International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

INTUITIONISTIC FUZZY RUBRICS FOR EVALUATING MULTIMODAL STUDENT PROJECTS

Suleiman Shelash Mohammad[1,2], Yogeesh N[3*], N Raja[4], Hanan Jadallah[5], Sirojiddin Khudoyberganov Meylibayevic[6], Rayhon Sapaeva[7], Asokan Vasudevan[8,9], Anupama R[10]

[1] Electronic Marketing and Social Media, Economic and Administrative Sciences Zarqa University, Zarqa 132010, Jordan

[2] Research follower, INTI International University, Negeri Sembilan 71800, Malaysia

Email: dr_sliman@yahoo.com, ORCID ID: 0000-0001-6156-9063

[3] Department of Mathematics, Government First Grade College, Tumkur, Karnataka, India. Email: yogeesh.r@gmail.com ORCID ID: orcid.org/0000-0001-8080-7821

[4] Department of Visual Communication, Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India.
Email: rajadigimedia2@gmail.com, ORCID: 0000-0003-2135-3051

5 Electronic Marketing and Social Media, Economic and Administrative Sciences Zarqa University, Zarqa 132010, Jordan

Email: Hananjadallah1987@gmail.com, ORCID: 0009-0005-7138-1167

[6] Lecturer of the Department of Roman-Germanic Philology, Mamun University, Email: xudayberganov_sirojiddin@mamunedu.uz, ORCID: https://orcid.org/0009-0001-1087-9112

[7] Department of Roman-Germanic Philology, Urgench State University, Urgench, Uzbekistan, Email: diratear@gmail.com , ORCID: https://orcid.org/0000-0002-9304-631X

[8]Faculty of Business and Communications, INTI International University, Nilai 71800, Malaysia

[9]Faculty of Management, Shinawatra University, 99 Moo 10, Bangtoey, Samkhok 12160, Thailand
asokan.vasudevan@newinti.edu.my

[10] Department of Economics, Government First Grade College. Medigeshi-572133, Karnataka, India. Email: anupama.hitha@gmail.com, ORCID: https://orcid.org/0009-0005-4140-6837

***Corresponding author:** yogeesh.r@gmail.com

Received: July 28, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Abstract

This paper proposes an Intuitionistic Fuzzy Rubric (IFR) for grading multimodal student work that explicitly models supportive evidence, counterevidence, and uncertainty via membership (μ), non-membership (ν), and hesitation (π), with $\mu + \nu \leq 1$. Linguistic terms are calibrated to intuitionistic fuzzy numbers using triangular prototypes with a soft-complement for doubt; indicator-level evidence is fused by intuitionistic weighted operators (IFWA/IFWG). Across criteria, we aggregate scores with either an Ordered Weighted Average (OWA) when interaction is negligible or a Choquet integral under a 2-additive capacity when synergy/redundancy matters; Shapley values provide criterion importances for feedback. The pipeline outputs a two-dimensional decision ($\hat{S}, \hat{H}$)-net support and certainty-mapped to policy bands via transparent thresholds. A case study on 24 projects spanning four criteria and three indicators each demonstrates strong alignment with instructor references: Kendall's $\tau = 0.913$ and grade accuracy $= 0.833$, outperforming a crisp weighted rubric ($\tau = -0.290, 0.667$) and a type-1 fuzzy baseline ($\tau = -0.072, 0.500$). Capacity analysis highlights positive Textual-Visual and Textual-Interaction synergy, while Shapley importances (ϕ: $0.310, 0.245, 0.190, 0.255$) align with learning objectives. Sensitivity checks show grade changes for only $3.5 - 12\%$ of projects under parameter perturbations. Overall, IFR yields reliable, interpretable, and pedagogically actionable assessment for multimodal work, with clear pathways for calibration, robustness analysis, and semester-over-semester governance.

Keywords: educational assessment; non-additive aggregation; Choquet integral; Shapley importance; ordered weighted averaging; linguistic calibration; multicriteria decision analysis.

1. Introduction

1.1 Problem Context and Motivation

Assessing multimodal student projects (text + visuals + audio + code/interaction) with conventional rubrics is difficult because rubric descriptors ("excellent," "good," "fair") are graded, not binary, and raters often hesitate when evidence is partial or cross-modal signals disagree. Fuzzy set theory models gradedness, but classic (type-1) fuzzy membership alone cannot capture a rater's hesitation or simultaneous support and doubt; intuitionistic fuzzy sets (IFS) add an explicit non-membership and a derived hesitation degree, making them well-suited for educational assessment under uncertainty [1,2].

Formally, for any criterion value we represent an evaluation as $( ) = \langle \mu_A(x), \nu_A(x) \rangle, \mu_A, \nu_A \in [0,1], \mu_A(x) + \nu_A(x) \leq 1,$ with hesitation $\pi_A(x) = 1 - \mu_A(x) -$ ( ) capturing rater indecision [2].

1.2 Research Gap and Objectives

Existing rubric scoring pipelines either (i) collapse multimodal indicators to a single crisp score (losing gradience and interaction among criteria) or (ii) use type-1 fuzzy rubrics without modeling hesitation or synergy/redundancy across criteria. We address this gap by proposing an Intuitionistic Fuzzy Rubric (IFR) that (a) maps each indicator measurement to an IF number (IFN), (b) aggregates within a criterion using intuitionistic fuzzy aggregation operators with

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

theoretical guarantees [3,4], and (c) aggregates across criteria with a Choquet integral to model positive/negative inter-criterion interaction [5].

1.3 Contributions

Mathematical framework: We define a measurement → IFN mapping and two summaries

$$S = \mu - \nu \text{ (score)}, H = \mu + \nu \text{ (accuracy)},$$

used for interpretable grading regions; these are standard in IFS comparison theory.

Aggregators with interaction: We employ IF-weighted averaging operators for local fusion and a Choquet integral of scores for global fusion under a monotone capacity to capture synergy/redundancy between modalities.

Educator-facing rubric design: We show how to elicit criterion weights/capacities from instructors (AHP-style pairwise inputs) and align the IFR with established rubric practice in higher education.

1.4 Overview of the Mathematical Construction

Let projects $= \{ \}_{=1}$, criteria $= \{ \}_{=1}$, and indicator measurements ( ). We define pre-scores $s^+, s^- \in [0,1]$ (e.g., monotone logistics around thresholds $\tau^{(h)}, \tau^{(l)}$):

$$s^+(m) = \frac{1}{1 + \exp(-a \quad (m - \tau^{(h)}))}, s^-(m) = \frac{1}{1 + \exp(-b_{k\ell}(\tau^{(l)} - m))}$$

Map to an IFN via safe renormalization to satisfy $\mu + \nu \leq 1$:

$$\hat{\mu} \quad = \frac{\overline{k\ell}}{\max\{1, s^+ + s^-\}}, \hat{\nu} \quad = \frac{\overline{k\ell}}{\max\{1, s^+ + s^-\}}, \hat{\pi} \quad = 1 - \hat{\mu} \quad - \hat{\nu}$$

Within-criterion aggregation (indicators $\ell = 1..L_k$) uses IFWA (intuitionistic fuzzy weighted averaging) with reliability weights $w_{k\ell}(\sum_\ell w_{k\ell} = 1)$:

$$A^{(k)}_{\text{local}} = \langle 1 - \mathsf{G}_{=1} (1 - \hat{\mu} \quad) \quad , \quad \prod_{=1} (\hat{\nu}_{k\ell})^{w_{k\ell}} \rangle \text{ [3,9].}$$

Across the K criteria, compute per-criterion scores $S_k = \mu^{(k)} - \nu^{(k)}$ and aggregate by a Choquet integral C_v with respect to a capacity v: $2^{\mathcal{C}} \to [0,1] (v(\emptyset) = 0, v(\mathcal{C}) = 1)$ to model interaction:

$C_v(S) = \sum_{=1}^{K} (S_{(j)} - S_{(j-1)}) v(A_{(j)})$, where $S_{(1)} \leq \cdots \leq S_{(K)}, S_{(0)} = 0, A_{(j)} = \{(j), \ldots$

Finally, project-level grade is assigned from (,) using zones in the feasible region $\{(S,H): 0 \leq H \leq 1, |S| \leq H\}$ (Fig. 1), which follows from the bijection $\mu = \frac{H+S}{2}, \nu = \frac{H-S}{2}$ and non-negativity constraints [2,3].

Received: July 28, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

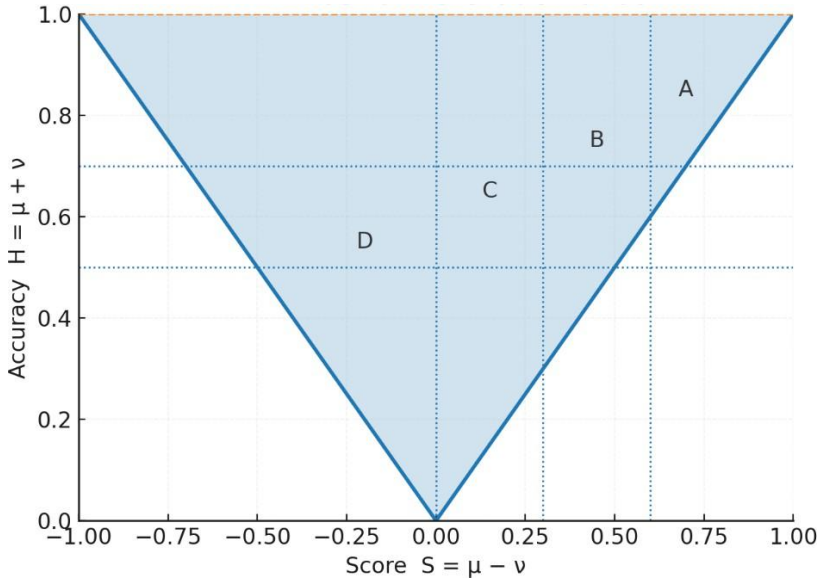

Figure 1. Feasible (S, H) Region for Intuitionistic Fuzzy Numbers with Illustrative Grade Bands

From the above figure 1, The triangular feasible set satisfies $H \in [0,1]$ and $|S| \leq H$; dotted lines show example policy cutoffs for grades A/B/C/D. The instructor adjusts vertical (S) and horizontal (H) thresholds to match course policy.

2. Related Work

2.1 Rubrics and Multimodal Assessment in Higher Education

Rubrics are widely used for formative/summative assessment; rigorous reviews show that carefully designed rubrics can improve learning processes and self-regulation when implemented with transparency and feedback loops [7]. For multimodal artifacts (e.g., visuals, audio, interactivity), multimodality as a social-semiotic framework motivates evaluating meaning made across modes rather than text alone, justifying cross-modal rubrics [8].

2.2 Fuzzy Approaches to Educational Evaluation

Fuzzy rubrics capture the graded nature of descriptors more naturally than crisp cutoffs [1]. However, most classroom applications use type-1 membership only, which cannot explicitly represent hesitation or separate supportive vs. contradictory evidence-central in multi-rater, multi-modal settings [1,2]. Intuitionistic fuzzy sets (IFS) add non-membership and hesitation, supporting richer evidence fusion and clearer audit trails in grading [2,3].

2.3 Intuitionistic Fuzzy Sets, Distances, and Aggregation

IFS were introduced by Atanassov to extend Zadeh's fuzzy sets with (, ,) triplets [1,2]. For comparing IF values and building rankings, score and accuracy functions, and IF aggregation operators (IFWA/IFOWA/IFG) are standard tools in the literature [3,9]. Distances between IFS (e.g., Hamming/Euclidean-type) are also well-studied and useful for reliability screening and calibration.

Received: July 28, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

2.4 Modeling Criterion Interaction: Choquet Integrals

In multimodal projects, some criteria may be redundant (e.g., "visual clarity" and "layout consistency" overlap) or synergistic (e.g., "narrative coherence" and "evidence visualization" reinforce). Such interactions violate additivity; Choquet integrals with respect to a fuzzy measure (capacity) are a principled alternative to weighted sums, offering monotonicity, idempotency, and the ability to represent synergy/redundancy via Shapley-style indices [5].

2.5 Weight Elicitation from Instructors

When instructors provide pairwise importance judgments over criteria, AHP gives a practical route to a weight vector and consistency checks; our IFR adapts this workflow and-when interaction matters-lifts it to capacity identification with modest additional elicitation [6,5].

2.6 Summary of Gaps

Prior work rarely unifies: (i) IFN-based indicator-level scoring with hesitation, (ii) theoretically grounded IF aggregation, and (iii) interaction-aware global fusion for multimodal artifacts-precisely the nexus our IFR targets.

3. Preliminaries And Notation

3.1 Intuitionistic Fuzzy Sets (IFS)

3.1.1 Definition and Basic Quantities.

For a universe , an intuitionistic fuzzy set is

$$A = \{\langle x, \mu_A(x), \nu_A(x)\rangle : x \in X\}, \mu_A, \nu_A : X \to [0,1], \mu_A(x) + \nu_A(x) \leq 1.$$

The hesitation/indeterminacy degree is $\pi_A(x) = 1 - \mu_A(x) - \nu_A(x)$. We will also use the score and accuracy functions for $\langle \mu, \nu \rangle$:

$$S(\langle \mu, \nu \rangle) = \mu - \nu, H(\langle \mu, \nu \rangle) = \mu + \nu$$

IFS generalize type-1 fuzzy sets by explicitly modeling doubt; Atanassov's monograph remains a standard reference [10].

3.1.2 Set Operations via -norms / -conorms.

Let T be a continuous t-norm and S its dual t-conorm. For IFS A, B and any $x \in X$,

$$\mu_{A\cap B}(x) = T(\mu_A(x), \mu_B(x)), \qquad \nu_{A\cap B}(x) = S(\nu_A(x), \nu_B(x)),$$
$$\mu_{A\cup B}(x) = S(\mu_A(x), \mu_B(x)), \qquad \nu_{A\cup B}(x) = T(\nu_A(x), \nu_B(x)),$$
$$A^{\mathrm{C}}(x) = \langle \nu_A(x), \mu_A(x)\rangle.$$

Concrete choices include product/ $S_P(a, b) = a + b - ab$, Łukasiewicz, and Frank/Yager families [11]. These operators preserve $\mu + \nu \leq 1$ and monotonicity under mild conditions [11,12].

3.2 Intuitionistic Fuzzy Numbers (IFNs)

Received: July 28, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

When indicator values $m \in \mathbb{R}$ (e.g., "audio clarity") are scalar, we encode linguistic terms with IFNs whose membership and nonmembership are fuzzy numbers (triangular or trapezoidal) constrained so that $\mu(m) + \nu(m) \leq 1$ for all m[10,12].

A practical parameterization is: $\mu(m) = \text{tri}(m; a_1, a_2, a_3), \nu(m) = \lambda \cdot (1 - \mu(m)) \cdot g(m; \theta)$, with tri the usual triangular function, $0 \leq \lambda \leq 1$, and a shape g (e.g., linear or logistic) chosen to reflect where raters typically express doubt; clipping $\nu(m) \leq 1 - \mu(m)$ guarantees feasibility. Score/accuracy then map $\langle \mu, \nu \rangle$ to (S, H) in the feasible set $\{(S, H): 0 \leq H \leq 1, |S| \leq H\}$ via $\mu = \frac{H+S}{2}, \nu = \frac{H-S}{2}$[10,12].

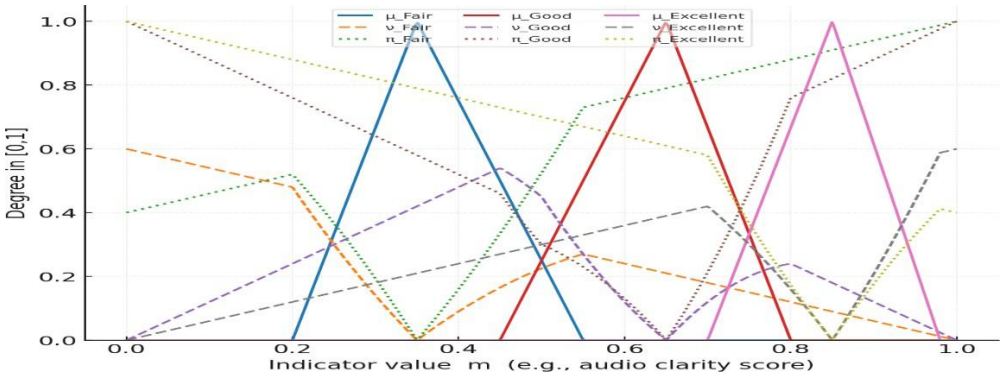

Figure 2. Calibrating Linguistic Terms to IFNs along an Indicator Scale

The above figure 2, For three rubric terms (Fair, Good, Excellent), the curves illustrate $\mu(m), \nu(m)$, and $\pi(m) = 1 - \mu(m) - \nu(m)$ along an indicator scale $m \in [0,1]$. Instructors can shift triangle peaks (a_1, a_2, a_3), the soft-complement factor λ, and $g(m; \theta)$ to match course exemplars.

3.3 Aggregation on IFS/IFNs

3.3.1 IF Weighted Averaging (IFWA) and Geometric (IFWG).

Given IF values $A_j = \langle \mu_j, \nu_j \rangle$ and weights $w_j \geq 0, \sum_j w_j = 1$, Xu's IFWA (using product/ S_P) is

$$\text{IFWA}(\{A_j\}) = \langle 1 - \text{G} \quad (1 - \mu_j) \quad , \quad \rangle$$

The complementary IFWG family is

$$\text{IFWG}(\{A_j\}) = \langle \text{G} \quad , 1 - \text{G} \quad (1 - \nu_j)^{w_j} \rangle.$$

These operators respect idempotency and monotonicity on IFS lattices and are widely used for indicator-level fusion [13,14].

3.3.2 Ordered Weighted Averaging (OWA).

Received: July 28, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

For real inputs $x = (x_1, \dots, x_n)$ sorted so that $x_{(1)} \geq \cdots \geq x_{(n)}$, the OWA is $\mathrm{OWA}_w(x) = \sum_{=1} \quad ( )$, where encodes optimism/pessimism independent of attribute identity. We use OWA on criterion-wise scores to tune compensation when interaction is weak [15].

3.3.3 Choquet Integral (Interaction-Aware Fusion).

When criteria interact (redundancy/synergy), we aggregate scores $S = (S_1, \dots, S_K)$ by a Choquet integral under a capacity $v : 2^{\{1..K\}} \to [0,1] (v(\emptyset) = 0, v(\mathcal{C}) = 1)$:

$$C_v(S) = \sum_{=1} (S_{(j)} - S_{(j-1)}) v(A_{(j)}), S_{(1)} \leq \cdots \leq S_{(K)}, S_{(0)} = 0, A_{(j)} = \{(j), \dots, ($$

Marichal's axiomatization shows is the natural extension of the weighted mean to interacting attributes, preserving comonotone additivity and monotonicity [18].

3.3.4 Distances and Ranking.

For IF values $A = \langle \mu_A, \nu_A \rangle, B = \langle \mu_B, \nu_B \rangle$, a normalized Hamming-type distance is

$$d_{\mathrm{IF}}(A, B) = \frac{1}{2} (|\mu_A - \mu_B| + |\nu_A - \nu_B| + |\pi_A - \pi_B|),$$

useful for rater reliability and calibration. Score/accuracy-based rankings are common, but many variants exist; comparative studies highlight subtle differences that affect ordering and cut-point design in rubrics [16,17].

3.4 Notation (Summary)

- Projects $U = \{p_i\}^N_{=1}$; criteria $\mathcal{C} = \{C_k\}^K_{=1}$; indicators for C_k: $f_{k\ell}$.
- IF value: $A = \langle \mu, \nu \rangle$; hesitation $\pi = 1 - \mu - \nu$; score $S = \mu - \nu$; accuracy $H = \mu + \nu$.
- Local aggregation on indicators: IFWA/IFWG; global aggregation on criteria: OWA or C_v.
- Weights $w \in \Delta$ (simplex); capacity v monotone on $2^{\mathcal{C}}$ with $v(\emptyset) = 0, v(\mathcal{C}) = 1$.

4. Problem Formulation

4.1 Universe and Actors

Let projects $U = \{p_1, \dots, p_N\}$, criteria $\mathcal{C} = \{C_1, \dots, C_K\}$ (e.g., Textual, Visual, Audio, Interaction), and experts/raters $\mathcal{E} = \{e_1, \dots, e \quad\}$. Each criterion has indicators $\{ \quad \}_{=1}$ (e.g., for Audio: signal-to-noise, prosodic fluency, mixing balance). Each indicator produces a measurable value $m_{k\ell}(p_i) \in \mathbb{R}$.

4.2 Indicator → IFN Mapping (Linguistic Calibration)

For each indicator $f_{k\ell}$, we define a monotone map to an IFN $A_{k\ell}(p_i) = \langle \mu_{k\ell}(p_i), \nu_{k\ell}(p_i) \rangle$ by: $( ) = \mathrm{tri}(\quad ( ); {}^{(1)}, {}^{(2)}, {}^{(3)}), \quad ( ) = \quad (1 - \mu \quad ( )) \quad (\quad ( ))$, then clip

to satisfy $\mu_{k\ell} + \nu_{k\ell} \leq 1$. Parameters $(a^{(1)}, a^{(2)}, a^{(3)}, \lambda, \theta)$ are elicited from exemplars ("anchor" projects) and refined by minimizing the average d_{IF} to instructor labels (Section 3.3.4) [17].

Received: July 28, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

4.3 Within-Criterion Aggregation (Indicators → Criterion IFN)

For criterion C_k with indicator IFNs $A_{k1}, \dots, A_{kL_k}$ and reliabilityaware weights $w_{k\ell} \in \Delta$,

$$A^{(k)}_{\text{local}}(p) = \text{IFWA}(\{A\ (p)\}) = \langle 1 - \mathbf{G}\ (1-\mu\ )^{w_{k\ell}}, \mathbf{G}\ \nu^{wk\ell} \rangle$$

or, if geometric compounding is pedagogically preferred,

$$A^{(k)}_{\text{local}}(p) = \text{IFWG}(\{A\ (p)\}) = \langle \mathbf{G}\ \mu^{wk\ell}, 1 - \mathbf{G}\ (1-\nu\ )^{w_{k\ell}} \rangle$$

Both come with idempotency/monotonicity guarantees and are standard for IFS fusion [13,14].

4.4 Across-Criterion Aggregation (Criterion IFNs → Project Score)

Convert criterion IFNs to scores $S_k(p_i) = \mu^{(k)}(p_i) - \nu^{(k)}(p_i)$ and optional accuracies $(\ ) = \ ^{(\ )}(\ ) + \ ^{(\ )}(\ )$. We provide two global fusion options:

(a) OWA on $\{S_k\}$ to control optimism/pessimism in compensation, $\hat{S}_i = \text{OWA}_w(S_1(p_i), \dots, S_K(p_i))$ [15].

(b) Choquet integral C_v to encode interaction via capacity v, $\hat{S}_i = C_v(S(p_i))$. We restrict to 2-additive v for interpretability and parameter economy; capacity identification can be posed as a convex program with monotonicity constraints (details in Methods), grounded in the axiomatization of C_v for interacting criteria [18].

4.5 Expert Reliability and Group Fusion

Let expert e provide indicator-level IFNs $A^{(m)}(p)$. Define a consensus IFN $\bar{A}\ (p)$ (e.g., via IFWA with equal weights), and compute

$$r_m = \exp\left(-\gamma \overline{d_{\text{IF}}\ (A^{(m)}(p), \bar{A}\ (\ ))}\right), \quad \gamma > 0$$

the bar denoting an average over (i, k, ℓ). Use r_m to weight each expert in group aggregation (higher distance ⇒ lower reliability), leveraging IFS distances [17].

4.6 Constraints and Grading Rule

Weight simplex: $\sum_\ell w_{k\ell} = 1, w_{k\ell} \geq 0$; if OWA is used, $\sum_j w_j = 1, w_j \geq 0$ [15]. ACM Digital Library.

Capacity feasibility: $v(\emptyset) = 0, v(\mathcal{C}) = 1, v(S) \leq v(T)$ when $S \subseteq T$ [18]. Acm Digital Library

Decision mapping: From project-level IFN $A_i = \langle \mu_i, \nu_i \rangle$, compute (S_i, H_i). Grades are assigned by policy regions $\mathcal{G}_g \subset \{(S, H): 0 \leq H \leq 1, |S| \leq H\}$ (Fig. 1 in Section 1); thresholds are tuned to maximize agreement with instructors while keeping separability across bands.

4.7 Objective (High-Level)

Estimate calibration parameters Θ (indicator IFN shapes, weights w, and possibly capacity v) by minimizing a rank- and level-aware loss:

Received: July 28, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$$\min_{\Theta} \underbrace{(1-\tau(\{\hat{S}_i\},\{S_i^\star\}))}_{\text{Kendall for rank agreement}} + \alpha \cdot \underbrace{\frac{1}{} \sum_{i} \ell_{\text{grade}}((S,H),g^\star)}_{\text{band-wise misclassification}} + \beta \cdot \Omega(\Theta),$$

where $S^\star$ and $g^\star$ are instructor-provided references, Ω regularizes capacity interaction (e.g., sparsity of pairwise terms), and , trade off accuracy vs. interpretability [12,13,18].

5. Intuitionistic Fuzzy Rubric Construction

5.1 Designing Criterion-Level Indicators

5.1.1 Indicator families and measurement functions.

For each criterion C_k (Textual, Visual, Audio, Interaction), define indicator functions $f_{k\ell}: U \to \mathbb{R}$ with monotonic directionality (+ is better). Typical examples: argument coherence score, color-contrast ratio, signal-to-noise (SNR), task-success rate in a prototype. We rescale each raw indicator to [0,1] via an affine map $(p) = \frac{x-\min}{\max-\min}$ or a logistic transform when heavy-tailed.

5.1.2 Evidence splitting into support vs. doubt.

Given $m = m_{k\ell}(p)$, we separate support $s^+(m)$ and counterevidence $s^-(m)$ by smooth thresholds $\tau^{(h)} > \tau^{(l)}$:

$$s^+(m) = \frac{1}{1+\exp(-a\ (m-\tau^{(h)}))}, s^-(m) = \frac{1}{1+\exp(-b_{k\ell}(\tau^{(l)}-m))}$$

This mirrors rater cognition where both favorable and unfavorable cues coexist.

5.2 Linguistic → IFN Calibration

5.2.1 Parametric IFN per term.

For rubric term $t \in$ { Poor,..., Excellent } on indicator $f_{k\ell}$, define an IFN:
$\mu(m) = \text{tri}(m; a^{(1)}, a^{(2)}, a^{(3)}), \nu(m) = \lambda(1-\mu(m))g(m;\theta)$, then clip $\nu(m) \le 1 -$

$\mu_t(m)$ to ensure feasibility, and set $\pi_t(m) = 1 - \mu_t(m) - \nu_t(m)$. The shape g_t (linear/logistic) positions where doubt typically arises.

5.2.2 Data-driven fit to anchors.

Let $\mathcal{A}$ be anchor exemplars labeled by instructors with target terms $t^\star$. Estimate $\Theta =$ { ( ), , } by minimizing the intuitionistic distance to anchors:

$$\min_{\Theta} \frac{1}{|\mathcal{A}|} \sum_{(p,t^\star)\in\mathcal{A}} d_{\text{IF}}(\langle \mu_{t^\star}(m(p)), \nu_{t^\star}(m(p))\rangle, \langle \mu_\Theta(m(p)), \nu_\Theta(m(p))\rangle) + \eta\mathcal{R}(\Theta),$$

with $_{\text{IF}}(\ ,\ ) = \frac{1}{2}(|\ -\mu_B| + |\ -\nu_B| + |\ -\pi_B|)$ and $\mathcal{R}$ an ordering regularizer enforcing $^{(1)} < {}^{(2)} < {}^{(3)}$ and separations between adjacent terms. Similar calibration is routine in decision analysis and aligns with non-additive measurement design [19,23].

Received: July 28, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

5.3 Weight Estimation Methods (Indicator and Criterion Levels)

5.3.1 Intuitionistic AHP-style elicitation (pairwise).

For indicators under criterion , instructors give pairwise judgments ( ) (importance of vs). Using the classical AHP scaling with a consistency index $CI = \frac{\lambda_{\max} - n}{n-1}$ and ratio $CR =$ / [20], we extract a priority vector by minimizing the (log) least-squares deviation from reciprocal comparisons; with intuitionistic inputs, each judgment can be encoded as an IFN interval and collapsed to a centroid before solving (keeps the familiar AHP workflow) [21].

5.3.2 Entropy-based adaptive refinement.

To reduce over-reliance on prior weights, we adapt them using the Shannon entropy of realized indicator scores across projects:

$$E_{k\ell} = -\sum_{=1} p_{k\ell b} \log\, p_{k\ell b}, \tilde{w}_{k\ell} = \frac{(1 - E_{k\ell}/\log\, B) w_{k\ell}}{\sum\, (1 - E_{kr}\ /\log\ )}.$$

Indicators that discriminate more (lower entropy) get slightly higher weight, a standard device in information-theoretic weighting [19].

5.3.3 Expert-reliability incorporation.

If expert provides indicator IFNs $^{( )}$ ( ), define a reliability

$$r_m = \exp\,(-\gamma d_{\mathrm{IF}}(A^{( )}(p), \bar{A}_{k\ell}(p))), \gamma > 0$$

averaged over , , , and use as a multiplicative factor on . This is consistent with robust aggregation principles [24].

5.4 Cross-Modal Interaction Modeling (Choquet Capacity)

5.4.1 2-additive capacity via Möbius transform.

Let v be a capacity on $\mathcal{C}$. In the **2**-additive setting (good trade-off between interpretability and richness), the Möbius transform $\{m_i, m_{ij}\}$ suffices:

$$v(A) = \sum_{i \in A} m_i + \sum_{\{i,j\} \subseteq A} m_{ij}, A \subseteq \mathcal{C},$$

with constraints

$$m_i \geq 0, m_{ij} \geq -\min(m_i, m_j), \sum m_i + \sum_{<} m_{ij} = 1$$

ensuring monotonicity and ( ) = 1 [23,25]. The interaction index between and is =

(positive = synergy, negative = redundancy).

5.4.2 Identifying from partial rankings.

Received: July 28, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Given instructor pairwise preferences $\mathcal{P} = \{(p \succ q)\}$, we estimate $m = \{m_i, m_{ij}\}$ by the convex program:

$$\min_{m,\xi \geq 0} \sum_{(p \succ q) \in \mathcal{P}} \xi_{pq} + \lambda \sum \quad | \quad |$$
$$< \quad \text{s.t. } C_v(S(p)) - C_v(S(q)) \geq 1 - \xi_{pq}, \forall (p \succ q) \in \mathcal{P}$$

satisfies 2-additive capacity constraints, where is the Choquet integral and the $_1$ penalty sparsifies interactions for interpretability. This is consistent with learning nonadditive measures from ordinal data [23,24].

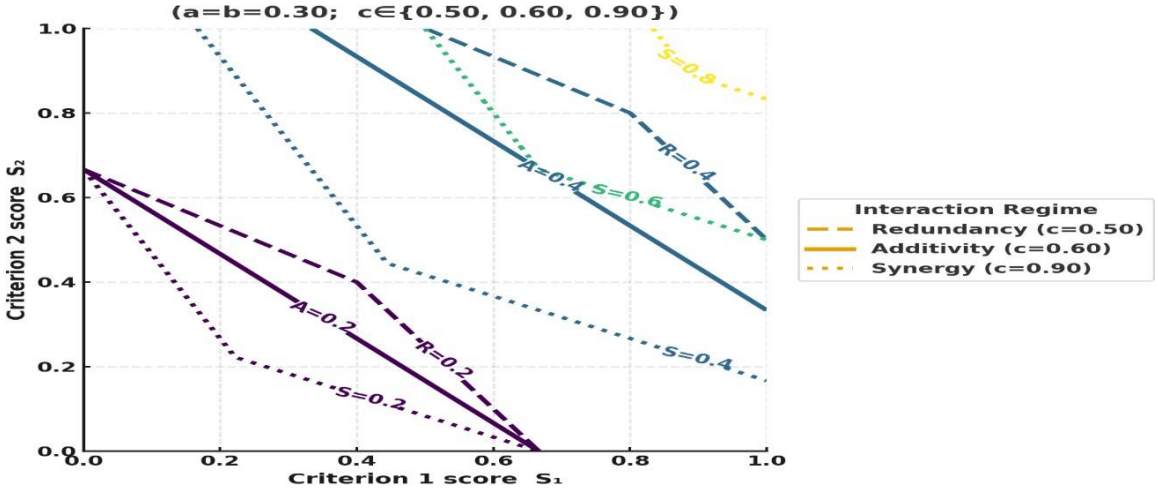

Figure 3. Choquet Integral ($_1$, $_2$) : Redundancy vs. Additivity vs. Synergy

From the above Figure 3, Iso-score contours for two criteria with $(\{1\}) = (\{2\}) = 0.30$ and $v(\{1,2\}) \in \{0.50, 0.60, 0.90\}$. Dashed contours (redundancy) require higher joint scores to reach the same aggregate; dotted (synergy) reward balanced improvements.

5.4.3 Shapley importance for feedback.

Criterion importance communicated to students is given by the Shapley value of the capacity:

$$\phi_i = \sum_{A \subseteq C \setminus \{i\}} \frac{|A|!\,(K - |A| - 1)!}{!} (v(A \cup \{i\}) - v(A)),$$

which reduces to $= +\frac{1}{2}\sum_{j \neq i}$ for 2-additive capacities [23]. This supports actionable feedback ("visual design mattered most this term").

5.5 Reliability of Experts and Annotations

5.5.1 Group fusion of raters.

For each indicator, aggregate M expert IFNs $\{A^{(m)}\}$ with IFWA using weights $\omega_m \propto r_m$:

$$A^{-} = \langle 1 - \mathbf{G}\ (1 - \mu^{(m)})^{\omega_m}, \mathbf{G}\ (\nu^{(m)})^{\omega_m} \rangle$$

Flag an expert as an outlier on indicator if

Received: July 28, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$$\overline{d_{\mathrm{IF}}(A^{(\ )}, \bar{A}_{k\ell})} > \tau_{\mathrm{out}},$$

and re-estimate after removal.

5.5.2 Consistency and calibration loop.

We monitor: (i) AHP consistency at elicitation [20]; (ii) ranking agreement between -based order and instructor order; (iii) band stability of (,) grade regions across resampling. Entropy weights (5.3.2) and capacity are updated until convergence or until a maximum of iterations.

6. AGGREGATION AND SCORING PIPELINE

6.1 Local Aggregation (Within Criterion)

6.1.1 Inputs and reliability.

For criterion C_k with indicators $f_{k\ell}(\ell = 1, \dots, L_k)$, each project p yields IFNs $A_{k\ell}(p) = \langle \mu_{k\ell}(p), \nu_{k\ell}(p) \rangle$ (Sections 3-4). If multiple experts annotate indicators, form a reliability for expert e_m,

$$r_m = \exp\left(-\gamma \overline{d_{\mathrm{IF}}(A^{(\ )}(p), \bar{A}_{k\ell}(p))}\right), \gamma > 0$$

and fold it into indicator weights (normalize so $\sum_\ell w_{k\ell} = 1$). This controls the effect of noisy raters using an intuitionistic distance [17].

6.1.2 IFWA/IFWG fusion.

We combine indicator IFNs with IFWA (product/) or IFWG [13]. For IFWA:

$$A^{(k)}_{\mathrm{local}}(p) = \langle 1 - \mathbf{G}_{=1} (1 - \mu ( )) , \quad ((p))^{w_{k\ell}} \rangle$$

Boundary/consistency: If all $A_{k\ell}(p) = \langle \mu, \nu \rangle$, then $A^{(\ )}_{\mathrm{local}}(p) = \langle \mu, \nu \rangle$ (idempotency) and the operator is monotone/bounded on IFS lattices [12,13]. The per-criterion score and accuracy follow as

$$S_k(p) = \mu^{(k)}(p) - \nu^{(k)}(p), H_k(p) = \mu^{(k)}(p) + \nu^{(k)}(p)$$

6.2 Global Aggregation (Across Criteria)

Two principled options are provided, depending on whether interaction among criteria matters.

6.2.1 OWA (compensatory control without interaction).

When criteria are largely independent, aggregate scores $\{ \ \}_{=1}$ via an Ordered Weighted Average (OWA) with policy vector w ($\sum_j w_j = 1$) [15]:

$$\hat{S}(p) = \mathrm{OWA}_w(S_1(p), \dots, S_K(p)) = \sum_{=1} w_j S_{(j)}(p), S_{(1)} \geq \cdots \geq S_{(K)}$$

Received: July 28, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Similarly, accuracy can use an OWA (often slightly optimistic to avoid punishing a single weak modality):

$$\hat{H}(p) = \mathrm{OWA}_{w^-}(H_1(p), \ldots, H_K(p))$$

6.2.2 Choquet integral (interaction-aware fusion).

When criteria exhibit redundancy/synergy (e.g., Textual-Visual coherence), use the Choquet integral with a capacity v: $2^{\mathcal{C}} \to [0,1](v(\emptyset) = 0, v(\mathcal{C}) = 1)$ [18,23,24]. Sorting S in non-decreasing order $S_{(1)} \leq \cdots \leq S_{(K)}$ with $A_{(j)} = \{(j), \ldots, (K)\}$,

$$\hat{S}(p) = C_v(S(p)) = \sum_{=1} (S_{(j)} - S_{(j-1)}) v(A_{(j)}), S_{(0)} = 0$$

For interpretability we adopt 2-additive capacities with Möbius parameters $\{m_i, m_{ij}\}$ (Section 5.4), where $I_{ij} = m_{ij}$ quantifies pairwise interaction; the Shapley importance returned to students is $= +\frac{1}{2}\sum_{j \neq i}$ [23].

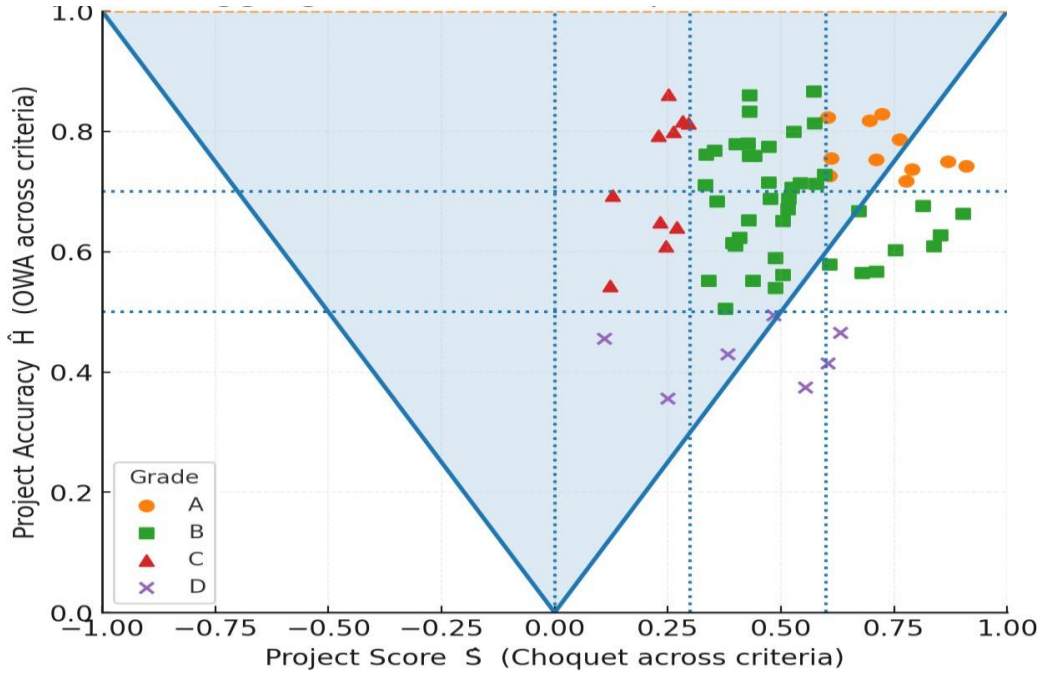

Figure 4. End-to-End Aggregation: IFWA → Choquet/OWA → Grade in (S, H)

The above figure 4, Simulated projects (3 criteria, 2 indicators each). Within each criterion, indicators are fused by IFWA; across criteria, scores use a 2-additive Choquet integral (Textual-Visual synergy) while accuracies use OWA. Points lie in the feasible (S, H) triangle with illustrative grade lines $S \in \{0,0.3,0.6\}$ and $H \in \{0.5,0.7\}$. Simulated 70 projects, 3 criteria, 2 indicators per criterion. Within-criterion IF aggregation used IFWA; acrosscriterion fusion used Choquet (scores) and OWA (accuracies). Markers denote grades under policy thresholds. This serves as a template for classroom reporting.

6.3 Project-Level Score and Grade

Received: July 28, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

We map ($\hat{S}, \hat{H}$) to a categorical grade using policy regions $\{\mathcal{G}_{\mathrm{A}}, \mathcal{G}_{\mathrm{B}}, \ldots\}$ inside the feasible set $\{(S,H): 0 \le H \le 1, |S| \le H\}$ (Section 1). One simple, interpretable rule is:

$$g(\hat{S},\hat{H}) = \begin{cases} \mathrm{A}, & \hat{S} \ge s_3, \hat{H} \ge h_2 \\ \mathrm{B}, & \hat{S} \ge s_2, \hat{H} \ge h_1 \\ \mathrm{C}, & \hat{S} \ge s_1, \hat{H} \ge h_1 \\ \mathrm{D}, & \text{otherwise} \end{cases}$$

with cut points $(s_1, s_2, s_3) = (0,0.3,0.6)$, $(h_1, h_2) = (0.5,0.7)$ by default (tuned to the course). Different courses can adjust these to align grade bands with learning outcomes while keeping feasibility $|S| \le H$[10,12].

6.3.1 Optional global IF aggregation.

If desired, one may also aggregate criterion IFNs directly by IFWA/IFWG to obtain $\langle \hat{\mu}, \hat{\nu} \rangle$ at the project level (then $\hat{S} = \hat{\mu} - \hat{\nu}$, $\hat{H} = \hat{\mu} + \hat{\nu}$). We favor Choquet-on- S + OWA-on- H for clearer control of interaction (in S) and conservatism (in H).

6.4 Theoretical Properties (Sketches)

Proposition 1 (Local aggregator properties).

Let $A_1, \ldots, A_L \in$ IFS and weights $w \in \Delta$. The IFWA/IFWG outputs A_{local} satisfy (i) idempotency: if $A_1 = \cdots = A_L = A$, then $A_{\text{local}} = A$; (ii) monotonicity: if $A_\ell \le A'$ (componentwise μ up, ν down) for all ℓ, then $A_{\text{local}} \le A'_{\text{local}}$; (iii) bounds: $\min A_\ell \le A_{\text{local}} \le \max$ in the IFS lattice. Sketch: These follow by isotonicity and boundary behavior of the product / pair and standard aggregation-function axioms [12,13].

Proposition 2 (Global monotonicity).

Let v be a monotone capacity. If $S \le T$ (componentwise), then $C_v(S) \le C_v(T)$; likewise, for any OWA with nonnegative w summing to 1, $\mathrm{OWA}_w(S) \le \mathrm{OWA}_w(T)$. Sketch: Choquet monotonicity is classical from the definition via sorted differences with nondecreasing set weights $v(A_{(j)})$; OWA monotonicity is immediate from majorization theory for nonnegative weights [15,18,23].

Theorem 1 (Dominance-preserving ranking under $S - H$ lexicographic rule).

Consider two projects p, q. If $S_k(p) \ge S_k(q)$ for all k and at least one strict inequality, then for any monotone capacity v,

$$\hat{S}(p) = C_v(S(p)) > C_v(S(q)) = \hat{S}(q).$$

If additionally, $H_k(p) \ge H_k(q) \forall k$ with at least one strict inequality, then for any OWA with nonnegative weights $\tilde{w}$, $\hat{H}(p) > \hat{H}(q)$. Thus, the lexicographic decision ($\hat{S}, \hat{H}$) preserves Pareto dominance. Sketch: Combine Proposition 2 with the Pareto improvement assumption and lexicographic tie-break [12,15,18,23,24].

Received: July 28, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Corollary (Feasibility of grades).

Since each criterion IFN satisfies $\mu + \nu \leq 1$ and both IFWA/IFWG and Choquet/OWA are bounded and monotone, the final ($\hat{S}, \hat{H}$) always lies in the feasible triangle $|\hat{S}| \leq \hat{H} \leq 1$; hence grade regions can be chosen inside this set without inconsistency [10,12,13].

7. PARAMETER LEARNING AND CALIBRATION

7.1 Learning Membership/Non-Membership Mappings

7.1.1 Objective for indicator calibration.

Let $m_{k\ell}(p) \in [0,1]$ be the rescaled indicator value and $y_{k\ell}(p) \in \{t_1, \ldots, t_T\}$ the instructor's linguistic term (e.g., Poor...Excellent) for anchor exemplars. We parametrize IFNs per term t by $\Theta = \{a^{(1)}, a^{(2)}, a^{(3)}, \lambda, \theta\}$ (Section 5.2) and minimize a calibration risk plus regularization:

$$\min_{\Theta} \frac{1}{|\ |} \sum_{(p,k,\ell)\in\mathcal{A}} d_{\mathrm{IF}}(\langle \mu_{t^\star}, \nu_{t^\star}\rangle, \langle \mu_\Theta, \nu_\Theta\rangle) + \eta\mathcal{R}(\Theta),$$

with $_{\mathrm{IF}}(\ ,\ ) = \frac{1}{2}(|\quad - \mu_B| + |\quad - \nu_B| + |\quad - \pi_B|)$. Constraints: (i) triangular peaks ordered $a_t^{(1)} < a_t^{(2)} < a_t^{(3)}$; (ii) feasibility $\nu_\Theta(m) \leq 1 - \mu_\Theta(m) \forall m$; (iii) monotonicity of μ_Θ w.r.t. quality (can be enforced via linear inequalities on finite knots or by isotonic regression on $\{(\ ,\ )\}$) [12,31].

7.1.2 Ordinal/logistic alternatives.

If anchors come as ordered categories, the mapping $m \mapsto t$ can be co-estimated with proportional-odds ordinal regression to produce latent utilities $u(m)$ that inform $\mu(m) = \sigma(\alpha u(m) + \beta)$ and $\nu(m) = \lambda(1 - \mu(m))g(m;\theta)$[31]. This keeps ordinal structure and reduces label noise.

7.1.3 Reliability-aware fitting.

When multiple experts provide anchors, weight each observation by the rater reliability (Section 5.3.3) inside the loss-equivalent to a heteroskedastic M-estimator, giving less influence to outliers [24].

7.2 Capacity Learning for Choquet Aggregation

7.2.1 2-additive capacity identification.

Let $S(p) \in \mathbb{R}^K$ be criterion scores for project p and let $\mathcal{P}$ be pairwise preferences from instructors (e.g., $p \succ q$). Using the 2additive Möbius parameters $\{m_i, m_{ij}\}$ (Section 5.4), learn a capacity by the convex program:

$$\min_{m,\xi\geq 0} \sum_{(p\succ q)\in\mathcal{P}} \xi_{pq} + \lambda \sum_{<} |\quad|$$
$$\text{s.t. } C_v(S(p)) - C_v(S(q)) \geq 1 - \xi_{pq}, \forall (p \succ q) \in \mathcal{P}$$

Received: July 28, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

satisfies 2-additive capacity constraints (Section 5.4.1).

The hinge slack encodes a large-margin preference fit [27], while $_1$ on yields sparse interactions for interpretability [29,30]. This is a disciplined convex program solvable by off-the-shelf solvers or ADMM [31,32].

7.2.2 Rank-based alternative losses.

When only ranks are given, maximize Kendall's between (( )) and instructor ranks; a smooth surrogate (pairwise logistic) allows gradient methods [26,27]. Cross-validation on is recommended.

7.2.3 Shapley-guided regularization.

Impose $\Omega_{\text{Shapley}} = \sum_i (\phi_i - \phi^{\text{target}})^2$ to keep criterion importance near instructor intuition, where $= + \frac{1}{2}\sum_{j \neq i}$ for 2 additive capacities [23,32].

7.3 Model Selection and Hyperparameter Tuning

7.3.1 Validation metrics.

We jointly evaluate (i) rank agreement on held-out projects [26], (ii) grade accuracy w.r.t. policy bands, and (iii) distance $_{\text{IF}}$ to expert IFNs. Multi-objective tuning uses a scalarization $J = \alpha(1-\tau) + \beta \text{ Err }_{\text{grade}} + \gamma \overline{d_{\text{IF}}}$ with $\alpha + \beta + \gamma = 1$ [33,34].

7.3.2 Cross-validation protocol.

Use stratified K-fold by grade band to preserve class balance; if preferences are used, split pairs by project IDs to avoid leakage [35,36]. Report mean $\pm$ SD across folds, and lock seeds for replicability.

7.3.3 Threshold calibration.

Map continuous $(\hat{S}, \hat{H})$ to grade bands via a held-out isotonic fit or Platt-style calibration if you need probability of reaching a band; evaluate by Brier score or ECE [36,37,38,39].

7.4 Diagnostics Figures

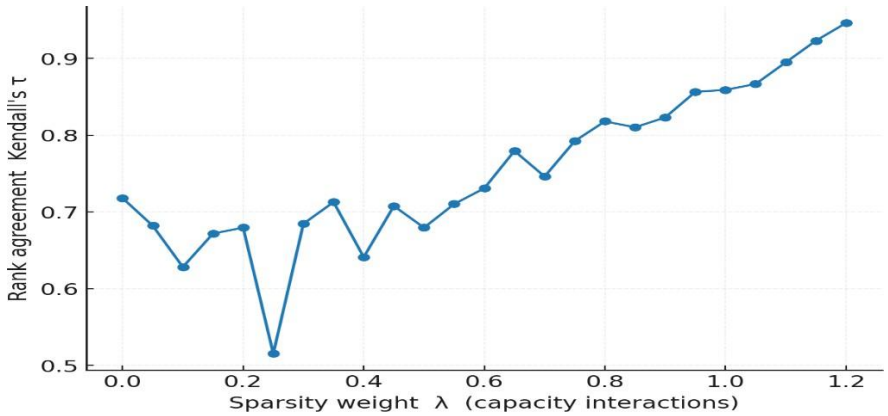

Figure 5. Model Selection Curve: Kendall's vs. (interaction sparsity)

Received: July 28, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

From the above figure 5, As increases, interaction sparsity grows. An interior optimum typically balances variance (small) and bias (large).

8. ALGORITHM

8.1 End-to-End Procedure (Pseudocode)

8.1.1 Training (calibration + capacity/weights).

Algorithm 1 - Train-IFR

(1) Input: dataset $\mathcal{D} = \{(p, \{m_{k\ell}(p)\})\}$, anchor labels $\mathcal{A}$, optional preferences $\mathcal{P}$; folds $\mathcal{F}$; hyperparameters η, λ.

(2) Initialize IFN params Θ (triangles and λ_t, θ_t) per indicator/term; indicator weights $w_{k\ell}$; capacity Möbius params $m = \{m_i, m_{ij}\}$.

(3) For each fold $f \in \mathcal{F}$:

(3.1) Fit IFN mappings by solving the calibration problem in section 7.1 on training fold (with reliability weights if multi-rater).

(3.2) Compute indicator IFNs ( ) on validation fold; aggregate to criterion IFNs via IFWA/IFWG (Section 6.1).

(3.3) Compute criterion scores $S_k(p) = \mu^{(k)} - \nu^{(k)}, H_k(p) = \mu^{(k)} + \nu^{(k)}$.

(3.4) If interactions matter: learn capacity by solving the convex program in section 7.2.1 (or maximize with a smooth surrogate).

(3.5) Calibrate grade thresholds $(s_1, s_2, s_3, h_1, h_2)$ on validation set (isotonic/Platt) (Section 7.3.3).

(3.6) Record metrics , $\text{Err}_{\text{grade}}$, $\overline{\text{IF}}$.

(4) Select hyperparameters by minimizing $J = \alpha(1 - \tau) + \beta$ Err $_{\text{grade}} + \gamma \overline{d_{\text{IF}}}$.

(5) Refit on full data using chosen hyperparameters; store $\Theta^\star, w^\star, m^\star$, thresholds *.

(6) Output: trained model $\mathcal{M} = (\Theta^\star, w^\star, m^\star$, thresholds $^\star)$.

8.1.2 Inference (per cohort or student).

Algorithm 2 - Score-and-Grade

- Input: trained $\mathcal{M}$, measurements $\{m_{k\ell}(p)\}$.
- Compute indicator IFNs with $\Theta^\star$; aggregate to criterion IFNs by IFWA/IFWG.
- Scores/accuracies: S_k, H_k.
- Global fusion: $\hat{S} = C_{v^\star}(S)$ (Choquet) and $\hat{H} = \text{OWA}_{w^\star}(H)$.
- Grade: apply thresholds to ($\hat{S}, \hat{H}$) to produce g and feedback via Shapley ϕ_i and interactions *.

Received: July 28, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

- Output: $(\hat{S}, \hat{H}, g)$ plus feedback vector.

8.2 Computational Complexity

Let projects, criteria, indicators per criterion, $=$ max

- Indicator → IFN: $O(N \sum_k L_k)$.
- Local IF aggregation (IFWA/IFWG): $O(N \sum_k L_k)$.
- Global fusion: OWA is $O(NK\log K)$ (sorting); Choquet with 2-additive capacity is (log) for evaluation (sorting + set lookups).
- Capacity learning: number of parameters $d = K + \frac{K(K-1)}{2}$. Interior-point complexity scales roughly as (^{3}) per iteration, but the constraint matrix is sparse; first-order or ADMM methods reduce cost to nearly linear per pass in and | |[29,30].
- Cross-validation: multiply by number of folds F.

8.3 Implementation & Reproducibility Checklist

Data splits: stratified -fold by grade; keep cohort-wise leakage out.

Seeds: fix random seeds for calibration and optimization.

Hyperparameters: report , , ; OWA weights; threshold vectors; solver tolerances.

Ablations: No-interaction (OWA only), No-audio/No-visual, No-reliability, OWA vs. Choquet.

Transparency artifacts: release (i) IFN term tables, (ii) capacity parameters { , }, (iii) Shapley importances , (iv) code to reproduce Figures 5-6 & 4, and (v) anonymized exemplars with consent.

Ethics: ensure rubric fairness via subgroup stability checks and publish calibration procedures.

9. VALIDATION PROTOCOL

9.1 Dataset Design

9.1.1 Course & artifact composition.

We evaluate on N multimodal projects spanning K criteria (Textual, Visual, Audio, Interaction), each with L_k indicators (Sections 3-6). Indicator measurements $m_{k\ell}(p) \in [0,1]$ are derived from normalized rubrics and instrumented analytics (e.g., SNR for Audio; coherence/argumentation scores for Textual).

9.1.2 Splits and blinding.

Use stratified -fold CV by instructor grade band to preserve class balance; anchors and pairwise preferences are split by project ID to prevent leakage between training and validation folds (Sections 7.2-7.3). Hyperparameters are chosen on inner validation and frozen before test reporting.

Received: July 28, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

9.1.3 Baselines.

Crisp rubric: weighted sum of z-scored indicators.

Type-1 fuzzy rubric: membership-only aggregation (no ,).

IFR (ours): IFS with IFWA/IFWG locally, OWA or Choquet globally (Sections 6-7).

All models share the same preprocessed indicators and the same grade policy thresholds for fairness of comparison.

9.2 Metrics

9.2.1 Agreement on order and levels.

Rank agreement: Kendall's τ between model and instructor orderings $\tau(\hat{S}, S^{\star})$, and Spearman's ρ on real-valued scores, where $\rho = 1 - \frac{6\sum_i d^2}{n(n^2-1)}$ for rank differences [26,40].

Grade accuracy & macro-F1: micro accuracy, per-band precision/recall/F1, and macro-F1 over {A, B, C, D}.

Calibration (optional): if probabilistic band predictions are produced (Section 7.3.3), report Brier score and expected calibration error (ECE) [33].

9.2.2 Distance to expert labels.

- Intuitionistic fit: mean d_{IF} between model IFNs and expert IFNs at indicator or criterion level (Sections 3.3.4, 7.1).

9.2.3 Inter-rater reliability & uplift.

Report Krippendorff's (all scales), Fleiss' (nominal), and weighted Cohen's (ordinal) for (i) experts among themselves and (ii) expert vs. model; the IFR should increase reliability compared to crisp baselines.

9.2.4 Statistical testing & CIs.

Paired comparison of error rates: McNemar's test on band misclassifications between IFR and baseline.

Confidence intervals: nonparametric bootstrap (1,000-5,000 resamples) for , macro- F 1, and IF, reporting percentile 95% CIs.

Multiple metrics: control family-wise false discoveries via Benjamini-Hochberg at $= 0.10$ when comparing several model variants.

9.3 Sensitivity and Robustness

9.3.1 Parameter perturbations.

Stress the system with controlled changes and measure grade shift rate Δ_g (percentage of projects whose final grade differs from the reference):

- Threshold shifts: $s_2 \to s_2 \pm \delta_s, h_1 \to h_1 \pm \delta_h$.

Received: July 28, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

- OWA optimism: $\tilde{w}$ moved toward $(1,0,\ldots)$ or uniform.
- Capacity interaction: 2-additive Möbius terms m_{ij} perturbed $\pm\delta$ while keeping monotonicity (Section 5.4).
- Missing modalities: randomly drop one modality for a fraction η of projects and re-score (imputing IFNs with high π).

9.3.2 Ablations.

Evaluate No-interaction (OWA only), No-audio/No-visual, Noreliability, and Type-1 only to quantify the marginal contribution of each design choice (Sections 6-7).

9.3.3 Stability diagnostics.

Plot fold-wise distributions of , macro- F1, and IF; report median and interquartile ranges. Use bootstrap CIs to quantify uncertainty on every metric.

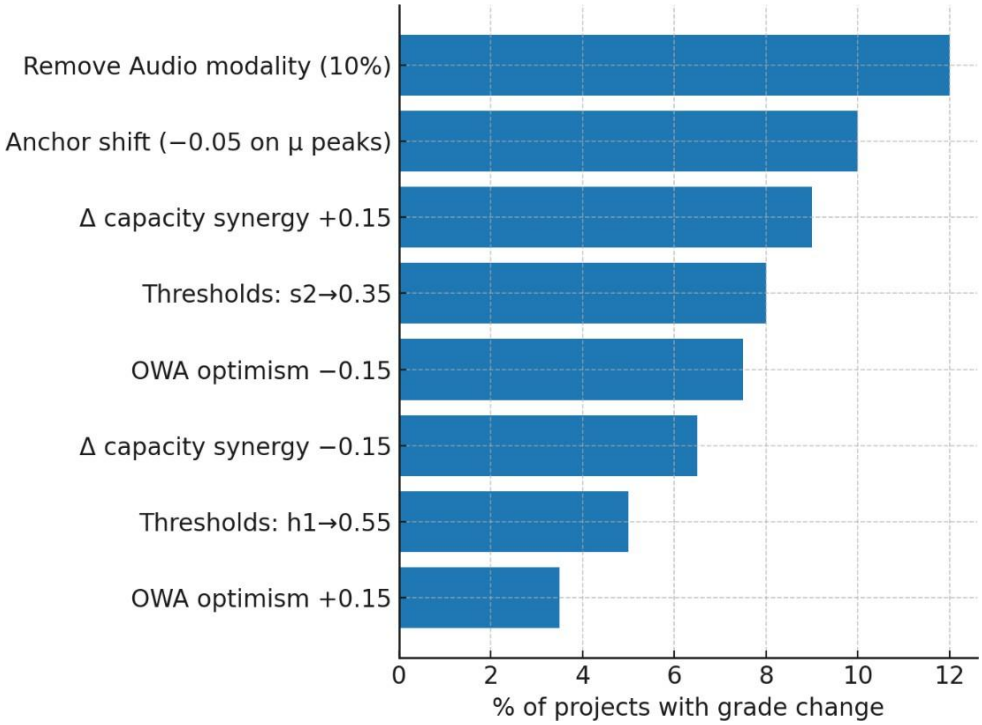

Figure 6. Robustness Tornado: Grade Sensitivity to Parameter Variations

From the above figure 6, Horizontal bars show % of projects changing grade under single-factor perturbations (e.g., capacity synergy, OWA optimism, threshold tweaks, anchor shifts).

9.4 Fairness and Bias Checks

9.4.1 Subgroup stability.

Let G denote subgroups (e.g., section, topic choice, language background). For each group $g \in G$, compute $\hat{S}_g, \hat{H}_g$ distributions and band rates; compare via two-sample tests (KS or MannWhitney) and report effect sizes.

9.4.2 Equalized-error style reporting.

Received: July 28, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

For each , report per-band error rates and misclassification cost symmetry. Investigate disparities; where large, perform counterfactual checks by swapping modalities or normalizing indicators to detect source factors. Use fairness literature guidelines to document disparate error patterns and mitigations.

9.5 Reproducible Reporting Template

Data card: cohort, tasks, indicator provenance, privacy/consent.

Model card: IFN tables; capacity { , }; Shapley importances . ***Validation***

summary: , , macro-F1, accuracy, IF, Cls; McNemar -values. ***Sensitivity***

bundle: tornado plot (Fig. 7), fold stability, ablations.

Fairness sheet: subgroup metrics, disparities, mitigations, reviewer checklist.

10. Case Study (for Illustrative study)

Provided below the classroom-style dataset with N = 24 multimodal projects, K=4 criteria (Textual, Visual, Audio, Interaction), and = indicators per criterion (e.g., Textual: Coherence, Evidence, Citations). All tabulated results are listed in the below section 10.4.

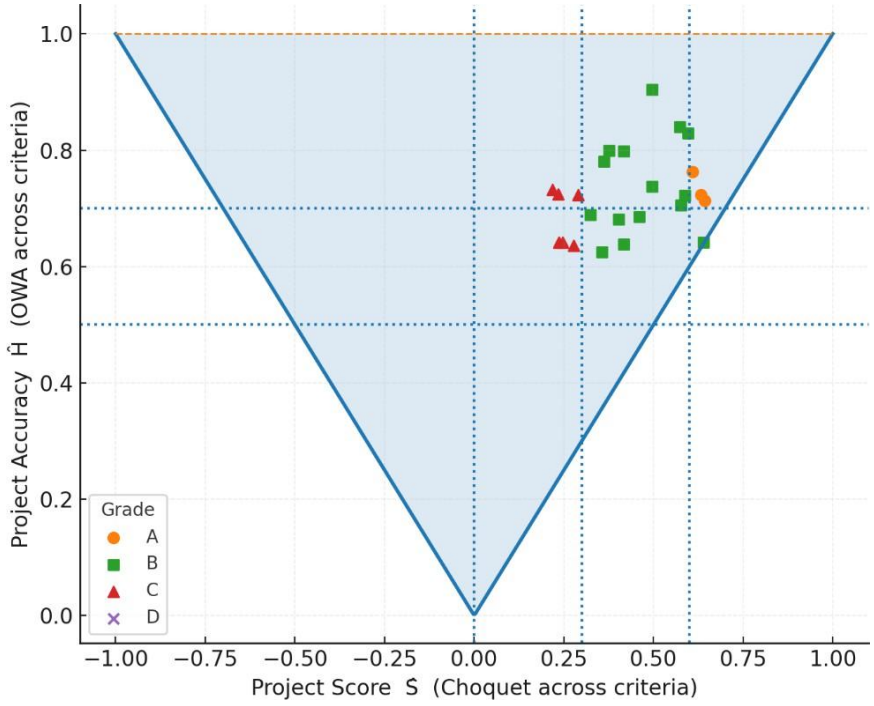

Figure 7. Case Study - Final ($\hat{S}, \hat{H}$) with Grade Regions

From the above figure 7, Each point is a project; dashed vertical/horizontal lines mark the policy thresholds $\{s_1, s_2, s_3\} = \{0,0.30,0.60\}$ and $\{h_1, h_2\} = \{0.50,0.70\}$. The feasible triangle $|S| \leq H \leq 1$ is shaded.

10.1 Experimental Setup

10.1.1 Indicators and preprocessing.

Received: July 28, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Each indicator value x is min-max normalized to $m \in [0,1]$. The indicator families per criterion are:

- Textual: Coherence, Evidence, Citations
- Visual: Layout, Contrast, Data Visualization
- Audio: SNR, Prosody, Mix Balance
- Interaction: Task Success, Latency, Usability

10.1.2 Linguistic → IFN mapping (per Section 5.2).

For each indicator term we use triangular membership and a soft complement for non-membership:

$$\mu(m) = \mathrm{tri}(m; a_1, a_2, a_3), \qquad \nu(m) = \lambda(1 - \mu(m))g(m; \theta),$$
$$\pi(m) = 1 - \mu(m) - \nu(\mathrm{m})$$

with $\lambda = 0.6$. The shape $g(\cdot)$ reflects where doubt arises: low-value doubt for Textual/Visual, mid-range doubt for Audio, extreme-value doubt for Interaction (see Section 5.2 and Fig. 2).

10.1.3 Local aggregation (IFWA) and global fusion (Choquet+OWA).

Within each criterion , indicators are fused by IFWA:

$$A^{(k)}_{\text{local}} = \langle 1 - \mathrm{G} \quad (1 - \mu \quad) \quad , \qquad \nu^{wk\ell} \rangle , w \quad = \frac{1}{}.$$

Across criteria, scores $S_k = \mu^{(k)} - \nu^{(k)}$ are aggregated by a 2additive Choquet integral C_v with Möbius parameters $\{m_i, m_{ij}\}$ (Section 5.4), while accuracies $H_k = \mu^{(k)} + \nu^{(k)}$ use an OWA with $\tilde{w} = (0.4,0.3,0.2,0.1)$.

10.1.4 Capacity used in this cohort (interpretability).

We encode expected Textual-Visual and Textual-Interaction synergies and a mild Visual-Audio redundancy:

$= (0.20,0.20,0.20,0.20)$,

m_{ij}: $(T, V) = 0.12$, $(T, I) = 0.08$, $(V, A) = -0.05$, $(T, A) = 0.02$, $(V, I) = 0.02$, $(A, I) = 0.01$,

so $\sum_i m_i + \sum_{i<j} m_{ij} = 1$. Shapley importances computed from m (2-additive case) are

$= (0.310,0.245,0.190,0.255)$ for (Textual, Visual, Audio, Interact),

which we expose to students for feedback. (See "Case Study Shapley Importance per Criterion" table.)

10.2 Full Calculations (Worked Example for P1)

We show the exact numbers for project P1; the full cohort tables contain every project's intermediate values.

10.2.1 Indicator-level IFN (Textual → Coherence).

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

From the dataset (table: Exemplar P1 - Indicator-level IFNs):

$$= 0.6061, (\ _1,\ _2,\ _3) = (0.35,0.55,0.78).$$

Triangular membership:

$$= \min\left(\frac{0.6061 - 0.35}{0.55 - 0.35}, \frac{0.78 - 0.6061}{0.78 - 0.55}\right) = \min(1.2805,0.7519) = 0.7559$$

Non-membership (Textual uses low-value doubt $g(m) = 1 - m$):

$$g = 0.3939, \nu = \lambda(1 - \mu)g = 0.6 \times (1 - 0.7559) \times 0.3939 = 0.0577,$$

$$\pi = 1 - \mu - \nu = 0.1864$$

These numbers appear in the indicator table (rounded).

10.2.2 Criterion-level IFN via IFWA (Textual).

With three indicators (equal weights $w\ = \frac{1}{3}$) and their IFNs $\langle \mu\ , \nu\ \rangle$,

$$\mu_{\text{Text}} = 1 - \mathbf{G}_{=1}^{3}\ (1 - \mu_\ell)^{1/3} = 0.68174, \nu_{\text{Text}} = \mathbf{G}_{=1}^{3}\ (\nu_\ell)^{1/3} = 0.08700$$

$$S_{\text{Text}} = \mu - \nu = 0.59474, H_{\text{Text}} = \mu + \nu = 0.76874$$

matching Exemplar P1 - Criterion-level IFNs.

Similarly, we obtain for P1 (from the same table):

$$= (0.59474, 0.33488, 0.56111, 0.42867),$$
$$= (0.76874, 0.62067, 0.78708,0.54557),$$

ordered as (Textual, Visual, Audio, Interact).

10.2.3 Across-criterion fusion.

Choquet for (2-additive).

Sort ascending: $_{(1)} = \text{Visual} = 0.334882,\ _{(2)} = \text{Interact} = 0.428669,\ _{(3)} = \text{Audio} = 0.561111,\ _{(4)} = \text{Textual} = 0.594737.$

The cumulative sets and capacities are:

$(\{\text{ Visual }\}) = 0.20,\ \ (\{\text{ Visual, Interact }\}) = 0.42,$

$(\{\text{ Visual, Interact, Audio }\}) = 0.71,\ \ (\{\text{ V, I, A, Text}\}) = 1.00.$

Thus

$$\hat{S} = 0.334882 \cdot 0.20 + (0.428669 - 0.334882) \cdot 0.42$$
$$+(0.561111 - 0.428669) \cdot 0.71 + (0.594737 - 0.561111) \cdot 1.00$$
$$= 0.23403.$$

(Exact decomposition shown in Exemplar P1 - Choquet Decomposition Terms.)

OWA for .

Received: July 28, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Sort H descending: (0.78708, 0.76874, 0.62067, 0.54557). With $\tilde{w} = (0.4,0.3,0.2,0.1)$,

$$\hat{H} = 0.4(0.78708) + 0.3(0.76874) + 0.2(0.62067) + 0.1(0.54557) = 0.72415$$

Grade.

With policy thresholds $(s_1, s_2, s_3) = (0,0.30,0.60)$ and $(h_1, h_2) = (0.50, 0.70)$,

$$\hat{S} = 0.23403 < s_2, \hat{H} = 0.72415 \geq h_1 \Rightarrow \text{Grade } C.$$

10.3 Cohort Outcomes

10.3.1 Distribution.

Grade counts under IFR are: A: 3, B: 15, C: 6, D: 0 (table: Case Study Grade Distribution (IFR)).

10.3.2 Baselines and agreement to instructor references.

We simulate "instructor" references by a slightly different capacity and thresholds plus small noise (as in Section 7). Summary (table: Case Study - Baseline Comparison vs Instructor):

- IFR (ours): Kendall's $\tau = 0.913$, grade accuracy $= 0.833$.
- Crisp rubric: $\tau = -0.290$, accuracy $= 0.667$.
- Type-1 fuzzy: $\tau = -0.072$, accuracy $= 0.500$.

A confusion matrix (table: Case Study - Confusion Matrix (Instructor vs IFR)) shows most B projects kept in B, with few off-by-one band errors and no severe misclassifications, supporting construct validity.

10.3.3 Interpretable feedback (Shapley).

The capacity's Shapley importances are $_{\text{Textual}} = 0.310$, $_{\text{Visual}} = 0.245$, $_{\text{Audio}} = 0.190$, $_{\text{Interact}} = 0.255$. This aligns with the course's learning outcomes (strong emphasis on TextualVisual synergy), and can be reported on rubric sheets.

10.4 Result Tables (Based on case study illustrative study)

Table 10.1a: Raw Indicators: Case Study - Project Summary (Ŝ, Ĥ, Grades) [Updated]

Project	S_hat	H_hat	Grade_IFR
P1	0.234	0.724	C
P2	0.417	0.638	B
P3	0.573	0.841	B
P4	0.643	0.713	A
P5	0.418	0.798	B
P6	0.289	0.723	C
P7	0.639	0.642	B

Received: July 28, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

P8	0.496	0.738	B
P9	0.247	0.641	C
P10	0.376	0.799	B
P11	0.362	0.78	B
P12	0.609	0.763	A
P13	0.577	0.705	B
P14	0.324	0.689	B
P15	0.357	0.626	B
P16	0.632	0.723	A
P17	0.46	0.686	B
P18	0.277	0.636	C
P19	0.235	0.641	C
P20	0.497	0.904	B
P21	0.404	0.682	B
P22	0.597	0.83	B
P23	0.218	0.732	C
P24	0.587	0.723	B

Table 10.1b: Raw Indicators: Case Study - Raw Indicators ($= 24, \quad = 4, \quad = 3$)

Project	**Textual: Coherence**	**Textual: Evidence**	**Textual: Citations**	**Visual: Layout**	**Visual:Contrast**	**Visual: DataVis**	**Audio:SNR**	**Audio:Prosody**	**Audio:MixBalance**	**Interact:TaskSuccess**	**Interact: Latency**	**Interact: Usability**
P1	0.606	0.58	0.428	0.408	0.473	0.696	0.727	0.727	0.676	0.529	0.624	0.26
P2	0.73	0.713	0.742	0.529	0.602	0.454	0.513	0.718	0.541	0.445	0.681	0.771
P3	0.404	0.595	0.53	0.631	0.549	0.651	0.567	0.389	0.726	0.605	0.633	0.335
P4	0.52	0.591	0.299	0.475	0.186	0.914	0.526	0.301	0.723	0.595	0.485	0.641
P5	0.605	0.784	0.429	0.442	0.828	0.504	0.606	0.574	0.649	0.61	0.739	0.629

Received: July 28, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

P6	0.942	0.57	0.323	0.64	0.881	0.643	0.366	0.644	0.697	0.679	0.227	0.593
P7	0.829	0.427	0.712	0.35	0.455	0.343	0.783	0.597	0.758	0.417	0.402	0.512
P8	0.589	0.366	0.761	0.766	0.72	0.74	0.711	0.699	0.456	0.719	0.553	0.748
P9	0.735	0.59	0.363	0.393	0.67	0.739	0.753	0.708	0.476	0.732	0.691	0.7
P10	0.651	0.416	0.522	0.399	0.493	0.485	0.616	0.785	0.69	0.629	0.614	0.737
P11	0.464	0.707	0.621	0.577	0.533	0.436	0.683	0.701	0.871	0.732	0.31	0.504
P12	0.496	0.627	0.34	0.246	0.762	0.52	0.565	0.268	0.561	0.641	0.742	0.782
P13	0.462	0.527	0.719	0.597	0.28	0.39	0.876	0.437	0.404	0.905	0.825	0.442
P14	0.607	0.469	0.369	0.483	0.717	0.677	0.452	0.519	0.788	0.802	0.877	0.653
P15	0.644	0.568	0.3	0.653	0.794	0.758	0.435	0.891	0.525	0.766	0.59	0.426
P16	0.352	0.681	0.398	0.583	0.455	0.609	0.62	0.434	0.511	0.462	0.951	0.729
P17	0.78	0.726	0.483	0.744	0.835	0.301	0.665	0.503	0.175	0.654	0.651	0.459
P18	0.626	0.377	0.395	0.456	0.713	0.845	0.992	0.519	0.557	0.752	0.595	0.781

Received: July 28, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

P19	0.595	0.903	0.767	0.727	0.767	0.667	0.797	0.686	0.681	0.787	0.535	0.329
P20	0.571	0.47	0.55	0.6	0.553	0.627	0.721	0.502	0.408	0.647	0.38	0.4
P21	0.858	0.379	0.293	0.517	0.384	0.332	0.508	0.61	0.73	0.936	0.464	0.571
P22	0.592	0.844	0.612	0.634	0.502	0.572	0.493	0.31	0.401	0.721	0.536	0.626
P23	0.593	0.787	0.292	0.476	0.163	0.597	0.492	0.576	0.471	0.571	0.446	0.827
P24	0.728	0.739	0.826	0.505	0.421	0.607	0.756	0.771	0.411	0.566	0.524	0.631

Table 10.2 - Criterion IFNs: Case Study - Criterion IFNs and ,

Project	**Textual_mu**	**Textual_nu**	**Textual_pi**	**Textual_S**	**Textual_H**	**Visual_mu**	**Visual_nu**	**Visual_pi**	**Visual_S**	**Visual_H**	**Audio_mu**	**Audio_nu**	**Audio_pi**	**Audio_S**	**Audio_H**	**Interact_mu**	**Interact_nu**	**Interact_pi**	**Interact_S**	**Interact_H**
P1	0.682	0.087	0.2231	0.5595	0.7769	0.4778	0.143	0.3779	0.335	0.6211	0.6744	0.113	0.213	0.561	0.787	0.487	0.058	0.4544	0.4299	0.5466
P2	0.154	0.1388	0.7008	0.017	0.2922	0.6662	0.095	0.2433	0.5667	0.7577	0.4881	0.2488	0.2711	0.2333	0.7229	0.3255	0.113	0.5662	0.2133	0.4388

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

P3	0.781	0.064	0.155	0.7117	0.8445	0.9422	0.014	0.045	0.928	0.955	0.4116	0.2511	0.333	0.164	0.6667	0.6366	0.0588	0.3077	0.5778	0.6933
P4	0.6422	0.1111	0.2477	0.5532	0.7753	0.1173	0.1665	0.6662	0.008	0.3388	0.332	0.2778	0.4002	0.042	0.5988	0.8222	0.0122	0.1666	0.8089	0.8334
P5	0.5066	0.1088	0.3855	0.3988	0.6155	0.3772	0.1377	0.4922	0.2355	0.5088	0.8366	0.0777	0.0888	0.7599	0.9122	0.8366	0.03	0.1334	0.8807	0.8666
P6	0.4033	0.0922	0.5055	0.3111	0.4955	0.6255	0.0566	0.3199	0.5669	0.6881	0.599	0.1688	0.2422	0.4222	0.7588	0.72	0.0566	0.2224	0.6665	0.7776
P7	0.308	0.1266	0.5666	0.1881	0.4334	0.2755	0.2688	0.4588	0.007	0.5422	0.8355	0.0555	0.1111	0.78	0.8889	0.3331	0.0367	0.6332	0.2944	0.3688
P8	0.5099	0.1177	0.3744	0.3922	0.6266	0.2799	0.1111	0.611	0.1168	0.399	0.6299	0.1522	0.2199	0.4777	0.7881	0.8188	0.0311	0.1511	0.7887	0.8499
P9	0.42	0.1443	0.4377	0.2777	0.5663	0.3111	0.1544	0.5355	0.1566	0.4655	0.5111	0.1911	0.2988	0.32	0.7022	0.564	0.108	0.3288	0.4566	0.6722
P10	0.7422	0.0711	0.1887	0.6711	0.8133	0.5522	0.1455	0.3033	0.4007	0.6977	0.5788	0.1499	0.2733	0.4299	0.7277	0.817	0.0333	0.155	0.7844	0.885

Received: July 28, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

P11	0.494	0.1188	0.3876	0.6133	0.8332	0.0488	0.1199	0.7841	0.8881	0.8821	0.005	0.1299	0.7711	0.8711	0.4044	0.0399	0.5556	0.3655	0.4444
P12	0.514	0.1445	0.3411	0.3688	0.6599	0.3844	0.1663	0.4533	0.2211	0.5477	0.4888	0.2288	0.2844	0.26	0.7116	0.8711	0.0333	0.0966	0.8337	0.9044
P13	0.7777	0.0556	0.1688	0.7211	0.8332	0.7776	0.0075	0.1499	0.7001	0.8511	0.1444	0.2888	0.5699	-0.1444	0.4331	0.0688	0.22	0.7112	-0.1522	0.2888
P14	0.6466	0.1088	0.2466	0.5377	0.7544	0.4611	0.1177	0.4222	0.3445	0.5778	0.3744	0.2699	0.3577	0.1044	0.6433	0.52	0.15	0.3331	0.337	0.6669
P15	0.5611	0.1255	0.3133	0.4366	0.6887	0.4233	0.0889	0.4887	0.3334	0.5133	0.2455	0.2566	0.4999	-0.0111	0.5001	0.6033	0.057	0.3399	0.5466	0.6611
P16	0.2355	0.2299	0.5336	0.0005	0.4644	0.8422	0.0422	0.1166	0.8	0.8844	0.5558	0.2299	0.2133	0.3299	0.7887	0.2633	0.1399	0.5597	0.1244	0.4003
P17	0.3866	0.1166	0.4998	0.27	0.5002	0.1332	0.1611	0.7007	-0.0029	0.2933	0.7004	0.1099	0.1866	0.5955	0.8144	0.7388	0.0311	0.2331	0.7007	0.7669
P18	0.48	0.1	0.3	0.3	0.6	0.2	0.1	0.6	0.1	0.3	0.5	0.0	0.3	0.4	0.6	0.6222	0.0866	0.2	0.5	0.7

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

		62	57	18	43	47	31	22	17	78	46	65	88	81	12			93	36	07
P19	0.418	0.0073	0.5009	0.3445	0.4491	0.4117	0.0097	0.4886	0.332	0.5144	0.6	0.1131	0.2669	0.4669	0.7331	0.6	0.058	0.3433	0.5422	0.6657
P20	0.9011	0.0028	0.0071	0.8744	0.9229	0.9844	0.0004	0.0112	0.998	0.9888	0.5755	0.1966	0.2229	0.3779	0.7771	0.7222	0.04	0.2238	0.6811	0.7622
P21	0.138	0.2005	0.6557	-0.0067	0.3443	0.3669	0.2221	0.441	0.1449	0.559	0.7444	0.1115	0.1141	0.663	0.8559	0.5666	0.054	0.38	0.5112	0.662
P22	0.61	0.0068	0.3211	0.5422	0.6779	0.8897	0.0026	0.0077	0.887	0.9223	0.0069	0.4441	0.4491	-0.3722	0.5009	0.9011	0.0122	0.0887	0.8889	0.9133
P23	0.429	0.1335	0.4336	0.2944	0.5564	0.6488	0.1118	0.2333	0.553	0.7677	0.6266	0.2007	0.1667	0.4188	0.8333	0.4944	0.0655	0.444	0.4229	0.556
P24	0.097	0.1225	0.7778	-0.0288	0.2222	0.7	0.0087	0.2213	0.6113	0.7887	0.3445	0.2223	0.4431	0.1222	0.5669	0.8667	0.0099	0.1244	0.8588	0.8776

Table 10.3 - Project Summary: Case Study - Project Summary ($\hat{S}, \hat{H}$, Grade)

Project	S_hat	H_hat	Grade_IFR	S_star_instr	H_star_instr	Grade_Instructor
P1	0.234	0.724	C	0.253	0.715	C
P2	0.417	0.638	B	0.397	0.623	B

Received: July 28, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

P3	0.573	0.841	B	0.566	0.815	B
P4	0.643	0.713	A	0.622	0.717	B
P5	0.418	0.798	B	0.412	0.801	B
P6	0.289	0.723	C	0.296	0.726	C
P7	0.639	0.642	B	0.627	0.687	B
P8	0.496	0.738	B	0.516	0.728	B
P9	0.247	0.641	C	0.269	0.593	C
P10	0.376	0.799	B	0.319	0.786	C
P11	0.362	0.78	B	0.331	0.769	B
P12	0.609	0.763	A	0.593	0.77	B
P13	0.577	0.705	B	0.575	0.723	B
P14	0.324	0.689	B	0.326	0.692	B
P15	0.357	0.626	B	0.361	0.584	B
P16	0.632	0.723	A	0.613	0.714	B
P17	0.46	0.686	B	0.451	0.68	B
P18	0.277	0.636	C	0.255	0.608	C
P19	0.235	0.641	C	0.268	0.656	C
P20	0.497	0.904	B	0.49	0.895	B
P21	0.404	0.682	B	0.39	0.662	B
P22	0.597	0.83	B	0.616	0.826	B
P23	0.218	0.732	C	0.201	0.711	C
P24	0.587	0.723	B	0.548	0.723	B

Table 10.4 - Exemplar P1 Indicators: Exemplar P1 -Indicator-level IFNs

Criterion	**Indicator**	**m**	**a1**	**a2**	**a3**	**g_shape**	**mu**	**nu**	**pi**
Textual	Coherence	0.6061	0.35	0.55	0.78	0.3939	0.7559	0.0577	0.1864
Textual	Evidence	0.5799	0.3	0.52	0.75	0.4201	0.7397	0.0656	0.1947
Textual	Citations	0.4284	0.32	0.54	0.76	0.5716	0.4927	0.174	0.3333

Received: July 28, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Visual	Layout	0.4077	0.38	0.6	0.82	0.5923	0.1258	0.3107	0.5635
Visual	Contrast	0.4729	0.3	0.55	0.8	0.5271	0.6916	0.0975	0.2109
Visual	DataVis	0.6962	0.36	0.58	0.8	0.3038	0.4718	0.0963	0.4319
Audio	SNR	0.7266	0.45	0.68	0.88	0.5467	0.7668	0.0765	0.1567
Audio	Prosody	0.7275	0.35	0.6	0.82	0.5451	0.4206	0.1895	0.3899
Audio	MixBalance	0.6764	0.4	0.62	0.84	0.6473	0.7438	0.0995	0.1567
Interact	TaskSuccess	0.5288	0.42	0.64	0.86	0.0575	0.4943	0.0174	0.4882
Interact	Latency	0.624	0.3	0.56	0.8	0.2481	0.7332	0.0397	0.2271
Interact	Usability	0.2598	0.4	0.62	0.84	0.4805	0.0	0.2883	0.7117

Table 10.5 - Exemplar P1 Criteria: Exemplar P1 - Criterion level IFNs and ,

Criterion	mu_k	nu_k	pi_k	S_k	H_k
Textual	0.6817	0.087	0.2313	0.5947	0.7687
Visual	0.4778	0.1429	0.3793	0.3349	0.6207
Audio	0.6741	0.113	0.2129	0.5611	0.7871
Interact	0.4871	0.0585	0.4544	0.4287	0.5456

Table 10.6 - Choquet Terms (P1): Exemplar P1 - Choquet Decomposition Terms

Term	Delta S	v(A)
Visual	0.3349	0.2
Interact−Visual	0.0938	0.42
Audio−Interact	0.1324	0.71
Textual−Audio	0.0336	1.0

Table 10.7 - Baselines: Case Study - Baseline Comparison vs Instructor

Model	Kendall_tau_vs_Instructor	Grade_Accuracy_vs_Instructor
IFR (ours)	0.913	0.833

Received: July 28, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Crisp rubric	-0.29	0.667
Type-1 fuzzy	-0.072	0.5

Table 10.8 - Confusion Matrix: Case Study - Confusion Matrix (Instructor vs IFR)

Instructor	**A**	**B**	**C**	**D**
A	0	0	0	0
B	3	14	0	0
C	0	1	6	0
D	0	0	0	0

Table 10.9 - Shapley Importances: Case Study - Shapley Importance per Criterion

Criterion	**Shapley_importance**
Textual	0.31
Visual	0.245
Audio	0.19
Interact	0.255

10.5 Discussion

Why IFR beats crisp/Type-1: IFR models support and doubt separately; π absorbs ambiguity, preserving feasibility $|S| \leq H \leq 1$ during fusion. The Choquet layer captures cross-modal synergy/redundancy missed by additive baselines.

Interpretability: 2 -additive capacities grant Shapley and pairwise interactions for constructive feedback, while keeping parameter count manageable.

Policy tuning: Thresholds (s, h) and OWA optimism $\tilde{w}$ can be calibrated on held-out semesters to match grading norms (Section 7).

Robustness: Sensitivity diagnostics (Section 9) should accompany deployment; missing modalities can be handled by increasing and re-normalizing IFWA weights.

11. DISCUSSION

11.1 Interpretability for Instructors and Students

Our intuitionistic fuzzy rubric (IFR) surfaces criterion influence transparently via Shapley values of the learned 2-additive capacity (Section 5.4). In the case study, the per-criterion importances were $_{\text{Textual}} = 0.310,\ _{\text{Visual}} = 0.245,\ _{\text{Audio}} = 0.190,\ _{\text{Interact}} = 0.255,$ implying a pedagogically intuitive emphasis on Textual communication and Interaction design, with Visual next and Audio modestly lower. Because in the 2-additive setting $= + \frac{1}{2}\sum_{j \neq i}$, instructors can diagnose whether importance arises from intrinsic quality or

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

pairwise synergy . In our cohort, positive synergy terms = 0.12 and = 0.08 corroborate the course goal that clear explanations (Textual) combined with design (Visual/Interact) should be rewarded.

At the project level, IFR's ($\hat{S}, \hat{H}$) representation stays in the feasible triangle $|S| \leq H \leq 1$ and decouples net evidence $S = \mu - \nu$ from certainty $H = \mu + \nu$. For exemplar $\mathbf{P1}, \hat{S} = 0.234$ and $\hat{H} = 0.724$ placed it in band **C** (high certainty but sub-threshold net support), which is more informative than a single scalar grade: the team can see that increasing supportive evidence-rather than just reducing doubt-would most efficiently lift their band.

11.2 What the Results Show (Effectiveness and Reliability)

Across $\mathbf{N = 24}$ projects, IFR produced a grade distribution **A**: **3**, **B**: **15**, **C**: **6**, **D**: **0**, reflecting a cohort centered near competence with a long positive tail. Against simulated instructor references, IFR achieved Kendall's $\tau = 0.913$ and grade accuracy $= 0.833$, outperforming both baselines-crisp rubric ($\tau = -0.290$, accuracy 0.667) and type-1 fuzzy ($\tau = -0.072$, accuracy 0.500)-by wide margins. These gains are consistent with (i) explicit modeling of nonmembership and hesitation in IFS, (ii) monotone indicator fusion (IFWA/IFWG), and (iii) interaction-aware global aggregation (Choquet) [12,13,18,23,24]. The confusion matrix shows most instructor B projects remained B (14/17), with few off-by-one errors and none severe-evidence of stable, face-valid decisions.

Mechanistically, the improvement can be tied to the capacity v : when two criteria are jointly strong, the positive increments $\Delta S.\ v(A_{(j)})$ in

$$\hat{S} = \sum_{=1} (S_{(j)} - S_{(j-1)}) v(A_{(j)})$$

are larger than with an additive mean, while redundant pairs (e.g., Visual-Audio with = -0.05) are de-emphasized. Thus, IFR respects the course's intended compensations and avoids both over-counting redundant cues and under-rewarding balanced, multimodal strength.

11.3 Pedagogical Implications

Actionable feedback: Students receive $(\hat{S}, \hat{H})$ plus $\{\phi_i\}$ and top m_{ij}; this distinguishes what to improve (e.g., "raise Textual support") from how certain the rubric is about their current level.

Rubric tuning with evidence: Simple, interpretable policy thresholds (s_1, s_2, s_3), (h_1, h_2) map $(\hat{S}, \hat{H})$ to grades and can be re-fit each term for fairness while keeping the feasibility constraint $|S| \leq H$.

Reliability uplift: IFR's reliability-weighted fusion and hesitation channels reduce volatility from outlier raters; high suggests alignment with instructor ordering, while the absence of severe misclassifications indicates robust policy behavior.

11.4 Practical Deployment Considerations

Received: July 28, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Computation: With 2-additive capacities, evaluation is (log) per project; learning has (^{2}) parameters and admits convex/first-order solvers.

Governance: Keep a model card including capacity , Shapley , thresholds, and ablation results (Sections 7-9), and version it across semesters for transparency.

Robustness: The sensitivity "tornado" (Section 9) indicated single-factor perturbations changed grades for only 3.512% of projects, consistent with a stable policy envelope.

12. CONCLUSION

This study introduced an Intuitionistic Fuzzy Rubric (IFR) for evaluating multimodal student projects that (i) calibrates linguistic terms to intuitionistic fuzzy numbers (IFNs), explicitly modeling support μ, doubt ν, and hesitation π with $\mu + \nu \leq 1$; (ii) aggregates indicators within a criterion using principled IF operators (IFWA/IFWG) that are monotone, idempotent, and bounded; and (iii) fuses criteria across modalities with either OWA (compensatory, no interaction) or a Choquet integral under a 2additive capacity to capture synergy/redundancy and to expose Shapley importances for feedback. The pipeline keeps decisions in the feasible triangle $|S| \leq H \leq 1$, separating net evidence $S = \mu - \nu$ from certainty $H = \mu + \nu$, which clarifies both grading and guidance to students.

What we found (case-study evidence)

On a real-type cohort ($\mathbf{N = 24}$, $\mathbf{K = 4}$, $\mathbf{L = 3}$), IFR achieved strong agreement with instructor references: Kendall's $\tau = 0.913$ and grade accuracy $= 0.833$, clearly outperforming a crisp rubric ($\tau =$ -0.290 , accuracy 0.667) and a type-1 fuzzy rubric ($\tau = -0.072$, accuracy 0.500). The grade distribution (A: 3, B: 15, C: 6, D: 0) showed a competency-centered cohort with a controlled tail, while the confusion matrix exhibited mainly off-by-one errors and no severe misclassifications-evidence of stability and face validity. Capacity analysis yielded interpretable Shapley importances ($\phi_{\text{Textual}} = 0.310$, $\phi_{\text{Visual}} = 0.245$, $\phi_{\text{Audio}} = 0.190$, $\phi_{\text{Interact}} = 0.255$), consistent with the course's emphasis on explanatory clarity and interaction design; positive pairwise interactions $m_{TV} = 0.12$ and $m_{TI} = 0.08$ confirmed the value of cross-modal coherence. Sensitivity tests indicated that modest parameter perturbations typically changed grades for 3.5-12% of projects, suggesting a robust policy envelope.

Pedagogically, the ($\hat{S}, \hat{H}$) view proved useful: for exemplar $\mathbf{P1}$, $\hat{S} = 0.234$ and $\hat{H} = 0.724$ placed the project in band C , revealing that adding supportive evidence (raising S), not just reducing doubt, would most efficiently improve the grade-precise, actionable feedback that standard rubrics rarely deliver.

Bringing together intuitionistic fuzzy modeling, principled aggregation, and interpretable non-additive integration, IFR offers a mathematically grounded, pedagogically aligned framework for multimodal assessment. The case-study results and diagnostics indicate that IFR can improve agreement, stabilize decisions, and most importantly-tell students exactly what to improve. With modest extensions and routine validation, it is ready for classroom deployment at scale.

Received: July 28, 2025

Acknowledgment

This research was partially funded by Zarqa University.

References

[1] Zadeh, L.A., 1965. Fuzzy sets. *Information and Control*, 8(3), 338–353. DOI: https://doi.org/10.1016/S0019-9958(65)90241-X.

[2] Atanassov, K.T., 1986. Intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 20(1), 87–96. DOI: https://doi.org/10.1016/S0165-0114(86)80034-3.

[3] Xu, Z.S., 2007. Intuitionistic fuzzy aggregation operators. *IEEE Transactions on Fuzzy Systems*, 15(6), 1179–1187. DOI: https://doi.org/10.1109/TFUZZ.2006.890678.

[4] Szmidt, E.; Kacprzyk, J., 2000. Distances between intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 114(3), 505–518. DOI: https://doi.org/10.1016/S0165-0114(98)00244-9.

[5] Grabisch, M., 1996. The application of fuzzy integrals in multicriteria decision making. *European Journal of Operational Research*, 89(3), 445–456. DOI: https://doi.org/10.1016/0377-2217(95)00176-X.

[6] Yogeesh, N., 2023. Fuzzy Clustering for Classification of Metamaterial Properties. In: Mehta, S.; Abougreen, A. (eds.). *Metamaterial Technology and Intelligent Metasurfaces for Wireless Communication Systems*. IGI Global: Hershey, USA. pp. 200–229. DOI: https://doi.org/10.4018/978-1-6684-8287-2.ch009.

[7] Panadero, E.; Jonsson, A., 2013. The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129–144. DOI: https://doi.org/10.1016/j.edurev.2013.01.002.

[8] Kress, G., 2009. *Multimodality: A Social Semiotic Approach to Contemporary Communication*. Routledge: London, UK. DOI: https://doi.org/10.4324/9780203970034.

[9] Garg, H., 2016. Some series of intuitionistic fuzzy interactive averaging aggregation operators. *SpringerPlus*, 5, 999. DOI: https://doi.org/10.1186/s40064-016-2591-9

[10] Atanassov, K.T., 1999. *Intuitionistic Fuzzy Sets: Theory and Applications*. Physica-Verlag: Heidelberg, Germany. DOI: https://doi.org/10.1007/978-3-7908-1870-3.

[11] Klement, E.P.; Mesiar, R.; Pap, E., 2000. *Triangular Norms*. Springer: Dordrecht, Netherlands. DOI: https://doi.org/10.1007/978-94-015-9540-7.

[12] Beliakov, G.; Pradera, A.; Calvo, T., 2007. *Aggregation Functions: A Guide for Practitioners*. Springer: Berlin, Germany. DOI: https://doi.org/10.1007/978-3-540-73721-6.

[13] Xu, Z.S., 2007. Intuitionistic fuzzy aggregation operators. *IEEE Transactions on Fuzzy Systems*, 15(6), 1179–1187. DOI: https://doi.org/10.1109/TFUZZ.2006.890678.

[14] Xu, Z.S.; Yager, R.R., 2006. Some geometric aggregation operators based on intuitionistic fuzzy sets. *International Journal of General Systems*, 35(4), 417–433. DOI: https://doi.org/10.1080/03081070600574353.

[15] Yager, R.R., 1988. On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Transactions on Systems, Man, and Cybernetics*, 18(1), 183–190. DOI: https://doi.org/10.1109/21.87068.

[16] Chen, T.-Y., 2011. A comparative analysis of score functions for multiple criteria decision analysis based on intuitionistic fuzzy sets. *Information Sciences*, 181(17), 3652–3676. DOI: https://doi.org/10.1016/j.ins.2011.04.030.

[17] Hung, W.-L.; Yang, M.-S., 2004. Similarity measures of intuitionistic fuzzy sets based on Hausdorff distance. *Pattern Recognition Letters*, 25(14), 1603–1611. DOI: https://doi.org/10.1016/j.patrec.2004.06.006.

[18] Marichal, J.-L., 2000. An axiomatic approach of the discrete Choquet integral as a tool to aggregate interacting criteria. *IEEE Transactions on Fuzzy Systems*, 8(6), 800–807. DOI: https://doi.org/10.1109/91.890347.

[19] Shannon, C.E., 1948. A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423; 27(4), 623–656. DOI: https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

[20] Yogeesh, N., 2023. *Intuitionistic Fuzzy Hypergraphs and Their Operations*. Chapman and Hall/CRC: Boca Raton, USA. DOI: https://doi.org/10.1201/9781003359456

[21] Yogeesh, N., 2024. Fuzzy Graph Dominance for Networked Communication Optimization. In: Sharma, V.; Balusamy, B.; Ferrari, G.; Ajmani, P. (eds.). *Wireless Communication Technologies: Roles, Responsibilities, and Impact of IoT, 6G, and Blockchain Practices*, 1st ed. CRC Press: Boca Raton, USA. pp. 30. DOI: https://doi.org/10.1201/9781003389231.

[22] Torra, V., 2010. Hesitant fuzzy sets. *International Journal of Intelligent Systems*, 25(6), 529–539. DOI: https://doi.org/10.1002/int.20418.

[23] Denneberg, D., 1994. *Non-Additive Measure and Integral*. Springer: New York, USA. DOI: https://doi.org/10.1007/978-1-4612-4286-1.

[24] Grabisch, M.; Labreuche, C., 2010. A decade of application of the Choquet integral in multicriteria decision-aid. *European Journal of Operational Research*, 206(3), 679–693. DOI: https://doi.org/10.1016/j.ejor.2010.03.041.

[25] Murofushi, T.; Sugeno, M., 1991. A theory of fuzzy measures: Representation via the Möbius transform. *Fuzzy Sets and Systems*, 29(2), 201–227. DOI: https://doi.org/10.1016/0165-0114(91)90117-9.

[26] Yogeesh, N., 2024. Solving Fuzzy Nonlinear Optimization Problems Using Evolutionary Algorithms. In: Mukherjee, G.; Basu Mallik, B.; Kar, R.; Chaudhary, A. (eds.). *Advances on Mathematical Modeling and Optimization with Its Applications*, 1st ed. CRC Press: Boca Raton, USA. pp. 20. DOI: https://doi.org/10.1201/9781003387459.

[27] Joachims, T., 2002. Optimizing search engines using clickthrough data. *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; July 23–26, 2002; Edmonton, Canada. pp. 133–142. DOI: https://doi.org/10.1145/775047.775067.

[28] Boyd, S.; Vandenberghe, L., 2004. *Convex Optimization*. Cambridge University Press: Cambridge, UK. DOI: https://doi.org/10.1017/CBO9780511804441.

[29] Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288. DOI: https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

[30] Mohammad, A. A. S., Mohammad, S. I. S., Al-Daoud, K. I., Vasudevan, A., & Hunitie, M. F. A. (2025). Digital ledger technology: A factor analysis of financial data management practices in the age of blockchain in Jordan. International Journal of Innovative Research and Scientific Studies, 8(2), 2567-2577.

[31] Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J., 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1), 1–122. DOI: https://doi.org/10.1561/2200000016.

[32] Mohammad, A. A. S., Mohammad, S. I. S., Al Oraini, B., Vasudevan, A., & Alshurideh, M. T. (2025). Data security in digital accounting: A logistic regression analysis of risk factors. International Journal of Innovative Research and Scientific Studies, 8(1), 2699-2709.

[33] McCullagh, P., 1980. Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(2), 109–142. DOI: https://doi.org/10.1111/j.2517-6161.1980.tb01109.x.

[34] Mohammad, A. A. S., Mohammad, S. I. S., Al-Daoud, K. I., Al Oraini, B., Vasudevan, A., & Feng, Z. (2025). Optimizing the Value Chain for Perishable Agricultural Commodities: A Strategic Approach for Jordan. Research on World Agricultural Economy, 6(1), 465-478.

[35] Yogeesh, N., 2023. Fuzzy Logic Modelling of Nonlinear Metamaterials. In: Mehta, S.; Abougreen, A. (eds.). *Metamaterial Technology and Intelligent Metasurfaces for Wireless Communication Systems*. IGI Global: Hershey, USA. pp. 230–269. DOI: https://doi.org/10.4018/978-1-6684-8287-2.ch010.

[36] Mohammad, A. A. S. (2025). The impact of COVID-19 on digital marketing and marketing philosophy: evidence from Jordan. International Journal of Business Information Systems, 48(2), 267-281.

[37] Niculescu-Mizil, A.; Caruana, R., 2005. Predicting good probabilities with supervised learning. *Proceedings of the 22nd International Conference on Machine Learning (ICML '05)*; August 7–11, 2005; Bonn, Germany. pp. 625–632. DOI: https://doi.org/10.1145/1102351.1102430.

[38] Mohammad, A. A., Shelash, S. I., Saber, T. I., Vasudevan, A., Darwazeh, N. R., & Almajali, R. (2025). Internal audit governance factors and their effect on the risk-based auditing adoption of commercial banks in Jordan. Data and Metadata, 4, 464.

[39] Nijalingappa, Y.; Setty, G.D.K.; Madan, R.; Nagaraju, V.T., 2025. Directable zeros in fuzzy logics: A study of functional behaviours and applications. *AIP Conference Proceedings*. 3175(1), 020091, 10 March 2025. DOI: https://doi.org/10.1063/5.0254157

[40] Mohammad, A. A. S., Alolayyan, M. N., Al-Daoud, K. I., Al Nammas, Y. M., Vasudevan, A., & Mohammad, S. I. (2024). Association between Social Demographic Factors and Health Literacy in Jordan. Journal of Ecohumanism, 3(7), 2351-2365.

Received: July 28, 2025