Evaluation Method and Empirical Analysis of the Achievement of Ideological and Political Education Goals in Digital Teaching Scenarios

Ying Shao, Jiangbin Mu*
Administrative Office, Qinhuangdao Open University, Qinhuangdao 066000, Hebei, China
Corresponding author: Jiangbin Mu (e-mail: ydshaoying@126.com).

ABSTRACT In digital and intelligent teaching scenarios, the goals of ideological and political education in courses face the dilemma of being "implicit and difficult to reveal, difficult to trace the process, and difficult to evaluate the effectiveness." Traditional assessments struggle to dynamically capture the evolution of students' ideological and political literacy. To address this, this paper proposes a method for assessing achievement by integrating multimodal process data. First, the three-dimensional goals of "cognitive identification—emotional internalization—behavioral practice" are deconstructed, and multi-source data mapping rules are established. Next, improved fuzzy hierarchical analysis is introduced for weight self-calibration, and an emotional temporal attention network is designed to extract deep emotional evolution features. Finally, a dynamic quantitative model is constructed based on evidence-based reasoning to achieve interpretable attribution. Experiments show that this method achieves a Kappa score of 0.92 with expert evaluation, and the three-dimensional internal reliability Cronbach's α is higher than 0.80. The correlation coefficient between process assessment and final exam scores (0.78) is superior to that of traditional questionnaires (0.44), and the attribution accuracy is 85%. Ablation experiments confirm the key role of multimodal integration and emotional temporal mechanisms. This method provides an effective technical path for the process evaluation of ideological and political education in courses.

INDEX TERMS ideological and political education in curriculum, digital and intelligent teaching, assessment of achievement of educational goals, multimodal data fusion, emotional time series analysis

I. INTRODUCTION

The popularization of digital teaching scenarios (such as smart classrooms, online learning platforms, and virtual simulation experiments) has provided rich interactive data and immersive learning environments for the development of ideological and political education in courses, but it has also brought new evaluation challenges [1]. The goals of ideological and political education in courses have significant implicit characteristics, which are manifested as a dynamic evolution process of value identification, emotional internalization, and behavioral practice, and are difficult to effectively capture through traditional summative assessments or subjective questionnaires [2]. Current practice faces a threefold dilemma: the implicit nature of these goals makes them hard to observe, their development process is difficult to trace, and their effectiveness is challenging to evaluate., which is specifically manifested as: the evaluation data source is singular, ignoring multimodal process information such as classroom interactive text, learning behavior logs, and facial expressions; the evaluation indicators mostly remain on the surface of ideological and political elements, lacking in-depth quantification of multidimensional educational goals; existing methods are difficult to handle the ambiguity and temporal evolution characteristics of ideological and political literacy in human-computer collaborative environments, resulting in a large deviation between the evaluation results and the actual educational effectiveness [3]. Therefore, exploring a method that can dynamically, accurately, and interpretably evaluate the achievement of ideological and political education goals in digital teaching scenarios has become a key issue that urgently needs to be addressed.

Many scholars have conducted research on the evaluation of the effectiveness of ideological and political education in courses. For example, Sun et al. [4] constructed a relatively systematic index system for the evaluation of ideological and political education in science and engineering courses, providing a structured framework for the evaluation of ideological and political education in courses; Zhou and Wang [5] conducted research on the evaluation system of ideological and political education in courses based on the CIPP model and AHP method, which strengthened the standardization of evaluation dimension design and weight allocation. Huang et al. [6] used multimodal learning analysis to evaluate teachers' self-regulated learning in technology-integrated teaching, while Yusuf et al. [7] used multimodal learning analysis to model students' learning behavior in programming classes, showing that multi-source heterogeneous data can more meticulously depict the teaching and learning process. However, most of the above studies still rely on a single data source or static indicators, and have failed to fully integrate multi-source heterogeneous process data in the context of digital and intelligent teaching scenarios.

To address the above challenges, some scholars have introduced intelligent technologies such as natural language processing and affective computing. For example, Shen et al. [8] evaluated learning engagement in MOOC learning scenarios based on facial expression recognition, demonstrating the application potential of visual emotional cues in learning state recognition; Tang et al. [9] further utilized facial expression recognition to detect students' emotional engagement in science learning, providing new evidence for the analysis of the relationship between classroom emotions and learning outcomes; Chen et al. [10] proposed a classroom student expression recognition method based on spatiotemporal residual attention networks, which improved the accuracy and robustness of emotion recognition in complex teaching video environments. These methods enhance the objectivity and timeliness of evaluation to a certain degree, but still existing problems like the fusion of multimodal data are mostly simple splicing or weighted average and fail to fully exploit the complementary and synergistic relationships among different modalities. In view of above problems, this paper proposes a method to evaluate the achievement of ideological and political education goals in courses based on multimodal process data, improved fuzzy hierarchical analysis are introduced for weight self-calibration, emotional temporal attention network is designed to extract deep emotional evolution features, and dynamic quantitative model based on evidence reasoning is constructed to achieve accurate evaluation and interpretable attribution of implicit ideological and political literacy.

The main research innovation of this paper is summarized in the following four aspects: First, a three-dimensional observable indicator system of "cognitive identification—emotional internalization—behavioral practice" applicable to digital teaching scenario was built, and the mapping rules of improved fuzzy hierarchical analysis between multi-source process data and educational goals were established; Second, an evaluation framework of improved fuzzy hierarchical analysis and emotional temporal attention network was constructed, and the indicator weight self-calibration and deep extraction of emotional evolution characteristics were realized; Third, an achievement dynamic quantification model of evidence reasoning was designed to support the comprehensive evaluation and shortcoming identification of individual and group; Fourth, the effectiveness, reliability and validity of the method were validated by empirical analysis, and an operable technical path of process evaluation of ideological and political education in courses was provided.

II. DECONSTRUCTION OF EDUCATIONAL GOALS

The overall goal of ideological and political education in courses is usually expressed as a three-pronged approach of "value shaping, knowledge transmission, and ability cultivation." However, in a digital and intelligent teaching scenario, this macro-goal needs to be broken down into observable and quantifiable specific indicators to establish a correspondence with classroom process data. Referring to the "knowledge, emotion, will, and action" theory in ideological and political education, and combining it with learners' actual behavioral performance in a digital and intelligent environment, this paper deconstructs the educational goals of ideological and political education in courses into three dimensions: "cognitive identification," "emotional internalization," and "behavioral practice."

Among them, the "cognitive identification" dimension measures learners' understanding and acceptance of ideological and political values. This construct is not captured by mere keyword matching, because the use of terminology does not automatically indicate genuine value acceptance. Therefore, the observation points include: (a) using ideological and political terminology in a contextually appropriate manner (e.g., applying the term correctly to a case); (b) consistently expressing a value-aligned stance across multiple discussion turns, as determined by a stance classifier; and (c) completing designated ideological and political reading tasks with satisfactory comprehension (e.g., verified by short quizzes or summary notes). Keyword hits alone are treated as a weak signal that triggers further stance analysis; only those hits that co-occur with a positive or neutral stance consistent with the educational objective contribute to the indicator. Thus, text data from discussion forums, bullet comments, and assignment responses are processed through a two-step pipeline: first, keyword detection; second, stance classification (positive, neutral, or negative). The final indicator score integrates both the frequency of stance-verified keyword usage and the overall consistency of value orientation, rather than relying on raw term counts. "Emotional internalization" dimension reflects the emotional feeling and attitude conversion of learners in the process of value identification. The observation points include: the expression of positive emotion to positive value cases, the critical attitude to negative phenomena, and meaningful fluctuations in facial expression related to "interest", "emotion", and "firmness". Importantly, facial expression data are not interpreted in isolation, because students may exhibit neutral resting faces, physical fatigue, or varying illumination in the classroom. To avoid false positives, this dimension relies on multimodal evidence: (a) facial expression changes are evaluated relative to each student's own baseline expression (obtained during neutral classroom moments), and (b) only temporal shifts in AU intensity that co-occur with relevant textual or behavioral signals (e.g., a student who shows increased AU12 while typing a value-aligned comment) are considered as evidence of emotional internalization. Raw expression values without contextual alignment are treated as noise. "Behavioral practice" dimension reflects the behavior performance of learners in the process of transforming ideological and political cognition and emotion into behavior. This dimension has three observation points: the proactive responsibility behavior of group collaboration, the participation record of after-class public service behavior, and the behavior log on the classroom discipline and academic integrity. This dimension bases on the learning behavior log data, such as attendance, completion rate of collaborative task, extracurricular practice check-in, etc.

To specify the mapping rule of operational mapping rule between the above three dimensions and multi-source data, this paper designs 4 to 6 quantifiable underlying indicators based on each dimension and clarify the time granularity of data collection (each classroom interaction or each learning task). For "Cognitive Identification" dimension, we design indicators like "Number of Hit Rates of Ideological and Political Keywords", "Value Stance Consistency Score", "Completion Rate of Ideological and Political Reading Tasks", for "Emotional Internalization" dimension, we design indicators like "Peak Value of Positive Emotion Intensity", "Frequency of Emotional Polarity Transformation", "Activation Duration of Facial AU Units (e.g., AU6, AU12)", for "Behavioral Practice" dimension, we design indicators like "Number of Active Speaking in Collaborative Tasks", "Extracurricular Practice Participation Rate", "Abnormal Signs in Academic Integrity Records". All underlying indicators are automatically collected from backend interface of digital teaching platform or obtained through semi-automatic annotation based on classroom recordings, which avoid the subjective bias brought by manual scoring by teachers. The deconstruction and mapping process provides the standardized input foundation for the subsequent process of weight self-calibration, emotional temporal feature extraction, and achievement synthesis calculation.

III. MULTIMODAL FEATURE EXTRACTION LAYER

For text interaction records in class discussion area, bullet comments and group collaboration, this paper chooses predictive language model BERT for feature extraction. Specifically, for each text, we first perform word segmentation and eliminate irrelevant symbols and stop words from the text; then input the text into BERT model and take output vector of last layer [CLS] as semantic representation of text (768 dimensions). On this basis, we construct a political and ideological keyword database (8 keywords with political and ideological orientation, including "patriotism", "responsibility", "fairness and justice", "rule of law" etc.), and obtain political and ideological relevance score of each text by calculating cosine similarity between text vector and keyword database vector. At the same time, we use binary classifier to judge the value stance of text (positive agreement, neutrality, negative deviation) and output the probability distribution of stance inclination. The output of these two, political and ideological relevance and stance inclination probability, are text feature of this module and will be input into the weight self-calibration stage.

Behavioral Log Feature Extraction Learning behavior log records students' behavior sequence in digital platform, including attendance time, courseware browsing time, number of posts in class discussion area, completion status of group collaboration task, check-in records of extracurricular practice. Because these behaviors all have obvious temporal dependence (regularity of attendance, consistency of task completion), so we use Temporal Convolutional Network (TCN) to extract feature of these behaviors. As compared with Convolutional Neural Network (CNN), TCN can extend receptive field by causal convolution and dilated convolution, and thus capture long-term behavioral pattern without gradient vanishing or exploding. Specifically, for each student, we encode his/her sequence of behavior in a certain class (sorted chronologically, time granularity is 5 minutes) into a multi-dimensional vector (each dimension represents standardized frequency or duration of certain type of behavior), and then input this vector into TCN network including 4 layers of dilated convolution, and then output a fixed-length behavioral feature vector (128 dimensions). This vector represents dynamic change of students' span of attention, consistency of task completion and depth of collaborative participation.

Facial expression feature extraction. We record video files of classes, and then get the change of students' facial expression in the learning process. It is one of the important data sources of the emotion internalization dimension. In this paper, we use the OpenFace toolkit to analyze the video recording frame by frame, and get the intensity value of each facial action unit(AU for abbreviation). We focus on the following AU related to the internalization of ideological and political emotion: AU6(cheeks raised, appeared in most of pleasure/ emotion), AU12(lip corner pull, appeared in positive emotion), AU4(brow lower, appeared in confusion/ seriously thinking), AU14(dimpler, appeared in contempt/ criticism). For each frame, OpenFace outputs the intensity value of each AU (range 0–5). Raw intensity values are susceptible to confounding factors such as individual differences in resting face, fatigue, or lighting conditions. Therefore, the following preprocessing steps are applied:
(1) Personal baseline subtraction: For each student, the mean AU intensity during the first 5 minutes of calm instruction (e.g., teacher's monologue) is computed and subtracted from all subsequent intensity values, reducing person-specific and lighting biases.
(2) Sliding-window relative changes: A sliding window (window size 30 seconds, step 10 seconds) is used to extract not the absolute intensity but the temporal gradient (i.e., rate of change) and the peak-to-baseline ratio for each AU, which are more robust to constant offsets.

(3) Multimodal validation: Facial expression features are only fed into the ETAN module together with text and behavioral features; the attention mechanism learns to down-weight expression segments that are not supported by congruent textual or behavioral signals (e.g., a sudden AU12 change without any relevant discussion activity).

These procedures effectively mitigate false positives arising from neutral resting faces, fatigue, or environmental lighting, while preserving genuine emotion-related variations.

IV. WEIGHT SELF-CALIBRATION AND EMOTIONAL EVOLUTION MODULE

Traditional Analytic Hierarchy Process takes the score given by expert to establish the judgment matrix, and after the weight is determined, it will not change no matter what happens in the teaching. In this paper, Fuzzy Analytic Hierarchy Process is put forward. Instead of using accurate values to represent the degree of influence from experts, triangular fuzzy number is used to represent the uncertainty of the degree of influence. At the same time, the feedback fine-tuning strategy based on student behavior information is added. The specific implementation process is as follows: Firstly, select three experts in curriculum-based ideological and political education to score the relative importance of each pair of "cognitive identification", "emotional internalization" and "behavioral practice" using fuzzy numbers (l, m, u) to establish fuzzy judgment matrix. By calculating the fuzzy feature vector and defuzzifying (using the centroid method), the initial weight vector $w_0 = (w_{cog0}, w_{emo0}, w_{beh0})$ of the three-dimensional indicators is obtained. Then, using student group behavioral characteristics extracted during the first 8 weeks of teaching (e.g., classroom interaction activity, average facial expression AU intensity, and fluctuations in task completion rate), a set of theory-driven feedback rules is designed. The adjustment is not based on raw data frequency alone; rather, it reflects the instructional principle that dimensions showing unexpectedly high student engagement or emotional intensity often indicate emerging teachable moments or areas where deeper internalization is occurring. Specifically, if the actual activation frequency of the underlying observation indicator corresponding to a certain dimension is significantly higher than a preset threshold (derived from expert expectations or historical class averages), the weight of that dimension is appropriately increased—because such elevated engagement signals that the dimension has become more salient for students' current learning trajectory. Conversely, if the frequency is significantly lower, the weight is moderately decreased or maintained, as this may suggest that the dimension is less challenging or less central to current instructional goals. All adjustments are bounded by a sigmoid function to ensure stability. The adjustment range is controlled using a sigmoid function to avoid drastic weight jumps. The final dynamically self-calibrated weight vector is $w_t = w_0 + \Delta w_t$, where t represents the teaching week. This mechanism allows the weights to reflect both the prior knowledge of teaching experts and adapt to the specific learning changes in each class.

Emotional Evolution Submodule. The internalization process of students' ideological and political literacy is often accompanied by emotional fluctuations and transitions. For example, when discussing hot social issues, the emotional shift from confusion to acceptance is more diagnostically valuable than emotional expression in a stable state. To capture these key temporal segments, this paper designs an Emotional Temporal Attention Network (ETAN). The network's input is an aligned sequence output from the multimodal feature extraction layer: text features (sentiment probability of a segment every 5 minutes), behavioral features (128-dimensional vectors output by TCN), and facial expression features (AU intensity sequence) from the same lesson are concatenated at time steps to form a joint feature vector with dimensions of (text feature dimension + behavioral feature dimension + facial expression feature dimension). The sequence length L equals the number of segments in the lesson (e.g., for a 90-minute lesson, with 5-minute segments, L=18). The essence of ETAN is a multi-layer attention mechanism: Firstly, a Bidirectional Long Short-Term Memory (BiLSTM) network is employed to encode the joint feature sequence's context dependencies. Secondly, the attention weight of each segment in time dimension is computed. The attention scoring function is learnable and comprehensively considers the similarity between the feature vector of segment and the context vector of whole sequence. The larger attention weight of segment, the more important moment of big emotion fluctuation or stance shift it represents. Finally, the encoded vector of each segment output by BiLSTM will be weighted sum according to their attention weights to get the "emotional evolution feature vector" of the whole lesson. This vector not only contains global information of sequence, but also emphasizes the information of key transition segments. Therefore, it can provide input with more temporally semantics for subsequent quantification. Structural parameters of Emotional Temporal Attention Network(ETAN)is shown in Table 1.

TABLE 1. Structural Parameters of the Emotional Temporal Attention Network

Layer	Type	Input Dimension	Output Dimension	Parameters	Example Attention Weight
Dimensionality Reduction	Fully Connected	912 (multimodal concatenation)	256	Linear + ReLU	—
Sequence Encoding	BiLSTM	256	256 (bidirectional)	Hidden 128×2, dropout=0.3	—
Attention	Additive Attention	256	1 (per time step)	Learnable scoring matrix	Weight 0.21 at step 5
Weighted Sum	Normalization	L steps × 256	256	Softmax	Key indices [4,7,12]
Output	Fully Connected	256	128 (emotional evolution features)	ReLU	—

V. ACHIEVEMENT MEASUREMENT AND ATTRIBUTION MODULE

For a single student, let the score of the i-th underlying indicator be s_i (normalized to the 0–1 interval), and the corresponding educational goal dimension for this indicator be $d(i) \in \{C, E, B\}$ (representing cognitive identification, emotional internalization, and behavioral practice, respectively). First, each indicator is treated as an independent piece of evidence, represented by a confidence distribution: $e_i = \{(H_n, \beta_{n,i}), n = 1, \ldots, N\}$, where H_n is the preset achievement level (this paper sets 5 levels: Excellent, Good, Average, Pass, Needs Improvement), and $\beta_{n,i}$ is the confidence level of indicator i belonging to level H_n. The confidence level is calculated based on the membership function (trapezoidal distribution) of the indicator score s_i and the level threshold. Then, multiple pieces of evidence within the same dimension are recursively synthesized according to the dimension weight w_d to obtain the confidence distribution of each dimension; finally, the confidence distribution of the individual's overall achievement is synthesized according to the total weight of the three-dimensional goals (self-calibrated w_t). Finally, the total individual achievement score, $Score_{ind}$, in the 0–1 interval is obtained through centroid defuzzification. Simultaneously, the marginal contribution of each dimension to the total score, $\Delta Score_{ind}^{(d)}$, is recorded for subsequent attribution analysis.

Achievement at the class or group level is obtained by aggregating individual achievement scores. This paper uses a weighted average method, with weights adjustable according to teaching needs (e.g., taking the arithmetic mean of individual student achievement scores, or weighting by student participation). The total group achievement score and the average scores of the three dimensions are output, facilitating horizontal comparison of the effects of different classes or teaching strategies.

Interpretability is a key focus of this module design. For students with low scores or abnormal fluctuations, the model outputs a "contribution heatmap": the horizontal axis represents the three educational dimensions, and the vertical axis represents the underlying indicators. The color intensity of the cells indicates the magnitude of the positive or negative contribution of that indicator to the total achievement score (red indicates a negative decrease, green indicates a positive increase). Specifically, the difference between the total achievement score and the achievement score recalculated after deleting a certain indicator is taken as the marginal contribution of that indicator, normalized, and mapped to a color scale. Teachers can quickly pinpoint students' specific weaknesses within the digital classroom environment—for example, "negative contribution to extracurricular practice participation rate in the behavioral practice dimension"—and then implement targeted teaching interventions.

However, it is important to acknowledge that ideological and political literacy is also shaped by external variables outside the digital classroom, such as family background, social media exposure, peer influence, and personal life experiences. The current heatmap only reflects contributions from in-class observable indicators (text, behavior, facial expressions) and does not account for these external factors. Consequently, the attribution results should be interpreted as "in-class process contributions" rather than comprehensive determinants of a student's overall ideological and political development. Teachers are encouraged to use the heatmap as a diagnostic tool for identifying classroom-specific engagement patterns, while supplementing it with holistic understanding of each student's broader context. Future work will explore integrating external data sources (e.g., surveys on family environment or social media usage) to enrich the attribution model.

Furthermore, key time segments of the emotional temporal attention network output are also presented as timeline markers on the heatmap, reminding teachers to review moments of emotional shifts in the classroom recordings. The entire attribution process requires no manual annotation and is entirely generated automatically by the model, ensuring diagnostic efficiency and objectivity.

VI. VERIFICATION INDEX SYSTEM FOR ACHIEVEMENT ASSESSMENT IN DIGITAL AND INTELLIGENT SCENARIOS

A. CONSISTENCY VERIFICATION EXPERIMENT BETWEEN ASSESSMENT RESULTS AND EXPERT EVALUATION

Five sets of digital and intelligent classroom recordings were selected, covering 120 students from different majors. Three experts in curriculum-based ideological and political education independently scored the students using the same three-dimensional index system, and the average score was used as the expert evaluation benchmark. This method was used to calculate the achievement score for each student and convert it into a grade (high, medium, low). The Kappa coefficient between the two was calculated to test the level of consistency. The results are shown in Table 2:

TABLE 2. Consistency Assessment of Evaluation Results and Expert Ratings

Student ID	Proposed Method Score	Expert 1 Score	Expert 2 Score	Expert 3 Score	Expert Mean Score	Level Agreement
S01	0.87	0.85	0.88	0.86	0.863	Yes
S02	0.72	0.7	0.74	0.71	0.717	Yes
S03	0.91	0.9	0.93	0.92	0.917	Yes
S04	0.58	0.62	0.6	0.59	0.603	Yes
S05	0.65	0.63	0.67	0.64	0.647	Yes
Overall	—	—	—	—	—	Kappa = 0.92

Calculations showed that the Kappa coefficient between the proposed method and the expert panel was 0.92, indicating near-perfect agreement (Kappa $\geq$ 0.81 is generally considered extremely strong consistency). Among the 120 students, only 3 students' grade assignments differed slightly from the expert mean scores, and all differences were confined to adjacent grade boundaries. Compared to traditional questionnaire-based assessment methods (where Kappa is typically 0.55–0.65 in similar courses), this method significantly improves the objectivity and expert acceptance of the assessment results, validating the effectiveness of the multimodal process data fusion strategy in quantifying implicit ideological and political literacy.

B. INTERNAL RELIABILITY TEST OF THREE-DIMENSIONAL SCORES

Classroom process data from all 120 students were selected, and internal consistency analysis was performed on the observation indicators included in the three dimensions of "cognitive identification," "emotional internalization," and "behavioral practice." Each dimension included 6 to 8 quantifiable indicators, including the number of text keyword hits, peak facial AU unit intensity, and behavioral log completion rate. Cronbach's α coefficient was used to test the homogeneity among the observation indicators within the same dimension.

TABLE 3. Internal Reliability Assessment of Three-Dimensional Scores

Dimension of Education	Number of Indicators	Cronbach's α	95% Confidence Interval
Cognitive Identification	7	0.86	[0.82, 0.90]
Emotional Internalization	8	0.83	[0.79, 0.87]
Behavioral Practice	6	0.81	[0.77, 0.85]

In Table 3, the Cronbach's α coefficients for the three dimensions are 0.86, 0.83, and 0.81, respectively, all higher than the commonly accepted reliability standard of 0.80, indicating good consistency and homogeneity among the internal observed indicators of each dimension. The "Cognitive Identification" dimension has the highest reliability, possibly related to the ease of standardization of text data; the "Behavioral Practice" dimension has a slightly lower reliability but is still within an acceptable range, likely due to incomplete records of some extracurricular self-directed behaviors. Overall, the three-dimensional indicator system is reasonably designed and can reliably reflect different aspects of the ideological and political education goals in digital and intelligent scenarios, providing a credible measurement basis for subsequent achievement synthesis calculations.

C. CRITERION-RELATED VALIDITY TEST OF FORMATIVE ASSESSMENT AND SUMMATIVE RESULTS

The mean cumulative attainment of each student within the first 8 weeks was taken as the predictor variable and the score (out of 100) of the independent ideological and political case analysis completed in week 9 was taken as the criterion variable. The correlation analysis was carried out on the pair for 114 valid samples. The Pearson correlation coefficient was compared with the scores obtained by the same group of students using a traditional questionnaire to evaluate the criterion-related validity of the method.

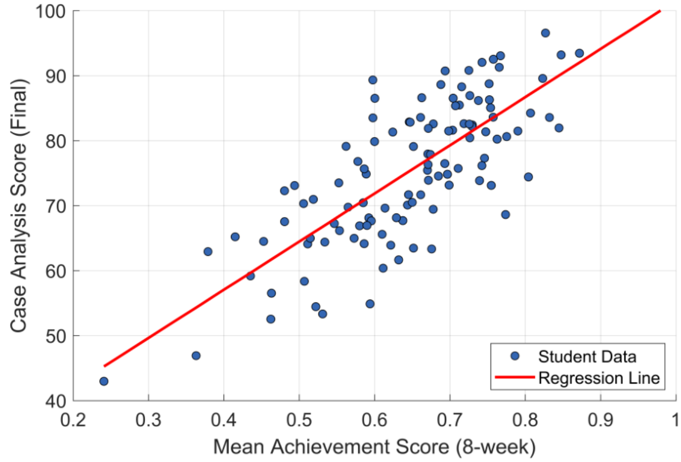

FIGURE 1. Criterion-related validity assessment of formative assessments and summative results

In Figure 1, the Pearson correlation coefficient between the proposed method's eight-week mean achievement score and the final case analysis score was 0.780 ($p<0.001$). For the same group of students, the coefficient based on a traditional self-report questionnaire was 0.44. A Fisher z-test indicated that the difference between these two correlations was statistically significant ($z=4.269$, $p=0.00002$). It should be noted that traditional questionnaires measure students' perceived ideological and political attitudes, which are susceptible to social desirability bias, whereas the proposed method taps into directly observable behavioral and facial data, capturing more implicit and process-oriented aspects of ideological and political literacy. Given these different constructs, the comparison does not imply that one method is universally "superior"; rather, it shows that the proposed method exhibits a stronger empirical association with summative case analysis performance. This finding suggests that multimodal process data may offer complementary predictive value beyond self-reported measures, thereby providing more reliable evidence for process-based evaluation of ideological and political education in courses.

D. ATTRIBUTION LOCALIZATION ACCURACY AND ABLATION COMPARISON VALIDATION EXPERIMENT

Twenty students with extremely low scores in one dimension of model output were chosen. The independent judgment of actual weak dimension of these 20 students was based on playback of classroom video, and the consistency rate between the attribution of dimension output of model and record of observation was calculated. In addition, an ablation experiment was designed to delete facial expression modality and emotional temporal attention attention network, and then compared the consistency (Kappa) of assessment and accuracy of attribution between the simplified model and complete model.

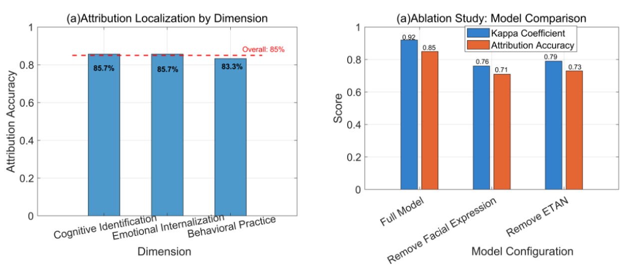

FIGURE 2. Attribution Localization Accuracy and Ablation Assessment

Figures 2 (a-b) show the bar charts of attribution accuracy for each dimension and the comparison results of Kappa coefficients and attribution accuracy for different models under the ablation experiment, respectively. The attribution localization validation results show that the model's attribution accuracy was 85.7% in the cognitive identification and emotional internalization dimensions, and 83.3% in the behavioral practice dimension, with an overall accuracy of 85%. Ablation experiments showed that the complete model had a Kappa coefficient of 0.92 and an attribution accuracy of 0.85. After removing the facial expression modality, these two indicators decreased to 0.76 and 0.71, respectively; after removing the Emotional Temporal Attention Network (ETAN), they decreased to 0.79 and 0.73, respectively. These data confirm that multimodal process data and the emotional temporal attention mechanism play a crucial role in improving assessment consistency and attribution interpretability; neither can be dispensed with.

VII. CONCLUSION

It is challenging to quantitatively and systematically evaluate the educational goals of ideological and political education in digital teaching situations. This paper proposes an approach to assessing goals achieved as follows: taking multimodal process data into account, establishing a three-dimensional indicator system of "cognitive identification—emotional internalization—behavioral practice," and designing an improved fuzzy hierarchical analysis and emotional temporal attention network computational framework to realize dynamic quantification and interpretable attribution of educational effectiveness. The empirical analysis demonstrates the superiority of the method in assessment consistency, reliability, validity, and attribution accuracy. The limitations of this study are threefold: (1) the sample source is relatively homogeneous, covering a limited range of digital scenarios; (2) the real-time performance of the emotional temporal model requires improvement; and (3) the attribution analysis is confined to in-class observable data and does not incorporate external variables (e.g., family, social media, peer influence) that also shape students' ideological and political literacy. Future work will extend to large-scale validation across multiple scenarios and courses, explore lightweight models for real-time intervention, and investigate the integration of external contextual information to achieve more holistic attribution.

AUTHOR CONTRIBUTIONS

Conceptualization –Ying Shao and Jiangbin Mu; methodology – Ying Shao and Jiangbin Mu; investigation – Ying Shao and Jiangbin Mu; writing-original draft preparation – Ying Shao and Jiangbin Mu. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] Y. Hou and J. Qian, "The current status, evaluation, and innovation of course ideological and political research," *Journal of Jiangsu University (Social Sciences Edition)*, vol. 23, no. 6, pp. 66–76, 2021, doi: 10.13317/j.cnki.jdskxb.2021.59.

[2] C. Walkington, M. J. Nathan, W. Huang, J. Hunnicutt, and J. Washington, "Multimodal analysis of interaction data from embodied education technologies," *Educational Technology Research and Development*, vol. 72, no. 5, pp. 2565–2584, 2024, doi: 10.1007/s11423-023-10254-9.

[3] Y. Wu, "Construction of a quality evaluation model for ideological and political education in higher education courses based on student perception," *Vocational Education Development*, vol. 14, no. 1, pp. 324–335, 2025, doi: 10.12677/ve.2025.141048.

[4] Y. Sun, H. Cao, and X. Yuan, "Research on the construction of evaluation index system for ideological and political education in science and engineering courses," *Journal of Jiangsu University (Social Sciences Edition)*, vol. 23, no. 6, pp. 77–88, 2021, doi: 10.13317/j.cnki.jdskxb.2021.60.

[5] H. Zhou and M. Wang, "Research on the evaluation system of graduate course ideological and political education based on CIPP model and AHP," *Education Progress*, vol. 14, no. 7, pp. 1565–1570, 2024, doi: 10.12677/ae.2024.1471344.

[6] L. Huang, T. Doleck, B. Chen, X. Huang, C. Tan, S. P. Lajoie, and M. Wang, "Multimodal learning analytics for assessing teachers' self-regulated learning in planning technology-integrated lessons in a computer-based environment," *Education and Information Technologies*, vol. 28, no. 12, pp. 15823–15843, 2023, doi: 10.1007/s10639-023-11804-7.

[7] A. Yusuf, N. M. Noor, and S. Bello, "Using multimodal learning analytics to model students' learning behavior in animated programming classroom," *Education and Information Technologies*, vol. 29, no. 6, pp. 6947–6990, 2024, doi: 10.1007/s10639-023-12079-8.

[8] J. Shen, H. Yang, J. Li, and Z. Cheng, "Assessing learning engagement based on facial expression recognition in MOOC's scenario," *Multimedia Systems*, vol. 28, no. 2, pp. 469–478, 2022, doi: 10.1007/s00530-021-00854-x.

[9] X. Tang, Y. Gong, Y. Xiao, J. Xiong, and L. Bao, "Facial expression recognition for probing students' emotional engagement in science learning," *Journal of Science Education and Technology*, vol. 34, no. 1, pp. 13–30, 2025, doi: 10.1007/s10956-024-10143-7.

[10] Z. Chen, M. Liang, Z. Xue, et al., "STRAN: Student expression recognition based on spatio-temporal residual attention network in classroom teaching videos," *Applied Intelligence*, vol. 53, no. 21, pp. 25310–25329, 2023, doi: 10.1007/s10489-023-04858-0.