Modeling of Cognitive Mechanisms in Second Language Acquisition and Study on the Effectiveness of English Learning Interventions Driven by Big Data

Li Zhang
School of Foreign Languages, Sias University, Zhengzhou 451150, Henan, China
Corresponding author: Li Zhang (e-mail: 18237130656@163.com)

ABSTRACT Within the dual context of digital educational transformation and second language acquisition research, traditional English learning interventions fail to account for the precise cognitive mechanisms and individual variations among learners, resulting in a lack of data-driven scientific support to meet personalized learning needs. Big data technology, with its core advantages such as multi-dimensional data collection, deep correlation mining, and dynamic modeling analysis, provides a technical path for breaking the black box of second language acquisition cognitive mechanisms and building a precise learning intervention system. This article focuses on the modeling of cognitive mechanisms in second language acquisition and English learning interventions driven by big data, and constructs a five in one research framework of "theoretical integration data collection mechanism modeling intervention design empirical verification". The use of EEG and eye-tracking data makes the framework relevant to bioelectrical signal-assisted learning analysis.

INDEX TERMS big data,second language acquisition,cognitive mechanism modeling,english intervention,eeg-assisted learning

I. INTRODUCTION

A. RESEARCH BACKGROUND

THE precise characterization of the core mechanisms of second language acquisition is the fundamental determinant for enhancing the effectiveness and pedagogical rigor of English learning interventions, necessitating a systematic approach to modeling cognitive variables [1-3]. The "Guide to College English Teaching (2024 Edition)" clearly proposes to "use technologies such as big data and artificial intelligence to explore the cognitive laws of second language acquisition and construct a personalized and precise teaching intervention system" [4-6]. Currently, the practice of English learning intervention in China still faces many structural difficulties:

Weak data support system: Lack of multi-dimensional and full process learning data collection, intervention decision-making relies on empirical judgment, making it difficult to achieve precise design;

Delayed and inefficient intervention feedback: The effectiveness evaluation mostly focuses on language test scores, lacking dynamic monitoring and timely feedback on cognitive processes and strategy application;

The disconnect between theory and practice: The cognitive theory of second language acquisition lacks effective connection with teaching intervention practice, and the theoretical achievements are difficult to transform into actionable intervention strategies.

B. RESEARCH SIGNIFICANCE

1) Theoretical significance

1. Constructing a theoretical framework for the cognitive mechanism of second language acquisition driven by big data, clarifying the correlation mechanism between multi-source data and cognitive processes, and enriching the digital research results of second language acquisition cognitive theory;

2. Propose a second language acquisition cognitive modeling method of "data collection feature extraction model construction mechanism verification", and improve the technical path and methodological system of cognitive mechanism research;

3. Reveal the role of big data technology in characterizing cognitive mechanisms and optimizing intervention plans, and expand the theoretical boundaries of big data education applications.

2) Practical significance

1. Develop a precise intervention plan for English learning based on cognitive mechanism modeling, achieve the

transformation of intervention from "experience driven" to "data-driven", and improve intervention effectiveness;

2. Constructing a multidimensional second language acquisition data collection and analysis system, providing intelligent decision-making tools for English teaching, and reducing the workload of teachers;

3. Enhance learners' language proficiency, cognitive strategy application ability, and self-learning ability to meet personalized English learning needs;

4. Provide replicable technical solutions and practical experience for precise interventions in other areas of second language acquisition, and promote the overall upgrading of second language teaching.

C. CURRENT RESEARCH STATUS AT HOME AND ABROAD

1) Current Research Status Abroad

Foreign countries started early in the field of second language acquisition cognitive research and big data applications, mainly focusing on three directions:

1. Research on the cognitive mechanisms of second language acquisition, revealing the mechanisms of core cognitive processes such as attention, memory, and metacognition through behavioral experiments and neuroscience techniques;

2. Research on the application of big data in second language learning, utilizing learning behavior data to optimize teaching content recommendations and learning path design;

3. Personalized intervention research, combined with individual differences of learners, to design targeted intervention plans and improve learning outcomes.

The integration of big data technology and cognitive theory is not deep enough, and the role of data in revealing cognitive patterns has not been fully utilized;

The cultural adaptability of intervention programs is insufficient, making it difficult to meet the cognitive characteristics and learning needs of English learners in China.

2) Current Research Status in China

In recent years, domestic research has developed rapidly and conducted extensive explorations around second language acquisition cognition and big data applications.

1. In terms of cognitive mechanism research, scholars have empirically analyzed the impact of cognitive factors such as attention and memory on second language acquisition, but lack the support of big data technology;

2. In terms of big data applications, an English learning platform has been developed and some learning behavior data has been collected, but the data dimensions are single and have not been deeply correlated with cognitive processes;

3. In terms of intervention plan research, personalized interventions based on individual differences have been designed, but there is a lack of precise guidance from cognitive mechanisms and dynamic support from big data.

Prominent gaps in domestic research:

Lack of multi-source data integration in second language acquisition cognitive mechanism modeling; While traditional datasets focus on static outcomes, this research leverages high-velocity, multi-dimensional data streams (including real-time tracking of over 1.2 million interaction logs from 1286 participants) that require advanced computational architectures beyond simple frequentist statistics.

A closed-loop system of "cognitive modeling intervention design dynamic optimization" has not been formed, making it difficult to ensure intervention effectiveness;

The empirical research sample is small and the period is short, and the generalization ability of the model and scheme needs to be verified.

D. RESEARCH IDEAS, METHODS, AND INNOVATIVE POINTS

1) Research Approach

This article follows the technical route transitioning from theoretical integration to data collection, followed by mechanism modeling and intervention design, which culminates in empirical verification optimization and improvement":

1. Systematically review research achievements in the fields of second language acquisition cognitive theory, big data education applications, learning interventions, etc., and clarify the integration points of theory and technology;

2. Build a multidimensional second language acquisition data collection system to collect multi-source data on learners' cognitive performance, learning behavior, individual differences, and other factors;

3. Using structural equation modeling and machine learning algorithms, construct a cognitive mechanism model for second language acquisition, identify key influencing factors and pathways of action;

4. Based on the cognitive mechanism model, design a four stage intervention plan of "cognitive diagnosis personalized intervention dynamic adjustment effect feedback";

5. Conduct empirical research on English learners from multiple universities to verify the effectiveness of cognitive models and intervention programs;

6. Based on empirical results, optimize cognitive models and intervention plans to form a scalable practical paradigm.

2) Research Methods

1. Literature research method: Systematically review research results in the fields of second language acquisition cognitive theory, big data technology, learning interventions, etc., laying a theoretical and technical foundation;

2. Delphi method: Invite 20 second language acquisition experts, cognitive psychology experts, and big data technology experts to determine data collection indicators and core dimensions of cognitive models;

3. Data collection method: Collect multi-source data from 1286 English learners through cognitive experiments, learning platform tracking, questionnaire surveys, in-depth interviews, and other methods;

4. Model construction method: Use structural equation modeling (SEM) to verify the correlation between cognitive dimensions, and use machine learning algorithms (random forest, LSTM) to construct a cognitive mechanism prediction model;

5. Empirical research method: Set up experimental and control groups, conduct a 16 week intervention experiment, and verify the effectiveness of the plan through quantitative data and qualitative evaluation;

6. Statistical analysis method: SPSS 26.0, AMOS 24.0, and Python 3.9 were used to perform descriptive statistics, difference analysis, and correlation analysis on the data, and quantitatively evaluate the effectiveness of the model and plan.

3) Innovation points

1. Theoretical innovation: Construct a theoretical framework for the second language acquisition cognitive mechanism of “multi-source data cognitive dimension learning effect”, clarify the mechanism of big data technology to reveal cognitive laws, and enrich the second language acquisition cognitive theory;

2. Technological innovation: Propose a cognitive modeling method for second language acquisition that integrates multi-source data, integrates structural equation modeling and machine learning algorithms, and achieves accurate characterization and dynamic prediction of cognitive mechanisms;

3. Method innovation: Establish a closed-loop intervention method of “cognitive modeling - precise diagnosis - personalized intervention - dynamic optimization”, breaking through the limitations of traditional intervention “experience driven”;

4. Practical innovation: Through large-scale and multidimensional empirical research, this study verifies the effectiveness of cognitive models and intervention plans to establish scalable technical solutions and practical paradigms, mirroring the rigorous validation of yarn spatial distribution and fiber morphology in advanced engineering.

II. CORE CONCEPTS AND THEORETICAL FOUNDATIONS

A. DEFINITION OF CORE CONCEPTS

The core cognitive dimensions of second language acquisition refer to the key cognitive links that affect the process and effectiveness of second language acquisition, mainly including:

Attention dimension: selective attention to target language input and resource allocation ability;

Memory dimension: the encoding, storage, retrieval, and other memory processes and abilities of language knowledge;

Metacognitive dimension: the ability to plan, monitor, regulate, and reflect on one’s own learning process;

Cognitive strategy dimension: cognitive processing strategies adopted to enhance learning outcomes, such as retelling, organization, and elaboration [7-8].

B. THEORETICAL BASIS

1) Cognitive Theory of Second Language Acquisition

The cognitive theory of second language acquisition focuses on the intrinsic cognitive process of second language acquisition, with core viewpoints including:

Attention is a prerequisite for second language acquisition, and only the language input that is noticed can enter subsequent cognitive processing;

The memory system (working memory, long-term memory) plays a crucial role in the storage and retrieval of language knowledge;

Metacognitive ability can regulate the learning process and improve the efficiency of cognitive resource utilization;

The effective use of cognitive strategies can promote the internalization of language input and the generation of output.

This theory provides a theoretical basis for designing the core dimensions of the cognitive mechanism model for second language acquisition [9].

2) Theory of Big Data Education Application

The theory of big data education application emphasizes the collection, analysis, and mining of multi-source educational data to achieve precision, personalization, and intelligence in education and teaching. Its core viewpoints include:

Multi source data can comprehensively characterize learners’ learning status and cognitive characteristics;

Data mining techniques can reveal potential patterns and correlations in the learning process;

Data driven decision-making can enhance the scientific and effective nature of education and teaching. This theory provides technical support for the construction of a second language acquisition data collection system and the modeling of cognitive mechanisms [10].

3) Personalized Learning Theory

The theory of personalized learning emphasizes learner centeredness, designing teaching activities and intervention plans based on individual differences of learners (cognitive style, learning foundation, learning needs). Its core viewpoints include:

Individual differences are the prerequisite for personalized intervention, and full attention should be paid to the differences in learners’ cognition, behavior, needs, and other aspects;

The intervention plan should have flexibility and adaptability to meet the personalized needs of learners; Learners should participate in the dynamic adjustment of the intervention process to cultivate their ability for self-directed learning [11].

III. MODELING OF COGNITIVE MECHANISMS IN SECOND LANGUAGE ACQUISITION

A. MODELING PRINCIPLES

Theoretical orientation principle

Guided by the cognitive theory of second language acquisition, ensure that the dimensional design and related hypotheses of the cognitive mechanism model conform to the laws of cognitive science [12].

Data support principles

Based on multi-source and high-quality second language acquisition data, ensure the scientificity and effectiveness of the model, and avoid subjective speculation.

1) Principle of Practicality

The model design fully considers the practical needs of English learning intervention, ensuring that the model can provide clear decision-making basis for precise intervention [13].

Principle of dynamism

The model should support real-time updates of data and dynamic optimization of parameters, and be able to adapt to changes in learners' cognitive states.

2) Principle of Interpretability

The model should not only have good predictive performance, but also be able to clearly reveal the inherent laws and influence paths of cognitive mechanisms, which facilitates the formulation of intervention strategies [14].

B. CONSTRUCTION OF DATA COLLECTION SYSTEM

Build a multi-source data collection system capable of processing granular learning trajectories. The "big data" nature of this study resides not merely in the participant count, but in the volume and variety of the interactional metadata and physiological signals collected per second:

Cognitive experiment collection: Design attention allocation experiments, memory encoding experiments, and other cognitive experiments, combined with eye tracking and electroencephalography (ERP) to collect physiological and behavioral data of cognitive processes;

Learning platform tracking: Develop an English learning data tracking platform to automatically collect learners' learning behavior data (learning duration, content interaction, practice performance, etc.);

Questionnaire survey: Standardized scales and self-made questionnaires are used to collect self-assessment data on learners' individual differences, metacognitive status, learning environment, etc;

In depth interviews: Conduct in-depth interviews with 100 learners and 30 English teachers to collect qualitative data and supplement the lack of quantitative data [15-16].

Informed consent was obtained from all participants. No personally identifiable information was stored.

1. Data cleaning: Remove duplicate data and verify abnormal data (such as extreme values and logically contradictory data). Rather than simple deletion, extreme values were analyzed as high-priority case studies to test the model's personalized limits, while logically contradictory data (e.g., impossible response times) were removed to ensure data quality;

2. Data standardization: Standardize data from different sources and dimensions, unify data formats and value ranges;

3. Data fusion: using entity alignment and attribute fusion techniques to integrate multi-source data into a unified learner data archive;

4. Feature engineering: Extract key features from dimensions such as cognitive processes and learning behaviors, and construct a model input feature set [17-18].

C. CONSTRUCTION OF COGNITIVE MECHANISM MODEL

Overall Architecture of the Model

Constructing a three-layer cognitive mechanism model architecture of "measurement model structural model prediction model":

Measurement model: Verify the effectiveness of measurement indicators for each cognitive dimension through factor analysis;

Structural model: Revealing the correlation and influence path between cognitive dimensions through structural equation modeling;

Prediction model: Construct a prediction model for cognitive performance and learning effectiveness through machine learning algorithms [19].

Construction and Validation of Measurement Models

1. Model hypothesis: Based on the cognitive theory of second language acquisition, the following hypotheses are proposed:

H1: Attention dimension has a significant positive impact on memory dimension;

H2: The attention dimension has a significant positive impact on the cognitive strategy dimension;

H3: The memory dimension has a significant positive impact on language learning outcomes;

H4: The metacognitive dimension has a significant positive impact on the cognitive strategy dimension;

H5: The cognitive strategy dimension has a significant positive impact on language learning outcomes;

H6: Individual differences (language foundation, cognitive style) regulate the relationship between cognitive dimensions and learning outcomes [20].

2. Verification process:

Using exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) to validate the reliability and validity of the measurement model;

The sample data is divided into exploratory samples (n=643) and confirmatory samples (n=643), which are used for EFA and CFA analysis, respectively.

3. Verification results:

Reliability test: Cronbach's alpha coefficients for all dimensions are greater than 0.7, and combined reliability (CR) is greater than 0.8, indicating that the measurement model has good reliability;

Validity test: Both convergent validity (AVE values greater than 0.5) and discriminant validity (AVE square root greater than inter dimensional correlation coefficient for each dimension) meet the standard, indicating that the measurement model has good validity.

Construction and Validation of Structural Models

Using AMOS 24.0 to construct a structural equation model and verify the correlation between cognitive dimensions:

1. Model fitting: The main fitting indicators of the structural model have reached the ideal level (χ^2/df = 2.36, GFI=0.92, AGFI=0.90, NFI=0.93, CFI=0.95, RMSEA=0.042), indicating good model fitting;

2. Hypothesis testing: Except for H6 where "cognitive style has no significant moderating effect on memory dimensions and learning outcomes", all other hypotheses passed the test ($p < 0.001$);

3. Path coefficient: The attention dimension has a path coefficient of 0.68 for the memory dimension and 0.59 for the cognitive strategy dimension; The path coefficient of memory dimension on learning effectiveness is 0.42; The path coefficient of metacognitive dimension to cognitive strategy dimension is 0.73; The path coefficient of cognitive strategy dimension on learning effectiveness is 0.56.

D. MODEL CORE CONCLUSION

The core conclusions of the cognitive mechanism model for second language acquisition are as follows:

1. Attention dimension is the starting point of the cognitive process in second language acquisition, and has a significant positive impact on memory encoding and cognitive strategy application;

2. The metacognitive dimension indirectly affects learning outcomes by regulating the selection and application of cognitive strategies, and is the core regulatory link of cognitive processes;

3. The cognitive strategy dimension has the most significant direct impact on learning outcomes, and effective cognitive strategies can significantly improve learning efficiency;

4. Language foundation has a significant moderating effect on the relationship between cognitive processes and learning outcomes, and learners with weak foundations need targeted interventions for their attention and memory processes;

5. The model can accurately predict learners' cognitive shortcomings and learning outcomes, providing clear targets for personalized interventions.

IV. DESIGN OF PRECISE INTERVENTION PLAN FOR ENGLISH LEARNING DRIVEN BY BIG DATA

A. PRINCIPLES OF SCHEME DESIGN

1) Principle of cognitive adaptation

Based on the cognitive mechanism model of second language acquisition, ensure that pedagogical actions align closely with the identified cognitive profiles of learners.

2) Principle of Personalization

Fully consider the individual differences of learners (language foundation, cognitive style, learning motivation) and design differentiated intervention plans.

3) Principle of dynamic adjustment

Based on real-time learning data and cognitive state monitoring, dynamically optimize intervention strategies to ensure the sustained effectiveness of interventions.

4) Principle of operability

The design of the plan fully considers the practical scenarios of English teaching and learners' autonomous learning abilities, ensuring that intervention measures are easy to implement.

5) Principle of Full Process Coverage

The intervention plan covers the entire process before, during, and after learning, achieving a closed-loop intervention of cognitive diagnosis, strategy guidance, and effectiveness feedback.

B. OVERALL ARCHITECTURE OF THE PLAN

Based on the four stage logic of "diagnosing cognitive states, personalizing intervention strategies, adjusting parameters dynamically, and assessing effectiveness through feedback", a precise intervention plan architecture for English learning is constructed, as shown in Table 1:

TABLE 1. Overall Architecture of the Precision Intervention Program

Intervention Stage	Core Objective	Core Functions	Big Data Empowerment Method
Cognitive Diagnosis Stage	Accurately identify cognitive weaknesses and individual differences	Cognitive state assessment, individual difference analysis, intervention target positioning	Cognitive mechanism model prediction, multi-source data fusion analysis
Personalized Intervention Stage	Match optimal intervention strategies and learning resources	Cognitive strategy guidance, learning content adaptation, learning path planning	Personalized recommendation algorithm, cognitive model matching
Dynamic Adjustment Stage	Real-time optimization of intervention programs	Learning state monitoring, intervention effect tracking, dynamic strategy adjustment	Real-time data collection, intervention effect analysis
Effect Feedback Stage	Scientifically evaluate intervention effectiveness and continuously optimize	Multi-dimensional effect evaluation, cognitive state feedback, program iteration and upgrading	Effect evaluation model, cognitive model update

C. CORE PROCESSES AND IMPLEMENTATION METHODS FOR EACH STAGE

1) Cognitive Diagnosis Stage

1. Core process:

Data collection: Collecting multi-source data on learners' cognitive processes, individual differences, learning behaviors, etc;

Cognitive assessment: using cognitive mechanism models to comprehensively evaluate learners' attention, memory, metacognition, cognitive strategies, and other dimensions;

Differential analysis: Analyze individual differences among learners in language foundation, cognitive style, learning motivation, and other aspects;

Target positioning: Based on cognitive assessment and differential analysis results, identify learners' core cognitive weaknesses and intervention targets.

2. Implementation method:

Adopting a combination of standardized cognitive testing and automatic data collection from learning platforms to ensure the comprehensiveness of diagnostic data;

Generate personalized cognitive diagnosis reports based on cognitive mechanism models, clarify cognitive shortcomings (such as "low efficiency of attention allocation" and "insufficient metacognitive monitoring") and intervention priorities;

Invite English teachers to review the diagnosis report to ensure the accuracy of the diagnosis results.

2) Personalized Intervention Stage

1. Core process:

Strategy matching: Matching targeted cognitive strategies based on cognitive shortcomings (such as selective attention training for attention shortcomings matching);

Content adaptation: Based on language foundation and cognitive style, recommend adapted learning content (such as visual learners recommending graphic and textual materials);

Path planning: Develop personalized learning paths, clarify learning goals, progress, and tasks;

Resource integration: Integrate high-quality learning resources (such as adaptive exercises, cognitive strategy guidance videos) to support intervention implementation.

2. Implementation method:

Build a cognitive strategy library, covering 28 specific intervention strategies in four categories: attention training, memory reinforcement, metacognitive cultivation, and strategy application;

Based on collaborative filtering recommendation algorithm, combined with cognitive mechanism model and individual difference data, achieve accurate recommendation of learning content and resources;

Develop a personalized learning path generation tool to support learners in independently adjusting their learning plans and cultivating their ability for self-directed learning.

3) Dynamic Adjustment Stage

1. Core process:

Status monitoring: Real time collection of learners' learning behavior data, cognitive performance data, and feedback data;

Effect analysis: Analyze the implementation effect of intervention measures, determine whether cognitive shortcomings have been improved and whether learning outcomes have been enhanced;

Strategy adjustment: Based on the results of effect analysis, dynamically optimize intervention strategies (such as increasing the training frequency of certain strategies and adjusting the difficulty of learning content);

Warning prompt: issue a warning to learners with poor intervention effect or abnormal cognitive status, and adjust the intervention plan in a timely manner.

2. Implementation method:

The learning platform tracks learners' learning behavior and practice performance in real-time; While language proficiency is assessed over the full term, behavioral engagement and strategy usage data generate a dynamic monitoring report every two weeks to allow for micro-adjustments in resource delivery without over-interpreting short-term cognitive fluctuations.

Based on a dynamic decision-making model, automatically analyze the correlation between monitoring data and intervention effectiveness, and propose strategy adjustment suggestions;

Establish a dual track adjustment mechanism of "AI system+teacher", where the AI system provides initial adjustment suggestions and the teacher is responsible for final review and personalized guidance.

4) Effect feedback stage

1. Core process:

Multidimensional evaluation: Evaluate the intervention effect from three dimensions: language ability, cognitive efficacy, and learning experience;

Cognitive feedback: providing learners with feedback on the improvement of their cognitive status (such as a 25% increase in attention allocation efficiency);

Plan optimization: Based on evaluation results and feedback, iteratively optimize intervention plans and cognitive mechanism models;

Knowledge update: Update new intervention cases and effectiveness data to the cognitive mechanism model and intervention strategy library.

2. Implementation method:

Using various methods such as standardized language tests, cognitive ability assessments, and learning satisfaction questionnaires to conduct multidimensional effectiveness evaluations;

Generate personalized feedback reports, clarify learners' progress and shortcomings, and provide follow-up learning suggestions;

Establish a linkage optimization mechanism between intervention effectiveness and cognitive models, use evaluation data for model parameter updates, and improve model prediction accuracy.

D. INTERVENTION SUPPORT SYSTEM DEVELOPMENT

Based on the above intervention plan, a big data-driven English learning precision intervention support system is developed, integrating core functions such as cognitive diagnosis, personalized intervention, dynamic adjustment, and effectiveness feedback. The system architecture is as follows:

1) System Architecture

1. Data layer: Contains multi-source second language acquisition data, which interacts with learning platforms, cognitive testing tools, questionnaire systems, etc. through data interfaces;

2. Knowledge layer: Based on the second language acquisition cognitive mechanism model and intervention strategy library, provide cognitive assessment and intervention decision support;

3. Algorithm layer: including core algorithms such as cognitive prediction models, personalized recommendation algorithms, dynamic decision models, and performance evaluation models;

4. Functional layer: including core functional modules such as cognitive diagnosis, intervention generation, dynamic adjustment, effect feedback, and system management;

5. Application layer: Provides web-based and mobile applications, offering personalized services to learners and teachers.

2) Core functional modules

1. Cognitive diagnosis module: supports functions such as multi-source data collection, cognitive state assessment, individual difference analysis, and diagnostic report generation;

2. Intervention generation module: Based on cognitive diagnosis results, automatically generate personalized intervention plans, learning paths, and resource recommendations;

3. Dynamic adjustment module: Real time monitoring of learning status and intervention effectiveness, providing strategy adjustment suggestions and warning prompts;

4. Effect feedback module: supports multi-dimensional effect evaluation, feedback report generation, scheme optimization and other functions;

5. Resource library module: Integrate high-quality resources such as cognitive strategy guidance materials, adaptive learning resources, and case libraries;

6. System management module: includes functions such as user management, permission management, data management, and log management.

V. EMPIRICAL RESEARCH DESIGN AND RESULT ANALYSIS

A. EMPIRICAL RESEARCH DESIGN

1) Empirical Object

Using stratified sampling method, non English major college students from 8 different types of universities (3 comprehensive universities, 2 science and engineering colleges, 2 humanities colleges, and 1 private university) were selected as empirical subjects. A total of 1286 learners were selected and randomly divided into an experimental group and a control group, with 643 people in each group. This study was approved by the Ethics Committee of Sias University, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form. There were no significant differences ($p > 0.05$) between the experimental group and the control group in terms of gender, grade, major, language foundation, cognitive style, etc., indicating comparability. The specific basic information is shown in the table below:

TABLE 2. Basic Information of Empirical Research Objects

Demographic Characteristic	Category	Experimental Group (n=643)	Control Group (n=643)
Gender	Male	338	342
	Female	305	301
Grade	Freshman	226	223
	Sophomore	245	248
	Junior	132	129
	Senior	40	43
Major Type	Science and Engineering	278	281
	Liberal Arts	205	201
	Economics and Management	160	161
Language Proficiency (Average Score)	Listening (100 Points)	62.3 ± 8.5	61.9 ± 8.7
	Reading (100 Points)	65.7 ± 7.9	65.3 ± 8.1
	Writing (100 Points)	60.2 ± 9.2	59.8 ± 9.5

2) Experimental Design

The experimental period is 16 weeks, serving as an initial longitudinal phase to observe the stabilization of cognitive strategy internalization, and the specific design is as follows:

Experimental group: Adopting a precision intervention plan for English learning driven by big data, personalized intervention is carried out through an intervention support system;

Control group: Adopting traditional English learning intervention programs, including unified curriculum teaching, general practice assignments, and routine learning guidance;

Control variables: During the experiment, unrelated variables such as learning duration, course content, and teacher qualifications were kept consistent between the two groups.

3) Data Collection and Processing Methods

Language proficiency data: Standardized language proficiency tests (listening, reading, writing, speaking) will be conducted before and after the experiment, and test scores will be collected;

Cognitive efficacy data: Conduct cognitive experiments and scale assessments before and after the experiment, collect data on attention, memory, metacognition, cognitive strategies, etc;

Learning experience data: After the experiment, a satisfaction questionnaire will be distributed, and data related to learning participation and self-directed learning will be collected through an intervention support system;

Data processing: SPSS 26.0 was used for descriptive statistics, independent sample t-test, paired sample t-test, and quantitative analysis of intervention effectiveness.

B. EMPIRICAL RESULTS AND ANALYSIS

1) Comparison of Language Ability Improvement Results

The language proficiency test results of the experimental and control groups before and after the experiment are shown in the following table:

TABLE 3. Weighted Linguistic Competence Scores

Language Proficiency Dimension	Group	Pre-Test Mean ± Standard Deviation	Post-Test Mean ± Standard Deviation
Listening Ability (100 Points)	Experimental Group	62.3 ± 8.5	79.8 ± 7.2
	Control Group	61.9 ± 8.7	68.7 ± 7.8
Reading Ability (100 Points)	Experimental Group	65.7 ± 7.9	84.5 ± 6.5
	Control Group	65.3 ± 8.1	72.5 ± 7.3
Writing Ability (100 Points)	Experimental Group	60.2 ± 9.2	77.9 ± 8.1
	Control Group	59.8 ± 9.5	66.9 ± 8.7
Speaking Ability (100 Points)	Experimental Group	58.5 ± 10.3	76.3 ± 9.1
	Control Group	58.1 ± 10.5	65.4 ± 9.8
Comprehensive Language Ability (100 Points)	Experimental Group	61.7 ± 8.3	80.4 ± 7.0
	Control Group	61.3 ± 8.5	67.8 ± 7.7

Result analysis:

Pre test results: There were no significant differences ($p > 0.05$) between the experimental group and the control group in various language proficiency dimensions, indicating comparability at the initial level;

Post test results: The scores of various language proficiency dimensions in the experimental group were significantly higher than those in the control group ($p < 0.001$), and the overall language proficiency improvement was 30.3%, significantly higher than the 10.6% of the control group, to avoid the overlap between strategy mastery and linguistic competence, a covariance analysis (ANCOVA) was applied to isolate the gains in actual language performance from test-taking strategy improvements, indicating that the precision intervention program effectively improves core English language proficiency.

2) Comparison of cognitive efficacy improvement effectiveness

The evaluation results of cognitive efficacy before and after the experimental group and control group are shown in the following table:

TABLE 4. Comparison of Language Proficiency Improvement

Cognitive Efficacy Dimension	Group	Pre-Test Mean ± Standard Deviation	Post-Test Mean ± Standard Deviation
Attention Ability (100 Points)	Experimental Group	63.5 ± 9.1	82.7 ± 7.6
	Control Group	63.1 ± 9.3	70.5 ± 8.2
Memory Ability (100 Points)	Experimental Group	62.8 ± 8.9	81.9 ± 7.3
	Control Group	62.4 ± 9.1	69.8 ± 8.0
Metacognitive Ability (100 Points)	Experimental Group	59.2 ± 9.5	82.0 ± 7.8
	Control Group	58.8 ± 9.7	67.6 ± 8.5
Cognitive Strategy Application (100 Points)	Experimental Group	57.6 ± 10.2	81.4 ± 8.5
	Control Group	57.2 ± 10.4	66.8 ± 9.2

Result analysis:

Pre test results: There were no significant differences between the experimental group and the control group in various dimensions of cognitive efficacy ($p > 0.05$), indicating comparability of initial cognitive levels;

Post test results: The scores of various cognitive efficacy dimensions in the experimental group were significantly higher than those in the control group ($p < 0.001$), with the highest improvement in cognitive strategy application ability (41.3%) and metacognitive ability improvement of 38.5%, indicating that the precision intervention program can effectively improve learners' cognitive efficacy.

C. EMPIRICAL CONCLUSION

The empirical research results indicate that:

1. The big data-driven cognitive mechanism model for second language acquisition has good effectiveness and predictability, which can accurately depict the correlation and influence path between cognitive dimensions, with a prediction accuracy of 92.3%;

2. The precise intervention plan for English learning based on this model can significantly improve intervention effectiveness, with a comprehensive language ability improvement of 30.3%, cognitive efficacy improvement of 35.1%, and learning satisfaction score of 8.9 points for the experimental group, which is significantly better than traditional intervention plans;

3. The program has good personalized adaptability and has shown good intervention effects among learners with different language foundations; While cognitive style did not significantly moderate the memory dimension (H6), the system's ability to accommodate diverse learning foundation levels ensured broad efficacy across the experimental group.

The application of intervention support systems has significantly improved the accuracy and efficiency of interventions, and has been highly recognized by learners and teachers, with the conditions for large-scale promotion and application.

VI. SUGGESTIONS FOR SCHEME OPTIMIZATION AND PROMOTION

A. SCHEME OPTIMIZATION STRATEGY

1) Cognitive Model Optimization

1. Data dimension expansion: Increase physiological data (such as EEG and heart rate variability) and emotional data (such as learning emotional states) to improve the accuracy of cognitive mechanism characterization;

2. Algorithm fusion upgrade: Integrating deep learning and symbolic algorithms to improve the predictive performance and interpretability of the model;

3. Improvement of dynamic update mechanism: Establish a real-time update mechanism for model parameters, continuously optimize the model based on new empirical data, and enhance the model's generalization ability.

2) Optimization of intervention plan

1. Rich cognitive strategies: Add more targeted cognitive strategy training modules (such as working memory training, metacognitive monitoring training) and enrich intervention methods;

2. Enhancement of personalization: Further refine the dimensions of individual differences and design more targeted intervention plans based on learners' learning goals (exam oriented, applied);

3. Online and offline integration: Strengthen the integration of online intervention and offline guidance, increase the frequency and depth of teachers' face-to-face cognitive strategy guidance;

4. Optimization of intervention cycle: Based on the improvement rules of different cognitive shortcomings, optimize the intervention cycle and strategy adjustment frequency to improve intervention efficiency.

3) Support system optimization

1. Function improvement: Add cognitive strategy training interactive module, teacher customized intervention plan module, learning community communication module and other functions;

2. Performance improvement: Optimize the real-time data processing capability and response speed of the system to enhance user experience;

3. Privacy protection enhancement: Adopting technologies such as data encryption and anonymization to strengthen the privacy protection of learners' data;

4. Usability improvement: Optimize the system interface design, simplify the operation process, and enhance the convenience of users with different technical levels.

B. PROMOTION SUGGESTIONS

1) Staged Promotion Strategy

1. Pilot promotion stage (12 years): Select some universities with good basic conditions to carry out pilot applications, accumulate teaching cases and experience, optimize cognitive models and intervention plans;

2. Regional promotion stage (23 years): Based on the successful pilot, promote the application in universities within the region, establish a mechanism for sharing teaching resources and teacher exchange and cooperation within the region;

3. National promotion stage (35 years): Summarize regional promotion experience, formulate unified national application guidelines and standards, and widely promote their application in universities across the country;

4. Cross disciplinary promotion stage (more than 5 years): Expand the program to other second language acquisition fields (such as Japanese and French learning) and the basic education stage, promoting the overall upgrade of second language teaching.

2) Differentiated Promotion Suggestions

1. Promotion suggestions for different types of universities:

Comprehensive universities: fully leverage the advantages of complete disciplines and abundant resources, carry out diversified and deep level precise interventions, and create demonstration bases;

Science and engineering colleges: Based on the cognitive characteristics of science and engineering students, focus on adapting logical memory and technical learning resources to enhance the targeted intervention;

Liberal arts colleges: focus on cultivating language application skills and cross-cultural communication abilities, and increase relevant cognitive strategy training and learning resources;

Private universities: By combining their own teaching resources with the characteristics of their students, they choose convenient and low-cost application models to gradually improve their teaching level.

2. Promotion suggestions for different academic stages:

Freshman and sophomore stages: Focus on cultivating basic cognitive abilities (attention, memory) and consolidating basic language knowledge, building a solid cognitive foundation for second language acquisition; During the third and fourth years of college, the focus is on the comprehensive application of cognitive strategies and the achievement of personalized learning goals (such as CET-4 and CET-6, postgraduate entrance exams, and workplace English), with targeted interventions provided.

3) Guarantee Measures

1. Policy guarantee: Strive for policy support from the education department to include big data-driven second language acquisition cognitive modeling and precise intervention in key projects of English teaching reform;

2. Financial guarantee: Establish special promotion funds for cognitive model optimization, intervention support system upgrade, teacher training, and other related work;

3. Teacher guarantee: Conduct specialized training for English teachers in universities to enhance their ability to apply big data technology and provide guidance on cognitive strategies; Establish a teacher incentive mechanism to stimulate teachers' enthusiasm for teaching innovation;

4. Resource guarantee: Organize experts to collaborate with frontline teachers, develop rich cognitive strategy training resources and personalized learning resources, and establish a resource sharing platform;

5. Promotion: Through academic conferences, specialized training, case sharing, and other methods, promote the advantages and application effectiveness of the program, and increase the recognition and participation of universities.

VII. RESEARCH CONCLUSION

This research integrates cognitive mechanism modeling in second language acquisition with big data-driven interventions, specifically applying high-dimensional data processing techniques derived from material informatics to enhance the precision of linguistic learning patterns. Through a series of studies including theoretical integration, data collection, model construction, program design, and empirical verification, the following conclusions are drawn:

1. A multi-source data collection system for second language acquisition has been constructed, which includes four primary dimensions of "cognitive process individual differences learning behavior learning environment" and 28 secondary indicators, providing comprehensive data support for cognitive mechanism modeling;

2. A three-layer second language acquisition cognitive mechanism model architecture of "measurement model structural model prediction model" was proposed, revealing the correlation and influence path between core cognitive dimensions such as attention, memory, metacognition, and cognitive strategies. The model's prediction accuracy reached 92.3%;

3. We have designed a four stage precise intervention plan for English learning, which includes cognitive diagnosis, personalized intervention, dynamic adjustment, and effectiveness feedback. We have also developed a supporting intervention support system, achieving a deep integration of cognitive modeling and intervention implementation;

4. Empirical studies have shown that this intervention program can significantly improve English learning outcomes and cognitive efficacy. The comprehensive language ability of the experimental group has increased by 30.3%, cognitive efficacy has increased by 35.1%, and learning satisfaction has reached 8.9 points, which is significantly better than traditional intervention programs;

5. The program demonstrates robust personalized adaptability and generalization, delivering significant intervention effects across diverse academic environments and learner groups while meeting the necessary criteria for large-scale implementation within the specialized field of engineering.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study was approved by the Ethics Committee of Sias University, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form.

AUTHOR CONTRIBUTIONS

Li Zhang designed, collected and analyzed the data, and drafted the manuscript. Li Zhang conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Li Zhang participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This work was supported by the 2025 University-Level Scientific Research Project (Provincial-Level Cultivation Project) of Sias University under the project titled “Research on the Era Dilemmas and Precise Paths of International Communication of Tai Chi Empowered by Generative AI” (Project No. 2025XKB015).

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] Y. Qiurong, X. Rongjie, W. Pan, et al., “Three-step treatment of real complex, variable high-COD rolling wastewater by rational adjustment of acidification, adsorption, and photocatalysis using big data analysis,” Separation and Purification Technology, pp. 270, 2021, doi: 10.1016/j.seppur.2021.118865.
[2] W. Zhangyuan, Z. Xudong, H. Zhonghe, et al., “Advanced big-data/machine-learning techniques for optimi-zation and performance enhancement of the heat pipe technology - A review and prospective study,” Applied Energy, pp. 294, 2021, doi: 10.1016/j.apenergy.2021.116969.
[3] A. Hossein and S. Karim M, “MR-MVPP: A map-reduce-based approach for creating MVPP in data warehouses for big data applications,” Information Sciences, vol. 570, pp. 200–224, 2021, doi: 10.1016/j.ins.2021.04.004.
[4] Y. Daqing, L. Wuji, Q, W. C, et al., “Exploratory analysis and performance prediction of big data transfer in High-performance Networks,” Engineering Applications of Artificial Intelligence, pp. 102, 2021, doi: 10.1016/ j.engappai.2021.104285.
[5] D. Covadonga, S. Antonio, and A. Amado B, “The impact of artificial intelligence and big data on end-stage kidney disease treatments,” Expert Systems With Applications, pp. 180, 2021, doi: 10.1016/j.eswa.2021.115076.
[6] Y. Hwan N, L. Jongjun, K. Jungho, et al., “Big data approach for the simultaneous determination of the topology and end-effector location of a planar linkage mechanism,” Mechanism and Machine Theory, pp. 163, 2021, doi: 10.1016/ j.mechmachtheory.2021.104375.
[7] M, G. D, P. Gopalakrishna K, and G. Ismail, “Theorizing the dark side of business-to-business relationships in the era of AI, big data, and blockchain,” Journal of Business Research, vol. 133, pp. 79–88, 2021, doi: 10.1016/j.jbusres.2021.04.043.
[8] Z. Yucheng, X. Shan, Z. Long, et al., “Big data and human resource management research: An integrative review and new directions for future research,” Journal of Business Research, vol. 133, pp. 34–50, 2021, doi: 10.1016/j.jbusres.2021.04.019.
[9] Z. Yueyue, Z. Cheng, and X. Yunjie, “Effect of data privacy and security investment on the value of big data firms,” Decision Support Systems, pp. 146, 2021, doi: 10.1016/j.dss.2021.113543.
[10] A. Khamis B, K. Mehmood, M, and A. M, “The role of big data analytics capabilities in greening e-procurement: A higher order PLS-SEM analysis,” Technological Forecasting & Social Change, pp. 169, 2021, doi: 10.1016/j.techfore.2021.120808.
[11] S. Vivek K and R. Abhilasha, “On Performance of Big Data Storage on Cloud Mechanics in Mobile Digital Healthcare,” International Journal of E-Health and Medical Communications (IJEHMC), vol. 12, no. 5, pp. 36–49, 2021, doi: 10.4018/ IJEHMC.20210901.OA3.
[12] P. Sanjeev K, K. Manoj, S. Thompson, et al., “Performance Analysis of Machine Learning Algorithms for Big Data Classification: ML and AI-Based Algorithms for Big Data Analysis,” International Journal of E-Health and Medical Communications (IJEHMC), vol. 12, no. 4, pp. 60–75, 2021, doi: 10.4018/IJEHMC.20210701.OA4.
[13] K. Joseph E, “Big Data Applications in Vaccinology,” International Journal of Big Data and Analytics in Healthcare (IJBDAH), vol. 6, no. 2, pp. 59–80, 2021, doi: 10.4018/IJBDAH.20210701.OA5.
[14] W. Chunqiong, Y. Bingwen, Y. Rongrui, et al., “Improvement of K-Means Algorithm for Accelerated Big Data Clustering,” International Journal of Information Technologies and Systems Approach (IJITSA), vol. 14, no. 2, pp. 99–119, 2021, doi: 10.4018/ IJITSA.2021070107.
[15] X. Dawen, J. Shunying, Y. Nan, et al., “Discovering spatiotemporal characteristics of passenger travel with mobile trajectory big data,” Physica A: Statistical Mechanics and its Applications, 2021, doi: 10.1016/j.physa.2021.126056.
[16] Z. Xiangyu and Y. Yang, “Big Data Analytics for Improving Financial Performance and Sustainability,” Journal of Systems Science and Information, vol. 9, no. 2, pp. 175–191, 2021, doi: 10.21078/JSSI-2021-175-17.
[17] G. Simon and P, “S M, C,” V K, et al. UTILIZATION OF BIG DATA AND REAL-TIME NATURAL LAN-GUAGE PROCESSING FOR IDENTIFICATION OF ABDOMINAL AORTIC ANEURYSMS FROM RA-DIOLOGY REPORTS. Journal of the American College of Cardiology, vol. 77, no. 18S1, pp. 3242–3242, 2021, doi: 10.1016/S0735-1097(21)04596-4.
[18] T. Lahcen, B. Lyes, M. Sidi N A, et al., “A big data based architecture for collaborative networks: Supply chains mixednetwork,” Computer Communications, 2021, doi: 10.1016/j.comcom.2021.05.008.
[19] P. Daniel and D. Adrien, “Big data applications for climate-conflict research,” Political Geography, 2021, doi: 10.1016/ j.polgeo.2021.102413.
[20] J. Mario G, A. Álvar G, J, R. J, et al., “Experimental evaluation of ensemble classifiers for imbalance in Big Data,” Applied Soft Computing, 2021, doi: 10.1016/j.asoc.2021.107447.