



A Study on the Construction of a Diversified Evaluation System for Ideological and Political Education in College English Courses Empowered by Digital Intelligence

Yue Yin^{ID}, Xiukui Song*

Department of Foreign Languages, Qilu Normal University, Jinan, China

Email: y1017716992@126.com, *songxiukui@163.com

How to cite this paper: Yin, Y. and Song, X.K. (2026) A Study on the Construction of a Diversified Evaluation System for Ideological and Political Education in College English Courses Empowered by Digital Intelligence. *Open Access Library Journal*, 13: e15723.

<https://doi.org/10.4236/oalib.1115723>

Received: July 1, 2026

Accepted: October 20, 2026

Published: October 23, 2026

Copyright © 2026 by author(s) and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

As a required undergraduate course, College English carries the dual task of language training and value guidance. Against the backdrop of educational digitalization, the evaluation of its ideological and political component has long suffered from single evaluators, rigid indicators and unquantifiable value outcomes. This paper reports an empirical case study that implemented, over a sixteen-week semester at Qilu Normal University, a digitally empowered diversified evaluation system built on the CIPP model and learning analytics. The system integrated a multi-modal intelligent evaluation platform, a four-dimensional indicator system whose weights were determined through expert consultation and the Analytic Hierarchy Process (AHP), and a four-subject scoring mechanism. Data were collected from 216 non-English-major freshmen in four intact classes, covering online discussion posts, cross-cultural reflection journals, oral presentation audio and cross-cultural writing tasks. Algorithmic scores showed strong agreement with teacher ratings (Pearson $r = 0.83$), and a pre-post comparison indicated a significant improvement in students' cultural identity and cross-cultural confidence ($p < 0.01$). The study specifies the observable variables feeding each CIPP dimension, justifies the weight allocation, describes how algorithmic scoring is validated and corrected, and discusses the boundaries and limitations of semantic and emotional analysis in capturing ideological growth. The findings offer an evidence-based reference for local normal universities seeking to reform their College English ideological and political evaluation.

Subject Areas

Foreign Language Education

Keywords

Digital Intelligence Empowerment, College English, Curriculum Ideological and Political Education, Diversified Evaluation System, Learning Analytics, CIPP Evaluation Model, Empirical Case Study

1. Introduction

As higher education enters the stage of comprehensive digital upgrading, big data, artificial intelligence and learning analytics have been reshaping the teaching logic, resource supply and quality evaluation of university general courses [1]. College English, as a compulsory course covering nearly all undergraduate majors, carries the important educational function of guiding students to form sound cross-cultural cognition, cultivate cultural confidence and tell China's stories in a foreign language. In recent years, Chinese universities have comprehensively promoted the construction of ideological and political education in College English, yet the supporting evaluation system has clearly lagged behind teaching reform practice, and planning, implementing and evaluating the ideological and political dimension of College English remains a complex process [2].

Traditional evaluation of the ideological and political dimension relies mainly on teachers' subjective scoring of final papers and classroom performance. It foregrounds the assessment of linguistic knowledge and neglects continuous tracking of the process of value shaping; evaluators are confined to course teachers, while evaluation indicators fail to quantify students' ideological awareness, cross-cultural emotional attitudes and moral practice. To reduce arbitrary judgement and establish a big-data-based, real-time monitoring mechanism, ecological and multi-dimensional evaluation frameworks for College English teaching have been proposed [3]. Local normal universities such as Qilu Normal University nevertheless face specific difficulties: large student enrolment, scarce full-time foreign-language teachers, scattered ideological-political teaching data, and limited capacity among teachers to operate digital evaluation tools [4].

2. Theoretical Framework and Research Design

2.1. The CIPP Evaluation Models

Stufflebeam's CIPP model divides educational evaluation into four interlocking dimensions—Context, Input, Process and Product—and emphasises whole-process, developmental and improvement-oriented evaluation rather than a single terminal judgement [5]. By breaking the limitation of summative evaluation, the model matches the diversified evaluation demands of College English ideological and political education, where the value growth process matters as much as its outcome, and supplies the structural skeleton for the multi-dimensional indicator system developed in Section 3.

2.2. Learning Analytics and the Student Value Portrait

Learning analytics relies on digital platforms to capture multi-dimensional learning behaviour and, through semantic analysis and emotion recognition, to mine the attitudes and value tendencies embedded in behavioural traces. Big-data portrait techniques have already been applied to precise curriculum ideological and political education, building label systems from behavioural and online-learning data so that teachers can track dynamic changes in students' thinking [6]. In this study, the same logic is used to move from subjective qualitative judgement to a data-informed, mixed quantitative and qualitative evaluation of otherwise invisible ideological and political literacy.

2.3. Digital-Intelligent Integration and the Smart Classroom

Digital-intelligent integration reshapes not only resources but also the teaching design of College English ideological and political courses, requiring value guidance to be embedded naturally in language tasks rather than added on mechanically [7]. Within the smart classroom, technology mediates interaction between language input, cultural experience and value guidance, making continuous process evaluation feasible at scale [8]. These two strands justify choosing a platform-and-data-driven design that treats evaluation as part of teaching rather than as a separate end-of-course act.

2.4. Research Design and Case Context

The study adopted an instrumental case-study design. The case was the College English course taught to first-year non-English majors at Qilu Normal University in the first semester of 2025-2026. Participants were 216 students in four intact classes taught by four College English instructors, taught over sixteen weeks. Data sources included the university intelligent learning platform logs, students' online discussion posts, cross-cultural reflection journals, recorded oral presentations and group debates, and final cross-cultural reports and English presentations on Chinese topics. A five-point Likert scale of cultural identity and cross-cultural confidence was administered before and after the semester. Algorithmic scores generated by the platform were compared with independent teacher ratings to estimate validity, and the pre-post scale scores were compared with paired-samples t-tests to examine value-added effect.

3. Construction of the Diversified Evaluation System

3.1. A Multi-Modal Intelligent Evaluation Platform

Taking the university's existing intelligent learning platform as the carrier, a dedicated evaluation module for College English ideological and political education was developed around three data-collection functions. First, text semantic analysis automatically identifies the value tendency in students' cross-cultural essays, reading-reflection logs and online discussion posts, extracts keywords related to cul-

tural confidence and national identity, and returns a quantitative ideological-literacy score. Second, audio and video behaviour tracking collects oral presentations, group debates and classroom speech, and analyses emotional polarity and cross-cultural cognitive attitude through speech-recognition algorithms. Third, a whole-cycle learning portrait integrates online study duration, resource-browsing preferences, ideological task completion and classroom interaction frequency into a personalised growth profile.

3.2. Four-Dimensional CIPP Indicators and Their Weighting

Combined with the characteristics of College English ideological and political teaching, a four-level indicator system covering Context, Input, Process and Product was built. The weights—Context 15%, Input 20%, Process 40% and Product 25%—were not assigned intuitively. Following the Delphi procedure and the Analytic Hierarchy Process (AHP) validated in CIPP-based curriculum ideological and political evaluation, thirteen experts (foreign-language teaching researchers, curriculum-ideology specialists and educational measurement specialists) completed two rounds of consultation; pairwise comparison matrices were constructed using Saaty's nine-point scale and all consistency ratios were below 0.10 [9]. The heavy weight on Process (40%) reflects the developmental nature of value formation, which unfolds over time rather than being captured by a single terminal task, and is consistent with expert consensus that process implementation is the core lever of curriculum ideological and political quality.

The observable variables behind each dimension are specified as follows. Context evaluation (15%) draws on teachers' ideological-political teaching designs and the matching degree of bilingual ideological resources with teaching objectives. Input evaluation (20%) measures students' pre-course ideological-cognition baseline, willingness to engage in digital learning, and cross-cultural preparation, captured by the entrance survey and pre-course tasks. Process evaluation (40%) is the core link and combines classroom interaction, ideological-reflection task completion, online cross-cultural exchange behaviour and collaborative group performance, scored jointly by the platform, peers and teachers. Product evaluation (25%) covers the final cross-cultural report, the English presentation on a Chinese topic, and an integrated score of language ability and ideological literacy.

3.3. A Four-Subject Collaborative Scoring Mechanism

To balance objectivity and subjectivity, the single-teacher model was replaced by a four-subject mechanism, in line with evidence that digitally mediated combinations of automated, teacher, peer and self-assessment jointly promote learning [10]. The intelligent algorithm provides a preliminary evaluation (40%) by automatically scoring objective data on learning behaviour, text semantics and audio emotion. The teacher provides a professional evaluation (30%), reviewing classroom performance, individual ideological growth and practical performance, and correcting machine bias. The student completes a digital self-evaluation (15%),

summarising changes in cross-cultural cognition and ideological gains. Peer mutual evaluation (15%) scores group members' collaborative tasks, cross-cultural communication attitude and cultural expression under a unified rubric. The four scores are aggregated into a composite ideological-literacy grade.

3.4. Validation and Correction of Algorithmic Scoring

Because algorithmic scoring is novel and carries bias risk, its validity was checked before full rollout. Two trained instructors independently rated a stratified sample of 120 cross-cultural essays and reflection journals on the same rubric used by the platform. Pearson correlation between algorithmic and mean teacher scores reached $r = 0.83$, and Cronbach's alpha across the two human raters was 0.87, indicating acceptable agreement. A two-stage correction mechanism then operated throughout the semester: the platform flagged any script whose algorithmic score deviated by more than one standard deviation from the peer/self-evaluation mean for that task for mandatory teacher review, and teachers annotated every score they overrode so that the scoring model could be recalibrated. To control bias, scoring was blind to student identity, and length, genre and dialect distributions were monitored each task; the teacher's 30% weight acts as the final human-in-the-loop safeguard.

4. Empirical Implementation and Preliminary Findings

4.1. Operational Sequence across the Semester

The system ran as a closed, repeatable sequence across the sixteen-week semester. At Week 1, the entrance survey established the Input baseline and introduced the evaluation rubric to students. From Week 2 to Week 14, continuous data collection fed the platform, which produced weekly preliminary algorithm scores. Each fortnight, teachers reviewed flagged scripts, released individual and class feedback through the platform, and adjusted subsequent ideological-political tasks. At Week 15, Product evaluation was completed through final reports and presentations. At Week 16, the exit survey measured value-added growth, and the semester evaluation report was generated. The sequence—data collection, platform scoring, teacher review, feedback release and teaching adjustment—thus formed an “evaluation-analysis-feedback-teaching reform” loop rather than a one-off terminal test.

4.2. Observable Variables per CIPP Dimension

For transparency, each CIPP dimension was tied to concrete observable variables. Context was measured by the completeness of ideological-political elements in the syllabus and the expert-rated matching of bilingual resources to objectives. Input was measured by the entrance cultural-identity scale, pre-course video completion and readiness quiz scores. Process was the richest dimension: it used platform log variables (study duration, resource access, discussion-post frequency and quality keywords), the length and value-tendency score of weekly reflection journals, au-

dio emotion polarity from oral tasks, and rubric scores for group collaboration. Product was measured by the final cross-cultural report, the English presentation on a Chinese topic, and an integrated language-and-value score. This specification prevents the evaluation from resting on vague judgement and makes every score traceable to a recorded behaviour or text.

4.3. Implementation Results

Three kinds of evidence emerged. First, algorithm-teacher agreement was stable at $r = 0.83$ across tasks, giving confidence in the automated component. Second, the paired pre-post comparison on the cultural-identity and cross-cultural confidence scale showed a significant improvement: the mean rose from 3.42 (SD = 0.61) to 4.01 (SD = 0.58), $t(215) = 11.74$, $p < 0.01$, with a medium-to-large effect. This value-added pattern is consistent with empirical work showing that process-oriented value-added evaluation of College English ideological and political education enhances students' learning self-efficacy, professional identity and patriotic attitude [11]. Third, teacher feedback indicated that continuous portraits reduced the invisible workload of tracking 216 students, while the four-subject mechanism made students more aware of their own value growth. These preliminary findings support the feasibility of the system at a local normal university of comparable scale.

5. Discussion: Boundaries, Limitations and Safeguards

5.1. Humanistic Orientation and Technological Boundaries

Digital intelligence remains an auxiliary evaluation tool and cannot replace teachers' humanistic care and value guidance. Semantic and emotion analysis reliably captures surface lexical and sentiment signals, but it may not fully capture reflective depth, tacit ideological growth, or context-specific meaning that depends on cultural nuance and teacher-student dialogue [12]. Scores are therefore treated as evidence to inform, not replace, teacher judgement; the evaluation process is designed around communication and guidance rather than simple data scoring, so that excessive quantification does not erode the humanistic nature of ideological and political education. A single case at one local normal university also limits generalisability; multi-site replication and longer longitudinal tracking are needed before the weights and agreement levels can be treated as stable norms.

5.2. Data Security and Institutional Safeguards

Because the system stores sensitive ideological-learning data, special management measures were formulated to clarify access authority, prohibit unauthorised disclosure of students' ideological evaluation records, and adopt encrypted storage and desensitisation. The Foreign Languages Department, Academic Affairs Office and Information Technology Center operate jointly: the Academic Affairs Office standardises the evaluation system, the Information Technology Center maintains technology and data security, and the Foreign Languages Department designs and

iterates the indicators. Teachers receive regular training in curriculum-ideology theory, platform operation and educational big-data analysis, and the system is reviewed on an academic-year cycle, with indicator weights and platform modules adjusted on the basis of teacher and student feedback.

6. Conclusion

This paper has reported an empirical case study rather than a purely conceptual proposal. Facing the separation of language and value assessment, the lack of whole-process tracking and the single evaluator that characterise College English ideological and political evaluation at local normal universities, the study used the CIPP model and learning analytics to construct and trial a digitally empowered diversified evaluation system—a multi-modal platform, an AHP-weighted four-dimensional indicator system with explicit observable variables, a four-subject scoring mechanism validated against teacher ratings, and a repeatable semester operational loop. The one-semester implementation at Qilu Normal University produced acceptable algorithm-teacher agreement and a significant pre-post gain in cultural identity and cross-cultural confidence, while the discussion has been explicit about the boundaries of semantic and emotional analysis. For comparable local normal colleges, the system offers a measurable, traceable and development-oriented basis for curriculum-ideology quality assessment, and a concrete path toward the deeper integration of digital technology with the fundamental task of fostering virtue through education.

Funding

2025 Research Incubation Project for Basic Education Teaching Reform at Qilu Normal University, “Research on Practical Strategies for Integrating Fine Traditional Chinese Culture into Primary and Secondary School English Textbooks” (Project Number: JP202404).

Author Contributions

Yue Yin was the principal investigator and first author. She conceived and designed the study, developed the theoretical framework (CIPP model and learning analytics), built the four-dimensional indicator system and its AHP-derived weights, designed the four-subject scoring mechanism and algorithm validation procedure, conducted the data collection and analysis, and drafted the manuscript.

Xiukui Song served as the supporting and corresponding author. She provided academic supervision, advised on the theoretical framework and indicator design, assisted in expert recruitment and system implementation, and reviewed and proofread the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Zhu, Z.T. and Hu, J. (2022) Analysis of the Essence and Research Prospect of Digital Transformation of Education. *China Educational Technology*, No. 4, 1-8, 25. (In Chinese).
<https://zdjy.cbpt.cnki.net/portal/journal/portal/client/paper/a3608d46d7902a2998c3710daa836356>
- [2] Huang, L.Y. (2022) Research on the Effect Evaluation of College English Curriculum Ideology and Politics Based on the CIPP Model. *Educational Academic Monthly*, No. 2, 57-63. (In Chinese).
https://www.zhangqiaokeyan.com/academic-journal-cn_detail_thesis/02012105892976.html
- [3] Wang, N. and Zheng, L.W. (2023) Construction of a College English Teaching Ecological Evaluation System Based on the DPSIR Model. *Journal of Inner Mongolia Normal University (Educational Science Edition)*, **36**, 89-99. (In Chinese).
<https://r.cnki.net/kmobile/Journal/detail/CFIG/NMJX202301013>
- [4] Zou, D.P. (2021) Opportunities and Challenges Faced by English Majors' Teachers in the Era of "Artificial Intelligence+." *Journal of Qilu Normal University*, **36**, 24-30. (In Chinese).
<https://sdjx.cbpt.cnki.net/portal/journal/portal/client/paper/0eb50ec8ab5004cbc63a2714725b69c6>
- [5] Li, F.Q. (2022) Reflections on a CIPP-Based Practical Teaching Evaluation System for Ideological and Political Theory Courses in Medical Colleges. *Journal of Nanjing University of Chinese Medicine (Social Science Edition)*, **23**, 59-63. (In Chinese).
<http://xbsk.njucm.edu.cn/cn/article/id/d787d206-d75a-4cc7-9482-a02e1dd261d9>
- [6] Niu, Y.F., Ye, J. and Xu, X.Z. (2024) Research on the Implementation Path of Precision Curriculum Ideological and Political Education Based on Big Data Portraits. *China Educational Informatization*, **30**, 95-102. (In Chinese).
<https://r.cnki.net/kmobile/Journal/detail/CFJC/YXXX202411010>
- [7] Ren, Z.F. and Dong, X.B. (2026) Research on the Integration Model of College English Curriculum Ideological and Political Education and National Discourse Competence Empowered by Digital Intelligence. *China Educational Technology*, No. 9, 130-137. (In Chinese).
<https://zdjy.cbpt.cnki.net/portal/journal/portal/client/paper/3099be293467b43c73e1b496ca6aca5>
- [8] Wang, M. (2024) Construction of College English Smart Classroom from the Perspective of Curriculum Ideology and Politics: Concepts and Practice. *Journal of Anhui University of Technology (Social Science Edition)*, **41**, 79-81. (In Chinese).
<https://xuebaosk.ahut.edu.cn/cn/article/id/43ff81ed-70d4-4789-9e2f-834e265c1db5>
- [9] Zhang, Y.J. (2025) Construction and Empirical Study of a Curriculum Ideology and Politics Teaching Evaluation System Based on the CIPP Model and the AHP Method. *Journal of Jiangxi Vocational and Technical College of Electricity*, **38**, 25-30. (In Chinese). <https://r.cnki.net/kmobile/Journal/detail/CFJC/DLZG202503005>
- [10] Wu, H.T. and Zhou, S.Q. (2026) Educational Digitalization Empowering Multiple Learning-Promoting Assessments in College English Teaching. *Journal of Hubei Open Vocational College*, **39**, 190-192. (In Chinese).
<https://inds3.cnki.net/kmobile/Journal/detail/MKZJ/HBHS202612062>
- [11] Mei, Q. (2026) Reform and Practice of Value-Added Evaluation of College English Curriculum Ideology and Politics Based on the OBE Concept. *Journal of Language Education*, **14**, 6-18, 129. (In Chinese).

<https://r.cnki.net/kmobile/Journal/detail/CFJC/YYZS202602002>

- [12] Miao, R.D. and Wang, Z. (2022) Technical Possibilities, Realistic Challenges and Countermeasures of Big Data Portrait Technology Boosting Precision Ideological and Political Education. *Studies in Ideological Education*, No. 7, 41-46. (In Chinese).
<https://sxjyyj.ustb.edu.cn/qikanlanmu/xuekejianshe/2022-10-26/1875.html>