

Sustainability of Digital International Chinese Literature Teaching via Fuzzy Analytic Hierarchy Process and Fuzzy Comprehensive Evaluation

Sihui Xiao¹

¹School of Literature and Media, Yulin Normal University

✉ Corresponding Author : Sihui Xiao <jennyxiaosihui@163.com>

Citation

Xiao, S. Sustainability of Digital International Chinese Literature Teaching via Fuzzy Analytic Hierarchy Process and Fuzzy Comprehensive Evaluation. *J. Vis. Exp.* (230), e69855, doi:10.3791/69855. (2026).

Abstract

This protocol describes a replicable method to evaluate the sustainability of digitally assisted international Chinese literature teaching using a hybrid FAHP–FCE model. The overall goal is to convert multidimensional sustainability judgments into a transparent, quantitative sustainability index (SI). The workflow constructs a four-dimensional indicator system comprising 19 indicators. Expert judgments derive indicator weights via FAHP, while learner and teacher questionnaire responses are mapped to membership degrees using the FCE procedure. Weighted fuzzy synthesis and defuzzification then yield the SI. Representative implementation across three universities in China (237 learners and 9 teachers) produced an overall SI of 78.2 (“Good”), with meaningful regional variation consistent with differences in cultural adaptation and institutional resources. The protocol is highly dependent on convening a credible expert panel. Furthermore, indicators must be operationalized consistently, as differing local policy definitions or curricular goals may complicate direct cross-site comparisons. Future applications should include longitudinal implementation across multiple semesters to test the stability of weights and membership distributions, alongside comparative studies that benchmark this fuzzy approach against alternative multi-criteria models.

Introduction

In recent years, the accelerating convergence of globalization and digitalization has reshaped international Chinese education, shifting from technology-assisted classrooms towards digitally native learning ecosystems^{1,2}. Within this shift, Chinese literature courses occupy a distinctive position because they do more than support language acquisition; they also cultivate cultural identity, aesthetic sensibility, and interpretive competence^{3,4}. These outcomes can be especially fragile in digital environments, where prevailing evaluation approaches often prioritize immediate, easily quantifiable cognitive gains while under-representing longer-horizon goals such as emotional resonance, sustained cultural engagement, and interpretive depth^{5,6}. Crucially, this protocol evaluates educational sustainability through the lens

of human behavior. By quantifying the cognitive responses, emotional engagement, and instructional behaviors of learners and teachers interacting with digital literature platforms, the framework translates subjective human experiences into measurable behavioral indicators.

Existing assessments remain limited in both scope and duration. Many studies still rely on simplified measures of knowledge mastery⁷, leaving complex learner experiences such as aesthetic pleasure, cultural empathy, and dialogic interpretation underspecified or treated as peripheral outcomes^{8,9}. Moreover, short observation windows—often shorter than a single semester—may miss cumulative effects, including how digital tools gradually shape learners’ interpretive routines and how teachers’ digital-pedagogical expertise develops through iterative practice^{10,11}.

Evaluation frameworks are also frequently fragmented: technical or pedagogical indicators are reported in isolation, while cultural and policy conditions that can constrain or enable sustainable implementation are rarely integrated into one coherent model¹². This fragmentation reduces transferability across sites, particularly in cross-cultural literature classrooms, where meaning-making is inherently open-ended and context-sensitive^{13,14}.

While previous research has extensively explored digital tool adoption and short-term cognitive gains in language acquisition, evaluation frameworks remain frequently fragmented. Existing literature lacks comprehensive models that integrate technical, pedagogical, cultural, and policy dimensions into a single measurable standard. This study needs to be conducted to address this critical gap by introducing a standardized FAHP-FCE framework, transitioning from subjective, short-term evaluations to a transparent, reproducible sustainability index that supports cross-site comparison.

To address these limitations, this protocol introduces a reproducible framework for evaluating the sustainability of digitally assisted international Chinese literature teaching. Building on fuzzy multi-criteria decision-making logic, the method integrates fuzzy analytic hierarchy process (FAHP) with fuzzy comprehensive evaluation (FCE)¹⁵ to operationalize sustainability across four linked dimensions: technical functionality, pedagogical effectiveness, cultural adaptability, and policy support. The approach is designed to offer three practical advantages over “crisp” scoring models: it transparently derives indicator weights from expert judgments while accommodating linguistic uncertainty; it converts questionnaire responses into membership degrees that better reflect gradated perceptions; and it yields an interpretable Sustainability Index that supports cross-site comparison and sensitivity analysis^{14,16,17}. By embedding interpretive transparency, cross-cultural adaptability, and long-term reliability into the evaluation workflow, the protocol moves beyond short-term outcome checks and provides a transferable assessment tool for other digitally mediated and cross-cultural teaching contexts¹⁸.

Protocol

Ensured the study complies with institutional human research ethics guidelines and applicable data-protection regulations^{19,20}. Written informed consent was obtained from each participant before any study procedure. Assigned each participant a study code, anonymized all datasets, and stored the linkage key (identifiers <> codes) in a separate, access-restricted file. Restricted raw data access to authorized personnel only, and document the retention period and deletion plan in the ethics application.

1. Preparation of evaluation context and materials

1. Define the instructional context (i.e., the specific educational environment, learner demographic profile, and pedagogical framework in which the digitally assisted teaching occurs).
 1. Specify the course type (international Chinese literature), target population, and delivery mode (in-person, blended, or online).
 2. Define the minimum learner proficiency threshold (e.g., Chinese Proficiency Test [HSK] level 5 or equivalent) and state the operational eligibility criteria (test certificate, placement result, or instructor confirmation)^{21,22}.
 3. Specify the term length (weeks), contact hours, platform/tooling constraints (without brand names), and the single evaluation objective: quantify teaching sustainability using a FAHP–FCE workflow^{23,24,25}.
2. Compile the materials list
 1. Record all supplies (forms, instruments, scripts, software, and storage) in the **Table of Materials**, including: (i) the 19-item learner questionnaire (**Supplementary File S1**), (ii) the expert pairwise-comparison form (**Supplementary File S2**), (iii) informed-consent forms, (iv) data-capture templates, and (v) analysis scripts and their versions.
 2. Assign file names and version identifiers for reproducibility (e.g., **Supplementary File S1_questionnaire_items.pdf**, **Supplementary File S2_expert_form.pdf**, **fahp_weighting.py**, **fce_membership.py**, **si_compute.py**, and **analysis_readme.txt**)^{26,27}.
3. Create the workflow diagram (**Figure 1**)
 1. Draw a one-page flowchart summarizing sections 1–10, and label inputs/outputs at each handoff (expert judgments > fuzzy matrices > weights; learner responses > membership matrices; synthesis > Sustainability Index).
 2. Mark PAUSE POINTS (Section 3.2, 4.2, 8.3), which indicate critical stages where the procedure can be safely halted and where specific data artifacts must be archived to ensure reproducibility, and indicate what artifacts must be saved at each point.

2. Construct the indicator framework

1. Build a three-level hierarchy.
 1. Map the evaluation objective to four dimensions (technical functionality, pedagogical effectiveness, cultural adaptability, policy support).
 2. Define the 19 indicators under the four dimensions (**Objective > Dimension > Indicator**), and ensure each

indicator is measurable via questionnaire response or documented evidence.

2. Define indicators and the response format.
 1. Write an operational definition for each indicator, specify whether it is positive (higher is better) or negative (lower is better), and record the indicator–dimension mapping in a data dictionary (**Table 1**).
 2. Specify the learner response scale as a 7-point Likert format with anchors: 1 = strongly disagree; 7 = strongly agree.

NOTE: Finalize and freeze the indicator set for reuse across cohorts and archive the data dictionary and **Table 1** version.

3. Assemble the expert panel and collect pairwise judgments (FAHP)

1. Recruit an expert panel.
 1. Recruit a cross-role panel (e.g., literature instructors, educational-technology specialists, assessment researchers, and administrators).
 2. Define eligibility criteria (years of experience, familiarity with digital instruction, and cross-cultural teaching exposure), and document the panel composition for transparency.
2. Collect fuzzy pairwise comparisons.
 1. Train experts with a short briefing on the dimensions/indicators and the meaning of triangular fuzzy numbers (TFNs)^{28,29,30}.
 2. Provide explicit anchors and examples to reduce variability (e.g., TFN [2, 3, 4] = “important”; TFN [4, 5, 6] = “very important”).
 3. Collect pairwise judgments for (i) the four dimensions and (ii) indicators within each dimension and verify completeness before closing the session.

4. Build fuzzy judgment matrices and test consistency

1. Construct fuzzy judgment matrices.
 1. Convert each expert’s pairwise responses into a fuzzy reciprocal matrix at each hierarchy level²⁵.
 2. Check for missing entries and enforce reciprocity rules before aggregation.
2. Test consistency and reconcile disagreements.
 1. Defuzzify each expert matrix to a crisp matrix using a prespecified rule, such as the centroid of a triangular

fuzzy number (TFN)^{31,32,33}. The crisp value is calculated using the following equation:

$$C = \frac{l + m + u}{3}$$

where C is the defuzzified crisp value, and l , m , and u represent the lower bound, modal (most likely) value, and upper bound of the TFN, respectively.

2. Compute the consistency ratio (CR) on the crisp matrix using the Analytic Hierarchy Process (AHP) procedure^{34,35}. This involves calculating the principal eigenvalue ($\lambda_{\max}$) of the matrix, determining the consistency index (

$$CI = \frac{\lambda_{\max} - n}{n - 1}$$

, where n is the matrix size), and dividing it by the appropriate random index (RI) to obtain the CR (

$$CR = \frac{CI}{RI}$$

). Set the acceptance threshold as $CR < 0.10$.

3. If $CR \geq 0.10$, return the inconsistent comparisons to the expert, reconcile disagreements using facilitated discussion, and recollect only the disputed judgments.

PAUSE POINT: Archive each matrix, CR report, and reconciliation log as replication artifacts.

5. Derive fuzzy weights (FAHP output)

1. Aggregate individual judgments.
 1. Aggregate expert matrices into a group-level fuzzy matrix using a specified operator (e.g., fuzzy geometric mean)^{17,36}.
 2. Record the aggregation operator and parameters in **Supplementary File S2**.
2. Defuzzify and normalize weights.
 1. Defuzzify the group-level fuzzy priorities to obtain crisp weights using the same prespecified defuzzification rule¹⁰.
 2. Normalize weights within each level (dimensions sum to 1; indicators within each dimension sum to 1) and output the final dimension and indicator weight vectors.
 3. Populate **Table 1** with final weights and lock the table version used for the subsequent FCE section.

6. Design and administer the learner questionnaire

1. Prepare the instrument.
 1. Provide the full 19 items verbatim in **Supplementary File S1**, each mapped to one indicator in **Table 1**.

2. Use a 7-point Likert scale with clear anchor text printed on every page/screen (1 = strongly disagree; 7 = strongly agree).
 3. Pilot-test the questionnaire with a small group (e.g., 10–15 learners) to confirm comprehension and completion time³⁷.
2. Administer the questionnaire and enforce one unique response per participant.
 1. Invite eligible learners (HSK 5+ or operational equivalent) and establish a standardized data collection protocol. Specify the exact administration window (e.g., a defined two-week period at the end of the term) and the completion mode (e.g., in-class supervised sessions lasting 15–20 min to ensure focused participation, or controlled online distribution via secure, timed survey links). Instructors or researchers must provide a standardized briefing before questionnaire distribution, clearly explaining the study's purpose, assuring data anonymity, and defining the 7-point Likert scale anchors to all respondents.
 2. Enforce one response per participant by issuing a single-use access token per learner (or an equivalent one-time code), disabling repeat submissions, and checking for duplicates using the token log.
 3. If anonymity is required, store tokens separately from responses and hash any login identifiers before analysis.
 4. Document all exclusion rules (e.g., incomplete responses, duplicate token entries) before data cleaning to prevent post hoc bias.
 3. Record metadata for subgroup analysis.
 1. Record comprehensive non-identifying metadata to clearly define the respondent profile for subgroup analysis. For learners, this must include demographic and academic characteristics such as native language group, age, gender, years of Mandarin study, current HSK proficiency level, academic major, course type, and study site. For the expert panel, explicitly record professional demographics including years of teaching experience, highest academic degree obtained, and specific domain expertise (e.g., literature instruction, educational technology).

CAUTION: Protect personally identifiable information and follow approved data-handling procedures.

7. Compute membership degrees (FCE)

1. Define evaluation grades and numeric scores.

1. Define a five-grade evaluation set: Excellent, Good, Moderate, Relatively Poor, Poor.
 2. Assign numeric scores for grade-to-index conversion (e.g., 90, 80, 70, 60, 50), and record the scoring scheme in **Supplementary File S3**³⁸.
2. Specify membership functions and build membership matrices.
 1. Specify membership functions for mapping Likert responses to grade memberships (e.g., trapezoidal functions)^{39,40}.
 2. Provide explicit parameters for each grade boundary in **Supplementary File S3**, and state whether the same parameters apply across all indicators.
 3. For each indicator, convert each learner response to a 5-element membership vector, then aggregate across learners (e.g., by averaging memberships) to form the indicator-level membership vector.
 4. Stack all indicator-level membership vectors to construct the indicator-by-grade membership matrix.

8. Perform fuzzy synthesis and derive the sustainability index (SI)

1. Apply a weighted fuzzy operator.
 1. Combine the indicator weight vector with the indicator membership matrix using a specified weighted fuzzy synthesis operator (e.g., weighted average operator) to obtain the comprehensive membership vector^{41,42}.
 2. Report the operator's name and formula in **Supplementary File S3** to ensure standalone reproducibility.
2. Defuzzify to compute SI on a 0–100 scale.
 1. Convert the comprehensive membership vector to a single SI value by weighted averaging of grade scores (as defined in section 8.1.2).
 2. Output dimension-level SI values by repeating section 9.1–9.2 within each dimension using dimension-specific indicator weights.
 3. Classify SI values using prespecified cutoffs: 85–90 = Excellent; 75–84 = Good; 65–74 = Moderate; 55–64 = Relatively Poor; ≤ 54 = Poor.

PAUSE POINT: Store intermediate vectors, matrices, and calculation sheets (or script outputs) with file hashes in a repository.

9. Optional qualitative validation (mixed methods)

1. Sample participants for interviews.
 1. Sample instructors and learners using a purposive scheme aligned with SI patterns (e.g., high vs. low cultural adaptability) and obtain additional consent for recording.
 2. Use a semi-structured interview guide focused on interpretability of scores, perceived drivers of sustainability, and contextual constraints (**Supplementary File S4**)⁴³.
2. Code and triangulate findings.
 1. Transcribe recordings verbatim and de-identify transcripts before coding.
 2. Code transcripts using a prespecified coding scheme, and document coder training, inter-coder agreement (if applicable), and resolution rules^{44,45}.
 3. Triangulate qualitative themes with dimension-level SI results to contextualize strengths and weaknesses without overriding the quantitative procedure.

10. Report and archive (reproducibility package)

1. Report on core outputs.
 1. Report dimension and indicator weights, membership matrices (or summaries), SI values, grade labels, and subgroup comparisons (e.g., site or proficiency groups).
 2. State all exclusion rules and the final analyzed sample sizes (ensure consistency with Abstract/Results).
2. Archive materials and code.
 1. Archive anonymized datasets, scripts, and spreadsheets used for FAHP weighting, membership computation, synthesis, and defuzzification.
 2. Provide software/runtime versions (e.g., programming language version; package versions) and fixed random seeds (if used)^{46,47}.
3. Run sensitivity analysis (recommended).
 1. Perturb indicator weights by $\pm 10\%$ (one-at-a-time or within dimension) while renormalizing, recompute SI, and summarize SI fluctuations to identify leverage points (**Table 2**).
 2. Interpret indicators with low weight but high sensitivity as potential “threshold risks,” and report these explicitly.

Results

Dimension weights and key indicators

Using the fuzzy analytic hierarchy process (FAHP), the dimension weights indicated that the technical dimension (T) contributed the largest share to sustainability (0.32), followed by the cultural dimension (C, 0.26) and the pedagogical dimension (P, 0.23), whereas the policy dimension (G) received the lowest weight (0.19)¹⁷. At the indicator level, the technical dimension prioritized functional iteration efficiency (T3, 0.25)¹ and algorithmic interpretability (T1, 0.22). The cultural dimension emphasized cross-cultural interpretive adaptability (C2, 0.29) and cultural-load word accuracy (C1, 0.27). Within the pedagogical dimension, emotional goal attainment (P3, 0.24) outweighed knowledge goal attainment (P1, 0.21), highlighting the importance of affective and interpretive outcomes in digitally mediated literature instruction⁴⁸. In the policy dimension, data privacy compliance (G1, 0.26) ranked highest, suggesting that governance-related criteria function as a high-salience condition even when overall policy weight is comparatively lower (see **Table 1**)^{19,49}.

Overall sustainability index and regional variation

The overall sustainability index (SI), computed using fuzzy comprehensive evaluation (FCE), was 78.2 and classified as “Good” (75–84). Dimension-level results showed stronger performance in the technical (82.5) and cultural (79.3) dimensions, whereas the pedagogical (76.8) and policy (73.1) dimensions were comparatively lower (**Tables 3** and **4**)³⁸. Regional comparisons showed statistically significant differences in SI ($F = 4.37$, $p < 0.05$)²⁵. East China (Shanghai) achieved the highest SI (81.7), followed by North China (Beijing) (79.5), while South China (Guangzhou) scored lowest (72.3; “Moderate”). Post-hoc testing indicated that the South China campus differed significantly from the other two regions ($p < 0.05$)⁵⁰. Learners’ native-language background was also associated with SI: English-native learners scored higher (80.1) than learners from less commonly taught languages (73.6; $t = 3.22$, $p < 0.01$), with Arabic learners showing the lowest SI (71.2)^{51,52}.

Sensitivity analysis

To test the robustness of the model, the authors conducted a sensitivity analysis (**Table 2**). This involved systematically adjusting the weight of each individual indicator by $\pm 10\%$ while keeping other relative weights proportional. This step-by-step perturbation helps identify key leverage points. Algorithmic interpretability (T1) produced the largest SI fluctuation ($\pm 6.3\%$), followed by cross-cultural adaptability (C2, $\pm 5.7\%$), emotional goal attainment (P3, $\pm 4.9\%$), and data privacy compliance (G1, $\pm 4.2\%$). These patterns indicate that interpretability and cross-cultural adaptability are high-impact levers for maintaining sustainability under uncertainty⁵³.

Qualitative themes (optional triangulation)

Qualitative coding of interview transcripts and classroom video records using qualitative analysis software identified three

convergent themes. First, emotional resonance dilution: digital tools improved efficiency but could reduce the perceived emotional depth of literary works. Second, teacher as curator: instructors increasingly filtered and contextualized system outputs to enhance interpretability and cultural appropriateness. Third, corpus scarcity in small languages: learners from less commonly taught languages reported more frequent mistranslations and culturally salient misinterpretations, consistent with lower SI patterns in those subgroups^{44,45}.

Model-related tests

The authors evaluated three a priori hypotheses regarding the sustainability dimensions and outcomes. First, Hypothesis 1

(H1) posited that the technical dimension would be the most critical component for sustainability; this was supported, as it carried the highest weight (0.32, $p < 0.05$). Second, Hypothesis 2 (H2) proposed that regions with varying economic and educational resources would show significantly different SI scores; this was partially supported. The South China campus scored significantly lower than the other regions, whereas the difference between Beijing and Shanghai was not statistically significant. Third, Hypothesis 3 (H3) stated that teachers' digital integration competence would positively influence overall sustainability; this was supported, as integration competence significantly moderated sustainability ($\beta = 0.31$, $p < 0.01$).

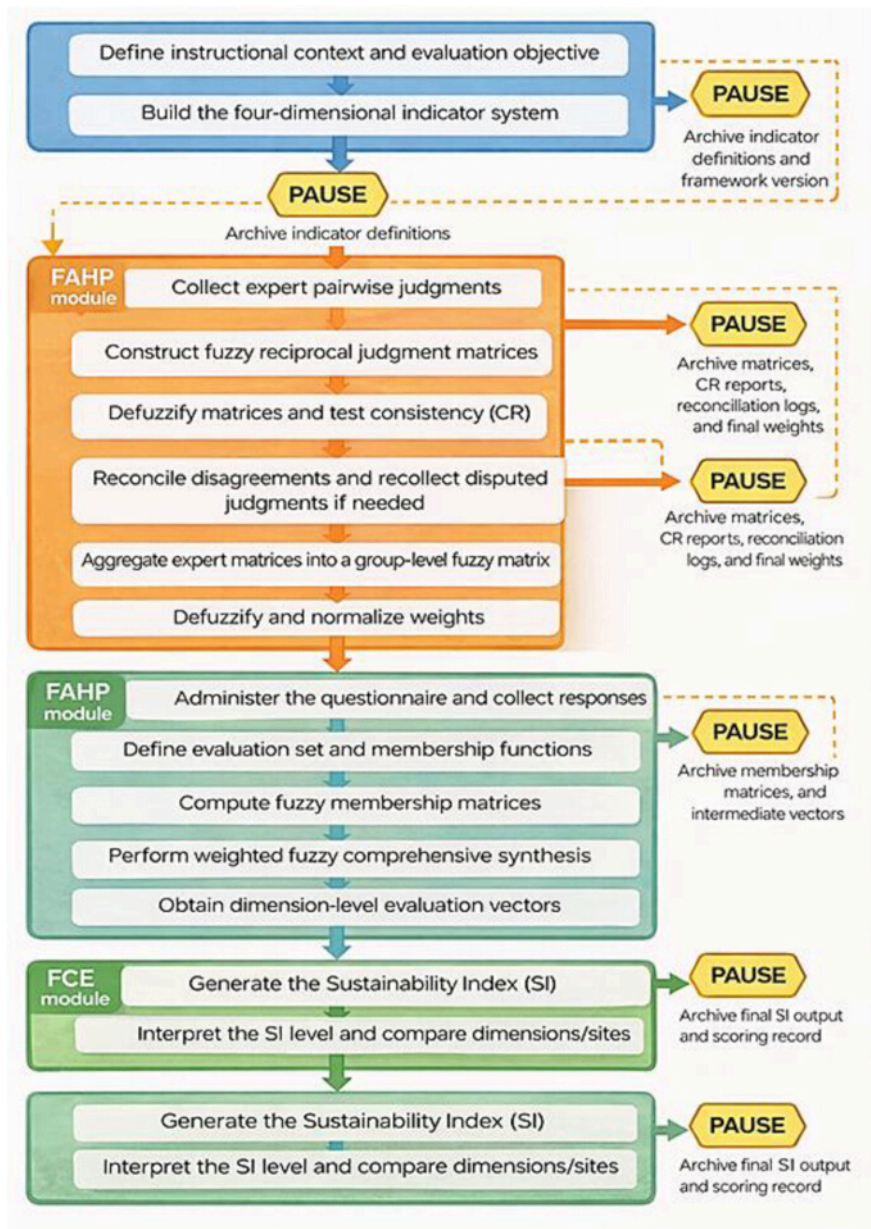


Figure 1: Workflow of the FAHP–FCE protocol for sustainability evaluation. The diagram summarizes the end-to-end workflow, from defining the instructional context and indicator system to collecting expert judgments (FAHP), computing membership matrices from questionnaire data (FCE),

and generating the sustainability index (SI). The figure also marks recommended pause points for archiving matrices, intermediate vectors, and calculation artifacts to support reproducibility. [Please click here to view a larger version of this figure.](#)

Dimension	Code	Indicator	Operational definition (summary)	Weight
Technology (T, 0.32)				
	T1	Algorithmic interpretability	Learners' ability to understand system outputs	0.22
	T2	System stability	Proportion of teaching time interrupted by failures	0.18
	T3	Functional iteration efficiency	Avg. time (days) from demand to update	0.25
	T4	Interface friendliness	Avg. minutes to complete independent operation	0.15
	T5	Multi-terminal adaptability	Integrity score across devices (1–7)	0.2
Pedagogy (P, 0.23)				
	P1	Knowledge objective attainment	Avg. test score improvement (%)	0.21
	P2	Teaching method integration	Teacher evaluation of strategy fit (1–7)	0.19
	P3	Emotional goal attainment	Learner-reported resonance (1–7)	0.24
	P4	Teacher–student interaction	Proportion of time in effective dialogue (%)	0.16
	P5	Learning motivation	Extra reading time (hours/week)	0.2
Culture (C, 0.26)				
	C1	Cultural load word accuracy	Rate of misinterpretation (%)	0.27

	C2	Cross-cultural interpretive adaptability	Teacher evaluation (1–5)	0.29
	C3	Imagery transformation	Correct understanding rate (%)	0.23
	C4	Multilingual coverage	Effectiveness for minority- language learners	0.21
Policy (G, 0.19)				
	G1	Data privacy compliance	Privacy policy integrity (1–7)	0.26
	G2	Resource input sustainability	3-year investment growth rate (%)	0.23
	G3	Teacher training system	Avg. annual training hours	0.2
	G4	Teaching evaluation adaptability	% of evaluations incl. digital tools	0.15
	G5	International standard alignment	Expert compliance rate (%)	0.16

Table 1: Hierarchical indicator system and FAHP-derived weights used in the FAHP–FCE sustainability evaluation framework. The table presents the four evaluation dimensions and 19 indicators, together with their codes, operational definitions, dimension assignments, dimension-level weights, and normalized indicator weights derived from FAHP.

Indicator Code	Indicator	Weight Perturbation	SI Fluctuation (%)	Impact Rank
T1	Algorithmic interpretability	±10%	±6.3%	1
C2	Cross-cultural interpretive adaptability	±10%	±5.7%	2
P3	Emotional goal attainment	±10%	±4.9%	3
G1	Data privacy compliance	±10%	±4.2%	4

Table 2: Sustainability index scores by dimension across the overall sample and regions. The table reports scores for Technology, Pedagogy, Culture, and Policy, together with the ensemble sustainability score (S) and grade categories, enabling comparison across the overall sample and regional groups.

Region / Sample	Technology (T)	Pedagogy (P)	Culture (C)	Policy (G)	Ensemble (S)	Grade
Overall	82.5	76.8	79.3	73.1	78.2	Good
East China (Shanghai)	84.1	77.5	81.3	74	81.7	Good
North China (Beijing)	83.2	77	79.8	74.2	79.5	Good
South China (Guangzhou)	74.2	72.5	68.9	71	72.3	Moderate

Table 3: Dimension-level sustainability index scores and grade categories. The table summarizes the SI scores for each evaluation dimension and the ensemble score, together with the corresponding grade labels based on predefined classification thresholds.

Dimension	SI Score	Grade
Technology (T)	82.5	Good
Pedagogy (P)	76.8	Good
Culture (C)	79.3	Good
Policy (G)	73.1	Moderate
Ensemble (S)	78.2	Good

Table 4: Sensitivity analysis of indicator weights on the sustainability index. The table reports SI fluctuations under $\pm 10\%$ perturbations of individual indicator weights, including the magnitude and rank order of impact.

Supplementary File S1: Learner questionnaire. The file contains the 19-item learner questionnaire developed for the sustainability evaluation framework, covering the four dimensions of Technology, Pedagogy, Culture, and Policy, together with response scales and indicator assignments used in the FCE analysis.[Please click here to download this file.](#)

Supplementary File S2: Expert pairwise-comparison form. The file contains the structured expert assessment form used in the FAHP stage to elicit pairwise judgments across dimensions and indicators, together with instructions and rating anchors for weight derivation.[Please click here to download this file.](#)

Supplementary File S3: Scoring scheme, membership functions, and calculation rules. This file provides the detailed scoring scheme and membership function parameters used in the fuzzy comprehensive evaluation (FCE) process. It includes the definition of evaluation grades, numerical score assignments, trapezoidal membership functions for converting 7-point Likert responses into fuzzy membership degrees, and the weighted synthesis and defuzzification rules used to compute the Sustainability Index (SI).[Please click here to download this file.](#)

Supplementary File S4: Semi-structured interview guide for qualitative validation. This file presents the semi-structured interview guide used for the optional mixed-methods validation component of the study. The guide includes thematic questions addressing technical functionality, pedagogical effectiveness, cultural adaptability, policy or institutional support, and suggestions for improving the long-term sustainability of digitally assisted international Chinese literature teaching.[Please click here to download this file.](#)

Discussion

This protocol demonstrates that a hybrid FAHP–FCE workflow can operationalize “teaching sustainability” in digitally assisted

international Chinese literature courses as a transparent, reproducible index rather than a loosely defined outcome. The results indicate that technical functionality functions as a prerequisite layer for sustainability, with algorithmic interpretability showing the highest sensitivity to sustainability index (SI) fluctuations³¹. Practically, this pattern is consistent with a staged logic: learners and instructors tend to require stable, explainable system behavior before they can sustain deeper cultural engagement and interpretive work⁵⁴.

A second leverage point was cultural adaptability, which was particularly consequential for learners from non-English and less commonly taught language backgrounds³⁴. The qualitative themes help interpret this quantitative pattern: when corpora and interface defaults are implicitly English-centered, learners who rely on other linguistic pathways may experience more frequent mistranslations or culturally salient misreadings, which in turn weakens perceived sustainability even if overall usability remains acceptable^{13,55}. Importantly, the protocol’s four-dimensional structure makes this trade-off visible: technical gains can coexist with cultural fragility, which would be obscured by single-score or purely cognitive outcome measures.

Several sections are critical for replicability and should be treated as non-negotiable checkpoints. First, the indicator system must remain stable within a cohort, with each item’s definition, polarity (positive/negative), and dimension mapping locked before data collection (**Table 1**). Second, expert judgments must be accompanied by an explicit consistency test and a documented reconciliation process when the threshold is exceeded; otherwise, weight vectors can become idiosyncratic and difficult to interpret across sites. Third, FCE computations must specify the evaluation set, membership functions, synthesis operator, and defuzzification rule in the protocol text and supplemental files; changing any of these choices without documentation can produce materially different SI values^{35,41,42}.

This technique also requires practical troubleshooting and minor modifications in real deployments. To reduce noise in expert inputs, provide anchor examples for triangular fuzzy numbers and run a brief calibration exercise before formal judgment collection. To protect data integrity in questionnaire administration, enforce one unique response per learner using single-use access codes and pre-specify exclusion rules for incomplete or duplicate entries. For the membership-function section, conduct a small pilot to verify that the chosen grade boundaries produce non-degenerate membership distributions (e.g., avoid cases where most responses collapse into a single grade), and adjust boundaries only before final analysis, while documenting all changes^{56,57}.

The protocol has several limitations that should guide appropriate use. It depends on a credible expert panel; if expert selection is narrow or not transparent, weights may reflect local preferences rather than transferable judgments. It also relies on an indicator set that can be operationalized consistently across contexts; when local policy definitions, platform affordances, or curricular goals differ substantially, direct cross-site comparison may require careful re-validation rather than simple reuse. Finally, policy support may appear low-weight while still acting as a “threshold risk” variable: compliance failure or funding discontinuity can collapse sustainability despite strong technical and pedagogical performance^{19,36,58}.

Relative to conventional crisp evaluation approaches, the hybrid FAHP–FCE method adds value by making weighting assumptions auditable, representing linguistic judgments under uncertainty, and producing outputs that can be archived for replication (weights, membership matrices, intermediate vectors, and SI). Beyond the present application, the workflow can be extended to other digitally mediated, cross-cultural teaching settings by (i) retaining the four-dimensional logic while adapting indicator wording, (ii) running sensitivity analysis to identify high-impact leverage points, and (iii) pairing SI outputs with qualitative triangulation to prevent over-interpretation of a single composite score. Future applications may also include longitudinal implementation across multiple semesters to test the stability of weights and membership distributions, and comparative studies that benchmark this fuzzy approach against alternative multi-criteria models under the same dataset^{59,60}.

Disclosures

The authors declare no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Artificial intelligence (AI) tools were used solely for language polishing and grammatical editing during the revision of this manuscript; the authors take full responsibility for the content and integrity of the published work.

Acknowledgments

The author thanks Yulin Normal University for institutional support. We are grateful to the participating teachers and students for their time and contributions, and to the expert panel members for providing judgments used in the FAHP procedures. We also thank colleagues in the School of Literature and Media for feedback that improved the clarity of the protocol.

Funding: This research is supported by the 2025 Guangxi Universities' Research Capacity Enhancement Program for Young and Middle-aged Teachers: Enhancing Cultural Identity of International Students in Guangxi through Telling Guangxi Stories Well (Project No. 2025KY0659). It is also funded by the 2025 Guangxi Undergraduate Teaching Reform Project: Reconstruction and Practice of the International Chinese Education Curriculum System Oriented Toward ASEAN under the New Liberal Arts Perspective (Project No. 2025JGB347).

References

1. Teräs, M. Education and technology: Key issues and debates. *Int. Rev. Educ.* **68** (4), 635-636 (2022).
2. Haleem, A., Javaid, M., Qadri, M. A., Suman, R. Understanding the role of digital technologies in education: A review. *Sustain. Oper. Comput.* **3**, 275-285 (2022).
3. Nussbaum, M. C. *Love's knowledge: Essays on philosophy and literature*. OUP USA, (1990).
4. Paran, A. The role of literature in instructed foreign language learning and teaching: An evidence-based survey. *Lang. Teach.* **41** (4), 465-496 (2008).
5. Biesta, G. Good education in an age of measurement: On the need to reconnect with the question of purpose in education. *Educ. Assess. Eval. Account.* **21** (1), 33-46 (2009).
6. Thelwall, M. *Word association thematic analysis: A social media text exploration strategy*. (2021).
7. Biggs, J., Tang, C., Kennedy, G. *Teaching for quality learning at university*. (2003).
8. Kidd, D. C. Castano, E. Reading literary fiction improves theory of mind. *Science*. **342** (6156), 377-380 (2013).
9. Paesani, K. Allen, H. W. Teacher development and multiliteracies pedagogy: Challenges and opportunities for postsecondary language programs. *Second Lang. Res. Pract.* **1** (1), 124-138 (2020).
10. Mishra, P. Koehler, M. J. Technological pedagogical content knowledge: A framework for teacher knowledge. *Teach. Coll. Rec.* **108** (6), 1017-1054 (2006).
11. Stockwell, G. 156-175, *Springer*. (2013).
12. Cohen, A., Soffer, T., Henderson, M. Students' use of technology and their perceptions of its usefulness in higher education: International comparison. *J. Comput. Assist. Learn.* **38** (5), 1321-1331 (2022).
13. Knowles, M. Kramsch, C. Context and culture in language teaching. *Mod. Lang. J.* **79** (4), 565 (1993).
14. Helm, F. The practices and challenges of telecollaboration in higher education in Europe. *Lang. Learn. Technol.* **19** (2), 197-217 (2015).

15. Peng, G., Han, L., Liu, Z., Guo, Y., Yan, J., Jia, X. An application of fuzzy analytic hierarchy process in risk evaluation model. *Front. Psychol.* **12**, 715003 (2021).
16. Saaty, R. W. The analytic hierarchy process—what it is and how it is used. *Math. Modell.* **9** (3), 161-176 (1987).
17. Liu, Y., Eckert, C. M., Earl, C. A review of fuzzy AHP methods for decision-making with subjective judgements. *Expert Syst. Appl.* **161**, 113738 (2020).
18. Dong, Z. Exploring cross-cultural communication content adaptability through advanced natural language processing and sentiment analysis. *Syst. Soft Comput.* **7**, 200290 (2025).
19. Slade, S. Prinsloo, P. Learning analytics: Ethical issues and dilemmas. *Am. Behav. Sci.* **57** (10), 1510-1529 (2013).
20. Bera. *Ethical guidelines for educational research*. British Educational Research Association, (2011).
21. Ministry of Education of the P. R. C. *Chinese proficiency grading standards for international Chinese language education (gf 0025–2021)*. Beijing Language and Culture University Press, (2021).
22. Koda, K. *Insights into second language reading: A cross-linguistic approach*. (Cambridge University Press, <https://doi.org/10.1017/CBO9781139524841> (2005).
23. Golden, B. L., Wasil, E. A., Harker, P. T. The analytic hierarchy process. *Appl. Stud.* **2** (1), 1-273 (1989).
24. Zadeh, L. A. Fuzzy sets. *Inf. Control.* **8** (3), 338-353 (1965).
25. Li, J., Xue, E., Guo, S. The effects of PISA on global basic education reform: A systematic literature review. *Humanit. Soc. Sci. Commun.* **12** (1), 106 (2025).
26. Sandve, G. K., Nekrutenko, A., Taylor, J., Hovig, E. Ten simple rules for reproducible computational research. *PLoS Comput. Biol.* **9** (10), e1003285 (2013).
27. Wilkinson, M. D. The fair guiding principles for scientific data management and stewardship. *Sci. Data.* **3** (1), 1-9 (2016).
28. Zadeh, L. A. The concept of a linguistic variable and its application to approximate reasoning—I. *Inf. Sci.* **8** (3), 199-249 (1975).
29. Klir, G. J., Yuan, B. Fuzzy sets and fuzzy logic: Theory and applications. *Possibility Theory versus Probab. Theory.* **32** (2), 207-208 (1996).
30. Shulman, L. Knowledge and teaching: Foundations of the new reform. *Harv. Educ. Rev.* **57** (1), 1-23 (1987).
31. Zhou, R. Chan, A. H. Using a fuzzy comprehensive evaluation method to determine product usability: A test case. *Work.* **56** (1), 21-29 (2017).
32. Ross, T. J. *Fuzzy logic with engineering applications*. John Wiley & Sons, (2005).
33. Opricovic, S. Tzeng, G. H. Defuzzification within a multicriteria decision model. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.* **11** (5), 635-652 (2003).
34. Ziegler, D. M. Fine-tuning language models from human preferences. *arXiv.* (2019).
35. Saaty, T. L. *The analytic hierarchy process: Planning, priority setting, resource allocation*. McGraw-Hill, (1980).
36. European, C. *Ethics guidelines for trustworthy AI*. (2021).
37. Artino Jr, A. R., La Rochelle, J. S., Dezee, K. J., Gehlbach, H. Developing questionnaires for educational research: Amee guide no. 87. *Med. Teach.* **36** (6), 463-474 (2014).
38. Zimmermann, H. J. *Fuzzy set theory—and its applications*. (2011).
39. Pedrycz, W., Gomide, F. *Fuzzy systems engineering: Toward human-centric computing*. John Wiley & Sons, (2007).
40. Mardani, A., Jusoh, A., Nor, K., Khalifah, Z., Zakwan, N., Valipour, A. Multiple criteria decision-making techniques and their applications—a review of the literature from 2000 to 2014. *Econ. Res.* **28** (1), 516-571 (2015).
41. Yager, R. R. On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Trans. Syst. Man Cybern.* **18** (1), 183-190 (2002).
42. Feng, S. Xu, L. D. Decision support for fuzzy comprehensive evaluation of urban development. *Fuzzy Sets Syst.* **105** (1), 1-12 (1999).
43. Creswell, J. W. Clark, V. L. P. *Designing and conducting mixed methods research*. Sage., (2017).
44. Saldaña, J. *The coding manual for qualitative researchers*. Sage. (2021).
45. O'Connor, C. Joffe, H. Intercoder reliability in qualitative research: Debates and practical guidelines. *Int. J. Qual. Methods.* **19**, 1609406919899220 (2020).
46. Stodden, V., Leisch, F., Peng, R. D. *Implementing reproducible research.* **546**, (2014).
47. Heil, B. J., Hoffman, M. M., Markowetz, F., Lee, S. I., Greene, C. S., Hicks, S. C. Reproducibility standards for machine learning in the life sciences. *Nat. Methods.* **18** (10), 1132-1135 (2021).
48. Yi, W., Huang, X., Kuzmin, S., Gerasimov, I., Luo, Y. Seekg: Sentiment analysis for e-learning evaluation incorporating knowledge graphs. *Educ. Inf. Technol.* **30** (12), 16291-16320 (2025).
49. Hoel, T. Chen, W. Privacy engineering for learning analytics in a global market: Defining a point of reference. *Int. J. Inf. Learn. Technol.* **36** (4), 288-298 (2019).
50. Field, A. *Discovering statistics using IBM SPSS Statistics*. (2024).
51. Cumming, G. Calin-Jageman, R. *Introduction to the new statistics: Estimation, open science, and beyond*. (2024).
52. Ruxton, G. D. The unequal variance t-test is an underused alternative to the Student's t-test and the Mann–Whitney U test. *Behav. Ecol.* **17** (4), 688-690 (2006).
53. Saltelli, A. *Global sensitivity analysis: The primer*. John Wiley & Sons, (2008).
54. Khosravi, H. Explainable artificial intelligence in education. *Comput. Educ. Artif. Intell.* **3**, 100074 (2022).
55. Blodgett, S. L., Barocas, S., Daumé III, H., Wallach, H. Language (technology) is power: A critical survey of "bias" in NLP. *arXiv.* (2020).
56. Mukhametzyanov, I. Specific character of objective methods for determining weights of criteria in MCDM problems: Entropy, critic, and SD. *Decis. Mak. Appl. Manag. Eng.* **4** (2), 76-105 (2021).
57. Dillman, D. A., Smyth, J. D., Christian, L. M. *Internet, phone, mail, and mixed-mode surveys: The tailored design method*. (2014).
58. Williamson, B., Piattoeva, N. 81–93, *Routledge*. (2020).
59. Singer, J. D. Willett, J. B. *Applied longitudinal data analysis: Modeling change and event occurrence*. (2003).
60. Kazimieras Zavadskas, E., Antucheviciene, J., Adeli, H., Turskis, Z. Hybrid multiple criteria decision making methods: A review of applications in engineering. *Sci. Iran.* **23** (1), 1-20 (2016).

数字化国际中文文学教学可持续性： 基于模糊层次分析法与模糊综合评价

肖思慧

玉林师范学院 文学与传媒学院

发表日期：2026年4月3日 DOI：10.3791/69855

摘要

本研究方案提出一种可复现的方法，采用FAHP-FCE混合模型评估数字技术辅助的国际中文文学教学的可持续性。总体目标是将多维度的可持续性判断转化为透明、量化的可持续性指数（SI）。研究流程构建了由4个维度、19项指标组成的指标体系；专家判断通过模糊层次分析法（FAHP）确定指标权重，学习者与教师问卷数据则通过模糊综合评价（FCE）映射为隶属度。随后进行加权模糊合成与去模糊化，得到SI。该方案在中国3所高校中的代表性实施（237名学习者、9名教师）得到总体SI=78.2，等级为“良好”；不同地区之间存在具有实际意义的差异，与文化适应性和院校资源差异相一致。该方案高度依赖可信的专家组，同时要求各指标在不同场域中保持一致的操作化定义；若地方政策定义或课程目标不同，跨场域直接比较可能受到限制。未来应开展跨多个学期的纵向实施，以检验权重和隶属度分布的稳定性，并将该模糊方法与其他多准则评价模型进行比较。

引言

近年来，全球化与数字化加速融合，重塑了国际中文教育，使其由“技术辅助课堂”逐步转向“数字原生学习生态”。在这一转型中，中文文学课程具有独特地位：它不仅服务于语言习得，也承担文化认同、审美感受力和阐释能力培养等功能。

这些目标在数字环境中尤其容易受到削弱。既有评价往往偏重即时、易量化的认知收益，而对情感共鸣、持续文化参与和深层阐释等长期目标关注不足。本方案从人的行为视角评价教育可持续性，通过量化学习者和教师在数字文学平台中的认知反应、情感投入与教学行为，把主观体验转化为可测量的行为指标。

现有评价在范围和持续时间上均存在局限。许多研究仍主要依赖知识掌握等简化指标，对审美愉悦、文化共情和对话式阐释等复杂学习体验界定不足；同时，短于一个学期的观察窗口可能无法捕捉数字工具对学习者的阐释习惯及教师数字教学能力的累积影响。

引言（续）

评价框架还经常呈碎片化状态：技术指标或教学指标被孤立报告，而可能制约或促进可持续实施的文化与政策条件很少被整合进同一模型。这降低了评价结果在不同场域间的可迁移性，尤其是在跨文化文学课堂中，意义建构本身具有开放性和情境依赖性。

既有研究虽然广泛探讨数字工具采用及语言学习中的短期认知收益，但仍缺乏把技术、教学、文化和政策四个维度整合为统一可测标准的综合模型。本研究因此引入标准化FAHP-FCE框架，使评价从主观、短期判断转向透明、可复现、可用于跨场域比较的可持续性指数。

本方案基于模糊多准则决策逻辑，将FAHP与FCE结合，在技术功能、教学有效性、文化适应性和政策支持四个相互关联的维度上实现可持续性的操作化。相较于传统“精确值”评分模型，该方法有三点优势：其一，在容纳语言判断不确定性的同时透明地获得专家权重；其二，将问卷回答转化为隶属度，更能反映渐进式感知；其三，得到可解释的可持续性指数，便于跨场域比较与敏感性分析。

研究方案

研究须符合机构人体研究伦理规范及适用的数据保护法规。所有参与者在研究程序开始前均应签署书面知情同意书；为每位参与者分配研究代码，对数据集进行匿名化处理，并将“身份信息—研究代码”对应表单独保存在受限文件中。原始数据仅授权人员可访问，并在伦理申请中明确保存期限与删除方案。

1. 准备评价情境与材料

- 1) 界定教学情境：明确数字化辅助教学发生的具体教育环境、学习者人口学特征与教学框架；明确课程类型为国际中文文学、目标群体及授课模式（线下、混合或在线）。
- 2) 规定最低学习者语言水平，例如HSK五级或同等水平，并明确资格判定方式（证书、分班测试或教师确认）。
- 3) 明确学期长度、课时、平台/工具限制（不写品牌名），并确定唯一评价目标：通过FAHP-FCE流程量化教学可持续性。

2. 汇总材料清单

记录全部表单、工具、脚本、软件和存储材料，包括：19项学习者问卷（补充文件S1）、专家两两比较表（S2）、知情同意书、数据采集模板以及分析脚本及版本。为所有文件分配明确文件名与版本号，以保证可复现性。

3. 绘制工作流程图（图1）

制作一页流程图概括第1—10部分，并在各步骤交接处标明输入/输出：专家判断→模糊矩阵→权重；学习者回答→隶属矩阵；模糊合成→可持续性指数。标注暂停点（3.2、4.2、8.3），说明可安全中止的位置及必须归档的数据材料。

2. 构建指标框架

建立三级层次结构：评价目标→四个维度（技术功能、教学有效性、文化适应性、政策支持）→19项指标。每项指标必须能够通过问卷回答或书面证据进行测量。

为每项指标撰写操作性定义，说明其为正向指标（越高越好）还是负向指标（越低越好），并在数据字典（表1）中记录“指标—维度”映射。学习者采用7点Likert量表：1=非常不同意，7=非常同意。指标体系一旦确定，应在不同队列重复使用时保持冻结，并归档数据字典和表1版本。

3. 组建专家组并收集两两判断（FAHP）

专家组应跨角色组成，例如文学教师、教育技术专家、评价研究者和管理人员。需预先规定专家资格，包括从业年限、数字教学熟悉程度及跨文化教学经验，并记录专家构成以保证透明度。

在正式判断前，对专家进行简短培训，解释维度、指标以及三角模糊数（TFN）的含义，并提供明确锚点和示例以降低判断差异，例如TFN [2,3,4]表示“重要”，[4,5,6]表示“非常重要”。收集四个维度之间以及各维度内部指标之间的两两比较，并在会话结束前核对完整性。

4. 构建模糊判断矩阵并检验一致性

将每位专家的两两比较转换为各层级的模糊互反矩阵，检查缺失项并执行互反规则。随后使用预先规定的去模糊化规则（例如三角模糊数重心法）把每个专家矩阵转换为精确矩阵。设三角模糊数为 (l,m,u) ，则去模糊值可按重心法计算。

按照AHP程序计算一致性比率CR：先求矩阵最大特征值 $\lambda_{\max}$ ，再求一致性指标 $CI=(\lambda_{\max}-n)/(n-1)$ ，最后以随机一致性指标RI计算 $CR=CI/RI$ 。接受阈值设为 $CR<0.10$ 。若 $CR\geq 0.10$ ，应把不一致的比较返回专家，通过引导讨论协调分歧，仅重新收集存在争议的判断。暂停点：归档每个矩阵、CR报告及协调记录。

5. 计算模糊权重（FAHP输出）

采用明确的聚合算子（如模糊几何平均）将个体专家矩阵合成为群体模糊矩阵，并在补充文件S2记录聚合方法和参数。随后采用同一预设规则进行去模糊化，得到精确权重；在各层级内归一化，使四个维度权重之和为1、每个维度内部指标权重之和为1。将最终权重填入表1并锁定版本，供后续FCE使用。

6. 设计并实施学习者问卷

在补充文件S1中逐字列出19个题项，每个题项对应表1中的一个指标。全程使用7点Likert量表，并在每页/每个屏幕明确标注1=非常不同意、7=非常同意。正式实施前可用10—15名学习者进行预测试，以确认题目理解度和完成时间。

邀请符合条件的学习者（HSK五级及以上或同等水平），建立标准化数据收集流程。明确问卷实施时间窗口（例如学期末连续两周）及完成方式，例如课堂监督下15—20分钟集中填写，或通过安全、限时的在线链接进行控制式发放。问卷发放前，教师或研究者须使用统一说明，解释研究目的、匿名原则及7点量表含义。

通过一次性访问令牌或等效一次性代码确保每位参与者只能提交一次；禁止重复提交，并通过令牌日志核查重复记录。如需匿名，应将令牌与问卷回答分开存储，并在分析前对登录标识进行哈希处理。数据清理前必须预先规定排除规则，如问卷未完成或令牌重复，以避免事后偏差。

记录用于亚组分析的非识别性元数据。学习者信息至少包括母语组、年龄、性别、汉语学习年限、当前HSK水平、专业、课程类型及研究地点；专家组则记录教龄、最高学历及专业领域（如文学教学、教育技术）。注意保护个人身份信息，并遵循获批的数据处理程序。

7. 计算隶属度（FCE）

定义五级评价集合：优秀、良好、中等、较差、差。为等级到指数的转换赋予数值，例如90、80、70、60、50，并在补充文件S3记录评分方案。

明确把Likert回答映射为等级隶属度的隶属函数（如梯形隶属函数），在S3中给出每个等级边界的具体参数，并说明所有指标是否采用同一参数。对每个指标，将每名学习者的回答转换为5元素隶属向量，再在学习者之间聚合（如取平均），得到指标层面的隶属向量；将全部指标向量堆叠成“指标×等级”的隶属矩阵。

8. 模糊合成并得到可持续性指数（SI）

采用明确的加权模糊合成算子（例如加权平均）把指标权重向量与指标隶属矩阵结合，得到综合隶属向量；在S3中报告算子名称和公式，以确保方案可独立复现。

通过等级分值的加权平均将综合隶属向量去模糊化，得到0—100分SI；在每个维度内部重复相同过程，计算维度SI。预设分类阈值：85—90=优秀；75—84=良好；65—74=中等；55—64=较差；≤54=差。暂停点：将中间向量、矩阵及计算表/脚本输出连同文件哈希值一并保存。

9. 可选的质性验证（混合方法）

按照SI模式进行目的性抽样，例如选择文化适应性高、低的教师和学习者，并为录音另行取得同意。使用半结构式访谈提纲，重点讨论分数的可解释性、可持续性的驱动因素及情境约束（补充文件S4）。

录音应逐字转写，并在编码前去除身份信息。采用预设编码方案进行编码，记录编码员培训、编码员间一致性（如适用）以及分歧解决规则。将质性主题与维度SI结果进行三角互证，用于解释优势与不足，但不得以质性结果替代定量程序。

10. 报告与归档（可复现性材料包）

报告维度权重、指标权重、隶属矩阵（或摘要）、SI、等级标签及亚组比较（如不同地点或语言水平组）。明确所有排除规则和最终分析样本量，并确保与摘要和结果部分一致。

归档用于FAHP权重计算、隶属度计算、模糊合成和去模糊化的匿名数据集、脚本和电子表格；提供软件/运行环境版本及固定随机种子（如使用）。推荐进行敏感性分析：将单个指标权重上下扰动10%，同时重新归一化其他权重，重新计算SI，并汇总SI波动以识别关键杠杆点。对“权重较低但敏感性较高”的指标，应视为潜在的阈值风险并明确报告。

结果

维度权重与关键指标

FAHP结果显示，技术维度(T)对可持续性的贡献最大，权重0.32；其次为文化维度(C,0.26)和教学维度(P,0.23)；政策维度(G)最低，为0.19。技术维度中，功能迭代效率(T3,0.25)和算法可解释性(T1,0.22)优先级较高；文化维度中，跨文化阐释适应性(C2,0.29)和文化负载词准确性(C1,0.27)最突出；教学维度中，情感目标达成(P3,0.24)高于知识目标达成(P1,0.21)，凸显数字文学教学中情感与阐释结果的重要性。政策维度中，数据隐私合规(G1,0.26)排名最高。

总体可持续性指数与地区差异

FCE计算得到总体SI=78.2，等级为“良好”。技术维度82.5、文化维度79.3表现较强，教学维度76.8、政策维度73.1相对较低。地区比较显示SI差异具有统计学显著性($F=4.37, p<0.05$)：华东（上海）最高81.7，华北（北京）79.5，华南（广州）最低72.3，为“中等”；事后检验表明华南校区与另外两地区差异显著($p<0.05$)。母语背景也与SI有关：英语母语学习者80.1，高于较少教授语种学习者73.6($t=3.22, p<0.01$)，其中阿拉伯语学习者最低71.2。

敏感性分析

研究通过将每个指标权重分别上下调整10%，并保持其他相对权重成比例，检验模型稳健性。算法可解释性(T1)导致最大SI波动($\pm 6.3\%$)，其次为跨文化适应性(C2, $\pm 5.7\%$)、情感目标达成(P3, $\pm 4.9\%$)和数据隐私合规(G1, $\pm 4.2\%$)。结果表明，在不确定条件下，可解释性和跨文化适应性是维持可持续性的高影响杠杆。

质性主题（可选三角互证）

对访谈文本和课堂视频记录进行质性编码后形成三个趋同主题。第一，情感共鸣稀释：数字工具提高效率，但可能降低文学作品被感知到的情感深度。第二，教师作为“策展者”：教师越来越多地筛选并情境化系统输出，以提升可解释性与文化适切性。第三，小语种语料稀缺：较少教授语种学习者报告更多误译和具有文化意义的误读，这与这些亚组较低的SI模式一致。

模型相关检验

研究检验了三个先验假设。H1认为技术维度是可持续性最关键的组成部分，结果支持该假设：技术维度权重最高(0.32, $p < 0.05$)。H2认为经济和教育资源不同的地区会出现显著不同的SI，该假设得到部分支持：华南校区显著低于其他地区，但北京与上海之间差异不显著。H3认为教师数字整合能力会正向影响总体可持续性，结果支持该假设：整合能力对可持续性具有显著调节作用($\beta = 0.31, p < 0.01$)。

图1：FAHP-FCE可持续性评价方案工作流程

该图概括从教学情境与指标体系界定、专家判断收集（FAHP）、根据问卷数据计算隶属矩阵（FCE），直至生成可持续性指数（SI）的全过程，并标出建议的暂停点，用于归档矩阵、中间向量和计算材料，以支持研究复现。

表1 FAHP-FCE可持续性评价框架的层级指标体系与权重

维度	代码	指标	操作性定义（摘要）	权重
技术(T,0.32)	T1	算法可解释性	学习者理解系统输出的能力	0.22
	T2	系统稳定性	因故障中断的教学时间比例	0.18
	T3	功能迭代效率	从需求提出到更新的平均天数	0.25
	T4	界面友好性	独立完成操作的平均分钟数	0.15
	T5	多终端适应性	跨设备完整性评分(1-7)	0.20
教学(P,0.23)	P1	知识目标达成	平均测试成绩提升率(%)	0.21
	P2	教学方法整合	教师对策略匹配度评价(1-7)	0.19
	P3	情感目标达成	学习者报告的共鸣程度(1-7)	0.24
	P4	师生互动	有效对话时间比例(%)	0.16
	P5	学习动机	额外阅读时间(小时/周)	0.20
文化(C,0.26)	C1	文化负载词准确性	误解率(%)	0.27
	C2	跨文化阐释适应性	教师评价(1-5)	0.29
	C3	意象转换	正确理解率(%)	0.23
	C4	多语种覆盖	对小语种学习者的有效性	0.21
政策(G,0.19)	G1	数据隐私合规	隐私政策完整性(1-7)	0.26
	G2	资源投入可持续性	3年投入增长率(%)	0.23
	G3	教师培训体系	年均培训小时数	0.20
	G4	教学评价适应性	包含数字工具的评价比例(%)	0.15
	G5	国际标准对接	专家合规率(%)	0.16

注：维度权重见括号；各指标权重为维度内部归一化权重。

表2 指标权重敏感性分析

指标代码	指标	权重扰动	SI波动(%)	影响排名
T1	算法可解释性	±10%	±6.3%	1
C2	跨文化阐释适应性	±10%	±5.7%	2
P3	情感目标达成	±10%	±4.9%	3
G1	数据隐私合规	±10%	±4.2%	4

表3 各地区/样本的维度可持续性指数

地区/样本	技术(T)	教学(P)	文化(C)	政策(G)	综合(S)	等级
总体	82.5	76.8	79.3	73.1	78.2	良好
华东（上海）	84.1	77.5	81.3	74.0	81.7	良好
华北（北京）	83.2	77.0	79.8	74.2	79.5	良好
华南（广州）	74.2	72.5	68.9	71.0	72.3	中等

表4 总体各维度SI与等级

维度	SI	等级
技术(T)	82.5	良好
教学(P)	76.8	良好
文化(C)	79.3	良好
政策(G)	73.1	中等
综合(S)	78.2	良好

补充文件

补充文件S1：学习者问卷。包含为可持续性评价框架开发的19项学习者问卷，覆盖技术、教学、文化和政策四个维度，并包括FCE分析所使用的作答量表与指标对应关系。

补充文件S2：专家两两比较表。包含FAHP阶段用于获取维度与指标两两判断的结构化专家评价表，以及权重推导所需的说明和评分锚点。

补充文件S3：评分方案、隶属函数与计算规则。给出FCE中的详细评分方案和隶属函数参数，包括评价等级、数值赋分、把7点Likert回答转换为模糊隶属度的梯形隶属函数，以及计算SI所使用的加权合成和去模糊化规则。

补充文件S4：质性验证半结构化访谈提纲。用于可选混合方法验证，问题涵盖技术功能、教学有效性、文化适应性、政策/机构支持，以及改善数字化国际中文文学教学长期可持续性的建议。

讨论

本方案表明，FAHP-FCE混合流程能够把数字化国际中文文学课程中的“教学可持续性”操作化为透明、可复现的指数，而非含义宽泛的结果变量。研究结果显示，技术功能构成可持续性的前提层，其中算法可解释性对SI波动最敏感。实践上，这符合一种阶段性逻辑：学习者和教师通常需要先获得稳定、可解释的系统行为，之后才可能持续开展更深层的文化参与和文学阐释。

第二个关键杠杆是文化适应性，对非英语母语及较少教授语种学习者尤为重要。质性主题有助于解释这一量化结果：当语料库和界面默认设置隐含地以英语为中心时，依赖其他语言路径的学习者更容易遇到误译或文化性误读，即使总体可用性尚可，也会降低其对可持续性的感知。四维结构的价值在于能够同时显示“技术提升”与“文化脆弱性”并存的权衡，而单一分数或纯认知结果指标可能掩盖这一点。

若要保证复现性，有三个环节必须视为不可妥协的检查点：第一，同一队列内指标体系必须保持稳定，数据收集前锁定每个题项的定义、正负向属性和维度映射；第二，专家判断必须进行明确的一致性检验，超过阈值时须记录协调过程；第三，FCE计算必须明确评价集合、隶属函数、合成算子和去模糊化规则，任何未经记录的改变都可能造成SI实质性变化。

讨论（续）

在真实部署中还需要故障排查和小幅调整。为减少专家输入噪声，应为三角模糊数提供锚点示例，并在正式判断前进行简短校准。为保护问卷数据完整性，应采用一次性访问代码确保每名学习者只提交一次，并预先规定不完整或重复记录的排除规则。对于隶属函数，应先进行小规模预测试，确认等级边界不会产生退化的隶属分布（例如大多数回答集中到单一等级）；若需调整，只能在最终分析前进行，并记录全部修改。

本方案存在若干局限。其一，依赖可信的专家组；若专家选择过窄或不透明，权重可能更多反映地方偏好，而非可迁移判断。其二，依赖可在不同情境中一致操作化的指标体系；当地方政策定义、平台功能或课程目标差异较大时，跨场域直接比较应先重新验证，而不能简单复用。其三，政策支持即使权重较低，也可能是“阈值风险”变量：一旦合规失败或资金中断，即使技术与教学表现良好，可持续性仍可能崩溃。

与传统精确值评价相比，FAHP-FCE的增益在于：权重假设可审计、能够表达不确定条件下的语言判断，并可归档权重、隶属矩阵、中间向量和SI等输出以支持复现。该流程还可扩展到其他数字媒介和跨文化教学场景：保留四维逻辑但调整指标表述；通过敏感性分析识别高影响杠杆；将SI与质性三角互证结合，避免对单一综合分数过度解释。未来还可开展跨多个学期的纵向实施，检验权重与隶属分布稳定性，并在同一数据集上与其他多准则模型比较。

利益披露

作者声明不存在可能影响本研究工作的竞争性经济利益或个人关系。人工智能（AI）工具仅用于稿件修订阶段的语言润色和语法编辑；作者对已发表作品的内容与完整性承担全部责任。

致谢

作者感谢玉林师范学院提供机构支持，感谢参与研究的教师和学生所投入的时间与贡献，并感谢专家组成员在FAHP程序中提供判断；同时感谢文学与传媒学院同事提出意见，使研究方案表述更加清晰。

基金项目

本研究获2025年度广西高校中青年教师科研基础能力提升项目“讲好广西故事提升广西留学生文化认同研究”（项目编号：2025KY0659）资助；同时获2025年度广西本科高校教学改革研究项目“新文科视域下面向东盟的国际中文教育课程体系重构与实践”（项目编号：2025JGB347）资助。

参考文献

为保证文献题名、期刊名、卷期、页码、DOI及作者信息的引用准确性，参考文献条目按原论文英文格式保留，不作中文化改写。以下文献编号与原文完全对应。

1. Ter ä s, M. Education and technology: Key issues and debates. *Int. Rev. Educ.* 68 (4), 635-636 (2022).
2. Haleem, A., Javaid, M., Qadri, M. A., Suman, R. Understanding the role of digital technologies in education: A review. *Sustain. Oper. Comput.* 3, 275-285 (2022).
3. Nussbaum, M. C. *Love's knowledge: Essays on philosophy and literature*. OUP USA, (1990).
4. Paran, A. The role of literature in instructed foreign language learning and teaching: An evidence-based survey. *Lang. Teach.* 41 (4), 465-496 (2008).
5. Biesta, G. Good education in an age of measurement: On the need to reconnect with the question of purpose in education. *Educ. Assess. Eval. Account.* 21 (1), 33-46 (2009).
6. Thelwall, M. Word association thematic analysis: A social media text exploration strategy. (2021).
7. Biggs, J., Tang, C., Kennedy, G. *Teaching for quality learning at university*. (2003).
8. Kidd, D. C. Castano, E. Reading literary fiction improves theory of mind. *Science*. 342 (6156), 377-380 (2013).
9. Paesani, K. Allen, H. W. Teacher development and multiliteracies pedagogy: Challenges and opportunities for postsecondary language programs. *Second Lang. Res. Pract.* 1 (1), 124-138 (2020).
10. Mishra, P. Koehler, M. J. Technological pedagogical content knowledge: A framework for teacher knowledge. *Teach. Coll. Rec.* 108 (6), 1017-1054 (2006).
11. Stockwell, G. 156-175, Springer. (2013).
12. Cohen, A., Soffer, T., Henderson, M. Students' use of technology and their perceptions of its usefulness in higher education: International comparison. *J. Comput. Assist. Learn.* 38 (5), 1321-1331 (2022).
13. Knowles, M. Kramsch, C. Context and culture in language teaching. *Mod. Lang. J.* 79 (4), 565 (1993).
14. Helm, F. The practices and challenges of telecollaboration in higher education in Europe. *Lang. Learn. Technol.* 19 (2), 197-217 (2015).
15. Peng, G., Han, L., Liu, Z., Guo, Y., Yan, J., Jia, X. An application of fuzzy analytic hierarchy process in risk evaluation model. *Front. Psychol.* 12, 715003 (2021).
16. Saaty, R. W. The analytic hierarchy process—what it is and how it is used. *Math. Modell.* 9 (3), 161-176 (1987).

17. Liu, Y., Eckert, C. M., Earl, C. A review of fuzzy AHP methods for decision-making with subjective judgements. *Expert Syst. Appl.* 161, 113738 (2020).
18. Dong, Z. Exploring cross-cultural communication content adaptability through advanced natural language processing and sentiment analysis. *Syst. Soft Comput.* 7, 200290 (2025).
19. Slade, S. Prinsloo, P. Learning analytics: Ethical issues and dilemmas. *Am. Behav. Sci.* 57 (10), 1510-1529 (2013).
20. Bera. Ethical guidelines for educational research. British Educational Research Association, (2011).
21. Ministry of Education of the P. R. C. Chinese proficiency grading standards for international Chinese language education (gf 0025 – 2021). Beijing Language and Culture University Press, (2021).
22. Koda, K. Insights into second language reading: A cross-linguistic approach. (Cambridge University Press, <https://doi.org/10.1017/CBO9781139524841> (2005).
23. Golden, B. L., Wasil, E. A., Harker, P. T. The analytic hierarchy process. *Appl. Stud.* 2 (1), 1-273 (1989).
24. Zadeh, L. A. Fuzzy sets. *Inf. Control.* 8 (3), 338-353 (1965).
25. Li, J., Xue, E., Guo, S. The effects of PISA on global basic education reform: A systematic literature review. *Humanit. Soc. Sci. Commun.* 12 (1), 106 (2025).
26. Sandve, G. K., Nekrutenko, A., Taylor, J., Hovig, E. Ten simple rules for reproducible computational research. *PLoS Comput. Biol.* 9 (10), e1003285 (2013).
27. Wilkinson, M. D. The fair guiding principles for scientific data management and stewardship. *Sci. Data.* 3 (1), 1-9 (2016).
28. Zadeh, L. A. The concept of a linguistic variable and its application to approximate reasoning—I. *Inf. Sci.* 8 (3), 199-249 (1975).
29. Klir, G. J., Yuan, B. Fuzzy sets and fuzzy logic: Theory and applications. Possibility Theory versus Probab. Theory. 32 (2), 207-208 (1996).
30. Shulman, L. Knowledge and teaching: Foundations of the new reform. *Harv. Educ. Rev.* 57 (1), 1-23 (1987).
31. Zhou, R. Chan, A. H. Using a fuzzy comprehensive evaluation method to determine product usability: A test case. *Work.* 56 (1), 21-29 (2017).
32. Ross, T. J. Fuzzy logic with engineering applications. John Wiley & Sons, (2005).
33. Opricovic, S. Tzeng, G. H. Defuzzification within a multicriteria decision model. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.* 11 (5), 635-652 (2003).
34. Ziegler, D. M. Fine-tuning language models from human preferences. *arXiv.* (2019).
35. Saaty, T. L. The analytic hierarchy process: Planning, priority setting, resource allocation. McGraw-Hill, (1980).
36. European, C. Ethics guidelines for trustworthy AI. (2021).
37. Artino Jr, A. R., La Rochelle, J. S., Dezee, K. J., Gehlbach, H. Developing questionnaires for educational research: Arme guide no. 87. *Med. Teach.* 36 (6), 463-474 (2014).
38. Zimmermann, H. J. Fuzzy set theory—and its applications. (2011).
39. Pedrycz, W., Gomide, F. Fuzzy systems engineering: Toward human- centric computing. John Wiley & Sons, (2007).
40. Mardani, A., Jusoh, A., Nor, K., Khalifah, Z., Zakwan, N., Valipour, A. Multiple criteria decision-making techniques and their applications – a review of the literature from 2000 to 2014. *Econ. Res.* 28 (1), 516-571 (2015).
41. Yager, R. R. On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Trans. Syst. Man Cybern.* 18 (1), 183-190 (2002).
42. Feng, S. Xu, L. D. Decision support for fuzzy comprehensive evaluation of urban development. *Fuzzy Sets Syst.* 105 (1), 1-12 (1999).
43. Creswell, J. W. Clark, V. L. P. Designing and conducting mixed methods research. Sage., (2017).
44. Saldaña, J. The coding manual for qualitative researchers. Sage. (2021).
45. O'Connor, C. Joffe, H. Inter coder reliability in qualitative research: Debates and practical guidelines. *Int. J. Qual. Methods.* 19, 1609406919899220 (2020).
46. Stodden, V., Leisch, F., Peng, R. D. Implementing reproducible research. 546, (2014).
47. Heil, B. J., Hoffman, M. M., Markowitz, F., Lee, S. I., Greene, C. S., Hicks, S. C. Reproducibility standards for machine learning in the life sciences. *Nat. Methods.* 18 (10), 1132-1135 (2021).
48. Yi, W., Huang, X., Kuzmin, S., Gerasimov, I., Luo, Y. Seekg: Sentiment analysis for e-learning evaluation incorporating knowledge graphs. *Educ. Inf. Technol.* 30 (12), 16291-16320 (2025).
49. Hoel, T. Chen, W. Privacy engineering for learning analytics in a global market: Defining a point of reference. *Int. J. Inf. Learn. Technol.* 36 (4), 288-298 (2019).
50. Field, A. Discovering statistics using IBM SPSS Statistics. (2024).
51. Cumming, G. Calin-Jageman, R. Introduction to the new statistics: Estimation, open science, and beyond. (2024).
52. Ruxton, G. D. The unequal variance t-test is an underused alternative to the Student's t-test and the Mann – Whitney U test. *Behav. Ecol.* 17 (4), 688-690 (2006).
53. Saltelli, A. Global sensitivity analysis: The primer. John Wiley & Sons, (2008).
54. Khosravi, H. Explainable artificial intelligence in education. *Comput. Educ. Artif. Intell.* 3, 100074 (2022).
55. Blodgett, S. L., Barocas, S., Daumé III, H., Wallach, H. Language (technology) is power: A critical survey of "bias" in NLP. *arXiv.* (2020).
56. Mukhametzhanov, I. Specific character of objective methods for determining weights of criteria in MCDM problems: Entropy, critic, and SD. *Decis. Mak. Appl. Manag. Eng.* 4 (2), 76-105 (2021).

57. Dillman, D. A., Smyth, J. D., Christian, L. M. Internet, phone, mail, and mixed-mode surveys: The tailored design method. (2014).
58. Williamson, B., Piattoeva, N. 81 – 93, Routledge. (2020).
59. Singer, J. D. Willett, J. B. Applied longitudinal data analysis: Modeling change and event occurrence. (2003).
60. Kazimieras Zavadskas, E., Antucheviciene, J., Adeli, H., Turskis, Z. Hybrid multiple criteria decision making methods: A review of applications in engineering. Sci. Iran. 23 (1), 1-20 (2016). Copyright © 2026,