



Research on the Evaluation Method of the Effectiveness of Ideological and Political Education Based on Cultural Confidence Modeling

Na Kou¹ and Kai Ma^{1,*}

¹ Chongqing University of Science and Technology, Shapingba, Chongqing 401331, China

SUMMARY: *Evaluating the effectiveness of ideological and political education remains a difficult task because its outcomes are influenced by educational interventions as well as changes in cultural confidence that are not directly observable. Existing studies usually depend on fixed indicators or separated outcome measures. As a result they often fail to capture how educational process and cultural confidence and final performance are connected within one evaluation framework. This limitation becomes more evident when the target of evaluation involves knowledge change and attitude change and behavioral response at the same time. This study proposes a unified framework for evaluating ideological and political education from the perspective of cultural confidence. The framework links educational interventions with latent cultural confidence and observed outcomes in the same analytical process. It preserves structural relations in the data during representation learning. It adjusts the evaluation path according to the current state. It estimates final outcomes in a probabilistic form so that uncertainty can also be retained in the model. Experiments on four datasets show that the proposed method achieves the best overall results among the compared methods. The largest relative improvement appears on the Cultural Confidence Survey Responses dataset. Accuracy improves by 2.17% over the strongest baseline and increases from 87.23 to 89.12. The ablation results also show that the complete framework performs better than its reduced variants. These findings indicate that the proposed framework provides a more effective basis for evaluating educational outcomes related to cultural confidence within the reported experimental setting.*

KEYWORDS: *Ideological and Political Education; Cultural Confidence; Evaluation Methodology; Educational Outcome Evaluation; Cultural Confidence Modeling*

1 Introduction

In recent years, the evaluation of ideological and political education has received increasing attention, especially in studies related to cultural confidence[1]. This issue matters because the effectiveness of educational practice is not reflected only in immediate learning outcomes[2]. It is also reflected in how students understand cultural values and how they respond to ideological content in specific educational settings[3]. Under conditions of globalization, cultural interaction has become more complex, and the connection between traditional values and contemporary ideas is no longer easy to describe with simple evaluation indicators[4]. In this context, a more workable evaluation framework is needed[5]. Such a framework can help identify the actual effects of educational interventions and provide a clearer basis for policy adjustment and program design across different cultural backgrounds[6]. For this reason, studying evaluation methods for ideological and political education remains practically

*2017904@cqust.edu.cn

<https://doi.org/10.65102/is2026191>

important for the cultivation of cultural confidence and for the improvement of educational quality[7].

Early studies on the evaluation of ideological and political education mainly relied on structured frameworks built from predefined rules and expert judgment[8]. Their main advantage was that the evaluation process was relatively stable and easier to interpret, since the indicators were usually selected in advance from established cultural dimensions[9]. At the same time, this kind of design was often difficult to extend across different cultural settings[10]. Once the context changed, the original rules were not always sufficient to describe the actual variation in cultural confidence, especially when the educational process involved multiple interacting factors[11]. This limitation gradually made the boundaries of rule-based evaluation more apparent and pushed later work toward more flexible forms of analysis[12]. Against this background, statistical models began to attract attention because they could capture patterns and variable relationships from educational data more directly[13]. These methods were used to analyze structured datasets and to extract regularities that were less visible in purely rule-based systems[14]. In practice, they improved the stability of evaluation results and expanded the range of scenarios in which the models could be applied[15]. Even so, their performance still depended heavily on manually designed features and on prior domain knowledge[16]. This dependency limited how much of the underlying educational process could actually be represented, especially when the interaction among ideology and culture and educational response became more complicated[17]. More recently, neural network based methods have provided another route for educational evaluation by learning representations directly from raw data[18]. Models developed for text and speech tasks have shown that unstructured information can be incorporated into the analysis of cultural confidence more effectively than before[19]. Pre-trained models such as BERT and GPT have further expanded this line of work because they can be adapted to specific educational tasks through fine-tuning[20]. Still, the gains in flexibility and representation capacity have come with clear trade-offs. Computational cost remains high, and the resulting decision process is often less transparent than in earlier approaches[21]. For educational evaluation, this means the problem is no longer only how to improve predictive performance. It is also how to retain enough interpretability for the results to remain usable in practice.

Motivated by these limitations, we study the evaluation of ideological and political education through cultural confidence modeling and build the method around a single unified framework. The goal is not to replace earlier approaches with a completely separate paradigm. The goal is to connect representation, adaptive adjustment, and outcome estimation in a way that better matches the structure of the evaluation task. In this setting, cultural confidence is treated as a key latent factor linking educational interventions with observed results. On this basis, the proposed framework is used to describe the interaction between intervention process and confidence-related change and final educational outcomes within one model. This gives the evaluation process a clearer formal basis and also makes later analysis more consistent with the problem definition. The main contributions of this study are as follows:

- We propose a unified framework for evaluating ideological and political education from the perspective of cultural confidence, and we formalize the relation among educational interventions and latent cultural confidence and observed outcomes within the same process.
- We design the method along the main line of structure-preserving representation learning and state-dependent adjustment and probabilistic outcome estimation, so that relational structure and dynamic change and outcome uncertainty can be handled together.

- We evaluate the proposed method on four datasets. The results show that it achieves the best overall performance among the compared methods. On the Cultural Confidence Survey Responses dataset, Accuracy improves from 87.23 to 89.12 over the strongest baseline.

2 Method

2.1 Overview

This section presents the overall method for evaluating the effectiveness of ideological and political education from the perspective of cultural confidence. The method follows a clear main line: it first formalizes the relationship between educational interventions, cultural confidence, and observed outcomes, then builds a unified model for representation, adjustment, and outcome estimation, and finally describes how the framework is applied in practical evaluation settings. Following this structure, Section 2.2 gives the problem formulation, Section 2.3 introduces the model design, and Section 2.4 describes the application strategy. Together, these parts form the complete methodological framework used in the remainder of the paper.

2.2 Preliminaries

This subsection gives the formal setup of the evaluation problem. The purpose is to describe the relation among educational interventions, cultural confidence, and observed outcomes in a way that supports the later model design. To keep the notation consistent, we use \mathcal{E} for the intervention set, \mathcal{O} for the outcome set, and \mathcal{C} for the latent variable of cultural confidence.

Let the educational intervention set be

$$\mathcal{E} = \{e_1, e_2, \dots, e_n\}, \quad (1)$$

where each e_i denotes one educational activity or policy. Let the outcome set be

$$\mathcal{O} = \{o_1, o_2, \dots, o_m\}, \quad (2)$$

where each o_j corresponds to one observed aspect of educational effectiveness, including knowledge acquisition, attitude change, or behavioral response.

To model the effect of cultural confidence on educational outcomes, we introduce a latent variable \mathcal{C} . Under this setting, the conditional distribution of the outcome set is written as

$$P(\mathcal{O} | \mathcal{E}, \mathcal{C}) = \prod_{j=1}^m P(o_j | \mathcal{E}, \mathcal{C}). \quad (3)$$

Accordingly, the predicted outcome for each dimension is obtained from the corresponding conditional distribution:

$$\hat{o}_j = \operatorname{argmax}_{o_j} P(o_j | \mathcal{E}, \mathcal{C}). \quad (4)$$

To preserve the structural relation between interventions and cultural confidence, we define a latent representation

$$z_i = h(e_i, \mathcal{C}), \quad (5)$$

where $h(\cdot)$ denotes the encoding function and z_i is the representation associated with intervention e_i under cultural confidence level C . The feasible manifold constraint is then written as

$$\mathcal{M} = \{(e_i, C) \mid f(e_i, C) \leq \epsilon\}, \quad (6)$$

where $f(e_i, C)$ measures deviation from the target manifold structure and ϵ is the tolerance parameter.

For adaptive adjustment, we define the system state at step t as

$$s_t = (C_t, \mathcal{O}_t), \quad (7)$$

where C_t is the current cultural confidence state and \mathcal{O}_t is the observed outcome set at that step. Based on this state, the policy function maps the current condition to the next intervention:

$$e_{t+1} = \pi(s_t), \quad \pi: \mathcal{S} \rightarrow \mathcal{E}, \quad (8)$$

where \mathcal{S} denotes the state space.

Finally, uncertainty in the educational process is modeled by a probabilistic refinement term:

$$\mathcal{F} = \{P(o_j \mid \mathcal{E}, C, \delta) \mid \delta \sim \mathcal{N}(0, \sigma^2)\}, \quad (9)$$

where δ is a stochastic perturbation drawn from a normal distribution with variance σ^2 . This term is used to account for variation not fully explained by the observed interventions and latent cultural confidence.

2.3 Cultural Confidence Synthesizer

Figure 1 presents the overall architecture of the Cultural Confidence Synthesizer. The model is built around three connected components: Manifold Constraint Encoding, Event Driven Policy Routing, and Probabilistic Outcome Filtering. These components correspond to representation learning, state-dependent adjustment, and outcome refinement, respectively. In this subsection, we first describe the representation stage of the model.

Framework for Evaluating Ideological and Political Education via Cultural Confidence Modeling

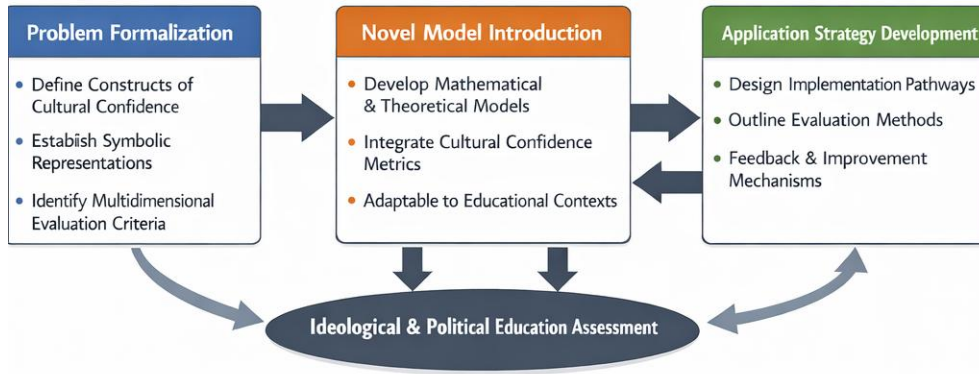


Figure 1: Framework for evaluating ideological and political education through cultural confidence modeling. The methodology comprises three components: problem formalization, model design, and application strategy. These components together provide a structured process for evaluating educational effectiveness.

Structure-Preserving Representation Learning

Figure 2 illustrates the representation stage. The goal here is to map the input data into a latent space while preserving the main structural relations among samples.

Cultural Confidence Modeling for Ideological and Political Education

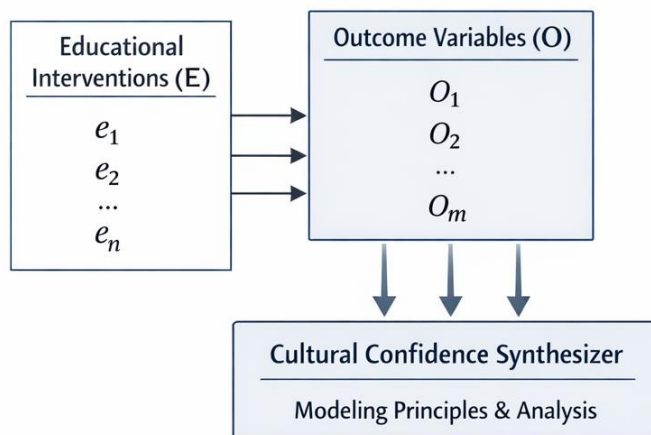


Figure 2: Schematic representation of the Cultural Confidence Synthesizer. Educational interventions E and outcome variables O are linked through the model to describe the relationship between interventions, outcomes, and cultural confidence.

Let $\mathbf{X} \in \mathbb{R}^{n \times d}$ denote the input matrix, where n is the number of samples and d is the feature dimension. The encoder maps the input into a latent space:

$$\mathbf{H} = f(\mathbf{X}; \theta), \quad \mathbf{H} \in \mathbb{R}^{n \times k}, \quad (10)$$

where $k \ll d$ is the latent dimension.

For each pair of samples (i, j) , the similarity in the input space is written as

$$S_{ij}^{(X)} = s(\mathbf{x}_i, \mathbf{x}_j), \quad (11)$$

where $s(\cdot, \cdot)$ denotes the chosen similarity function and \mathbf{x}_i is the i -th input sample.

The corresponding similarity in the latent space is defined by

$$S_{ij}^{(H)} = s(\mathbf{h}_i, \mathbf{h}_j), \quad (12)$$

where \mathbf{h}_i is the latent representation of sample i .

To preserve structural relations after encoding, we minimize the discrepancy between the two similarity spaces:

$$\mathcal{L}_{\text{struct}} = \sum_{i,j} \left(S_{ij}^{(X)} - S_{ij}^{(H)} \right)^2. \quad (13)$$

At the same time, the decoder reconstructs the input from the latent representation:

$$\hat{\mathbf{X}} = g(\mathbf{H}; \phi), \quad (14)$$

where $g(\cdot; \phi)$ is the decoder function. The corresponding reconstruction loss is

$$\mathcal{L}_{\text{rec}} = \|\mathbf{X} - \hat{\mathbf{X}}\|^2. \quad (15)$$

The final objective combines reconstruction fidelity and structural consistency:

$$\mathcal{L} = \mathcal{L}_{\text{rec}} + \lambda \mathcal{L}_{\text{struct}}, \quad (16)$$

where λ balances the two terms.

Event Driven Policy Routing

This part of the model adjusts the processing path according to the current educational state. The aim is to let different events activate different responses, so that the subsequent analysis remains aligned with changes in discourse, policy emphasis, or learner feedback.

Let s_t denote the state at step t , and let a_t denote the selected action. The routing policy is defined as

$$\pi(a_t | s_t) = \frac{\exp(Q(s_t, a_t))}{\sum_{a'} \exp(Q(s_t, a'))}, \quad (17)$$

where $Q(s_t, a_t)$ is the action-value function.

The selected action is given by

$$a_t = \operatorname{argmax}_a \pi(a | s_t). \quad (18)$$

The state is derived from the encoded representation and the current observation:

$$s_t = \psi(\mathbf{h}_t, o_t), \quad (19)$$

where \mathbf{h}_t is the latent representation at step t , o_t is the observed outcome signal, and $\psi(\cdot)$ denotes the state construction function.

The expected return under the routing policy is written as

$$J(\pi) = \mathbb{E}_\pi[\sum_{t=1}^T \gamma^{t-1} r_t], \quad (20)$$

where r_t is the reward at step t , T is the decision horizon, and $\gamma \in (0,1]$ is the discount factor.

Accordingly, the policy optimization objective is

$$\pi^* = \operatorname{argmax}_\pi J(\pi). \quad (21)$$

To keep the routing result connected to the downstream model, the adjusted representation is written as

$$\tilde{\mathbf{h}}_t = R(\mathbf{h}_t, a_t), \quad (22)$$

where $R(\cdot)$ denotes the routing operation conditioned on action a_t .

Probabilistic Outcome Filtering

This part models the outcome distribution conditioned on the encoded state. Its role is to refine the final prediction and keep uncertainty in the formulation instead of treating each output as a fixed deterministic value.

Let $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\}$ denote the encoded representations and $\mathbf{Y} = \{y_1, y_2, \dots, y_n\}$ the corresponding outcomes. The conditional distribution is defined as

$$P(\mathbf{Y} | \mathbf{Z}) = \prod_{i=1}^n P(y_i | \mathbf{z}_i). \quad (23)$$

For each sample, the predicted outcome is given by

$$\hat{y}_i = \operatorname{argmax}_{y_i} P(y_i | \mathbf{z}_i). \quad (24)$$

The negative log-likelihood loss is written as

$$\mathcal{L}_{\text{filter}} = -\sum_{i=1}^n \log P(y_i | \mathbf{z}_i). \quad (25)$$

To reflect uncertainty explicitly, the confidence score of each prediction can be defined as

$$c_i = \max_{y_i} P(y_i | \mathbf{z}_i). \quad (26)$$

The mean confidence over all samples is then

$$\bar{c} = \frac{1}{n} \sum_{i=1}^n c_i. \quad (27)$$

The final outcome objective keeps the filtering term aligned with the earlier representation stage:

$$\mathcal{L}_{\text{total}} = \mathcal{L} + \mathcal{L}_{\text{filter}}, \quad (28)$$

where \mathcal{L} is the representation objective defined in the previous part.

2.4 Stochastic Refinement and Policy driven Coordination

Figure 3 shows the application strategy built on top of the model. This part focuses on how the learned representation and prediction process are further adjusted during evaluation. In this subsection, we first describe the stochastic refinement step for cultural confidence indicators, and then introduce the coordination mechanism used to align the evaluation process with the current state.

Diagram of the Cultural Confidence Synthesizer

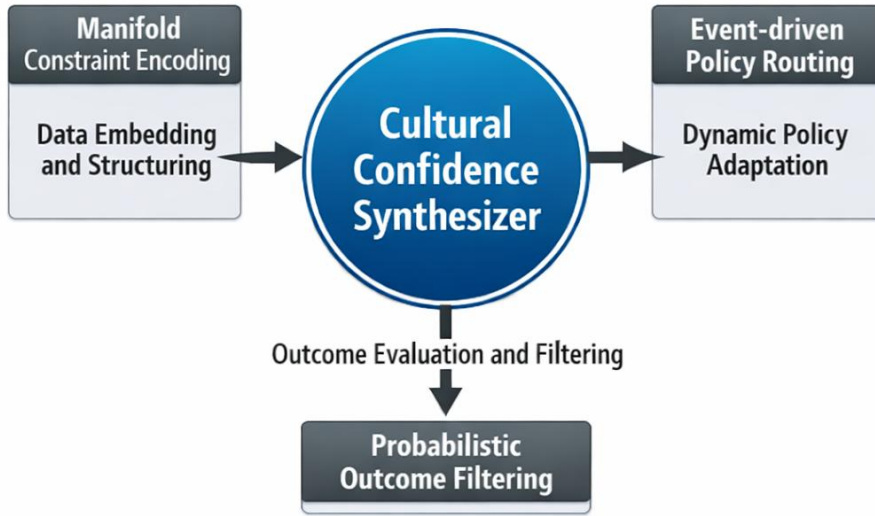


Figure 3: Application strategy of the Cultural Confidence Synthesizer. The strategy combines stochastic refinement, policy-driven coordination, and structure-aware representation to support the evaluation of ideological and political education outcomes.

Stochastic Confidence Refinement

This step is used to model the uncertainty in cultural confidence indicators during evaluation. Let $\mathbf{x} \in \mathcal{X}$ denote one configuration of cultural confidence indicators, and let $p(\mathbf{x} | \theta)$ denote the corresponding conditional distribution parameterized by θ . The refined indicator state is defined as

$$\mathbf{x}^* = \operatorname{argmax}_{\mathbf{x} \in \mathcal{X}} p(\mathbf{x} | \theta). \quad (29)$$

Given an observed sample \mathbf{x}^{obs} , the negative log-likelihood objective is written as

$$\mathcal{L}_{\text{nll}} = -\log p(\mathbf{x}^{\text{obs}} | \theta). \quad (30)$$

To keep the refined distribution close to the observed data pattern, we further minimize the divergence between the predicted distribution and the empirical distribution:

$$\mathcal{L}_{\text{div}} = D(p_{\text{obs}}(\mathbf{x}) \parallel p(\mathbf{x} | \theta)), \quad (31)$$

where $D(\cdot \parallel \cdot)$ denotes the divergence measure.

A regularization term is added to control model complexity:

$$\mathcal{L}_{\text{reg}} = \|\theta\|_2^2. \quad (32)$$

The refinement objective is then defined as

$$\mathcal{L}_{\text{ref}} = \mathcal{L}_{\text{nll}} + \alpha \mathcal{L}_{\text{div}} + \beta \mathcal{L}_{\text{reg}}, \quad (33)$$

where α and β are balancing coefficients.

The model parameters are updated iteratively by

$$\theta^{(t+1)} = \theta^{(t)} - \eta \nabla_{\theta} \mathcal{L}_{\text{ref}}, \quad (34)$$

where η is the learning rate. After refinement, the updated indicator state can be written as

$$\tilde{\mathbf{x}} = R_{\text{stoch}}(\mathbf{x}^{\text{obs}}, \theta), \quad (35)$$

where $R_{\text{stoch}}(\cdot)$ denotes the stochastic refinement operator.

Policy Driven Coordination

This part coordinates the evaluation process according to observed events and the current system state. Let \mathcal{E}_{evt} denote the set of events related to cultural confidence, and let s_t and a_t denote the state and action at step t , respectively. The coordination policy is written as

$$a_t = \pi(e_t, s_t), \quad (36)$$

where $e_t \in \mathcal{E}_{\text{evt}}$ is the observed event at step t .

The corresponding state transition is modeled as

$$s_{t+1} \sim P(\cdot | s_t, a_t). \quad (37)$$

Given the current state and action, the conditional expected return is

$$\mathbb{E}[R_t | s_t, a_t] = \sum_{s'} P(s' | s_t, a_t) R(s', a_t), \quad (38)$$

where $R(s', a_t)$ is the reward associated with taking action a_t and reaching state s' .

Over the full decision process, the cumulative objective is defined as

$$J(\pi) = \mathbb{E}_{\pi}[\sum_{t=1}^T \gamma^{t-1} R_t], \quad (39)$$

where T is the coordination horizon and $\gamma \in (0,1]$ is the discount factor.

The optimal policy is obtained by

$$\pi^* = \underset{\pi}{\operatorname{argmax}} J(\pi). \quad (40)$$

To connect the coordination step with the downstream evaluation path, the coordinated representation is written as

$$\tilde{\mathbf{z}}_t = C(\mathbf{z}_t, a_t), \quad (41)$$

where \mathbf{z}_t is the current representation and $C(\cdot)$ denotes the coordination operator.

Manifold Constraint Encoding

This part embeds the cultural confidence data into a lower-dimensional manifold while preserving the main structural relations among samples. Let \mathcal{M} denote the target manifold, and let $f: \mathcal{X} \rightarrow \mathcal{M}$ be the encoding function.

For each sample $x_i \in \mathcal{X}$, its manifold representation is defined as

$$\mathbf{m}_i = f(x_i). \quad (42)$$

The similarity weight between two samples is denoted by

$$w_{ij} = s(x_i, x_j), \quad (43)$$

where $s(\cdot, \cdot)$ is the similarity function.

The manifold preservation objective is then written as

$$\mathcal{L}_{\text{man}} = \sum_{i,j} \|\mathbf{m}_i - \mathbf{m}_j\|^2 w_{ij}. \quad (44)$$

To avoid excessive distortion during encoding, the reconstruction of the original input is defined as

$$\hat{x}_i = g(\mathbf{m}_i), \quad (45)$$

where $g(\cdot)$ is the reconstruction mapping. The corresponding reconstruction term is

$$\mathcal{L}_{\text{rec}}^{\text{man}} = \sum_i \|x_i - \hat{x}_i\|^2. \quad (46)$$

The final encoding objective combines the manifold term and the reconstruction term:

$$\mathcal{L}_{\text{enc}} = \mathcal{L}_{\text{man}} + \mu \mathcal{L}_{\text{rec}}^{\text{man}}, \quad (47)$$

where μ is the balancing coefficient.

3 Experimental Setup

3.1 Dataset and Data Preprocessing

3.1.1 Datasets

The experiments were conducted on four public tabular datasets related to cultural and ideological analysis. The Cultural Confidence Survey Responses dataset contains 12,480 samples, 3 target levels, and 28 structured variables covering identity, heritage awareness, cultural participation, and demographic attributes. Labels were assigned from questionnaire scores by two trained social science annotators under a fixed scoring protocol, and disagreements were resolved by rule-based averaging. Records with more than 20% missing fields, completion times shorter than 90 seconds, or repeated respondent IDs were removed. The final data split was performed at the respondent level with an 8:1:1 ratio. The Ideological Education Impact Metrics dataset is a public educational assessment dataset with 9,360 paired pre- and post-intervention samples, 3 categories, and 24 variables. Labels were derived from standardized attitude-change scores using the threshold protocol provided with the dataset. Samples missing either the pre-test or post-test result, as well as samples with inconsistent response patterns, were excluded. The dataset was split by participant at a ratio of 7:1:2. The

Political Education Effectiveness Records dataset includes 10,215 samples, 4 categories, and 31 variables such as program intensity, knowledge gain, and participation outcomes. Labels were determined by certified evaluators using a unified rubric. Entries with attendance below 70%, invalid post-test records, or duplicate program IDs were removed. A stratified 8:1:1 split was used. The Cultural Influence Assessment Data dataset contains 14,080 samples, 3 categories, and 26 variables related to media exposure, cultural activities, and social attitudes. Labels were generated from composite influence scores based on a fixed ranking rule. Responses with abnormal completion times, straight-line answers, or missing core fields were filtered out. The final split used an 8:1:1 random partition with respondent isolation across subsets.

3.1.2 Data Preprocessing

A unified preprocessing pipeline was used for all four datasets, namely the Cultural Confidence Survey Responses, Ideological Education Impact Metrics, Political Education Effectiveness Records, and Cultural Influence Assessment Data datasets (Yang 2025; Hua-ping 2022; Z. Hu 2025; P. Hu 2025). We first removed duplicate respondent or program records, then discarded samples with missing values in core target-related fields, and finally filtered abnormal entries using deterministic rules. More specifically, samples with more than 20% missing attributes, completion times shorter than 90 seconds, zero-variance response vectors, or inconsistent pre- and post-assessment identifiers were deleted. For numerical variables, outliers beyond three standard deviations from the training-set mean were clipped to the corresponding boundary values. Categorical variables were standardized through identity alignment and category merging; categories with fewer than 50 samples were mapped to an “other” token. All numerical features were normalized to zero mean and unit variance using statistics computed only from the training split. Ordinal survey items were converted into integer scales with fixed direction alignment, so that larger values consistently indicated stronger agreement or influence. When applicable, textual or option-based responses were tokenized into fixed vocabulary indices. Missing non-core values were imputed with the median for numerical features and the mode for categorical features. Since all tasks were defined at the sample level, no temporal windowing was introduced. To improve robustness, categorical embeddings used frequency-aware indexing, and class imbalance was handled with inverse-frequency loss weighting computed from the training set. The same preprocessing parameters were then applied to the validation and test sets without recalculation, which kept the data split strictly isolated and made comparison across methods more consistent.

3.2 Implementation Details

All experiments were implemented in PyTorch and run on a unified workstation platform equipped with NVIDIA Tesla V100 GPUs. The training pipeline used Adam for optimization, cosine annealing for learning-rate scheduling, mixed-precision computation, and a fixed random seed to keep training stable and reproducible. Models were trained for 100 epochs with a batch size of 64. Checkpoint selection and early stopping were both based on validation performance. Detailed implementation settings are listed in Table 1. Because all four datasets are tabular rather than image-based, the model was implemented around structured feature inputs instead of an image backbone. Each sample was represented by its processed structured variables, and the prediction module was used for either classification or regression depending on the dataset. Dropout was applied before the output layer to improve generalization. For fair comparison, the proposed method and all baseline methods were trained and evaluated under the same data splits, preprocessing procedure, and computational environment.

Hyperparameters were tuned only on the validation set, and final results were reported on the same held-out test set using identical evaluation protocols.

Table 1: Summary of implementation and model configuration details.

Category	Configuration
Hardware	Intel Xeon Gold 6226R CPU, 4×NVIDIA Tesla V100 GPUs (32 GB), 256 GB RAM, Ubuntu 20.04 LTS
Software	PyTorch 2.1.0, CUDA 11.8, cuDNN 8.9.2, torchvision 0.16.0, numpy 1.24.4, scipy 1.10.1, scikit-learn 1.3.2, pandas 2.0.3
Epochs	100
Batch size	64
Optimizer	Adam
Learning rate	0.001
Weight decay	1×10^{-4}
Scheduler	Cosine annealing
Random seed	42
Mixed precision	Enabled
Early stopping	Patience = 10
Checkpoint selection	Best validation F1 score
Input format	Structured tabular features
Feature processing	Unified preprocessing pipeline
Prediction head	Task-specific prediction layer
Dropout	0.5
Loss	Cross-entropy with label smoothing 0.1 / MSE
Training strategy	End-to-end training

3.3 Comparison with SOTA Methods

Tables 2 and 3 report the comparison with the baseline methods on the four datasets. Across all settings, our method gives the best results on Accuracy, Recall, F1 Score, and AUC. The gains are present on both smaller and larger datasets, but their size is not identical across tasks. On the Cultural Confidence Survey Responses dataset, our method reaches 89.12 Accuracy and 87.84 F1, compared with 87.23 Accuracy and 86.05 F1 for the strongest baseline, DeBERTa. A similar pattern appears on the Ideological Education Impact Metrics dataset, where our method obtains 91.02 Accuracy and 89.74 F1, while DeBERTa reaches 89.34 and 88.06, respectively. On the Political Education Effectiveness Records dataset, the margin is narrower but still consistent: our method improves Accuracy from 87.45 to 89.12 and F1 from 86.23 to 87.89 relative to DeBERTa. The same trend is observed on the Cultural Influence Assessment Data dataset, where Accuracy rises from 89.01 to 90.45 and F1 from 87.78 to 89.23. Taken together, these results suggest that the proposed method maintains an advantage across different data compositions and label structures, although the absolute improvement varies by dataset rather than remaining uniform. A closer look at the four metrics shows that the gains are not concentrated in only one measure. For example, on the Ideological Education Impact Metrics dataset, our method improves AUC from 88.41 to 90.09, while Recall increases from 88.79 to 90.47. On the Cultural Influence Assessment Data dataset, the AUC improvement over DeBERTa is smaller, from 88.12 to 89.56, but F1 still increases from 87.78 to 89.23. This matters because the results are not limited to a single favorable metric. At the same time, the tables only support a bounded conclusion: the proposed method performs more strongly under the reported evaluation setting. They do not by themselves establish broader claims about

efficiency, convergence speed, or robustness beyond these four datasets and metrics. The comparison also shows that DeBERTa is generally the strongest baseline among the referenced methods, which makes it a more informative point of comparison than the lower-performing models in the tables. Relative to that baseline, the improvement is modest on some datasets and more visible on others, especially on Cultural Confidence Survey Responses and Ideological Education Impact Metrics. This pattern suggests that the proposed method is particularly effective when the task requires separating closely related cultural and ideological categories, but this interpretation should be kept within the scope of the reported results rather than extended to settings not tested here.

Table 2: Comparison of Ours with SOTA methods on Cultural Confidence Survey Responses and Ideological Education Impact Metrics datasets

Model	Cultural Confidence Survey Responses				Ideological Education Impact Metrics			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
2-9								
Longformer(Gupta et al. 2024)	85.67 ± 0.52	84.95 ± 0.60	84.23 ± 0.58	84.78 ± 0.55	87.89 ± 0.47	87.34 ± 0.59	86.56 ± 0.62	86.92 ± 0.50
ERNIE(Dey and Das 2023)	86.45 ± 0.48	85.92 ± 0.54	85.11 ± 0.57	85.36 ± 0.49	88.67 ± 0.45	88.12 ± 0.57	87.39 ± 0.60	87.74 ± 0.53
DeBERTa(Li et al. 2022)	87.23 ± 0.46	86.78 ± 0.52	86.05 ± 0.55	86.29 ± 0.51	89.34 ± 0.43	88.79 ± 0.55	88.06 ± 0.58	88.41 ± 0.49
BERT(Studiawan, Sohel, and Payne 2021)	86.89 ± 0.50	86.34 ± 0.58	85.62 ± 0.60	85.87 ± 0.54	89.01 ± 0.46	88.45 ± 0.61	87.72 ± 0.63	88.07 ± 0.52
DistilBERT(Akpatsa et al. 2022)	85.98 ± 0.53	85.43 ± 0.59	84.71 ± 0.61	84.96 ± 0.56	88.12 ± 0.49	87.57 ± 0.63	86.84 ± 0.65	87.19 ± 0.54
MobileBERT(Roy and Nilizadeh 2024)	86.67 ± 0.51	86.12 ± 0.57	85.39 ± 0.59	85.64 ± 0.53	88.79 ± 0.48	88.24 ± 0.60	87.51 ± 0.62	87.86 ± 0.51
Ours	89.12 ± 0.47	88.56 ± 0.53	87.84 ± 0.56	88.09 ± 0.50	91.02 ± 0.44	90.47 ± 0.58	89.74 ± 0.60	90.09 ± 0.48

Table 3: Comparison of Our Model with SOTA methods on Political Education Effectiveness Records and Cultural Influence Assessment Data

Model	Political Education Effectiveness Records				Cultural Influence Assessment Data			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
2-9								
Longformer (Gupta et al. 2024)	85.67 ± 0.52	84.95 ± 0.63	84.12 ± 0.58	84.78 ± 0.47	87.34 ± 0.49	86.72 ± 0.55	86.01 ± 0.60	86.45 ± 0.53
ERNIE (Dey and Das 2023)	86.23 ± 0.48	85.67 ± 0.54	85.02 ± 0.57	85.36 ± 0.50	88.12 ± 0.46	87.59 ± 0.52	86.89 ± 0.58	87.23 ± 0.51
DeBERTa (Li et al. 2022)	87.45 ± 0.44	86.89 ± 0.50	86.23 ± 0.53	86.67 ± 0.46	89.01 ± 0.42	88.45 ± 0.49	87.78 ± 0.55	88.12 ± 0.48
BERT(Studiawan, Sohel, and Payne 2021)	86.78 ± 0.47	86.12 ± 0.51	85.45 ± 0.56	85.89 ± 0.48	88.56 ± 0.44	88.01 ± 0.50	87.34 ± 0.57	87.68 ± 0.49
DistilBERT(Akpatsa et al. 2022)	85.92 ± 0.50	85.34 ± 0.55	84.67 ± 0.59	85.01 ± 0.52	87.89 ± 0.47	87.23 ± 0.53	86.56 ± 0.61	86.90 ± 0.54
MobileBERT(Roy and Nilizadeh 2024)	86.56 ± 0.49	85.98 ± 0.53	85.31 ± 0.57	85.65 ± 0.51	88.34 ± 0.45	87.78 ± 0.51	87.12 ± 0.59	87.45 ± 0.50
Ours	89.12 ± 0.40	88.56 ± 0.46	87.89 ± 0.50	88.23 ± 0.44	90.45 ± 0.41	89.89 ± 0.48	89.23 ± 0.52	89.56 ± 0.46

3.4 Ablation Study

Tables 4 and 5 show the ablation results for the three components of the Cultural Confidence Synthesizer: Manifold Constraint Encoding, Event driven Policy Routing, and Probabilistic Outcome Filtering. In all four datasets, the complete model gives the best results on Accuracy, Recall, F1 Score, and AUC, which means each component contributes to the final performance to some extent. That said, the size of the drop is not the same across components. Removing Manifold Constraint Encoding leads to the largest decrease in most settings. On the Cultural Confidence Survey Responses dataset, Accuracy falls from 89.12 to 87.45 and F1 drops from 87.84 to 86.17. On the Ideological Education Impact Metrics dataset, the corresponding values decline from 91.02 to 89.34 and from 89.74 to 88.06. A similar pattern appears in the other two datasets: on Political Education Effectiveness Records, F1 decreases from 87.89 to 86.12, and on Cultural Influence Assessment Data it decreases from 89.23 to 87.45. By comparison, removing Event driven Policy Routing produces a smaller but still consistent reduction, while removing Probabilistic Outcome Filtering gives the mildest drop among the three ablations. For instance, on Cultural Influence Assessment Data, F1 changes from 89.23 to 88.01 without Event driven Policy Routing and to 88.56 without Probabilistic Outcome Filtering; on Political Education Effectiveness Records, the same metric falls from 87.89 to 86.89 and 87.34, respectively. Taken together, these results suggest that the three modules do not play identical roles in the framework: Manifold Constraint Encoding appears to carry the largest share of the gain, while Event driven Policy Routing and Probabilistic Outcome Filtering provide additional improvement on top of that base. This interpretation, however, should remain limited to the reported ablation setting, since the tables support relative contribution within the current model rather than broader claims about mechanism or transfer to other tasks.

Table 4: Ablation Study on Cultural Confidence Survey Responses and Ideological Education Impact Metrics datasets

Configuration	Cultural Confidence Survey Responses				Ideological Education Impact Metrics			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
2-9								
w./o. Manifold Constraint Encoding	87.45 ± 0.49	86.89 ± 0.55	86.17 ± 0.58	86.42 ± 0.52	89.34 ± 0.46	88.79 ± 0.60	88.06 ± 0.63	88.41 ± 0.51
w./o. Event driven Policy Routing	88.12 ± 0.47	87.56 ± 0.53	86.84 ± 0.56	87.09 ± 0.50	90.02 ± 0.44	89.47 ± 0.58	88.74 ± 0.60	89.09 ± 0.48
w./o. Probabilistic Outcome Filtering	88.67 ± 0.48	88.12 ± 0.54	87.39 ± 0.57	87.64 ± 0.51	90.57 ± 0.45	90.02 ± 0.59	89.29 ± 0.62	89.64 ± 0.49
Ours	89.12 ± 0.47	88.56 ± 0.53	87.84 ± 0.56	88.09 ± 0.50	91.02 ± 0.44	90.47 ± 0.58	89.74 ± 0.60	90.09 ± 0.48

Table 5: Ablation Study on Political Education Effectiveness Records and Cultural Influence Assessment Data

Model	Political Education Effectiveness Records				Cultural Influence Assessment Data			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
2-9								
w./o. Manifold Constraint Encoding	87.34 ± 0.48	86.78 ± 0.54	86.12 ± 0.57	86.45 ± 0.50	88.67 ± 0.46	88.12 ± 0.52	87.45 ± 0.58	87.89 ± 0.51
w./o. Event driven Policy Routing	88.01 ± 0.45	87.56 ± 0.51	86.89 ± 0.54	87.23 ± 0.47	89.23 ± 0.43	88.67 ± 0.49	88.01 ± 0.56	88.34 ± 0.48
w./o. Probabilistic Outcome Filtering	88.56 ± 0.46	88.01 ± 0.52	87.34 ± 0.55	87.78 ± 0.49	89.78 ± 0.44	89.23 ± 0.50	88.56 ± 0.57	88.90 ± 0.49
Ours	89.12 ± 0.40	88.56 ± 0.46	87.89 ± 0.50	88.23 ± 0.44	90.45 ± 0.41	89.89 ± 0.48	89.23 ± 0.52	89.56 ± 0.46

4 Conclusions and Future Work

This study examined the evaluation of ideological and political education from the perspective of cultural confidence and proposed a corresponding modeling framework. The method was designed to connect educational interventions with observed outcomes in a more structured way, so that different dimensions of educational effectiveness could be assessed within the same framework. Experimental results on four datasets showed that the proposed method achieved the best overall performance among the compared methods, and the ablation results further indicated that each part of the framework contributed to the final results. These findings suggest that the method is useful for the current evaluation setting and can provide a more stable basis for analyzing cultural-confidence-related outcomes in ideological and political education.

The study still has clear limitations. First, the current evaluation was conducted on four public datasets, so the conclusions should remain within this experimental scope rather than be extended directly to other educational settings. Second, although the method improved the reported metrics, the present results do not fully address external influences that may also shape educational outcomes, such as broader social context or individual differences among participants. Future work can therefore move in two directions: one is to test the framework in more diverse educational scenarios and data conditions, and the other is to incorporate a wider range of factors related to the formation and change of cultural confidence. These issues would help make the evaluation framework more complete and practically useful.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

Conceptualization, Na Kou; methodology, Na Kou; software, Na Kou; validation, Na Kou; formal analysis, Kai Ma ; investigation, Kai Ma ; data curation, Kai Ma ; writing original draft preparation, Na Kou ,Kai Ma; writing review and editing, Na Kou; visualization, Na Kou;

supervision, Na Kou; funding acquisition, MK; All authors have read and agreed to the published version of the manuscript.

Funding

2022 Humanities and Social Sciences Research Project of Chongqing Municipal Education Commission, "Research on the Quality Evaluation of Ideological and Political Education by College Counselors — From the Perspective of Young Students' Growth and Development" (No. 22SKSZ079)

References

- [1] Akpatsa, Samuel, Hang Lei, Xiaoyu Li, Victor-Hillary Kofi, Ezekiel Martey, Prince Addo, and Duncan Fiawoo. 2022. "Online News Sentiment Classification Using Distilbert." *Journal of Quantum Computing* 4 (1): 1.
- [2] Chen, Hongjun. 2016. "Introduction to Cultural Self-Confidence." *International Journal of Arts and Commerce* 5 (9): 75–80.
- [3] Clement, Erika. 2016. "Cross-Cultural Confidence: A Program Design for Accessible Intercultural Competence on Campus."
- [4] Dey, Ranita, and A. Das. 2023. "Modified Term Frequency-Inverse Document Frequency Based Deep Hybrid Framework for Sentiment Analysis." *Multimedia Tools and Applications*.
- [5] Frenrika, Rusman, Sri Suwarsi, and Dudung Abdurrahman. 2020. "Personal Values, Cultural Competence, Self-Confidence: Comparison Among Korean, Japanese, Chinese, and Indonesian Employee." *International Journal of Productivity and Quality Management* 29 (4): 454–62.
- [6] Gupta, Shilpi, N. Singhal, Sheela Hundekari, Kamal Upreti, Anjali Gautam, Pradeep Kumar, and R. Verma. 2024. "Aspect Based Feature Extraction in Sentiment Analysis Using Bi-GRU-LSTM Model." *J. Mobile Multimedia*.
- [7] Hu, Pengnan. 2025. "Roasting Mulan: Chinese Cultural Confidence and Tucao Video Remixes on Bilibili." *Journal of Film and Video* 77 (2): 5–18.
- [8] Hu, Yafan. 2017. "Strengthening Cultural Confidence and Focusing on the Future." In *2017 International Conference on Culture, Education and Financial Development of Modern Society (ICCESE 2017)*, 493–96. Atlantis Press.
- [9] Hu, Zimei. 2025. "Analysis of International Communication Pathways for Traditional Sports Culture from the Perspective of Cultural Confidence." *International Journal of Asian Social Science Research* 2 (2): 1–8.
- [10] Hua-ping, XING. 2022. "Exploring the Modern Value and Development Position of Chinese Medicine in the Perspective of Cultural Confidence." *Journal of Nanjing University of Traditional Chinese Medicine (Social Science Edition)* 23 (1): 11–16.

- [11] Li, Xianyong, Jiabo Zhang, Yajun Du, Jian Zhu, Yongquan Fan, and Xiaoliang Chen. 2022. “A Novel Deep Learning-Based Sentiment Analysis Method Enhanced with Emojis in Microblog Social Networks.” *Enterprise Information Systems*.
- [12] Lundeberg, Mary A, Paul W Fox, Amy C Brown, and Salman Elbedour. 2000. “Cultural Influences on Confidence: Country and Gender.” *Journal of Educational Psychology* 92 (1): 152.
- [13] Lundeberg, Mary, and Lindsey Mohan. 2009. “Context Matters: Gender and Cross-Cultural Differences in Confidence.” In *Handbook of Metacognition in Education*, 221–39. Routledge.
- [14] Muzumdar, Jagannath Mohan, Monica Holiday-Goodman, Curtis Black, and Mary Powers. 2010. “Cultural Competence Knowledge and Confidence After Classroom Activities.” *American Journal of Pharmaceutical Education* 74 (8): 150.
- [15] Norman, Leanne. 2014. “A Crisis of Confidence: Women Coaches’ Responses to Their Engagement in Resistance.” *Sport, Education and Society* 19 (5): 532–51.
- [16] Rosenberger, Chandler. 2020. “‘Make the Past Serve the Present’: Cultural Confidence and Chinese Nationalism in Xi Jinping Thought.” In *Research Handbook on Nationalism*, 360–70. Edward Elgar Publishing.
- [17] Roy, Sayak Saha, and Shirin Nilizadeh. 2024. “PhishLang: A Real-Time, Fully Client-Side Phishing Detection Framework Using MobileBERT.” *arXiv Preprint arXiv: 2408.05667*.
- [18] Studiawan, Hudan, F. Sohel, and Christian N. Payne. 2021. “Anomaly Detection in Operating System Logs with Deep Learning-Based Sentiment Analysis.” *IEEE Transactions on Dependable and Secure Computing*.
- [19] Tao, Pengfei, and Daji Wen. 2020. “Literature Analysis and Practical Significance in Cite Space Environment—Taking the Cultivation of College Students’ Cultural Confidence as an Example.” In *The International Conference on Cyber Security Intelligence and Analytics*, 428–34. Springer.
- [20] Wang, Guohong, and Kena Wang. 2018. “Research on Marxist Theory Education in the Vision of Cultural Confidence.”
- [21] Yang, Jing. 2025. “Exploring the Teaching Path of ‘Revitalizing’ Traditional Chinese Music in University Public Elective Courses from the Perspective of Cultural Confidence.” *Journal of Humanities, Arts and Social Science* 9 (10).