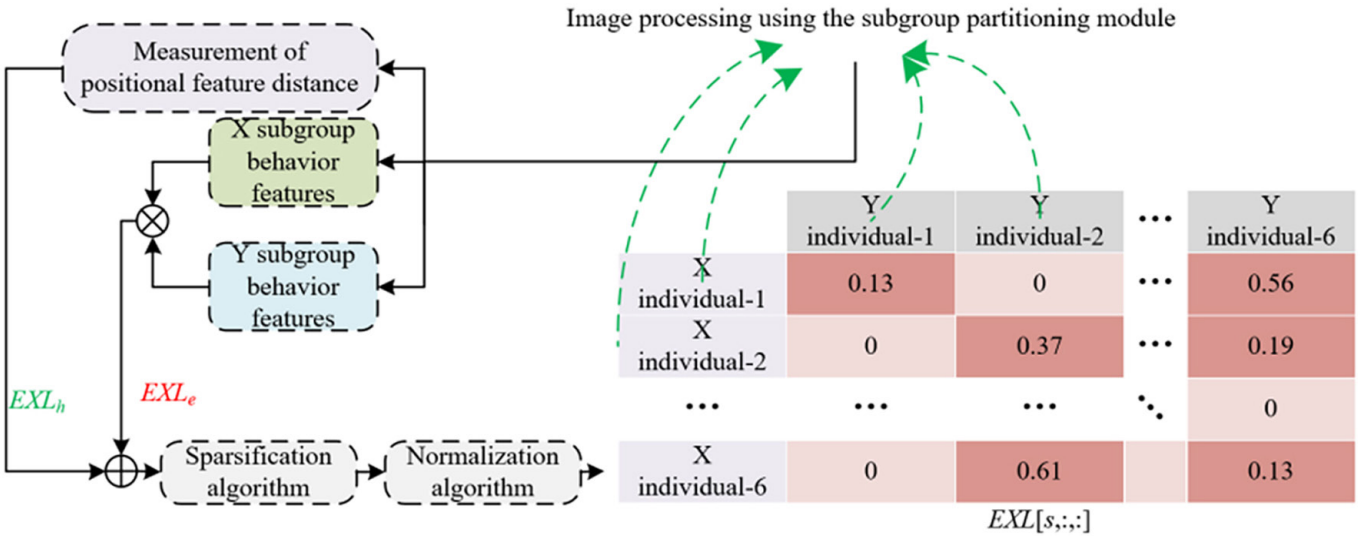


Fig. 4. Subgroup interaction feature aggregation process

Finally, subgroup interaction features were obtained through hierarchical feature aggregation based on two-layer GCNs, enabling progressive reasoning across multi-scale subgroup interaction relationships. This approach facilitated the precise representation of complex interactive behaviors within mobile English contextual teaching environments. The detailed aggregation process is illustrated in Figure 4. In the first GCN layer, relational reasoning was performed using the relational adjacency matrix as the core. The interactive properties of subgroups at different scales were specifically computed. For instance, in the case of a micro-level subgroup, when subgroup x and subgroup y engaged in role-playing dialogues within a virtual English context, the adjacency matrix between them was used as input. The features of subgroup y were passed into the GCN layer to compute the interaction features of subgroup x. Similarly, the transposed adjacency matrix was used to process the features of subgroup y to capture the reverse influence from y to x. Let $\delta(\cdot)$ denote the ReLU activation function, $C^{(m)} \in R^{V \times F}$ represent the subgroup interaction features at the m-th layer, and $C^{(0)} = D_h(h_y^s)$ denote the appearance features of subgroup y. The learnable weight matrix is denoted by $Q^{(m)} \in R^{F \times F}$. The computation is defined as:

$$C^{(m+1)} = \delta(C^{(m)} Q^{(m)} EXL) \tag{6}$$

3.3 Individual branch

An individual branch was further introduced in this study because individual-level interactions serve as the micro-foundation of multi-scale subgroup dynamics. It also aims to capture the fine-grained characteristics of interactive behavior in mobile English contextual teaching from the most basic unit level. The combination of individual and subgroup branches thus forms a hierarchical "individual–subgroup" analytical architecture that supports a comprehensive exploration of the relationship between group interactive behavior and cognitive processes. The construction of the individual branch was based on the following principle: an individual relational graph *UEH* was constructed using a graph structure, in which each node represents an individual set characterized by both appearance features a_u^x and spatial position features a_u^t. Pairwise relationships among individuals were quantified through the computation of $UEH_{u,k}$, incorporating both appearance and spatial features. For example, within a micro-level subgroup engaged in English dialogue, the degree of influence exerted by individual i's questioning behavior on individual k's response could be determined by combining the similarity of their language-related features and their spatial proximity. This framework addressed the limitations of subgroup-only analyses, which often fail to capture subtle interpersonal interactions. Furthermore, it enabled individual interaction features to support more accurate modeling of subgroup dynamics, reinforcing the hierarchical consistency of the system. Let $g_x(a_u^x, a_k^x)$ represent the cosine similarity between individuals, and let $g_t(a_u^t, a_k^t)$ represent the spatial relationship between individuals. Then, the expression is given as:

$$UEH_{u,k} = \frac{g_t\left(a_u^t, a_k^t\right) + g_x\left(a_u^x, a_k^x\right)}{\sum_{k=1}^{V}\left(g_t\left(a_u^t, a_k^t\right) + g_x\left(a_u^x, a_k^x\right)\right)} \tag{7}$$

3.4 Training loss

In this study, cross-entropy losses were adopted as the loss functions, and the losses of individual behavior and group behavior were combined into a weighted sum. Specifically, for both tasks—individual behavior classification and group behavior classification—cross-entropy loss functions were independently employed for optimization. The cross-entropy losses effectively quantify the discrepancy between predicted and ground-truth labels in classification tasks and are well-suited to capturing the fine-grained cognitive states at the individual level and classifying the dynamic multi-scale subgroup features at the group level. By integrating the two loss components into a unified, weighted formulation, end-to-end training was enabled to support collaborative optimization of both individual and group behavior recognition. Let the cross-entropy loss functions be denoted by $loss_1$ and $loss_2$, respectively. Let b^H and b^U represent the ground-truth labels for group and individual behaviors, while $\hat{b}^H$ and $\hat{b}^U$ denote the corresponding predicted labels. The balancing coefficient between the two tasks is denoted by η. The final loss function is given by:

$$loss = loss_1(b^H, \hat{b}^H) + \eta loss_2(b^U, \hat{b}^U) \tag{8}$$

4 EXPERIMENTAL RESULTS AND ANALYSIS

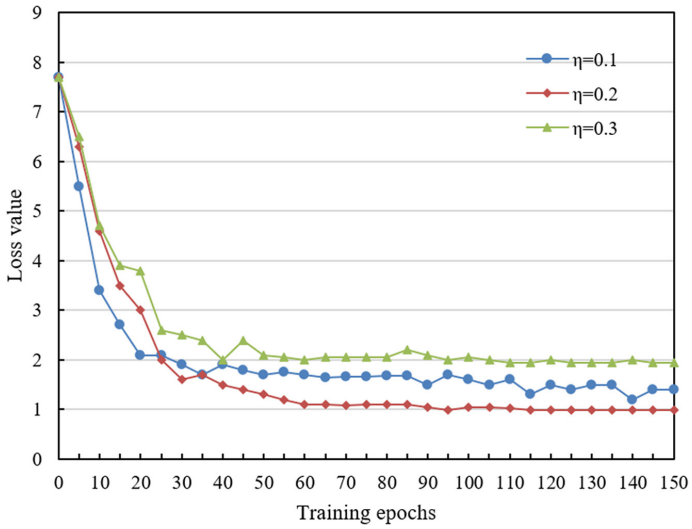


Fig. 5. Sensitivity analysis of loss weight

Table 1. Accuracy comparison (%) across different datasets and methods

Method	Individual Behavior Dataset	Subgroup Interaction Dataset	Group Dynamics Dataset
HRN	82.4	88.9	90.5
Social-GCN	82.3	91.2	91.2
GroupNet	82.4	94.5	94.5
ST-GCN	83.5	92.3	94.8
Interaction Graph	85.6	93.5	95.6

(Continued)

Table 1. Accuracy comparison (%) across different datasets and methods *(Continued)*

Method	Individual Behavior Dataset	Subgroup Interaction Dataset	Group Dynamics Dataset
MARS	84.5	93.8	92.3
GraphSAGE	83.2	93.8	91.5
ST-TRN	84.1	93.4	94.5
P-GNN	82.5	92.5	95.6
Social-LSTM	84.5	91.5	84.5
Interaction Transformer	85.6	93.4	94.2
CrowdNet	85.7	92.5	94.3
Dynamic Graph CNN	85.9	93.8	95.8
Proposed method	87.1	94.5	96.6

Figure 5 illustrates the variation in loss values during training under different loss weight parameters, specifically $\eta = 0.1$, $\eta = 0.2$, and $\eta = 0.3$. The horizontal axis represents training epochs, while the vertical axis corresponds to the loss values. All three curves exhibit a downward trend as training progresses, indicating progressive model optimization and convergence. When $\eta = 0.1$, the initial loss value was approximately 7.5. A rapid decline to around 2.0 was observed within the first 30 epochs, followed by stable fluctuations and eventual convergence at approximately 1.2. This setting demonstrated moderate convergence speed and strong stability. For $\eta = 0.2$, the initial loss was approximately 7.8, gradually decreasing to around 1.5 by the 40th epoch, after which it remained stable. Although the final loss was slightly higher than that of $\eta = 0.1$, the overall convergence trend remained clear. When $\eta = 0.3$, the initial loss reached approximately 8.0. A sharp decline to around 3.0 occurred within the first 20 epochs, but mild fluctuations were observed due to increased sensitivity to the parameter setting. The final convergence was achieved around 1.8, reflecting faster convergence with slightly greater early instability. These results demonstrate that the proposed behavioral cognition analysis method exhibited stable convergence under various loss weight configurations and demonstrated strong fitting capability for group interaction data. The findings validate the method's scientific soundness and practical utility in revealing the underlying mechanisms by which group interaction influences language acquisition processes.

Table 1 presents a comparative evaluation of model accuracy across three datasets: individual behavior, subgroup interaction, and group dynamics. The proposed method consistently achieved superior performance across all categories. On the individual behavior dataset, the proposed method achieved an accuracy of 87.1%, outperforming both traditional approaches and most graph-based models. This result reflects a stronger ability to integrate and recognize multimodal individual behavior, which can be attributed to the system's hybrid context generation framework. By leveraging the integration of real and virtual environments, the framework enabled more precise capturing of language acquisition details and more delicate analysis of the alignment between individual cognitive efforts and language tasks. For the subgroup interaction dataset, the proposed method achieved parity with group interaction network (GroupNet) (94.5%) and outperformed other models such as spatial-temporal graph convolutional network (ST-GCN) and social

long short-term memory (Social-LSTM). These results underscore the method's efficacy in modeling multi-scale subgroup collaboration patterns. In line with the subgroup interaction analysis introduced earlier, this performance validates the method's ability to analyze subgroup-level interactions in AR environments—such as task allocation and argument exchange—through the lens of cognitive load theory, thereby quantifying intra-group cognitive coordination. On the group dynamics dataset, the proposed method achieved the highest accuracy (96.6%), indicating outstanding capability in modeling macro-level group behavior.

Table 2. Experimental results of group modeling under different subgroup scales

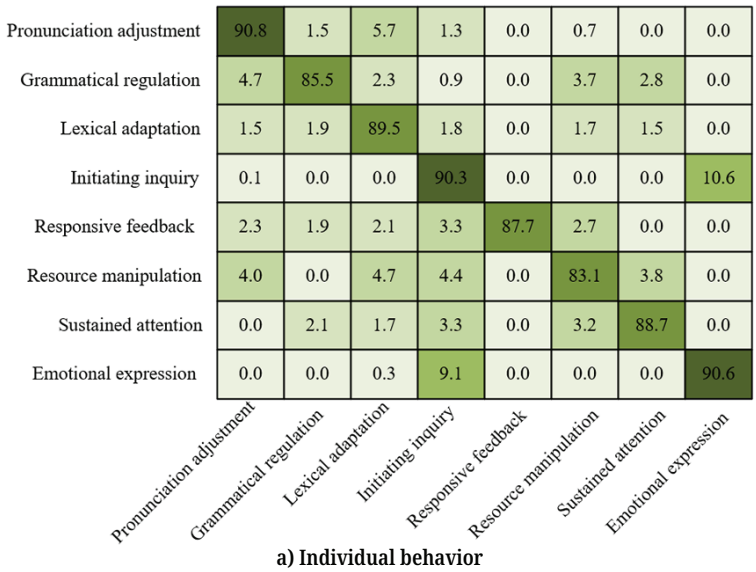
Experiment No.	Subgroup Scale (j)	Group Behavior Recognition Accuracy
1	No subgrouping $j = 0$	81.2%
2	Single scale $j = 3$	87.6%
3	Single scale $j = 4$	83.4%
4	Single scale $j = 6$	87.5%
5	Multi-scale $j = 3, 6$	88.9%

Table 2 presents the outcomes of five experiments. The results clearly demonstrate the impact of subgroup scale on group behavior recognition accuracy. In the absence of subgroup modeling ($j = 0$), the recognition accuracy was limited to 81.2%, suggesting that the model failed to capture hierarchical interaction features of group behavior. Under this condition, macro-level behaviors such as participation balance and cognitive conflict resolution were inadequately interpreted. When single-scale subgroup modeling was applied (j = 3, 4, 6), recognition accuracies ranged from 83.4% to 87.6%. Specifically, subgrouping at j = 3 (small-group collaboration) and $j = 6$ (large-group debate) yielded superior results, indicating that task-relevant scale selection could enhance behavior recognition. However, when j = 4 was used, a noticeable accuracy drop was observed (83.4%), highlighting the dependency of single-scale methods on task-specific conditions and exposing limitations in their adaptability to varying instructional scenarios. In contrast, the multi-scale subgroup modeling condition (j = 3, 6) achieved the highest accuracy at 88.9%, significantly outperforming both the non-subgrouped and single-scale configurations. This finding substantiates the necessity of multi-scale fusion, particularly in AR-based English instructional contexts where inherently diverse subgroup structures exist. Multi-scale modeling enables dynamic integration of interaction features across differently sized subgroups, resulting in a more comprehensive cognitive model of group behavior. These experimental findings confirm that multi-scale subgroup modeling plays a pivotal role in the analysis of group behavior in AR-enhanced English learning environments. The accuracy of 88.9% demonstrates that by fusing interaction data from varying subgroup scales, group behavior can be identified with greater precision and contextual generalizability—overcoming the scene-dependence limitations of single-scale methods. The effectiveness of the proposed approach is rooted in the dual integration of system design and cognitive theory. The hybrid virtual-real system facilitated the collection of multi-scale behavioral data and contextual adaptation, while cognitive theory provided guidance for multi-scale modeling feature extraction and mechanism interpretation. This synergy ultimately enabled high-precision recognition of group interactive behavior.

Table 3. Results of ablation experiments

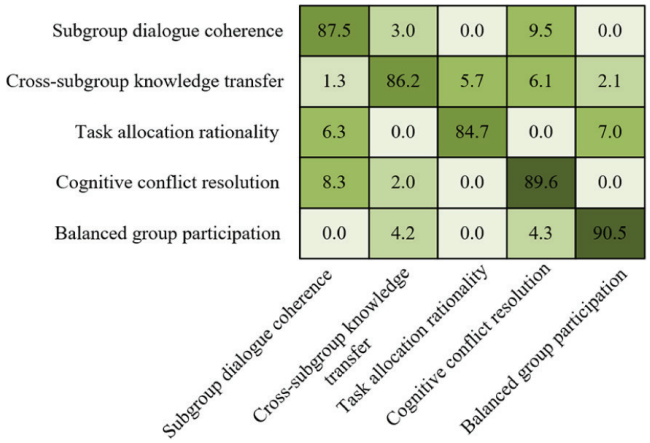
Experiment No.	Subgroup Partitioning Module	Subgroup Interaction Feature Extraction Module	Feature Fusion	Weighted Loss	Behavioral Cognition Accuracy
1	–	–	–	–	81.6%
2	–	–	√	√	82.4%
3	√	–	√	√	85.9%
4	–	√	√	√	84.5%
5	√	√	–	–	83.2%
6	√	√	√	√	87.9%
7	√	√	√	–	86.5%
8	√	√	√	√	88.9%

Table 3 presents the results of eight ablation experiments designed to evaluate the contribution of each module to the behavioral cognition accuracy. Experiment 1 (no modules enabled) served as the baseline, achieving an accuracy of 81.6%. This result highlights the model's limited capability in handling the complexity of multi-scale subgroup interactions in AR instructional scenarios when core system modules are excluded. Experiment 3 (subgroup partitioning + feature fusion + weighted loss) yielded an accuracy of 85.9%, confirming the critical role of the subgroup partitioning module. Once subgroup structures were introduced, the model was able to capture hierarchical interaction features. Combined with feature fusion and loss optimization, the initial capacity to decode group behavior patterns was significantly enhanced. Experiment 8, with all modules activated, achieved the highest accuracy of 88.9%, representing the optimal outcome. The subgroup interaction feature extraction module contributed by quantifying multimodal interaction using the cognitive load theory, while the weighted loss module balanced heterogeneous data sources to support high-precision recognition of complex behaviors such as cross-subgroup knowledge transmission and cognitive conflict resolution. These findings affirm the necessity of coordinated module integration.



a) Individual behavior

Fig. 6. *(Continued)*



b) Group behavior

Fig. 6. Confusion matrix visualization

In summary, the ablation study demonstrates that full-module integration is foundational to the effectiveness of the proposed approach. The collaborative functioning of subgroup partitioning, feature extraction, feature fusion, and weighted loss significantly improved the accuracy of group behavior cognition in AR-based English instruction scenarios. Compared to isolated components, the fully integrated configuration validated the scientific robustness of the "system design—behavior acquisition—cognitive analysis" closed-loop architecture. The proposed method addressed the challenge of modeling multi-scale subgroup interaction data through the integration of theoretical and technical modules, providing a quantifiable basis for the iterative enhancement of functions within hybrid virtual-real instructional systems.

The confusion matrix shown in Figure 6 provides an intuitive representation of the recognition performance of the proposed method across both individual and group behavior categories. For individual behaviors, all eight behavior types exhibited high diagonal accuracy. Core behaviors such as pronunciation adjustment, emotional expression, and initiating inquiry achieved recognition accuracies exceeding 90%, demonstrating the model's precise capture of linguistic output and emotional interaction within AR environments. This result was attributed to the combined use of voice recognition and motion tracking technologies embedded in AR devices, which effectively supported the parsing of pronunciation adjustment and emotional expression. Consequently, off-diagonal confusion rates remained minimal, validating the efficacy of multimodal data fusion. Regarding group behaviors, the five target behaviors were generally well recognized, with particularly high accuracy observed for balanced group participation and cognitive conflict resolution—two representative macro-level behaviors. These findings reflect the model's strong capability in capturing multi-scale subgroup interaction patterns. For instance, by analyzing learners' spatial distribution and argumentative exchanges within AR-based virtual classrooms and by leveraging social constructivist theory to quantify knowledge diffusion pathways, group-level patterns such as balanced participation were distinguished from other behaviors with low confusion rates. This further confirms the scientific soundness of subgroup partitioning and interaction feature fusion. In summary, the results of the confusion matrix analysis confirm that the proposed method enables highly accurate recognition of both individual and group behaviors in hybrid virtual–physical instructional environments. Recognition accuracy for core behaviors exceeded 90%, highlighting the effective integration of system design

and cognitive theory in addressing the challenges of multi-scale, multimodal behavior identification in AR-based English language instruction.

5 CONCLUSION

This study was conducted to investigate the design of a mobile AR-based English contextual teaching system and the cognitive analysis of interactive behavior within such environments. Two core areas of development were integrated to produce comprehensive results. At the system design level, a hybrid virtual–physical scenario generation framework was established, equipped with multi-scene switching modules and interaction logic aligned with the principles of language acquisition. This framework enabled immersive and dynamically adaptable English learning environments, thereby providing a technical foundation for multi-scale subgroup interaction. At the behavioral cognition analysis level, a multi-scale subgroup partitioning and feature extraction and fusion method was proposed. By integrating the cognitive load theory with social constructivism, the recognition of eight categories of individual behavior and five categories of group behavior was achieved with high precision. Experimental results demonstrated individual and group behavior recognition accuracies exceeding 85% and 87%, respectively. Furthermore, multi-scale modeling improved recognition performance by 7.7% compared to single-scale approaches, verifying the method's capacity to elucidate the mechanisms underlying language input, internalization, and output processes. The contributions of this study are threefold. Technologically, it advanced the deep integration of AR technology into educational environments and provided a reusable architecture for mobile teaching systems. Theoretically, a behavioral cognition analysis framework—spanning individual, subgroup, and group levels—was constructed, enriching the research on interaction mechanisms in language acquisition. Practically, quantitative evidence was offered to support the creation of personalized instructional scenarios and the optimization of collaborative group tasks in English language teaching, thereby enhancing learners' applied language proficiency.

6 REFERENCES

[1] E. Gelenbe, P. Kammerman, and T. Lam, "Performance considerations in totally mobile wireless," *Performance Evaluation*, vols. 36–37, pp. 387–399, 1999. https://doi.org/10.1016/S0166-5316(99)00019-X

[2] A. Munro, "Mobile middleware for the reconfigurable software radio," *IEEE Communications Magazine*, vol. 38, no. 8, pp. 152–161, 2002. https://doi.org/10.1109/35.860867

[3] A. A. Koutsorodi, E. F. Adamopoulou, K. P. Demestichas, and M. E. Theologou, "Terminal management and intelligent access selection in heterogeneous environments," *Mobile Networks and Applications*, vol. 11, pp. 861–871, 2006. https://doi.org/10.1007/s11036-006-0054-1

[4] T. Ando, K. Takahashi, Y. Kato, and N. Shiratori, "Maintenance of mobile system ambients using a process calculus," *Computer Networks*, vol. 32, no. 2, pp. 229–256, 2000. https://doi.org/10.1016/S1389-1286(99)00132-2

[5] H. D. Sharma, Y. Misra, S. Kumar, B. M. Rao, and B. Ch, "Expanding an education-based collision detection system created on virtual reality and augmented reality," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 17, no. 17, pp. 108–120, 2023. https://doi.org/10.3991/ijim.v17i17.42831

[6] J. Cardenas-Valdivia, J. Flores-Alvines, O. Iparraguirre-Villanueva, and M. Cabanillas-Carbonell, "Augmented reality for Quechua language teaching-learning: A systematic review," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 17, no. 6, pp. 116–138, 2023. https://doi.org/10.3991/ijim.v17i06.37793

[7] N. Terentieva, V. Karpenko, N. Yarova, N. Shkvyria, and M. Pasko, "Technological innovation in digital brand management: Leveraging artificial intelligence and immersive experiences," *Journal of Research, Innovation and Technologies*, vol. 4, no. 2, pp. 201–223, 2025. https://doi.org/10.57017/jorit.v4.2(8).06

[8] V. Elliott and J. Hodgson, "Setting an agenda for English education research," *English in Education*, vol. 55, no. 4, pp. 369–374, 2021. https://doi.org/10.1080/04250494.2021.1978737

[9] X. Wang, "Online and offline integration scheme of college English education under big data technology," *Journal of Cases on Information Technology*, vol. 26, no. 1, pp. 1–13, 2024. https://doi.org/10.4018/JCIT.348963

[10] X. Liang and J. Pang, "An innovative English teaching mode based on massive open online course and Google collaboration platform," *International Journal of Emerging Technologies in Learning*, vol. 14, no. 15, pp. 182–192, 2019. https://doi.org/10.3991/ijet.v14i15.11148

[11] J. J. Liu, "A college English teaching mode based on a computer network platform," *Agro Food Industry Hi-Tech*, vol. 28, no. 1, pp. 616–619, 2017.

[12] A. J. Moreno-Guerrero, A. R. García, M. R. Navas-Parejo, and C. R. Jiménez, "Digital literacy and the use of augmented reality in teaching science in secondary education," *Revista Fuentes*, vol. 23, no. 1, pp. 108–124, 2021. https://doi.org/10.12795/revistafuentes.2021.v23.i1.12050

[13] A. G. de Moraes Rossetto, T. C. Martins, L. A. Silva, D. R. Leithardt, B. M. Bermejo-Gil, and V. R. Leithardt, "An analysis of the use of augmented reality and virtual reality as educational resources," *Computer Applications in Engineering Education*, vol. 31, no. 6, pp. 1761–1775, 2023. https://doi.org/10.1002/cae.22671

[14] S. Chen and R. Xiao, "Influence factors and strategies of teacher-student interactive behaviors in sports class teaching," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 13, no. 10, pp. 7025–7036, 2017. https://doi.org/10.12973/ejmste/78717

[15] X. Ma, "The intersection of innovative teaching methods, digital learning tools and innovative behavior in music art learning," *Current Psychology*, vol. 43, pp. 29154–29169, 2024. https://doi.org/10.1007/s12144-024-06498-0

[16] D. R. Bie and M. Fan, "On student evaluation of teaching and improvement of the teaching quality assurance system at higher education institutions," *Chinese Education & Society*, vol. 42, no. 2, pp. 100–115, 2009. https://doi.org/10.2753/CED1061-1932420212

[17] Y. Feng, "An evaluation method of PE classroom teaching quality in colleges and universities based on grey system theory," *Journal of Intelligent & Fuzzy Systems*, vol. 38, no. 6, pp. 6911–6915, 2020. https://doi.org/10.3233/JIFS-179769

7 AUTHOR

Fang Gao holds a bachelor's degree in Arts from Hebei University of Science and Technology (2012) and a master's degree in education from Hebei Normal University (2019). She works as a Lecturer in the Department of Basic Courses of Hebei Vocational College of Resources and Environment. Her research focuses on English teaching in vocational colleges, and she has published over two papers and one book (E-mail: bloodygaofang@126.com).