Digital Intelligence PRESS

Journal of Digital Frontier

https://doi.org/10.67541/jdf2611

Research Article **Volume 1, Issue 2 (Sep 2026)** *Open Access*

Adaptive AHP-Entropy Combined Weighting Model for Teachers' Digital Literacy Assessment

Jie Zhang *

College of Education, Namseoul University, Cheonan, Republic of Korea

***Corresponding author:** Jie Zhang. Email: 19820632185@163.com

Abstract

To address the widespread reliance on equal-weight assumptions in the assessment of teachers' digital literacy and the lack of a mechanism for diagnosing disagreements between subjective and objective weighting, this study proposes a conflict-aware and adaptively balanced combined weighting model (CAW-AE). The model jointly measures structural conflicts between subjective and objective weights using a geometrically weighted combination of cosine similarity and rank displacement. Weighting states are classified into low-, medium-, and high-conflict intervals according to the degree of conflict. Differentiated indicator-level reconciliation coefficients incorporating a conflict penalty are then constructed. Their reconciliation intensity is determined by an analytical solution that minimizes overall deviation and is subject to truncation constraints ensuring a convex combination. Empirical results based on assessment data from 347 teachers and 20 indicators show a global conflict degree of 0.271, with conflict concentrated in digital social responsibility. CAW-AE has a Herfindahl–Hirschman index of 0.066 and a coefficient of variation of 0.334, occupying a justified position between the individual methods and a fixed-coefficient compromise. Tests show a rank reversal rate of 8.2% under ±20% perturbations of expert judgments, lower than the 14.8% of AHP alone and the 9.7% of game-theoretic combination. Its group discrimination (Kruskal-Wallis H = 33.76) is the highest among the six schemes. Its rank correlation with the external criterion is 0.428, significantly higher than those of equal weighting and entropy weighting alone. These results indicate that CAW-AE improves ranking stability and differentiation between groups without sacrificing the reasonable contribution of domain knowledge.

Keywords

Teachers' digital literacy; Combined weighting; Conflict diagnosis; Adaptive reconciliation; Ranking stability

Received: 25 Aug 2026; Revised: 12 Sep 2026; Accepted: 19 Oct 2026; Published: 29 Oct 2026
Copyright: ©2026 The Author(s). Published by Digital Intelligence Press Limited. This is an open access article under the CC BY 4.0 license.

1. INTRODUCTION

Research on the assessment of teachers' digital literacy is undergoing a subtle but far-reaching shift in emphasis [1],[2]. Over the past decade and more, academic and policy attention has focused mainly on "what to assess" and "what to use for assessment." The former has generated numerous competency frameworks and dimensional classifications, while the latter has produced diverse tools, including self-assessment scales, certification examinations, and behavioral observation. However, as assessment frame proliferates and tools proliferate, a less visible methodological issue has emerged: these tools generally rely on equal-weight assumptions or a single subjective weighting method when assigning indicator weights. Self-assessment scales often simply sum or average dimension scores; certification

examinations frequently substitute item counts or point allocations for a justification of weights; and behavioral observation often relies on observers' experience to allocate attention [3]. Arbitrary weight assignment contrasts sharply with the sophistication of assessment frameworks. Assessment "precision" remains confined to indicator selection and has not extended to the structure of weights.

The essence of this weighting dilemma is that equal weighting systematically distorts assessment validity rather than merely reducing numerical precision. Different dimensions of teachers' digital literacy contribute heterogeneously to overall literacy performance. The behavioral observability and variability of digital application capabilities are generally substantially greater than those of digital awareness or legal and ethical norms [4],[5]. Yet the latter dimensions have a conceptually foundational status and should not simply be downgraded. The equal-weight assumption conceals substantive inequity behind formal equality: it mistakes "equally important conceptually" for "equally contributive in measurement," pushing assessment conclusions toward the mean and weakening their ability to identify differences within the teacher population [6],[7],[8]. When assessment results inform teacher development diagnosis, training needs identification, or professional development planning, an inappropriate weight structure propagates downstream through the assessment process, causing subsequent decisions to rest on a smoothed literacy profile. In this sense, combined weighting is no longer a technical detail but a methodological bottleneck in moving the assessment of teachers' digital literacy from existing to being effective.

The evolution of combined weighting in multicriteria decision-making offers methodological resources for addressing this bottleneck, but its development also reveals a clear gap in application contexts. Research combining AHP and entropy weighting has accelerated since 2021, yet it remains heavily concentrated in engineering selection, energy system evaluation, and management performance ranking [9],[10],[11]. These fields share relatively stable indicator systems and assessment objects whose heterogeneity is expressed primarily through indicator values rather than indicator structures. Combined weighting can therefore be satisfied with a single global compromise coefficient. Educational assessment differs in indicator dimensions embody explicit value orientations, expert judgments incorporate domain knowledge, and sample data contain variation that framework designers may not have anticipated. Disagreements between the two are often substantive structural conflicts rather than random noise. Nevertheless, most educational applications remain at the level of simple linear weight, adding subjective and objective weights using a fixed coefficient without diagnosing the nature of disagreement or differentiating combination strategies. Such "spliced" combinations evade the central question: when expert judgments the combination should contradictory signals in a dimension, toward which source should the combination shift, by how much, and on what grounds?

This study addresses that gap through CAW-AE, whose distinguishing feature is the progression from static weighting to conflict-driven adaptive reconciliation. Rather than seeking global balance between subjective and objective weights, the model uses joint conflict diagnosis to distinguish numerical deviation from ranking misalignment and identify the uneven distribution of conflict across indicators. It then determines differentiated indicator-level reconciliation coefficients, allowing combined weights in conflict-intensive dimensions to shift moderately toward the more informative source. Correspondingly, the study constructs a hierarchical indicator system for teachers' digital literacy, using the five-dimensional framework in the Ministry of Education standard as its foundation and introducing ethical responsibility and data-driven decision-making as cross-cutting components. "Dual-channel signal identifiability" serves as the screening criterion, excluding items that provide no effective differentiation through either expert judgments or data variation. Indicator system construction serves the input requirements of the combined weighting model rather than conceptual exhaustiveness.

For validation, expert judgment perturbation experiments, bootstrap resampling, group discrimination tests, and external criterion association tests assess CAW-AE against existing combined weighting methods in ranking stability, differentiation, and criterion-related validity. Confirmatory evidence is explicitly distinguished from directional evidence, and rankings on a single statistic do not replace cautious judgments about the strength of support.

The paper follows progression from indicators to the model and from weighting to validation. Section 2 constructs the indicator system and describes data sources and preprocessing, providing model inputs that satisfy dual-channel identifiability. Section 3 develops CAW-AE through basic weighting submodules, quantitative diagnosis of subjective-objective weight conflicts, adaptive reconciliation, and model property analysis. Conflict diagnosis and differentiated indicator-level reconciliation coefficients constitute its central innovations. Section 4 applies the model to real data, presents weighting results, compares weight structures with equal weighting, AHP alone, entropy weighting alone, fixed-coefficient combination, and game-theoretic combination, and interprets the educational assessment implications of reconciliation. Section 5 organizes evidence around ranking stability, differentiation, and external criterion association to determine whether CAW-AE is more suitable for assessing teachers' digital literacy than existing combined weighting methods. Section 6 summarizes model contributions, applicability boundaries, and limited implications for assessment practice, and identifies dynamic weighting and transferability tests of the conflict measure as future directions. "Weight ability" connects the sections: indicator screening establishes whether weighting is possible, model construction establishes justified weighting, validation establishes effective weighting, and conclusions establish its boundaries.

2. CONSTRUCTION OF THE ASSESSMENT INDICATOR SYSTEM AND DATA ACQUISITION

2.1 Indicator sources and screening criteria

The indicator system is based on the five-dimensional framework established by the Ministry of Education's education industry standard, Teachers' Digital Literacy: digital awareness, digital technology knowledge and skills, digital application, digital social responsibility, and professional development [12],[13]. Its strengths include a clear hierarchy, standardized terminology, and broad recognition within education digitalization policy. However, its secondary and tertiary dimensions tend to describe capabilities rather than measurable constructs. For example, "digital resolve" is defined as "the belief in overcoming difficulties when facing issues in education digitalization." This formulation can differentiate expert judgments but provides almost no stable variation extractable from existing assessment records in the sample-data channel. Similarly, "legal and ethical norms" under digital social responsibility is conceptually indispensable, yet teacher scores on this dimension often cluster at the high end, producing an extremely low coefficient of variation and leaving entropy weighting unable to extract useful information.

To address these structural shortcomings, this section adds two cross-cutting components from international frameworks to supplement weakly quantifiable areas in the Ministry standard. The first operationally extends ethical responsibility. Drawing on the EU DigCompEdu framework's definition of teachers' professional responsibility for learners' digital well-being, it translates this responsibility into observable behavioral indicators. The second is data-driven decision-making, drawing on the UNESCO AI competency framework for teachers, which positions data literacy as a prerequisite for

interpreting learning analytics and adjusting teaching [14],[15],[16],[17]. Both components can generate meaningful differences in expert preferences because they involve value trade-offs in professional judgment. They also provide sufficient cross-sample variation because teachers differ observably in their use of data and frequency of ethical decision-making in actual teaching.

Screening uses dual-channel signal identifiability as its core operational standard, implemented through two exclusion rules. First, an indicator is excluded for lacking subjective-channel differentiation if more than 80% of experts rate it as "equally important" in pilot pairwise comparisons. Second, an indicator is excluded from failing to provide useful information for entropy calculation if its sample coefficient of variation is below 0.15 or its absolute kurtosis exceeds 2.0, indicating excessive concentration or extreme clustering. This dual-channel screening reduces the initial pool from 31 candidate indicators to 20 retained indicators, a reduction of 35.5%. Of the 11 excluded indicators, 7 lack subjective-channel differentiation, concentrated in the digital cognition and legal and ethical norms subdimensions, while 4 have insufficient objective-channel variation, concentrated in digital willingness. This result is itself a noteworthy methodological signal: the conceptually most foundational dimensions of teachers' digital literacy, awareness and responsibility, are precisely those from which weighting models find it most difficult to extract effective weights.

Table 1. Indicator distributions before and after dual-channel screening

Criterion dimension	Candidate indicators	Retained indicators	Main reason for exclusion
Digital awareness	6	3	No subjective-channel differentiation (4 items)
Digital technology knowledge and skills	5	4	Insufficient objective-channel variation (1 item)
Digital application	8	7	Insufficient objective-channel variation (1 item)
Digital social responsibility	6	3	No subjective-channel differentiation (3 items)
Professional development	4	2	Insufficient objective-channel variation (2 items)
Cross-cutting items (ethical responsibility + data-driven)	2	1	No subjective-channel differentiation (1 item)

Table 1 shows that exclusions are unevenly distributed across the five criterion-level dimensions. Digital applications have the highest retention rate, 7/8 = 87.5%, whereas digital awareness and digital social responsibility each retain only 50%. This imbalance reflects substantive differences in weight ability rather than screening bias. Application indicators naturally exhibit observable variation in behavioral frequency and quality levels, while awareness and responsibility indicators tend to converge toward normative expectations in standardized assessment.

2.2 Establishment of the hierarchical indicator system

The final system has three hierarchical levels. The goal is the comprehensive assessment of teachers' digital literacy; the criterion level contains six dimensions, comprising the Ministry's five dimensions and an integrative cross-cutting dimension; and the indicator level contains 20 operationalizable items. The sixth criterion-level dimension, ethics-data literacy, merges the retained cross-cutting indicators of data-driven instructional adjustment and care for learners' digital ethics. The

merger is justified by their significant positive correlation in pilot data, Spearman $\rho = 0.47$, $p < 0.01$, and by stronger consistency of expert agreement after their combination than when they were considered separately.

Each indicator-level item has an operational definition and a specified scoring direction. Definitions follow behavioral anchoring: observable or retrospectively traceable events, rather than abstract capability statements, anchor measurement. For example, the frequency of producing technology-integrated teaching plans under digital instructional design is defined as "the number of lesson plans produced in the past semester whose instructional design documents explicitly specify stages involving digital tools and expected learning outcomes." It is a benefit-type indicator: larger values indicate stronger literacy performance. The number of student data exposure risk incidents under digital security protection is defined as "the number of incidents in the past academic year in which teachers' improper actions made students' personal information visible or accessible in unauthorized settings." It is a cost-type indicator: smaller values indicate stronger literacy performance. Of the 20 indicators, 17 are benefit-type and 3 are cost-type. All cost-type indicators fall under digital social responsibility, consistent with this dimension's emphasis on minimum compliance rather than higher-level excellence.

Table 2. Correspondence between criterion dimensions and indicators and their scoring directions

Criterion level	Indicators	Benefit-type	Cost-type	Representative indicator
Digital awareness	3	3	0	Frequency of proactive attempts to use technology in teaching
Digital technology knowledge and skills	4	4	0	Proficiency rating for subject-specific tools
Digital application	7	7	0	Frequency of producing technology-integrated teaching plans
Digital social responsibility	3	0	3	Student data exposure risk incidents
Professional development	2	2	0	Cross-school dissemination of digital teaching research outcomes
Ethics-data literacy	1	1	0	Recorded instructional adjustments based on learning data

The scoring directions in Table 2 constrain subsequent weighting. Cost-type indicators must be reversed during min-max normalization. Applying the same formula to cost- and benefit-type indicators without distinguishing them would systematically overestimate this dimension's information entropy in entropy weighting.

2.3 Data collection and preprocessing

The study requires two independent data sources that share the same indicator structure. Neither involves interventional measurement of human participants or collection of identifiable personal information. The subjective channel obtains pairwise comparisons for AHP through consultation with domain experts [18]. The objective channel performs secondary analysis of existing indicator-level

score records held by a regional teacher development institution.

The expert consultation group was established through the disciplinary platform of the researchers' institution. Following the principle of domain coverage rather than simply maximizing numbers, 12 professionals were invited as methodological consultants: 4 educational technology researchers, all with experience in digital literacy assessment; 3 teacher training administrators responsible for regional digital literacy improvement program design; 3 frontline teaching administrators experienced in advancing digital teaching; and 2 educational measurement researchers providing methodological calibration of judgment matrix consistency. The group size of 12 reflects the degrees-of-freedom constraints of AHP pairwise comparison matrices: consultation must adequately cover sources of heterogeneity in criterion-level judgments without introducing so many participants that the consistency ratio of the aggregated matrix deteriorates. Experts provided formative judgments on indicators' relative importance in their professional capacity. Their views formed construct weights rather than measurements of any individual. No identifiable expert information was collected, and judgments entered the analysis as anonymous aggregates. Each expert compared the six criterion dimensions pairwise, producing 15 comparisons, and compared indicators within each criterion group. The Saaty 1–9 scale was used. Individual judgment matrices were aggregated by geometric averaging, after which consistency ratios were calculated. CR values for the six criterion-level judgment matrices ranged from 0.021 to 0.067, all below the acceptance threshold of 0.10. Digital application had the lowest CR, 0.021, indicating the strongest expert preference consensus. Digital social responsibility was the highest, 0.067, reflecting tolerable but clearly greater disagreement than in other dimensions over how much weight cost-type indicators should receive. This disagreement pattern is consistent with the anticipated direction of conflict diagnosis in subsequent combined weighting.

Objective data consists of an existing indicator score matrix from a regional assessment of teachers' digital literacy. The sample comprises n = 347 teachers across primary, junior secondary, and senior secondary education, with 2-31 years of teaching experience. These are not questionnaire data collected specifically for this study, but indicator-level archives produced during routine assessment by a regional teacher development institution and accessed for secondary analysis. Before delivery, data were de-identified and contained no teacher names, identity codes, or information traceable to individuals. The research team had no direct contact with the data subjects. Consequently, these data do not involve the common method bias and self-report inflation inherent in questionnaire surveys, and their generation was independent of this study's hypotheses. Among the 20 indicators, the 17 benefit-type indicators have continuous or quasi-continuous raw scores, with frequency counts natural-log transformed before use; the 3 cost-type indicators are event counts.

Preprocessing proceeds in three steps. First, indicator directions are unified: the 3 cost-type indicators are negated and shifted into a nonnegative range so that higher values consistently indicate better performance across all 20 indicators. Second, the directionally unified indicators undergo standard min–max normalization:

$$y_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)} \tag{1}$$

Here, x_{ij} is the directionally unified value of teacher i on indicator j; $\min(x_j)$ and $\max(x_j)$ are its minimum and maximum across all 347 samples; and $y_{ij} \in [0, 1]$ is the normalized score. The third step detects and handles outliers. The interquartile range (IQR) of normalized scores is calculated separately for each indicator, and observations outside $[Q, 1 - 3IQRQ3 + 3IQR]$ are classified as

extreme. Each of the 3 cost-type indicators has 1-2 extreme high observations, corresponding to unusually frequent risk incidents, for a total of 5 flagged observations. These are winsorized rather than deleted, because deletion would discard valid information on the teacher's other 19 indicators. Extreme values are replaced with the boundary values at $Q3 + 3IQR$ for the corresponding indicator. After winsorization, absolute skewness is below 1.2 for all indicators, meeting the basic distributional requirements of subsequent entropy calculations.

Table 3. Key distributional characteristics after preprocessing

Indicator category	Indicators	Normalized mean range	Coefficient of variation range	Winsorized observations
Benefit-type (frequency)	8	0.41–0.63	0.28–0.51	0
Benefit-type (rating)	9	0.52–0.71	0.19–0.38	0
Cost-type (event count)	3	0.73–0.89	0.42–0.58	5

Table 3 reveals a distributional characteristic directly relevant to weighting: normalized means for cost-type indicators are systematically high, 0.73–0.89, and their coefficients of variation, 0.42-0.58, are the highest among the three indicator categories. Entropy weighting will therefore assign relatively high objective weights to the three cost-type indicators in digital social responsibility because of their lower entropy and larger between-sample differences. In AHP, however, subjective weights for this dimension may be suppressed by disagreement over whether minimum requirements should receive the same weight as excellent performance.

3. DEVELOPMENT OF THE CAW-AE COMBINED WEIGHTING MODEL

3.1 Basic weighting submodules

The subjective AHP submodule takes the three-level hierarchy established in Section 2 as input. Suppose there are m criterion dimensions and n indicators under a criterion. Experts compare elements at the same level pairwise to construct judgment matrix $A = (a_{kl})_{n\times n}$, where a_{kl} expresses the importance of element k relative to element l. Judgments use the Saaty 1–9 scale and satisfy $a_{lk} = 1/a_{kl}$ and $a_{kk} = 1$. After individual matrices are aggregated by geometric averaging, the eigenvalue equation is solved:

$$Aw^A = \lambda_{\max} w^A \tag{2}$$

Here, $w^A = \left(w_1^A, w_2^A, \ldots, w_n^A\right)^{\mathsf{T}}$ is the normalized eigenvector associated with the largest eigenvalue $\lambda_{\max}$, giving the subjective AHP weights, and satisfies $\sum_{k=1}^{n} w_k^A = 1$ and $w_k^A > 0$. Consistency is tested using the consistency ratio:

$$CR = \frac{CI}{RI},\ CI = \frac{\lambda_{\max} - n}{n-1} \tag{3}$$

Here, RI is the random consistency index, obtained from a table according to matrix order n.

When $CR < 0.10$, matrix consistency is acceptable and the eigenvector provides a valid subjective weighting input. Otherwise, experts must revise the matrix. The criterion-level CR values reported in Section 2, 0.021–0.067, all satisfy this requirement. Accordingly, consistency testing details are not repeated here; consistency serves as a prerequisite for valid subjective weights.

The objective entropy weighting submodule takes the preprocessed normalized score matrix $Y = (y_{ij})_{N\times n}$ from Section 2 as input. Here, N is the number of teachers, n is the number of indicators, and $y_{ij} \in [0,1]$ is the min-max normalized score of teachers i on indicator j. First, each sample's score is expressed as a proportion of the total score for indicator j:

$$p_{ij} = \frac{y_{ij}}{\sum_{i=1}^{N} y_{ij}} \tag{4}$$

The information entropy of indicator j is then calculated:

$$e_j = -\frac{1}{\ln N}\sum_{i=1}^{N} p_{ij} \ln p_{ij} \tag{5}$$

By convention, $p_{ij} \ln p_{ij} = 0$ when $p_{ij} = 0$. Information entropy satisfies $e_j \in [0,1]$; smaller values indicate larger between-sample differences and greater information content. The divergence coefficient $d_j = 1 - e_j$ is derived from entropy and normalized to obtain objective weights:

$$w_j^E = \frac{d_j}{\sum_{k=1}^{n} d_k} \tag{6}$$

Here, $w^E = (w_1^E, w_2^E, \ldots, w_n^E)^\top$ is the entropy weight vector, satisfying $\sum_{j=1}^{n} w_j^E = 1$ and $w_j^E > 0$. The two basic submodules thus output subjective weight vector w^A and objective weight vector w^E, which form the input pair for subsequent conflict diagnosis.

3.2 Quantitative diagnosis of subjective–objective weight conflicts

Conflict diagnosis does not determine whether AHP or entropy weighting is more correct. It identifies the indicators on which weights substantively disagree and whether the disagreement concerns numerical values or rankings. This section defines conflict through two components: cosine similarity captures directional deviation between weight vectors, while rank displacement captures structural misalignment in indicator importance rankings. Cosine similarity alone cannot sufficiently distinguish numerical fine-tuning disagreements from ranking-reversal disagreements. Directional deviation may be small, yet ranking changes can substantially affect assessment conclusions. Conversely, when weights differ numerically but produce identical rankings, their conflict has limited substantive implications for assessment practice [19]. A joint measure is therefore necessary. The cosine similarity component is defined as:

$$S_c = \frac{w^A \cdot w^E}{\| w^A \|_2 \cdot \| w^E \|_2} \tag{7}$$

Here, $\|\cdot\|_2$ is the Euclidean norm. $S_c \in (0,1]$, with values closer to 1 indicating greater directional agreement. The rank displacement component uses descending ranks of the n indicators

under the two weight vectors. Let r_j^A and r_j^E be the descending subjective and objective ranks of indicator j, respectively, where smaller ranks indicate larger weights. Normalized rank displacement is defined as:

$$S_r = 1 - \frac{1}{\frac{n(n^2-1)}{6}} \sum_{j=1}^{n} \left(r_j^A - r_j^E\right)^2 \tag{8}$$

This expression is effectively equivalent to Spearman rank correlation in the context of weight rankings. The denominator $n(n^2-1)/6$ is the maximum possible sum of squared rank deviations, so $S_r \in [0,1]$; values closer to 1 indicate more consistent ranking structures. Joint conflict is defined through the weighted geometric meaning of the two components:

$$C = 1 - (S_c)^{\alpha}(S_r)^{1-\alpha} \tag{9}$$

Here, $\alpha \in (0,1)$ denotes the relative emphasis on directional consistency versus ranking consistency. We set $\alpha = 0.5$, giving equal emphasis to numerical deviation and ranking misalignment in conflict diagnosis. $C \in [0,1]$, with larger values indicating more severe structural conflict. Compared with cosine distance alone, the joint measure identifies a special case: numerical weights are close, with high S_c, but particular indicators undergo critical rank reversals, with low S_r. Joint conflict still provides a strong signal, whereas cosine similarity alone would misclassify it as low conflict.

Conflict thresholds are calibrated through simulation rather than arbitrarily chosen from experience. Many weight pairs satisfying normalization are generated randomly in a weight space with $n = 20$. Each pair's joint conflict C is calculated together with the rate at which it produces different assessment conclusions in simulated evaluation settings. A Kendall rank correlation below 0.90 between rankings of weighted scores defines a substantive difference in conclusions. When C is below 0.18, fewer than 5% of weight pairs produce such differences. When C exceeds 0.32, the proportion rises sharply above 40%. Accordingly, two boundaries define three intervals. $C \leq 0.18$ indicates low, reconcilable conflict, where fixed-coefficient combination can be used without a conflict penalty. $C > 0.32$ indicates high conflict requiring correction, where adaptive reconciliation must be activated. Between them, $0.18 < C \leq 0.32$ is the medium-conflict interval, in which reconciliation intervenes moderately. This quantitative classification follows turning points in the simulated relationship between conclusion differences and conflict, rather than a priori assumption.

Table 4. Conflict intervals and corresponding reconciliation strategies

Conflict interval	Interval label	Simulated conclusion difference rate	Reconciliation strategy	Initial reconciliation intensity
$C \leq 0.18$	Low conflict	<5%	Fixed-coefficient combination without penalty	$\gamma = 0$
$0.18 < C \leq 0.32$	Medium conflict	5%–40%	Moderate reconciliation; linear penalty intervention	$\gamma \in (0, 0.5]$
$C > 0.32$	High conflict	>40%	Strong reconciliation; penalty dominates shift direction	$\gamma \in (0.5, 1]$

The ranges of reconciliation intensity parameter γ in Table 4 correspond to conflict intervals.

3.3 Adaptive reconciliation mechanism

Traditional game-theoretic combined weighting treats the combination of subjective and objective weights as a game between two players and uses linear combination coefficients satisfying Nash equilibrium as its solution [20],[21],[22]. It implicitly assumes equal influence for the two vectors and seeks balance between them. This assumption is not held in the assessment of teachers' digital literacy. Structural conflict in a particular dimension itself conveys information about which source is more credible. If conflict occurs in indicators with substantial data variation but weak expert consensus, objective weights contain more information. If it occurs where expert judgments are highly consistent, but sample variation is insufficient, the domain knowledge underlying subjective weights is more credible. CAW-AE therefore does not seek equal balance. When conflict exceeds a threshold, it introduces a conflict penalty that shifts combined weights moderately toward the more informative source.

Let $w^C = \left(w_1^C, \ldots, w_n^C\right)^\top$ be the combined weight vector, whose elements are weighted combinations of subjective and objective weights:

$$w_j^C = \lambda_j w_j^A + \left(1 - \lambda_j\right) w_j^E \tag{10}$$

Unlike traditional fixed-coefficient combination, λ_j is an indicator-level reconciliation coefficient rather than a global constant. It depends jointly on the indicator's local conflict state and global conflict. Define the local conflict contribution of indicator j as:

$$c_j = \frac{\left|w_j^A - w_j^E\right|}{\sum_{k=1}^{n} \left|w_k^A - w_k^E\right|} \tag{11}$$

c_j describes indicator j's share of total conflict and satisfies $\sum_{j=1}^{n} c_j = 1$. Reconciliation coefficient λ_j incorporates a conflict penalty $\gamma \cdot C \cdot c_j$ of the form:

$$\lambda_j = \lambda_0 + \gamma \cdot C \cdot c_j \cdot \operatorname{sgn}\left(w_j^E - w_j^A\right) \tag{12}$$

Here, λ_0 is the baseline reconciliation coefficient, set to $\lambda_0 = 0.5$, giving equal subjective and objective contributions in the absence of conflict; γ is the reconciliation intensity parameter; C is the joint conflict defined in Section 3.2; and $\operatorname{sgn}(\cdot)$ is the sign function. When indicator j has a larger objective than subjective weight, $w_j^E > w_j^A$, the penalty increases λ_j and shifts the combined weight toward the objective source. The opposite case shifts it toward the subjective source. The shift magnitude depends jointly on global conflict C, local conflict contribution c_j, and reconciliation intensity γ. When $C \leq 0.18$, $\gamma = 0$ is used, giving $\lambda_j \equiv \lambda_0 = 0.5$; the model reduces to an equally weighted linear combination without conflict penalization.

Reconciliation intensity γ is not specified subjectively. Instead, a constrained optimization problem minimizes the overall deviation of combined weights from the two basic vectors. Define the overall deviation function:

$$D(\gamma) = \sum_{j=1}^{n} \left(w_j^C(\gamma) - w_j^A\right)^2 + \sum_{j=1}^{n} \left(w_j^C(\gamma) - w_j^E\right)^2 \tag{13}$$

Substituting the expression for $w_j^C(\gamma)$ and rearranging gives $D(\gamma)$ as a quadratic function of γ.

Differentiating with respect to γ and setting the derivative to zero yields the analytical solution:

$$\gamma^* = \frac{\sum_{j=1}^{n} \left(w_j^A - w_j^E\right) \left[\lambda_0 w_j^A + (1-\lambda_0) w_j^E - \frac{w_j^A + w_j^E}{2}\right]}{\sum_{j=1}^{n} \left(w_j^A - w_j^E\right)^2 \cdot C \cdot c_j \cdot \mathrm{sgn}\left(w_j^E - w_j^A\right)} \tag{14}$$

This solution shows that γ^* depends on the direction and magnitude of subjective–objective weight differences, local conflict distribution, and global conflict rather than being an arbitrarily assigned constant. To keep reconciliation coefficients λ_j within feasible domain $[0, 1]$, truncation constraints are imposed on γ^*:

$$\gamma^{**} = \min\{\max\{\gamma^*, 0\}, \gamma_{\max}\} \tag{15}$$

Here, $\gamma_{\max}$ is the upper bound inferred from $\lambda_j \in [0, 1]$. Truncation keeps combined weights within the convex combination of the subjective and objective weights, avoiding extrapolation risk.

The algorithm takes the subjective AHP vector, objective entropy vector, indicator count n, and baseline reconciliation coefficient λ_0 as inputs. First, the cosine similarity, rank displacement, and joint conflict C are calculated using Section 3.2. Second, the interval containing C determines the initial range of reconciliation intensity: low conflict directly outputs a fixed-coefficient combination and terminates, while medium and high conflict continue. Third, each indicator's local conflict contribution c_j is calculated. Fourth, overall deviation minimization yields γ^*, and truncation gives γ^{**}. Fifth, indicator-specific coefficients λ_j are calculated, followed by combined weights w^C. Sixth, normalization and nonnegativity constraints are checked before results are output. Computational complexity is of order $O(n)$, excluding AHP eigendecomposition and entropy preprocessing, and is negligible relative to the scale of subsequent assessment applications.

3.4 Model properties

Bounded reconciliation coefficients are a prerequisite for practical weighting. The truncation constraints in Section 3.3 restrict γ^{**} to $[0, \gamma_{\max}]$, while the choice of $\gamma_{\max}$ ensures that $\lambda_0 + \gamma^{**} \cdot C \cdot c_j \cdot \mathrm{sgn}(\cdot)$ remains within $[0, 1]$ for any indicator j. Because $C \in [0, 1]$, $c_j \in [0, 1]$, and $\sum_j c_j = 1$, local conflict contribution has an upper bound of 1, attained when all conflict is concentrated in one indicator. The corresponding minimum of $\gamma_{\max}$ is $0.5/(C \cdot c_j)$. This upper bound ensures that combined weight w^C is always a convex combination of w^A and w^E; that is, w^C lies on or within the line segment connecting the two basic weight vectors and cannot be driven outside the convex combination region by penalization.

Robustness to perturbations in basic weights can be described at two levels. In the low-conflict interval, $\gamma = 0$, combined weights reduce to a fixed-coefficient combination with $\lambda_0 = 0.5$. Sensitivity to perturbations in w^A or w^E is directly bound by the fixed coefficient, and perturbations propagate linearly. Under medium or high conflict, coefficient λ_j itself depends on the difference between basic weights. Perturbations affect combined weights directly by changing the values being combined, and indirectly by changing conflict C, local contribution c_j, and consequently the reconciliation coefficient. This indirect path makes the response self-regulating rather than a simple linear amplification. When perturbations enlarge subjective–objective differences, conflict and

penalization increase, strengthening the shift toward the more informative source and partially offsetting imbalance in the weight structure. This self-regulation defines the conditions for CAW-AE's robustness advantage over fixed coefficients: it is pronounced when perturbations mainly affect high-conflict indicators but weakens when all indicators are perturbed uniformly without changing the conflict structure, in which case behavior approaches fixed-coefficient combination.

The fundamental distinction from fixed-coefficient linear combination can be illustrated by a limiting case. A fixed-coefficient method expresses combined weights as:

$$w^C = \lambda_0 w^A + (1 - \lambda_0) w^E \tag{16}$$

Here, λ_0 is a global constant that varies with neither indicator nor conflict state. In CAW-AE, setting joint conflict $C \equiv 0$, corresponding to identical subjective and objective weights, makes the penalty zero and gives $\lambda_j \equiv \lambda_0$ for every j. The combined weight expression then reduces to the fixed-coefficient form. Fixed-coefficient linear combination is therefore a degenerate special case of CAW-AE when conflict is identically zero. The converse does not hold. When $C > 0$, indicator-level coefficients λ_j vary with c_j and sign direction, and the combined weights are not equivalent to a linear combination with any single fixed coefficient.

4. EMPIRICAL WEIGHTING AND ANALYSIS OF RESULTS

4.1 Experimental setup

The dataset is the regional teacher digital literacy indicator score matrix described in Section 2, with sample size $N = 347$ and indicator count $n = 20$. It covers primary, junior secondary, and senior secondary education, with teaching experience ranging from 2 to 31 years. After min-max normalization and winsorization, all 20 indicators lie within $[0, 1]$; the 17 benefit-types and 3 cost-type indicators have been directionally unified. The sample contains 142 primary, 118 junior secondary, and 87 senior secondary teachers. These unequal group sizes reflect the actual structure of the regional teaching workforce and were not artificially balanced.

Five comparison methods are used: equal weighting, with $1/n$ for every indicator; AHP alone; entropy weighting alone; fixed-coefficient linear combination with $\lambda = 0.5$; and game-theoretic combined weighting. The game-theoretic approach is the most directly comparable existing combination method. Its combination coefficient is obtained from first-order conditions minimizing deviation between subjective and objective weights, but it includes neither conflict diagnosis nor adaptive indicator-level adjustment. All five methods and CAW-AE share the same indicator system and normalized data matrix; only their weight generation mechanisms differ.

Two descriptive measures characterize weight distributions. Concentration is measured by the Herfindahl–Hirschman index (HHI) [23],[24]:

$$HHI = \sum_{j=1}^{n} \left(w_j\right)^2 \tag{17}$$

Here, w_j is the weight of indicator j. Under equal weighting, $HHI = 1/n = 0.05$; larger HHI values indicate greater concentration. Dispersion is measured by the coefficient of variation:

$$CV = \frac{\sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(w_j - \bar{w}\right)^2}}{\bar{w}} \tag{18}$$

Here, $\bar{w} = 1/n$ is the mean weight. Larger CV values indicate greater dispersion around the mean. The two measures describe different aspects of weight structure. HHI is sensitive to clustering among high-weight indicators, whereas CV captures overall dispersion. Together, they distinguish structures dominated by a few indicators from those with broadly differentiated weights.

4.2 Basic AHP and entropy weighting results

Among the six criterion dimensions, digital application receives the highest subjective weight, followed by knowledge and skills. Digital awareness and professional development occupy intermediate positions, while digital social responsibility and ethics–data literacy receive relatively low subjective weights. This pattern is consistent with experts' domain experience: application capabilities are regarded as the core expression of teachers' digital literacy, while responsibility dimensions are considered necessary but nondiscrimination. At the indicator level, all three highest subjective weights belong to digital application, while two of the three lowest belong to digital social responsibility.

Objective entropy weighting produces a markedly different distribution. Two of the three highest-weight indicators are cost-type indicators in digital social responsibility, and the third is the frequency of cross-school dissemination in professional development. Cost-type indicators receive high objective weights because they have the highest normalized coefficients of variation, lowest information entropy, and largest divergence coefficients. Several digital application indicators instead receive low-to-medium objective weights because their sample score distributions are relatively concentrated. Although teachers generally integrate digital technology into teaching plans infrequently, between-teacher variation is smaller than variation in risk incident counts under responsibility. Ranking disagreement is most pronounced in digital social responsibility: this dimension ranks low under AHP but shifts upward overall under entropy weighting because of the high variability of its cost-type indicators.

Table 5. Comparison of subjective and objective weight distributions across criterion dimensions

Criterion dimension	AHP weight	Entropy weight	Weight difference (A–E)	Rank change
Digital awareness	0.148	0.112	+0.036	Unchanged
Digital technology knowledge and skills	0.196	0.143	+0.053	Unchanged
Digital application	0.287	0.171	+0.116	Unchanged
Digital social responsibility	0.094	0.238	−0.144	Down 3 places
Professional development	0.131	0.187	−0.056	Up 1 place
Ethics–data literacy	0.144	0.149	−0.005	Unchanged

Table 5 shows that four of the six criterion dimensions retain the same ranks under subjective and

objective weighting, with disagreement concentrated in digital social responsibility and professional development. Digital social responsibility has the largest absolute weight difference, 0.144, and shifts down three ranks, making it the principal source of structural conflict. Digital applications have a substantial difference of 0.116, with AHP exceeding entropy weighting, but its rank is unchanged. This is a difference in magnitude rather than structure.

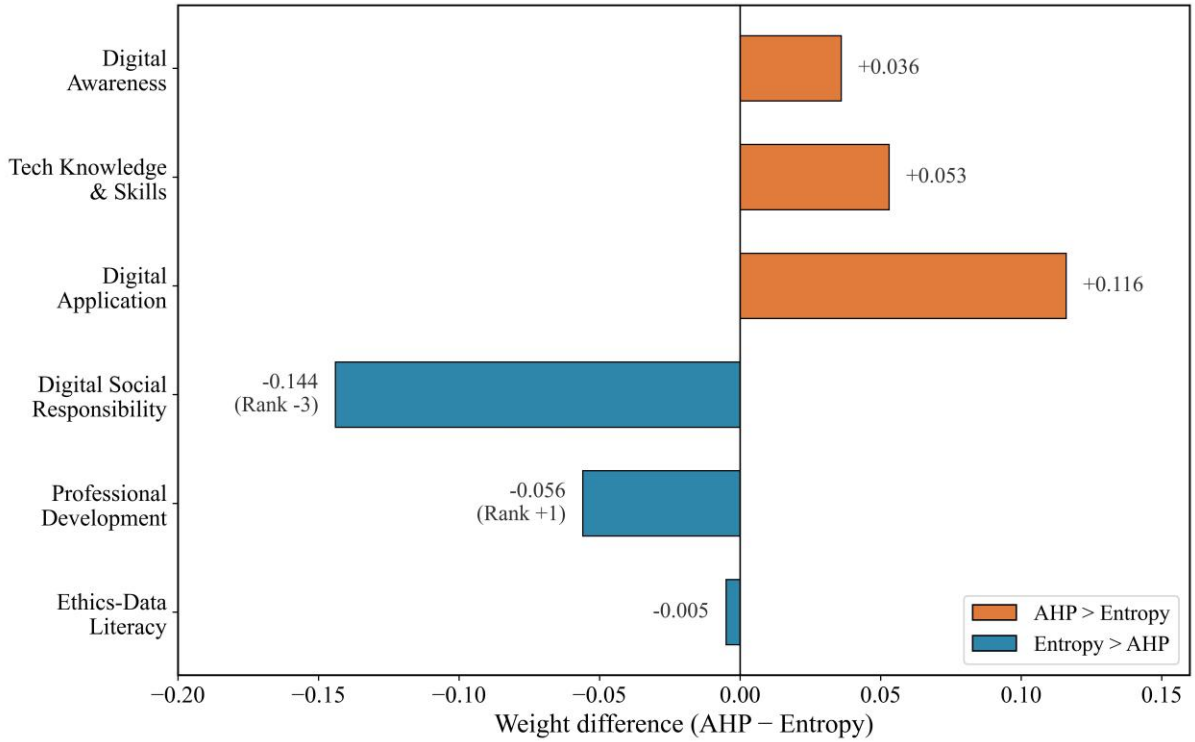


Figure 1. Decomposition of subjective–objective weight deviations across criterion dimensions

In Figure 1, the two long bars for digital application and digital social responsibility point in opposite directions and have similar lengths, identifying the strongest tension between subjective and objective weights. The other four dimensions have clearly shorter bars and moderate weight deviations.

4.3 CAW-AE combined weighting results

Global joint conflict C is 0.271, within the medium-conflict interval, $0.18 < C \leq 0.32$. Structural conflict has therefore not reached the level requiring strong reconciliation but is no longer negligible, so moderate intervention is appropriate. Cosine similarity is $S_c = 0.891$ and the rank displacement component is $S_r = 0.744$. Their comparison identifies the nature of the conflict: numerical directional deviation is relatively mild, while structural ranking misalignment is more pronounced. The joint measure captures the latter's contribution. If cosine similarity alone were used, $S_c = 0.891$ might be misinterpreted as low conflict. The rank displacement component explicitly reveals its actual source, the downward rank shift of digital social responsibility.

Criterion-level conflict contributions are highest for digital social responsibility, followed by professional development. Digital applications have an intermediate contribution despite their large numerical difference because their rank does not change. Each of the other three dimensions contributes less than 0.10. Reconciliation coefficients λ_j correspond to this distribution. All three cost-type indicators in digital social responsibility have coefficients below 0.5, indicating shifts toward objective

weights, with risk incidents having the lowest coefficients. Several digital application indicators have coefficients above 0.5, shifting toward subjective weights because their objective weights are systematically depressed by concentrated sample scores. The estimated intensity parameter γ^{**} lies within the range corresponding to medium conflict. Truncation is not triggered, indicating that the analytical solution already satisfies convex combination feasibility.

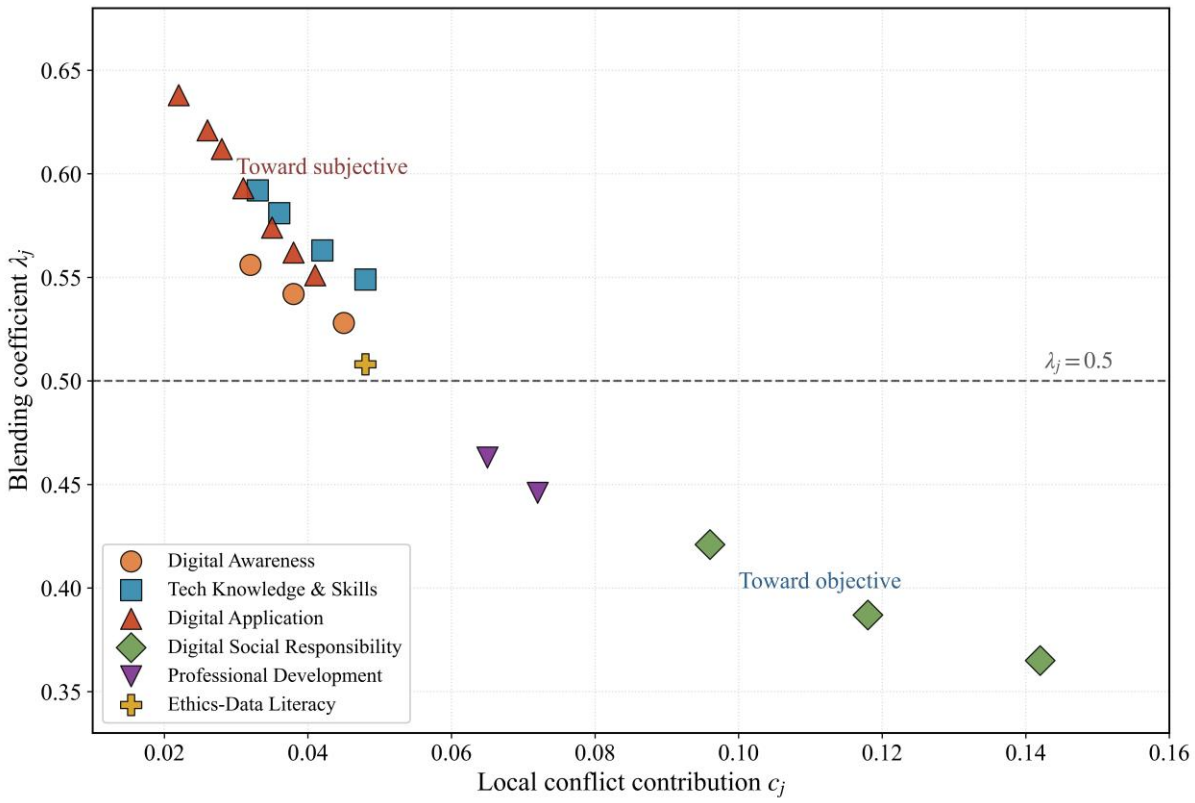


Figure 2. Relationship between indicator-level reconciliation coefficients and local conflict contributions

Figure 2 labels the indicator counts and their membership in the six criterion dimensions, showing clustering within dimensions. The three digital social responsibility points cluster tightly in the lower right, the two professional development points occupy the lower-middle region, and the seven digital application points extend from the upper left toward the center. Local conflict contributions thus form dimension-level blocks rather than being randomly distributed across indicators. This indicator-level evidence supports interpreting conflict diagnosis at the dimension level.

Table 6. Reconciliation coefficients and weight shifts for representative indicators

Indicator (dimension)	AHP weight	Entropy weight	Reconciliation coefficient	Combined weight	Shift direction
Frequency of technology-integrated teaching plans (application)	0.071	0.038	0.612	0.058	Toward subjective
Proficiency in subject-specific tools (knowledge and skills)	0.054	0.041	0.563	0.048	Toward subjective
Student data exposure risk incidents (social responsibility)	0.029	0.091	0.387	0.067	Toward objective
Records of digital ethics compliance (social responsibility)	0.033	0.078	0.421	0.059	Toward objective
Cross-school dissemination of digital teaching research (professional development)	0.048	0.082	0.446	0.067	Toward objective
Recorded learning-data-based instructional adjustments (ethics–data)	0.052	0.061	0.508	0.057	Near balance

Table 6 shows that indicator-level changes in reconciliation coefficients follow local conflict contributions, but their magnitude remains moderate because it is constrained by global conflict $C = 0.271$. The student data exposure risk incident indicator has the largest shift: its combined weight of 0.067 lies between its subjective weight of 0.029 and objective weight of 0.091, closer to the latter. The coefficient for learning data-based instructional adjustment records is close to 0.5, reflecting a small subjective-objective difference and low conflict contribution. In the final combined ranking, digital applications retain the highest criterion-level weight. Digital social responsibility rises substantially relative to AHP without exceeding its original entropy weight. Ethics-data literacy occupies an intermediate position and deviates least from the two basic weights.

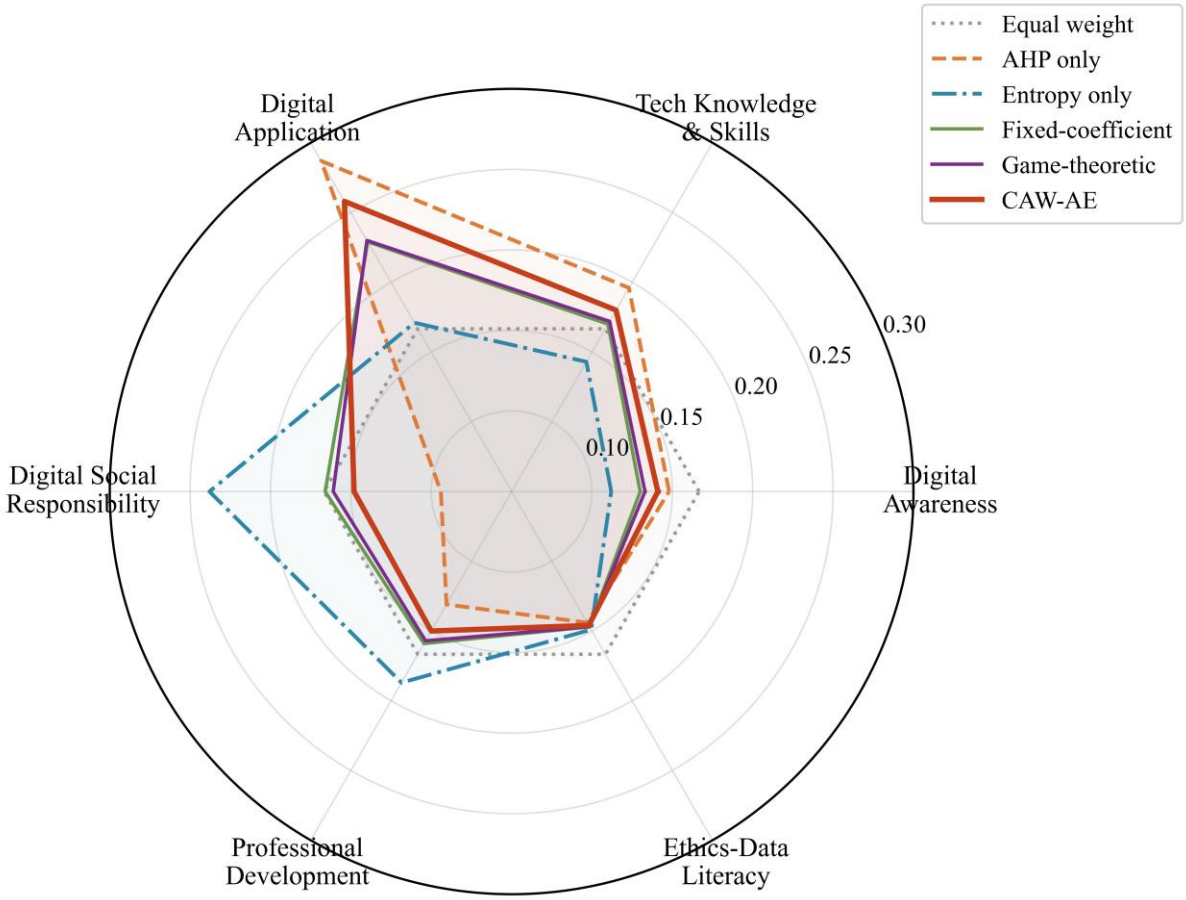


Figure 3. Radar comparison of criterion-level weights under six weighting schemes

In Figure 3, equal weighting forms a regular hexagon, with equal weights across dimensions, and serves as reference. AHP alone bulges outward toward digital application and inward toward digital social responsibility. Entropy weighting alone shows the opposite pattern. Fixed-coefficient and game-theoretic combinations lie between them and have similar, relatively symmetric compromise profiles.

4.4 Comparison of weight structures with alternative methods

Comparing the six weight distributions, equal weighting has HHI = 0.050 and CV = 0, defining lower-bound references for concentration and dispersion. AHP alone has HHI = 0.063 and CV = 0.312, while entropy weighting alone has HHI = 0.071 and CV = 0.398. The latter is more concentrated and dispersed, reflecting the amplification of high-variability indicators under objective weighting. Fixed-coefficient linear combination has HHI = 0.058 and CV = 0.241; its concentration and dispersion lie between the two basic methods and tend toward averaging, an inevitable outcome of a fixed-coefficient compromise. Game-theoretic combination has HHI = 0.061 and CV = 0.287, close to but slightly higher than the fixed-coefficient method because its coefficient is determined by deviation minimization rather than fixed at 0.5. CAW-AE has HHI = 0.066 and CV = 0.334. Its concentration exceeds fixed-coefficient and game-theoretic combinations but is below entropy weighting alone; its dispersion exceeds those two combinations but is below both individual methods.

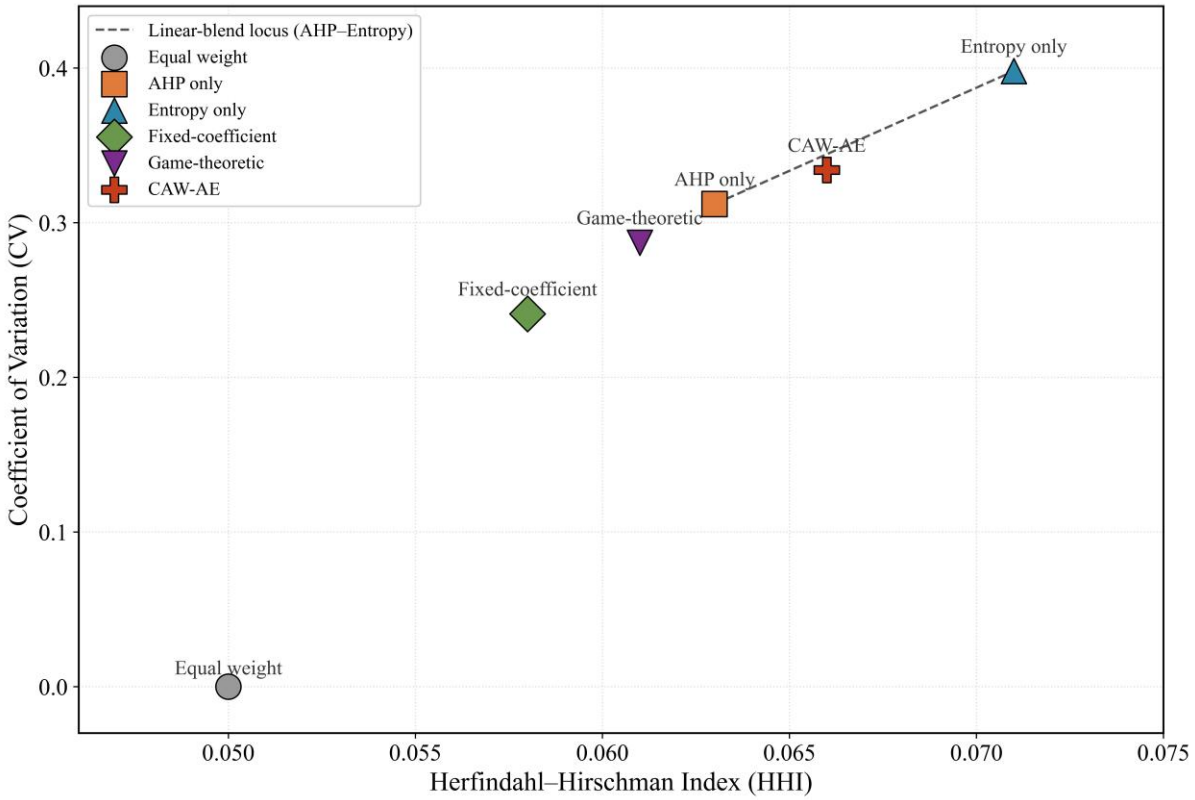


Figure 4. Distribution of weight structure characteristics across weighting schemes

Equal weighting lies in the lower-left corner of the plot, HHI = 0.050 and CV = 0, serving as the origin reference. AHP alone and entropy weighting alone occupy the upper-middle and upper-right regions, forming endpoints in the two basic directions. Fixed-coefficient and game-theoretic combinations cluster close together in the middle-left region, showing limited structural differences between their compromise and equilibrium solutions. CAW-AE lies in the middle-right region, horizontally between fixed-coefficient combination and entropy weighting alone and vertically above fixed-coefficient and game-theoretic combinations.

Table 7. Comparison of weight structure characteristics across weighting schemes

Weighting scheme	HHI	CV	Dimension of highest-weight indicator	Minimum-to-maximum weight ratio
Equal weighting	0.050	0.000	—	1.00
AHP alone	0.063	0.312	Digital application	0.31
Entropy weighting alone	0.071	0.398	Digital social responsibility	0.22
Fixed-coefficient combination	0.058	0.241	Digital application	0.38
Game-theoretic combination	0.061	0.287	Digital application	0.34
CAW-AE	0.066	0.334	Digital application	0.29

The ratio of minimum to maximum weight in Table 7 further characterizes distribution tails. CAW-

AE's ratio is 0.29, below AHP alone, 0.31, and game-theoretic combination, 0.34, but above entropy weighting alone, 0.22. CAW-AE thus compresses extreme weight differences less strongly than the fixed-coefficient method, 0.38, and the game-theoretic method, 0.34, but more strongly than entropy weighting alone. This is consistent with moderate reconciliation in the medium-conflict interval: the mechanism corrects the undervaluation of digital social responsibility without flattening the structure to the degree produced by fixed-coefficient averaging.

The Kendall rank correlation matrix shows that CAW-AE correlates 0.812 with AHP alone, 0.694 with entropy weighting alone, 0.883 with game-theoretic combination, and 0.847 with fixed-coefficient combination. Agreement is strongest with the game-theoretic method, but the two are not identical. For the three cost-type indicators in digital social responsibility, CAW-AE ranks are closer to the original entropy ranks than those of the game-theoretic method; for digital application, they are closer to the original AHP ranks.

5. MODEL TESTING AND VALIDATION

5.1 Validation framework

The validation framework follows a controlled design: identical data and assessment procedures, with only the weight generation mechanism differing. All six schemes-equal weighting, AHP alone, entropy weighting alone, fixed-coefficient combination, game-theoretic combination, and CAW-AE-use the normalized score matrix from Section 2 and weighting results from Section 4. Comprehensive assessment scores are uniformly calculated as:

$$S_i = \sum_{j=1}^{n} w_j \, y_{ij} \tag{19}$$

Here, S_i is the comprehensive score of teachers i, w_j is the scheme-specific weight of indicator j, and y_{ij} is the normalized score. The weighting scheme is the only variable. Score calculation, grouping, and testing are identical, ensuring that between-scheme differences are attributable to weighting.

Ranking stability is operationally defined as a scheme's ability to preserve its indicator weight ranking under input perturbations relative to the unperturbed baseline. Agreement between perturbed and baseline rankings is measured using Kendall rank correlation:

$$\tau = \frac{n_c - n_d}{n(n-1)/2} \tag{20}$$

Here, n_c and n_d are the numbers of concordant and discordant indicator pairs, respectively, and $n = 20$. Values of τ closer to 1 indicate stronger ranking preservation. The rank reversal rate is the proportion of 500 perturbation trials in which τ falls below 0.90; a lower rate indicates greater stability. The threshold of 0.90 comes from the substantive conclusion-difference boundary calibrated in Section 3's simulations and follows the same logic as the conflict thresholds.

Differentiation is tested through comprehensive score distributions and group discrimination. Distributional shape is described by the coefficient of variation, skewness, and kurtosis, with skewness and kurtosis estimated using moments [25],[26]:

$$SK = \frac{\frac{1}{N}\sum_{i=1}^{N}(S_i - \bar{S})^3}{\left[\frac{1}{N}\sum_{i=1}^{N}(S_i - \bar{S})^2\right]^{3/2}} \tag{21}$$

$$KU = \frac{\frac{1}{N}\sum_{i=1}^{N}(S_i - \bar{S})^4}{\left[\frac{1}{N}\sum_{i=1}^{N}(S_i - \bar{S})^2\right]^{2}} - 3 \tag{22}$$

Here, $\bar{S}$ is the mean score and $N = 347$. Group discrimination is tested by examining whether quartile groups based on comprehensive scores differ significantly on an independent performance indicator, using Kruskal-Walli's statistic [27]:

$$H = \frac{12}{N(N+1)}\sum_{g=1}^{4} n_g \bar{R}_g^2 - 3(N+1) \tag{23}$$

Here, g identifies the quartile group, n_g is the sample size of group g, and $\bar{R}_g$ is the mean rank of group g on the independent performance indicator. Larger H values indicate clearer group separation on that independent indicator and thus stronger identification of literacy differences.

External criterion association is measured using Spearman rank correlation between comprehensive scores and the criterion:

$$\rho = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N(N^2-1)} \tag{24}$$

Here, d_i is the difference between teacher i's comprehensive score rank and criterion rank. Differences in correlation coefficients between schemes are tested using Fisher's z transformation:

$$Z = \frac{z_1 - z_2}{\sqrt{\frac{1}{N-3} + \frac{1}{N-3}}},\ z = \frac{1}{2}\ln\frac{1+\rho}{1-\rho} \tag{25}$$

Here, z_1 and z_2 are the Fisher-transformed values for CAW-AE and the comparison scheme, respectively, and Z follows a standard normal distribution. This test determines whether CAW-AE's criterion association gain exceeds sampling fluctuation rather than merely comparing numerical correlations.

5.2 Sensitivity tests

Random perturbation experiments apply uniform random perturbations of ±10% and ±20% to all off-diagonal entries of the six criterion-level expert judgment matrices. Each level is repeated 500 times, with full AHP eigenvector estimation and subsequent combined weighting rerun after every perturbation. Table 8 reports reversal rates. Equal weighting is invariant to inputs and always has a zero-reversal rate. It provides only a lower reference bound and is excluded from comparative stability judgments. AHP alone has reversal rates of 6.2% and 14.8%, indicating sensitivity to expert judgment perturbations and the inherent vulnerability of purely subjective weighting. Entropy weighting alone does not depend on expert judgments and has zero reversal at both levels, but this "stability" entirely sacrifices domain knowledge and cannot directly be interpreted as a methodological advantage. Fixed-coefficient and game-theoretic combinations lie between the two individual methods, with respective

rates of 4.8% and 11.6%, and 4.1% and 9.7%. CAW-AE has rates of 3.4% and 8.2%, the lowest among all schemes relying on expert judgments.

Table 8. Rank reversal rates under perturbations of expert judgments

Weighting scheme	Reversal rate at ±10% perturbation (%)	Reversal rate at ±20% perturbation (%)	Increase in reversal rate (percentage points)
Equal weighting	0.0	0.0	0.0
AHP alone	6.2	14.8	8.6
Entropy weighting alone	0.0	0.0	0.0
Fixed-coefficient combination	4.8	11.6	6.8
Game-theoretic combination	4.1	9.7	5.6
CAW-AE	3.4	8.2	4.8

The increase in reversal rate column in Table 8 shows how rapid stability deteriorates when perturbation rises from ±10% to ±20%. CAW-AE increases by 4.8 percentage points, less than AHP alone, 8.6, fixed-coefficient combination, 6.8, and game-theoretic combination, 5.6. Its stability advantage therefore concerns both absolute reversal rates and the rate of degradation as perturbations intensify.

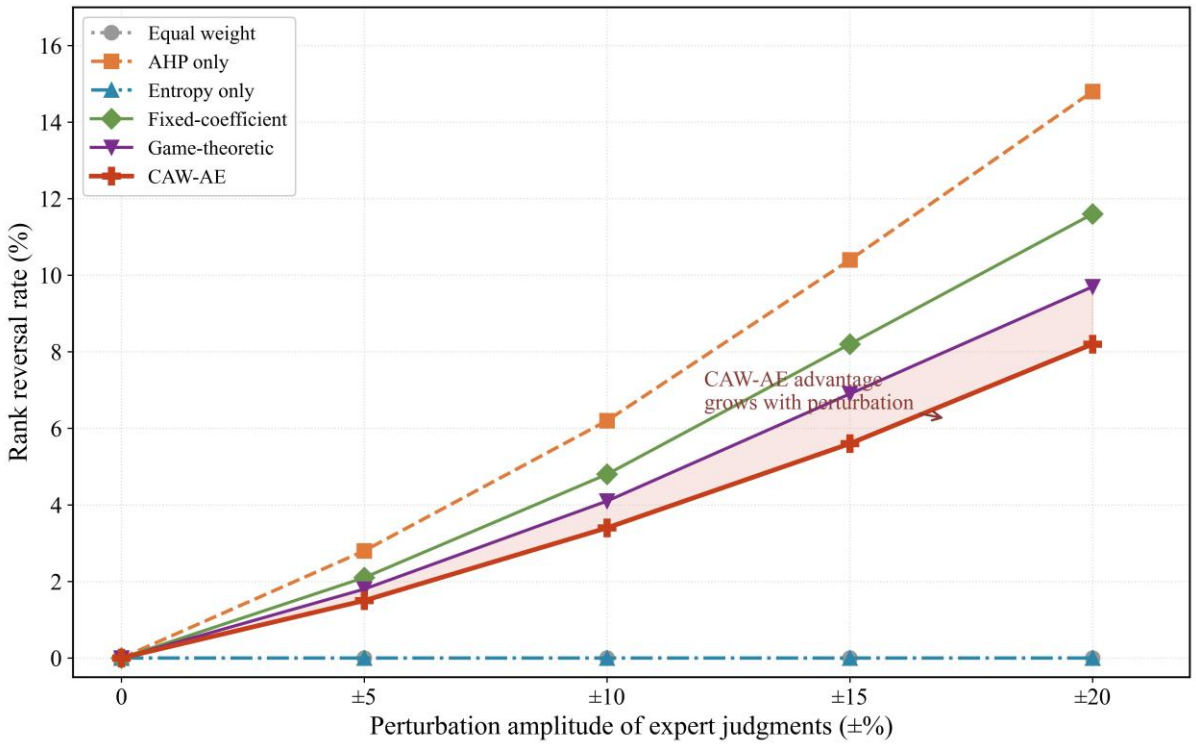


Figure 5. Rank reversal rates by perturbation amplitude of expert judgments

In Figure 5, equal weighting and entropy weighting alone form horizontal lines near the horizontal axis, providing reference boundaries. AHP alone has the steepest slope and reaches the highest point at

±20%. Fixed-coefficient and game-theoretic combinations lie below AHP with similar shapes, nearly overlapping under low perturbation and separating slightly at high perturbation. CAW-AE lies below all expert-dependent schemes at every perturbation level. Its curve is flat under low perturbation and rises more slowly under high perturbation. Shading highlights the reversal-rate gap between CAW-AE and game-theoretic combination.

Bootstrap resampling of objective data evaluates weight estimation stability from the sample side. The 347 observations are resampled with replacement 1,000 times. After each resampling, entropy weights and the entire combination procedure are recalculated, and the coefficient of variation of each estimated indicator weight is recorded. Table 9 reports the mean and maximum coefficients across the 20 indicators for all six schemes. Equal weighting does not depend on data and has zero variation. Entropy weighting alone has the highest mean coefficient, 0.187, reflecting sensitivity to sample composition: resampling alters the proportion of highly variable indicators in the sample and destabilizes objective weight estimates. AHP alone also has zero variation because it does not depend on data, but this stability similarly sacrifices the data channel. Fixed-coefficient and game-theoretic combinations have mean coefficients of 0.094 and 0.087, respectively. CAW-AE has 0.071, the lowest among all schemes using both information channels.

Table 9. Coefficients of variation in weight estimates under bootstrap resampling

Weighting scheme	Mean coefficient of variation	Maximum coefficient of variation	Indicators with coefficient of variation above 0.15
Equal weighting	0.000	0.000	0
AHP alone	0.000	0.000	0
Entropy weighting alone	0.187	0.312	11
Fixed-coefficient combination	0.094	0.178	3
Game-theoretic combination	0.087	0.164	2
CAW-AE	0.071	0.139	0

Under entropy weighting alone, 11 indicators have weight estimate coefficients of variation above 0.15, concentrated among cost-type and low-frequency indicators. Their sample variability is strongly affected by individual observations, producing large weight fluctuations under resampling. CAW-AE has no such indicators: all 20 remain below 0.15, indicating that reconciliation suppresses objective-channel sampling fluctuations.

5.3 Differentiation tests

After comprehensive scores are calculated for all 347 teachers under each scheme, Table 10 reports their distributional characteristics. Equal weighting yields a coefficient of variation of 0.183, skewness of −0.142, and kurtosis of −0.318. Its nearly symmetric, light-tailed distribution directly reflects equal-weight smoothing. AHP alone yields a coefficient of variation of 0.241, skewness of 0.316, and kurtosis of 0.487. Its right-skewed, slightly heavy-tailed distribution reflects how subjective amplification of highly weighted application indicators clusters some teachers' scores toward the high end. Entropy weighting alone has the highest coefficient of variation, 0.312, and the largest skewness, 0.584, and

kurtosis, 1.126. Concentrated amplification of highly variable indicators produces clear right skew and heavy tails, separating some teachers substantially because of extreme cost-type performance. Fixed-coefficient and game-theoretic combinations fall between the individual methods and tend toward averaging, with coefficients of variation of 0.226 and 0.238. CAW-AE has a coefficient of variation of 0.267, above those two combinations but below entropy weighting alone. Its skewness, 0.372, and kurtosis, 0.641, likewise occupy upper-intermediate positions.

Table 10. Comprehensive score distributions and group discrimination across schemes

Weighting scheme	Coefficient of variation	Skewness	Kurtosis	Kruskal–Wallis H	Significance of differences between quartile groups
Equal weighting	0.183	−0.142	−0.318	18.42	p<0.01
AHP alone	0.241	0.316	0.487	27.65	p<0.001
Entropy weighting alone	0.312	0.584	1.126	31.08	p<0.001
Fixed-coefficient combination	0.226	0.208	0.274	24.37	p<0.001
Game-theoretic combination	0.238	0.251	0.352	26.14	p<0.001
CAW-AE	0.267	0.372	0.641	33.76	p<0.001

CAW-AE's H value is 33.76, the highest of all six schemes, exceeding entropy weighting alone, 31.08, and AHP alone, 27.65, and more substantially exceeding fixed-coefficient combination, 24.37, and game-theoretic combination, 26.14. This ordering has methodological significance: differentiation does not simply increase with weight dispersion. Entropy weighting alone has a higher coefficient of variation, 0.312, than CAW-AE, 0.267, but a lower H value.

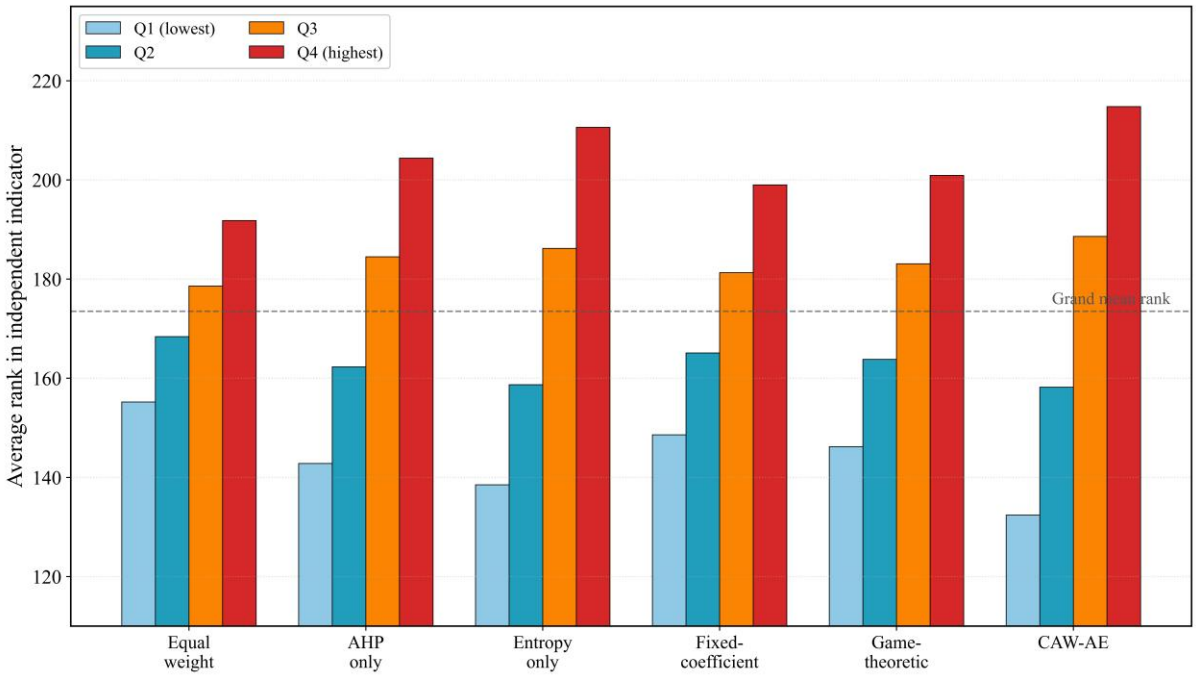


Figure 6. Mean ranks on the independent performance indicator by comprehensive score

quartile group and scheme

In Figure 6, equal weighting has the smallest differences among its four bars and a limited mean rank gap between the first and fourth groups, consistent with its lowest H value of 18.42 in Table 10. Entropy weighting alone shows higher values at the extremes and lower values in the middle. Its fourth group is prominent, but the second and third groups do not clearly separate from the first. This nonmonotonic pattern explains why its H value, although high, remains below CAW-AE's. CAW-AE shows the clearest monotonic increase: the first group has the lowest mean rank and the fourth the highest, with relatively uniform differences between adjacent groups. Separation between its second and third groups is clearly better than under fixed-coefficient and game-theoretic combinations.

5.4 External criterion association tests

The criterion is chosen according to available data. The independent performance indicator is an aggregate of recorded frequencies of teachers' digital instructional behaviors, drawn from routine regional teaching management platform data. It is independent of the indicator score matrix in collection channel, timing, and recorded content and does not duplicate any indicator in the system. Association is measured by Spearman rank correlation, and between-scheme correlation differences are tested using Fisher's z transformation. Table 11 reports criterion correlations and differences from CAW-AE. Correlations are 0.312 for equal weighting, 0.387 for AHP alone, 0.341 for entropy weighting alone, 0.364 for fixed-coefficient combination, 0.379 for game-theoretic combination, and 0.428 for CAW-AE. CAW-AE is highest, exceeding the next-highest scheme, AHP alone, by 0.041 and game-theoretic combination by 0.049.

Table 11. Spearman correlations between comprehensive scores and the external criterion and tests of differences

Weighting scheme	Criterion correlation	Difference from CAW-AE	Fisher Z	Significance of difference
Equal weighting	0.312	−0.116	2.87	p<0.01
AHP alone	0.387	−0.041	1.34	p=0.180
Entropy weighting alone	0.341	−0.087	2.26	p<0.05
Fixed-coefficient combination	0.364	−0.064	1.98	p<0.05
Game-theoretic combination	0.379	−0.049	1.71	p=0.087
CAW-AE	0.428	—	—	—

CAW-AE's criterion association gains over equal weighting, entropy weighting alone, and fixed-coefficient combination are statistically significant, $p < 0.05$. Its gains over AHP alone, $p = 0.180$, and game-theoretic combination, $p = 0.087$, do not reach the 0.05 significance level.

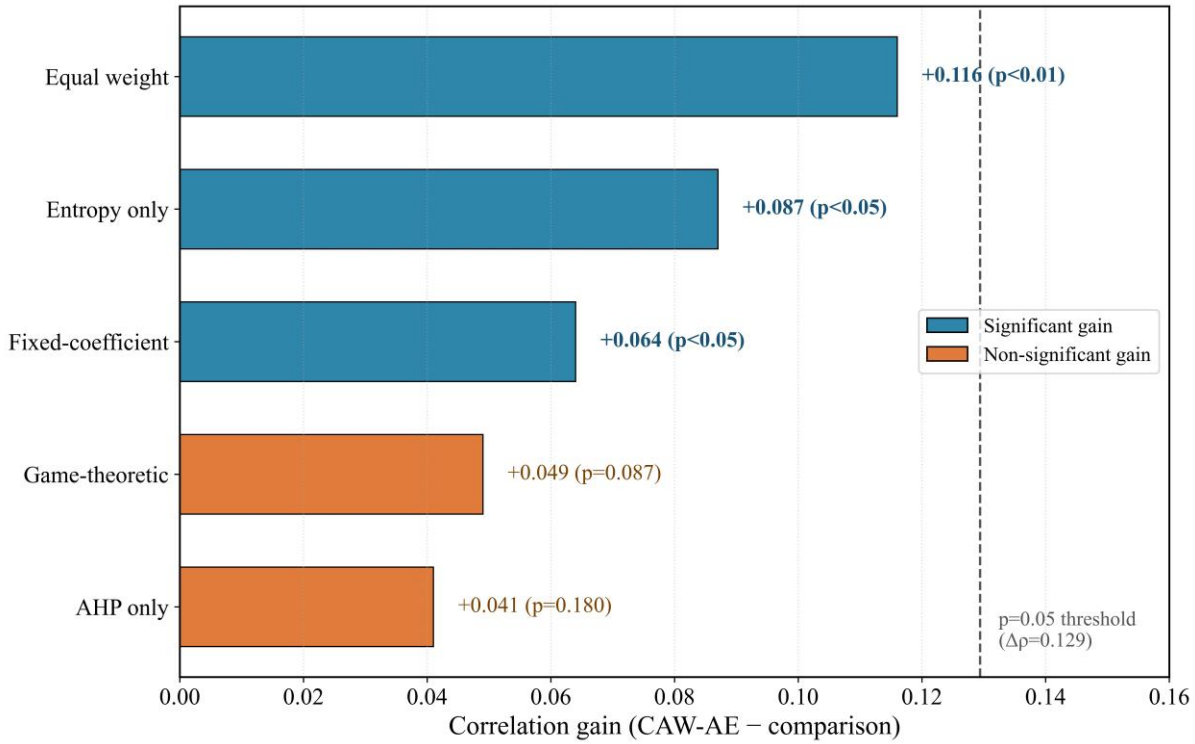


Figure 7. Decomposition of CAW-AE's criterion association gains over comparison schemes

All bars in Figure 7 extend rightward, indicating positive gains over every comparison scheme. End markers distinguish significance levels: gains over equal weighting, entropy weighting alone, and fixed-coefficient combination are marked significant, whereas those over AHP alone and game-theoretic combination are marked nonsignificant.

6. CONCLUSIONS

6.1 Main work and findings

This study's methodological contribution is to advance combined weighting from static weighting to conflict-driven adaptive reconciliation. Traditional approaches, whether fixed-coefficient linear combinations or game-theoretic equilibrium combinations, share an implicit assumption: subjective–objective disagreements should be resolved into a global compromise coefficient rather than treated as informative objects. CAW-AE starts from the opposite position. Structural conflict on a particular indicator reveals asymmetric credibility between expert judgments and data variation. Combination should identify this asymmetry and use it to determine the shift direction rather than smooth it away. Joint conflict combines cosine similarity and rank displacement geometrically, diagnosing more than numerical deviation alone. In this study, digital application has a large weight difference but stable ranking, while digital social responsibility has both numerical deviation and downward rank displacement. The joint measure identifies the latter as the main conflict source, whereas cosine similarity alone would misclassify it as low conflict. Differentiated indicator-level coefficients make each shift's direction and magnitude traceable to local conflict, giving the resulting weights interpretable source attribution.

The main empirical finding is that subjective–objective conflicts in teachers' digital literacy are

not uniformly distributed but form blocks at the criterion-dimension level. They concentrate on digital social responsibility and professional development. Digital social responsibility combines numerical deviation and ranking misalignment and is the principal structural conflict source. Digital application, despite its large numerical deviation, reflects magnitude rather than structural disagreement because its rank remains stable. This pattern has substantive implications: awareness and responsibility, conceptually viewed as foundational, are precisely the dimensions from which stable weights are hardest to extract. Experts consider responsibility indicators necessary but nondiscrimination, while their cost-type variability produces high objective weights. Reconciled weights are asymmetric: application shifts toward subjective judgment to preserve the domain knowledge framework, while responsibility shifts toward objective weights to incorporate data-driven correction. Overall concentration and dispersion occupy a justified position between individual methods and fixed-coefficient compromise and cannot be reproduced by any globally fixed linear combination ratio.

Validation conclusions require distinctions in evidential strength. For ranking stability, CAW-AE has the lowest reversal rate and weight estimation variation among dual-channel schemes under expert judgment perturbations and bootstrap resampling. Its advantage grows as perturbations intensify, consistent with the self-regulation inferred from model properties. For differentiation, CAW-AE has the strongest group discrimination, and this is not simply due to inflated weight dispersion: entropy weighting alone has greater dispersion but weaker discrimination. CAW-AE achieves the best discrimination with moderate dispersion. For external criterion association, CAW-AE has the highest correlation, but gains over AHP alone and game-theoretic combination are not statistically significant. Only gains over equal weighting, entropy weighting alone, and fixed-coefficient combination are significant. This pattern defines the limits of model effectiveness: confirmatory advantages are concentrated in ranking stability and group discrimination, while criterion association gains provide directional rather than decisive support. Numerical ordering of correlations should not replace significance testing.

6.2 Applicability boundaries

CAW-AE depends on several prerequisites; outside them, its behavior can depart from design expectations. The first is expert judgment quality. Subjective AHP weights are an input to conflict diagnosis. If the consistency ratio approaches or exceeds the 0.10 acceptance threshold, those weights are unreliable. Apparent subjective–objective disagreement may then mainly reflect expert judgment noise rather than substantive asymmetry between information sources. Reconciliation toward subjective weights would amplify rather than correct errors. CR values in this study range from 0.021 to 0.067, supplying adequate subjective inputs, but this condition may not persist when expert composition changes or judgment tasks become more complex.

The second prerequisite concerns sample size. Entropy weight estimates have high variance with insufficient observations. Bootstrap tests show 11 entropy-only indicators with weight estimation coefficients of variation above 0.15, reduced to zero by CAW-AE. However, this suppression itself requires an adequate sample. With only a few dozen observations, objective sampling fluctuations may exceed the mechanism’s compensatory capacity, destabilizing conflict diagnosis. A third boundary is the context sensitivity of thresholds. The low- and high-conflict boundaries of 0.18 and 0.32 were calibrated from turning points in simulated conclusion-difference rates and depend on indicator count and weight-space dimensionality. When the number of indicators differs substantially from 20 or tolerance for rank reversals differs, thresholds require recalibration. Direct transfer may trigger reconciliation when it is unnecessary or leave it inactive when it is needed.

CAW-AE is also unsuitable when indicators exhibit severe multicollinearity. Local conflict contributions are calculated independently by indicator, implicitly assuming that disagreements do not interfere across indicators. Highly correlated indicators may receive similar AHP weights but very different entropy weights because of minor distributional differences. Correlation structure can then artificially inflate local conflict, causing shifts in reconciliation coefficients to stop reflecting genuine asymmetry between sources. This limitation is not unique to CAW-AE, but its indicator-level conflict-driven mechanism makes it more sensitive to multicollinearity than global compromise methods.

6.3 Limited implications for assessment practice

Practical implications are directional indications rather than policy recommendations. The first concerns the methodological risks of habitual equal weighting in assessing teachers' digital literacy. Equal weighting is a convenient default without a weighting rationale but systematically smooths away heterogeneous contributions across dimensions. Here, it has the lowest or nearly lowest group discrimination and criterion association of all schemes. This suggests that equal-weight assumptions may conceal substantive dimensional differences when assessments diagnose teacher development or identify training needs, pushing conclusions toward the mean and weakening differentiation. This indication is limited to this study's indicator system and sample and should not be generalized into a rejection of all equal-weight practice.

Combined weights can inform identification of teacher training priorities, but only indirectly. CAW-AE's reconciled structure shows that digital social responsibility contains substantially more objective information than its subjective weight suggests. Actual teacher differences in this dimension are therefore much larger than experts' anticipated differentiation. If training needs are identified from the dimensional structure of assessment scores, increasing responsibility weights may shift emphasis moderately from application capability improvement toward data security and ethical norms. This inference is mediated by weight structure and does not substitute weighting results for needs assessment. Weights express relative importance within an assessment context and are not equivalent to priorities for allocating training resources; those also depend on trainability, costs, and current levels.

6.4 Future research

Two extensions are possible. The first is dynamic weighting. Current CAW-AE coefficients use full-sample conflict states, implicitly assuming homogeneous source credibility across teachers. However, professional development stages may systematically differ in literacy weaknesses and data variability. Novice teachers may vary more in technical operations, whereas experienced teachers may differ more in data-driven decision-making. Conditioning coefficients on development stage or school level would allow weights to adapt to group characteristics. Conflict diagnosis would need to move from a static full-sample measure to conditional group measures. The methodological challenge is balancing the effect of reduced group sample sizes on entropy stability against reconciliation's compensatory capacity.

The second extension is testing the transferability of conflict measures. The weighted geometric combination of cosine similarity and rank displacement was designed for assessments where numerical deviation and ranking misalignment can be separated. Applicability to other educational settings depends on whether subjective–objective disagreements there have the same separability. With strong expert consensus and sufficient data variation, the components may covary strongly, limiting the added information from their combination. With very few indicators or highly skewed weights, rank

displacement may have insufficient resolution. Cross-context simulations and empirical comparisons are needed to examine the measure's behavior. Although beyond this study's scope, they are necessary for extending CAW-AE from a single-context model to a general combined weighting framework.

Abbreviations

CAW-AE, conflict-aware and adaptively balanced combined weighting model;

IQR, interquartile range;

HHI, Herfindahl–Hirschman index;

CV, coefficient of variation;

CR, consistency ratio;

RI, random consistency index;

Supplementary Material

Not applicable.

Appendix

Not applicable.

Ethics approval and consent to participate.

This study did not involve human participants, animal subjects, or any data requiring ethical approval. Therefore, ethics approval and consent to participate are not applicable.

Acknowledgements

The authors would like to thank the editors of this journal and all the anonymous reviewers who provided valuable comments on this work.

Competing interests

The authors declare that they have no financial or personal relationships that may have inappropriately influenced them in writing this article.

Author contributions

All authors have read and agreed to the published version of the manuscript. The author's

contributions are specified as follows: **J.Z.:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration.

Funding information

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability

The data that support the findings of this study are available upon request from the corresponding authors, **J.Z.**

Disclaimer

The views and opinions expressed in this article are those of the authors and are the product of professional research. It does not necessarily reflect the official policy or position of any affiliated institution, funder, agency, or that of the publisher. The authors are responsible for this article's results, findings, and content.

Declaration of AI and AI-assisted Technologies in the Writing Process

During the writing of this article, the author used ChatGPT for spelling and grammar checking. After using this tool, the author reviewed and edited the content as needed and assumes full responsibility for the final published content.

REFERENCES

[1] Bachtiar, B. (2025). Preparing citizens for the future of digital literacy and AI: With a focus on Indonesian EFL teachers. In Digital citizenship and the future of AI engagement, ethics, and privacy (pp. 405-440). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-9015-3.ch015

[2] Liu, C., Gao, Z., Tu, Y., Zeng, S., Li, S., Chen, H., & Chiou, W. K. (2026, July). Research on Potential Risks and Governance Mechanisms of Journalism and Communication Education Enabled by Generative Artificial Intelligence. In International Conference on Human-Computer Interaction (pp. 230-245). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-29903-1_14

[3] Heinsen, R. S. (2024). Gaze behavior reveals automaticity and attention allocation during music teaching vs. observing. Journal of Eye Movement Research, 17(2), 9. https://doi.org/10.16910/jemr.17.2.3

[4] Angus, D., Hayden, L., Obeid, A. K., Tan, X. Y., Carah, N., Burgess, J., ... & Bagnara, J. (2024). Computational methods for improving the observability of platform-based advertising. Journal of

advertising, 53(5), 661-680. https://doi.org/10.1080/00913367.2024.2394156
[5] Pham Thi, T. D., Pham, V. K., & Duong, N. T. (2025). Understanding M-learning app adoption: an integrated model for college students in Vietnam. Interactive Learning Environments, 33(10), 5847-5869. https://doi.org/10.1080/10494820.2025.2488144
[6] Kim, T., & Lee, J. (2020). Measuring teacher effectiveness for equity: value-added model of teachers' distributive effects on classroom achievement gaps. Educational Research and Evaluation, 26(1-2), 30-52. https://doi.org/10.1080/13803611.2021.1906708
[7] Triventi, M. (2020). Are children of immigrants graded less generously by their teachers than natives, and why? Evidence from student population data in Italy. International Migration Review, 54(3), 765-795. https://doi.org/10.1177/0197918319878104
[8] Charalambous, C. Y., Hill, H. C., Chin, M. J., & McGinn, D. (2020). Mathematical content knowledge and knowledge for teaching: exploring their distinguishability and contribution to student learning: CY Charalambous et al. Journal of Mathematics Teacher Education, 23(6), 579-613. https://doi.org/10.1007/s10857-019-09443-2
[9] Bohra, S. S., & Anvari-Moghaddam, A. (2022). A comprehensive review on applications of multicriteria decision-making methods in power and energy systems. International Journal of Energy Research, 46(4), 4088-4118. https://doi.org/10.1002/er.7517
[10] Sarkar, A., & Goswami, S. S. (2026). A comprehensive review: The novel weighting methods for multi-criteria decision-making (MCDM). Spectrum of Operational Research, 1-26. https://doi.org/10.31181/sor202781
[11] Zhao, H., Song, S., Lv, X., & Bao, Y. (2025). A multi-dimensional evaluation model for power enterprise procurement performance based on fuzzy analytic hierarchy process and TOPSIS integration. Scientific Reports, 15(1), 41290. https://doi.org/10.1038/s41598-025-25042-z
[12] Gallego-Arrufat, M. J., García-Martínez, I., Romero-López, M. A., & Torres-Hernández, N. (2024). Digital rights and responsibility in education: A scoping review. Education Policy Analysis Archives, 32. https://doi.org/10.14507/epaa.32.7899
[13] Lan, H. (2025). Sustainable development of social responsibility in universities in the digital era: Perceptions and advances. Information Development, 02666669241303124. https://doi.org/10.1177/02666669241303124
[14] Fan, J., Wang, H., & Gu, X. (2025). Development and validation of a framework and scale for primary and secondary school teachers' data-artificial intelligent competence. Frontiers in Psychology, 16, 1756893. https://doi.org/10.3389/fpsyg.2025.1756893
[15] Yadav, S. (2026). Addressing Sustainable Development Goals Through Competency-Based Education: Artificial Intelligence Innovation and Teachers' Competency Development. In Addressing Sustainable Development Goals Through Competency-Based Education (pp. 1-26). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3373-3720-3.ch001
[16] Panagiotou, N. (2025). AI literacy and the future of education: A framework for ethical and inclusive learning models. Journal of Interdisciplinary Knowledge, 8, e01635-e01635.

https://doi.org/10.37497/jik.v8iknowledge.1635
[17] Ochieng, P. A. (2027). Artificial Intelligence in Teacher Professional Development for Instructional Leadership Capacity. In Harnessing AI for Education Reforms: Frameworks, Best Practices and Considerations for Quality Education (pp. 75-96). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3373-8740-6.ch004
[18] Liu, Z., & Yang, Y. (2025). Fundamental usability evaluation in visualization: Integrating Nielsen's principles and the analytic hierarchy process. International Journal of Human–Computer Interaction, 41(20), 13012-13040. https://doi.org/10.1080/10447318.2025.2467455
[19] Kovari, A. (2026). Generative AI, performance, and learning: A framework for comparing interventions across assessment regimes. Computers, 15(9), 633. https://doi.org/10.3390/computers15090633
[20] Yu, Z., Sun, G., Lin, S., Hu, H., & Xu, J. (2025). Operationalizing game-theoretic weighting in public hospital cost control: an implementation framework from Chinese tertiary hospitals. Cost Effectiveness and Resource Allocation, 23(1), 11. https://doi.org/10.1186/s12962-025-00616-w
[21] Zhang, C., Wang, Z., Luo, L., & Zhan, W. (2026). Sustainable Flood Emergency Material Demand Grading and Routing Re-Optimization: A Game-Theoretic Combined Weighting and Grey-TOPSIS Approach in Jiaozuo, China. Sustainability. https://doi.org/10.3390/su182010617
[22] Jiang, F., Zhu, Z., Li, S., & Zhang, J. (2026). Navigating Sustainable Energy Transitions and Environmental Constraints in PV Siting. Sustainability, 18(8), 3915. https://doi.org/10.3390/su18083915
[23] Kozáková, M., & Kmety Barteková, M. (2020). Analysis of market concentration in creative industry. In SHS Web of Conferences (Vol. 83, p. 01035). EDP Sciences. https://doi.org/10.1051/shsconf/20208301035
[24] Evren, A., Tuna, E., Ustaoglu, E., & Sahin, B. (2021). Some dominance indices to determine market concentration. Journal of Applied Statistics, 48(13-15), 2755-2775. https://doi.org/10.1080/02664763.2021.1963421
[25] Bastianin, A. (2020). Robust measures of skewness and kurtosis for macroeconomic and financial time series. Applied Economics, 52(7), 637-670. https://doi.org/10.1080/00036846.2019.1640862
[26] Chen, Q., & Xiao, Z. (2025). Spread regression, skewness regression, and kurtosis regression with an application to the US wage structure. Journal of Applied Econometrics, 40(3), 325-340. https://doi.org/10.1002/jae.3105
[27] Jin, L., Dong, Y. Y., Xu, J. P., Chen, M. S., Zeng, R. X., & Guo, L. H. (2025). Relationship between the laboratory test-based frailty index and overall mortality in critically ill patients with acute pancreatitis: a retrospective study based on the MIMIC-IV database. Frontiers in Medicine, 12, 1524358. https://doi.org/10.3389/fmed.2025.1524358

APA Citation

Zhang, J. (2026). Adaptive AHP-Entropy Combined Weighting Model for Teachers' Digital Literacy Assessment. *Journal of Digital Frontier, 1*(2), 92-123. **https://doi.org/10.67541/jdf2611**