

Comparison of Single Expert and Fuzzy Group Decision Making in Fuzzy SAW

Eka Putri Rachmawati^{1*}, Muhammad Basyier Ardima², Mochamad Idris³, Ardian Fakhru Rosyad⁴

^{1,2,3,4}Universitas Semarang, Central Java, Indonesia

ARTICLE INFO

Article history:

Received: 19 July 2026

Revised: 13 September 2026

Accepted: 12 October 2026

DOI:

10.61100/tacit.v4i2.399

Keywords:

Fuzzy-SAW, Fuzzy Group Decision Making, Single Expert, Student Ranking, Decision Support System

ABSTRACT

This study compares single-expert and group-based criterion weighting in a Fuzzy Simple Additive Weighting (Fuzzy-SAW) decision support system for student ranking. Data from 116 students across four classes at a Nahdlatul Ulama-affiliated Madrasah Aliyah were evaluated using academic, attendance, and extracurricular criteria. The Single Expert (SE) approach was compared with Fuzzy Group Decision Making (FGDM), involving four equally weighted experts. Linguistic judgments were converted into Triangular Fuzzy Numbers, while missing attendance values were imputed as 75. Both approaches used identical data and normalization procedures. Results showed near-perfect ranking agreement between SE and FGDM (Spearman's $r_s = 0.9964$; MAE = 0.35 positions). Correlations with academic report rankings were 0.8063 and 0.8092, respectively. These findings demonstrate limited ranking changes following expert aggregation, rather than proven improvements in accuracy or bias reduction. The study provides a controlled methodological comparison, although its generalizability remains limited by the single-school setting.



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

ABSTRAK

Penelitian ini membandingkan pembobotan kriteria oleh pakar tunggal dan kelompok pakar dalam Sistem Pendukung Keputusan (SPK) berbasis Fuzzy Simple Additive Weighting (Fuzzy-SAW) untuk pemeringkatan siswa berprestasi. Data 116 siswa dari empat kelas di Madrasah Aliyah yang berafiliasi dengan Nahdlatul Ulama dianalisis berdasarkan kriteria akademik, kehadiran, dan kegiatan ekstrakurikuler. Pendekatan Single Expert (SE) dibandingkan dengan Fuzzy Group Decision Making (FGDM) yang melibatkan empat pakar dengan bobot penilaian setara. Penilaian linguistik dikonversi menjadi Triangular Fuzzy Numbers, sedangkan data kehadiran yang hilang diimputasi menggunakan nilai 75. Kedua pendekatan menggunakan data dan prosedur normalisasi yang sama. Hasil penelitian menunjukkan kesesuaian peringkat yang hampir sempurna antara SE dan FGDM (Spearman $r_s = 0,9964$; MAE = 0,35 posisi). Korelasi dengan peringkat berdasarkan nilai rapor akademik masing-masing sebesar 0,8063 dan 0,8092. Temuan ini menunjukkan bahwa penggabungan penilaian beberapa pakar hanya menghasilkan perubahan peringkat yang relatif kecil, bukan membuktikan peningkatan akurasi atau pengurangan bias. Penelitian ini memberikan kontribusi melalui perbandingan metodologis yang terkontrol, meskipun generalisasi hasilnya masih terbatas karena penelitian hanya dilakukan pada satu sekolah.

1. INTRODUCTION

Education plays a fundamental role in developing competent human resources capable of responding to future challenges (Hermansyah et al., 2023; Ramadhan, 2024). In contemporary educational systems, student assessment is expected to encompass cognitive, affective, and psychomotor dimensions (Syafi'i et al., 2025). However, assessment practices often remain heavily dependent on academic achievement, while non-academic dimensions receive comparatively limited consideration. This imbalance raises concerns about whether conventional academic rankings adequately represent students' broader competencies, particularly when discipline, attendance, and extracurricular participation are also considered important indicators of

* Corresponding author, email address: ekapu2tt@gmail.com

achievement. Consequently, multidimensional assessment requires an approach that integrates academic performance with relevant non-academic criteria (Mahbub et al., 2022; Witanto et al., 2020).

The integration of multiple assessment criteria presents methodological challenges because educational indicators may be measured using different scales and may involve subjective judgments. Academic achievement is generally represented by numerical scores, whereas extracurricular participation and other non-academic indicators may initially be expressed through categorical assessments. Such differences complicate the comparison of student performance and may reduce transparency when rankings are determined manually (Cahyani et al., 2024; Melati & Triyono, 2020). Decision Support Systems (DSS) offer a structured approach to addressing these challenges through Multi-Criteria Decision Making (MCDM). One commonly applied method is Simple Additive Weighting (SAW), which calculates an alternative's overall preference score by aggregating normalized criterion values according to their assigned weights (Yusnira & Saputri, 2023). Nevertheless, the reliability of SAW-based rankings depends not only on the assessment data but also on how criterion weights are determined. Different weighting judgments may produce different student rankings, even when the underlying dataset remains unchanged.

Fuzzy logic provides an additional framework for representing uncertainty in qualitative assessments. Unlike conventional binary classifications, fuzzy sets express degrees of membership within the interval $[0, 1]$, allowing linguistic judgments to be represented mathematically (Ramadhan, 2024; Wijaya & Ipriadi, 2025). The integration of fuzzy logic with SAW, commonly referred to as Fuzzy-SAW, enables the combination of quantitative indicators and linguistically expressed assessments within a structured decision-making process (Rahman et al., 2023; Yusnira & Saputri, 2023). However, converting qualitative judgments into fuzzy numbers does not automatically eliminate subjectivity. Subjectivity may remain embedded in the assignment of criterion importance, particularly when the weighting process depends on the preferences of a single decision-maker. Therefore, the selection and aggregation of expert judgments constitute important methodological considerations in Fuzzy-SAW applications.

Previous studies have demonstrated the applicability of SAW and Fuzzy-SAW to educational decision-making. Cahyani et al. (2024), for example, examined student achievement selection using SAW, illustrating the usefulness of structured multi-criteria evaluation in educational settings. Lestari et al. (2020) investigated Fuzzy-SAW in student admission selection, while Christioko et al. (2017) applied fuzzy multi-attribute decision-making to the selection of outstanding university students. These studies provide a foundation for understanding how multi-criteria methods can support educational selection and ranking. Nevertheless, their reported applications primarily emphasize the implementation of decision-making methods and the resulting selection outcomes. They do not, within the aspects discussed in the present manuscript, establish how replacing a single expert's criterion weights with aggregated judgments from multiple experts affects ranking consistency and individual students' positions. This distinction is important because the effectiveness of a decision-support method cannot be evaluated solely through its ability to generate a ranking; the consequences of alternative weighting arrangements must also be examined.

A related methodological issue concerns the aggregation of expert judgments. Sudakov and Sivakova (2020) discuss approaches for combining expert assessments, providing a methodological basis for examining group-based decision-making. In a single-expert approach, criterion weights reflect the judgments of one evaluator. In contrast, Fuzzy Group Decision Making (FGDM) combines assessments from multiple decision-makers to construct a collective weighting structure. Although group aggregation introduces a broader range of expert perspectives, it should not automatically be assumed to produce more accurate or less biased decisions. Its practical contribution must be assessed by examining whether the resulting weights meaningfully alter decision outcomes and whether those changes are consistent across the alternatives being evaluated. Accordingly, the specific research gap addressed in this study is the limited comparative evidence, within the educational Fuzzy-SAW applications reviewed, concerning the effects of single-expert versus fuzzy group-based criterion weighting on student ranking agreement and positional changes.

This issue is particularly relevant to Madrasah Aliyah (MA), where student assessment may involve general academic subjects, religious education, specialization subjects, local curriculum content, attendance, and extracurricular participation. These dimensions represent different aspects of student performance and may receive different levels of importance depending on the evaluator's educational responsibilities and professional judgment. A comparative weighting analysis is therefore necessary to determine whether incorporating multiple expert perspectives changes the prioritization of assessment criteria and, subsequently, the

resulting student rankings. Rather than assuming that group-based weighting is inherently superior, the present study evaluates the extent to which its results differ from those generated through a single-expert approach.

To address this gap, the study applies Fuzzy-SAW to assessment data from 116 eleventh-grade students distributed across four classes (XI-A to XI-D) at a Nahdlatul Ulama-based Madrasah Aliyah. Five criteria are considered: the average scores of religious and general subjects, specialization subjects, local content subjects, attendance, and extracurricular participation. Two criterion-weighting approaches are compared. The Single Expert (SE) model uses assessments provided by the school's Vice Principal for Curriculum, whereas the FGDM model aggregates assessments from four experts, including the internal expert and three experts from comparable educational institutions. Expert judgments are expressed through linguistic variables, converted into Triangular Fuzzy Numbers (TFNs), aggregated using a fuzzy averaging approach, and subsequently defuzzified to obtain criterion weights (Sudakov & Sivakova, 2020). The normalized decision matrix is then combined with the respective weighting structures to generate student rankings.

The principal contribution of this study lies in its direct comparison of two expert-weighting arrangements within the same Fuzzy-SAW assessment framework and student dataset. Specifically, the study examines (1) differences in criterion weights generated by single-expert and group-based assessments, (2) the degree of agreement between the resulting student rankings, and (3) the magnitude of individual ranking changes across the four classes. Spearman's rank correlation coefficient is employed to evaluate ranking agreement, while Mean Absolute Error (MAE) is used to quantify positional differences. Rankings derived from academic report scores are also included as a reference for examining how the multidimensional assessment models differ from academic-only rankings. By separating differences in expert weighting from differences in ranking outcomes, the study contributes empirical evidence on the practical implications of incorporating multiple expert judgments into educational decision-support systems. The findings are expected to inform the design of more transparent and methodologically accountable student assessment procedures, without presuming that multi-expert aggregation necessarily guarantees superior decision accuracy.

2. THEORETICAL FRAMEWORK AND HYPOTHESES

A. Previous Research

The Fuzzy Simple Additive Weighting (FSAW) method has been proven effective in improving decision-making accuracy across various sectors because of its ability to handle qualitative data ambiguities while offering simple, stable, and precise computations. In the education field, FSAW able to outperform the standard SAW method in new student admissions selection with an accuracy rate of up to 85% because of its ability to handle the ambiguity of human language (Lestari et al., 2020), high success in selecting outstanding students according to faculty regulations (Christioko et al., 2017), and selection of Pioneering International Standard School students with a very significant accuracy level reached 95.8% (Rachmad et al., 2018). This success aligns with its flexible application in HR management, this method helps transform manual evaluations into a more objective and measurable system for permanent employee appointments (Putri, 2018). In public services plays a role in minimizing nepotism and subjectivity, like in the selection of village meeting participants (Wardani et al., 2023) and in the distribution of aid for uninhabitable houses, which can achieve up to 100% targeting accuracy (Arzian et al., 2020). In industrial and infrastructure risk analysis like determining maintenance strategies for factory equipment (Sagar et al., 2013), allocating budgets for ICT infrastructure (Yulanda & Utama, 2024), managing risk in water projects in the oil sector (Khazaee & Naseri, 2025), and even evaluating the effectiveness of laws for renewable energy investments (Eti & Yüksel, 2024). Comparatively, FSAW is considered more stable and efficient than Fuzzy TOPSIS because its process is fast and not sensitive to uncertainties in criteria weights (Tafazzoli et al., 2024).

Even so, most previous research—including the initial studies conducted by the author—has been limited to presenting final results in the form of static ranking numbers. This study fills that research gap by developing a Multidimensional Competency Analysis. The novelty offered lies in mapping students' potential into subject groups and attendance criteria simultaneously, which is then integrated with graphical visualization to support more comprehensive strategic decision-making for schools.

B. Expert Evaluation Methodology

According to research by V. Sudakov & T. Sivakova (2020), the main problem in combining assessments from a group of experts to make a valid final decision is:

1. Weakness of Early Defuzzification: Converting experts' fuzzy values into crisp numbers at the start of the process often loses valuable information, like the experts' level of doubt or degree of confidence.
2. Influence of Biased/Extreme Experts: Ordinary averaging methods are very prone to distortion if one expert gives a very high or very low value (outlier).

The mathematical approach to aggregating assessments from multiple experts is as follows:

1. Fuzzy Weighted Sum

The Fuzzy Weighted Sum (FWS) approach combines expert evaluations (X_i) while taking into account the relative competence of each decision-maker, represented through fuzzy confidence weights (W_i). This method has the advantage of accommodating differences in qualifications and authority among experts, so the evaluations aren't just averaged without considering the credibility of the assessors. Group assessment aggregation using FWS is formulated as follows:

$$X^* = \sum_{i=1}^n W_i X_i.$$

where X^* is the result of consensus aggregation, W_i represents the i expert qualification weight, and X_i is the fuzzy rating matrix/value given by the i expert ($i = 1, 2, \dots, n$).

2. Pairwise Comparison of Fuzzy Rank

Pairwise Comparison of Fuzzy Ranks the second approach adopts pairwise comparison by converting fuzzy ratings into a matrix of fuzzy binary relations $R = [r_{ij}]$. This matrix serves to evaluate the degree of preference whether an alternative or object A_i is superior to A_j . To ensure the validity of the group assessment, this approach adapts the mathematical foundation of the Saaty Analytic Hierarchy Process (AHP) through the calculation of the maximum eigenvalue ($\lambda_{\max}$). The calculation aims to measure the degree of mismatch or inconsistency ratio (C) of the fuzzy preference relationship given by experts.

3. Average Fuzzy Ranking

The third approach is the Median Kemeny Fuzzy-based Average Fuzzy Ranking, which is specifically designed to overcome data limitations caused by extreme values or outliers. This algorithm works by constructing a "Fuzzy Median" value (R^v) that is closest to the majority of expert preferences, thus preventing isolated expert opinions from skewing the group aggregation results. The differences in preferences between experts are mathematically measured using the Manhattan Distance. Then, this formulation is transformed into a Linear Programming model so that the consensus computation process can be done quickly, accurately, and efficiently.

Among those three approaches, this study adopts the consensus-based fuzzy average aggregation principle to form a group of experts (FGDM) to directly compare its performance against the Single-Expert model.

C. Fuzzy Set Theory and Linguistic Variables

Fuzzy Set Theory and Linguistic Variables Fuzzy set theory was introduced by Lotfi A. Zadeh to deal with the uncertainty and ambiguity of information in human thought. In contrast to the classic two-value logic (0 or 1), fuzzy logic measures the degree of membership of an element in the range of $[0, 1]$. In real applications, qualitative assessments are expressed using linguistic variables, which are variables whose values are words in natural language (such as "Very Low", "Medium", to "Very High"). The linguistic variables are transformed into mathematical representations in the form of Triangular Fuzzy Numbers (TFN) defined by three parameters of the membership function (a, b, c), where $a \leq b \leq c$.

D. Fuzzy Simple Additive Weighting (FSAW)

Fuzzy Simple Additive Weighting (FSAW) is a hybrid method that integrates fuzzy set theory into the computational structure of Simple Additive Weighting (SAW). This method is designed to solve Fuzzy Multi-

Attribute Decision Making (FMADM) problems when the evaluation criteria are qualitative or ambiguous. Through this approach, subjective assessments are converted into structured numerical values to produce more objective and precise decisions. Systematically, the computation steps of the FSAW method are carried out as follows:

1. Determination of Linguistic Criteria and Variables: Determine m criteria (C_j ; $j=1, 2, \dots, m$) and n of expert/decision-maker groups E_k ; $k=1, 2, \dots, n$). Each subjective assessment is mapped into a linguistic variable in TFN format.
2. Aggregation of Assessment and Weight Defuzzification: Calculates the average fuzzy score of the whole-expert assessment (A_{jk}) for each j criterion:

$$A_{jk} = \frac{f_{j1}^k + f_{j2}^k + \dots + f_{jn}^k}{n}$$

The aggregated fuzzy value (a, b, c) is then converted into a crisp value through the Center of Area (COA) defuzzification method:

$$e = \frac{(a+b+c)}{3}$$

The normalized weight (W_j) is obtained by dividing the defuzzification value of criterion j by the total defuzzification value of the entire criterion.

3. Decision and Normalization Matrix Construction: Converts the alternative data (A_i ; $i=1, 2, \dots, n$) into a fuzzy decision matrix, and then normalizes it to a matrix $R = [r_{ij}]$ according to the criterion attribute:

$$r_{ij} = \begin{cases} \frac{x_{ij}}{\max_i(x_{ij})} & \text{if } j \text{ is benefit} \\ \frac{\min_i(x_{ij})}{x_{ij}} & \text{if } j \text{ is cost} \end{cases}$$

4. Total Score Calculation and Ranking: Calculating the total preference score (TS_i) for each alternative by multiplying the normalized matrix by the criteria weights (W_j):

$$TS_i = \sum_{j=1}^m r_{ij} \cdot W_j$$

The alternative with the highest TS value takes the top spot in the decision-making system.

3. RESEARCH METHOD

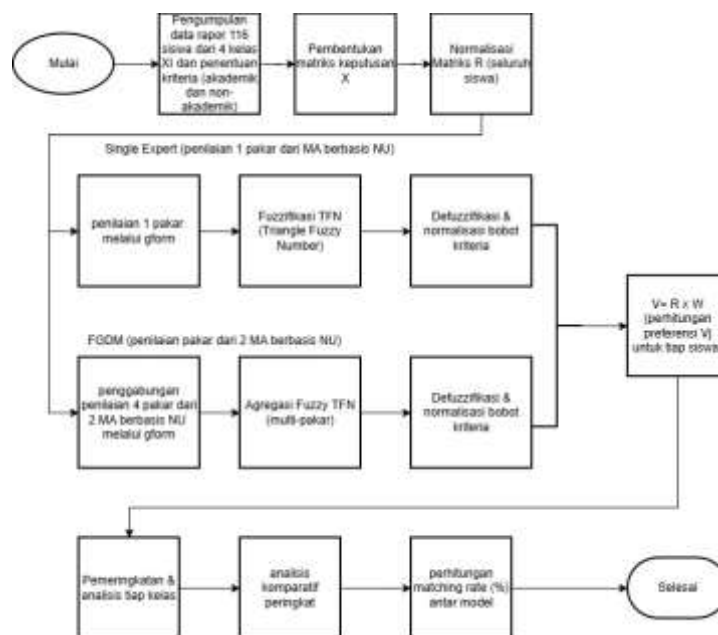


Figure 1. Research Flow

Explanation of Figure 1 is as follows:

1. Formation of the Decision Matrix and Normalization (F-SAW)

The decision matrix X is formed from the real student score data. Next, normalization (R) is carried out according to the nature of the criteria (benefit or cost):

$$r_{ij} = \begin{cases} \frac{x_{ij}}{\max_i(x_{ij})} & \text{if } j \text{ is benefit} \\ \frac{\min_i(x_{ij})}{x_{ij}} & \text{if } j \text{ is cost} \end{cases} \quad (1)$$

2. Single Expert Model Weighting and FGDM Model

Language Conversion to Triangular Fuzzy Scale (TFN):

Each expert k provides linguistic variables that are converted into TFN numbers $\tilde{W}_j^k = (l_{jk}, m_{jk}, u_{jk})$

Group Weight Aggregation (Fuzzy Average):

$$\tilde{W}_j = \left(\frac{\sum l_{jk}}{K}, \frac{\sum m_{jk}}{K}, \frac{\sum u_{jk}}{K} \right) \quad (2)$$

Defuzzification (Center of Area Method):

$$w_j = \frac{l_j + m_j + u_j}{3} \quad (3)$$

Weight normalization:

$$w_j^* = \frac{w_j}{\sum w_j} \quad (4)$$

3. Final Score Calculation (Preference V_i)

The calculation of the preference value of the i -th student alternative (V_i) is done with the formula:

$$V_i = \sum_{j=1}^n w_j^* \cdot r_{ij} \quad (5)$$

4. Model Evaluation Validation Using Spearman Rank Correlation and Mean Absolute Error (MAE)

To evaluate the reliability, consistency, and stability of the Decision Support System (DSS) model being developed, a double comparative validation approach is used. This approach combines macro correlation analysis through Spearman Rank Correlation (r_s) and micro deviation measurement using Mean Absolute Error (MAE). Using these two types of validation is crucial because each metric measures different statistical dimensions and complements each other.

a. Spearman's Rank Correlation (r_s) to Measure Hierarchy Alignment

Spearman's Rank Correlation (r_s) is used as a non-parametric metric based on rank-order analysis to measure the degree of alignment and consistency of alternative order hierarchies between macro decision models (Wisniewski & Brannan, 2024). This metric evaluates whether the monotonic relationship and the main decision structure are maintained without being distorted when processed through different aggregation algorithms (Salabum & Urbaniak, 2020). To ensure the objectivity of interpretation, the correlation strength reference refers to the Schober et al. (2018) standard, where the value of $r_s \geq 0,80$ indicates a "very strong" correlation level (Schober et al., 2018). Correlation strength level by Schober et al. (2018) can be seen in table 1.

Table 1. correlation strength level by Schober et al.(2018)

Correlation coefficient	Interpretation
0,00-0,10	Neglibible correlation
0,10-0,39	Weak correlation
0,40-0,89	Moderate correlation
0,70-0,89	Strong correlation
0,90-1,00	Very strong correlation

However, the Spearman metric has a limitation (blind spot) because it only focuses on the orientation of the relative hierarchy without measuring the distance or magnitude of the physical shift in real position (d_i). Therefore, r_s requires complementary metrics such as MAE to prove the precision level of the shift in a micro-sense.

b. Mean Absolute Error (MAE) to Measure the Magnitude of Position Shift

As a complementary metric, Mean Absolute Error (MAE) is applied to quantify the average shift distance per individual in a linear way based on the absolute value ($|d_i|$) (Karunasingha, 2022). The main advantage of MAE lies in its naturally measured and proportional assessment without giving extra weight to outlier values, making it effective at dissecting both systematic and random shift components when testing model stability (Robeson & Willmott, 2023). However, MAE has its limits since it cannot indicate overall direction or hierarchical order. Therefore, combining r_s and MAE forms a holistic evaluation framework, where Spearman measures macro-level trend consistency, while MAE measures the magnitude of position shift deviations on a micro level (Karunasingha, 2022; Salabum & Urbaniak, 2020).

Combining these two metrics creates a holistic evaluation framework that reinforces itself in assessing model outputs:

- a. Macro Analysis (r_s): Shows the consistency of logic and trend tendencies of group expert recommendations (Fuzzy Group Decision Making) compared to single expert preferences or pure academic values (Academic Report).
- b. Micro Analysis (MAE): Shows the model's precision and stability by measuring how subtle the shifts in average positions produced by the aggregation algorithm are.

The synergy between a high correlation coefficient ($r_s \rightarrow 1.00$) and a low mean absolute error (MAE $\rightarrow 0$) provides solid empirical evidence that the FGDM-based decision model not only maintains macro-level hierarchy patterns but also is very stable and precise on a micro level.

B. Dataset and Assessment Criteria

The subjects of this study included 116 eleventh-grade students divided into four classes (XI-A, XI-B, XI-C, and XI-D) from a Nahdlatul Ulama-based high school (NU based high school), with the research objects being manual report card summaries and student assessment criteria data. The evaluation of achievements and determination of outstanding students in this study involved five multidimensional criteria covering academic and non-academic aspects (C_1 to C_5), detailed as follows:

1. C_1 (Average Score of Religious and General Subjects): Measures academic achievement in mandatory core subjects.
2. C_2 (Average Score of Elective/Specialization Subjects): Measures students' academic mastery in specific fields of study.
3. C_3 (Average Score of Local Content Subjects): Measures learning achievement in the institution-specific curriculum (including NU-related studies).
4. C_4 (Attendance Level/Presence): A non-academic discipline indicator measured based on student absence summaries.
5. C_5 (Extracurricular Activity): Indicator of non-academic activity based on student participation in self-development activities at school.

Details of the Subjects used in C_1 , C_2 , and C_3 based on the subject groups at high school can be seen in Table 1.

Table 2. Details of subjects for criteria C_1 , C_2 , C_3

Kriteria	Kelompok Mata Pelajaran	Daftar Mata Pelajaran
C_1	Religion and General Subjects	Quran hadith, Aqidah and morals, fiqh, Islamic cultural history, Pancasila and citizenship education, Indonesian language, Arabic language, mathematics, national history, English, arts and culture, physical and spiritual education and health, crafts
C_2	Elective Subject	Biology, chemistry, physics, advanced mathematics, sociology
C_3	local content	NU-ness, Javanese language, informatics, takrib/targhib, multimedia

The academic criteria data (C_1 , C_2 , and C_3) come from the average report card grades (on a 0–100 scale). Meanwhile, the discipline criterion (C_4) is converted from the student's absenteeism into an equivalent crisp score, as shown in Table 2.

Table 3. Absence Criteria

Absence condition	Crisp Value	Academic Information
Full Attendance (0 Sick, 0 Excused, 0 Unexcused)	100	Very High Discipline
Light Sick/Excused (Total 1-4 days, 0 Unexcused)	85	Good Discipline (Integrated Permission)
Serious Sick/Excused (Total >4 days, 0 Unexcused)	75	Moderate Discipline (Health/Urgent Factor)
Light Unexcused (1-3 days Unexcused)	60	Lacking Discipline (Without Explanation)
Moderate Unexcused (4-10 days Unexcused)	40	Undisciplined (Homeroom Teacher Warning)
Severe Unexcused (>10 days Unexcused)	20	Very Undisciplined (Special Evaluation)

Setting crisp values for the attendance criterion (C_4) uses a 0–100 scale to stay in line with the academic criteria (C_1 – C_3). The scoring scale is designed based on the level of discipline and the rational impact of absences:

1. Sick and Excused (Scores 85 & 75): Classified as conditions beyond the student's control. Score reductions are kept minimal to maintain fairness, with a score of 75 also serving as the neutral value limit (Fuzzy Neutral Imputation).
2. Absent / Unexcused (Scores 60, 40, & 20): Classified as disciplinary violations. Penalties are applied in a tiered and steep manner (20-point differences) to reflect academic consequences for intentional truancy.

Criterion C_5 , which is participation in extracurricular activities, has conversion scores that can be seen in Table 3.

Table 4. C_5 Criteria

Original Data Condition	Crisp Values	Information
Participating in 2 extracurriculars (mandatory + optional)	100	Shows high activity
Participating in 1 extracurricular (only mandatory)	85	Fulfills self-development obligations according to school standards

4. DATA ANALYSIS AND DISCUSSION

A. Formation of the Decision Matrix (X)

The values for criteria C_1 , C_2 , C_3 , C_4 , and C_5 are processed using predefined values. The criteria conversion data is presented with alternative codes for students from class XIA to XID, totaling 116 students, and the conversion numbers per criterion can be seen in Table 4.

Table 5. Raw data of criteria for the Decision matrix (X)

Alternative	C_1	C_2	C_3	C_4	C_5
A1	83,49038	85,15	81,7	75	85
A2	83,85577	85,35	82	75	85
A3	85,89423	86,15	83,3	75	85
A4	82,875	84,7	81,9	75	100
A5	85,43269	87,7	83,3	75	85
A6	79,80769	69,5	80,8	75	85
A7	84,25	86,65	82,9	75	85
A8	84,86538	86,2	83,1	75	85
A9	85,98077	85,45	82,4	75	85
A10	84,55449	84,65	82,3	75	85
A11	84,72756	87,5	83,2	75	85
A12	84,5609	85,25	83,2	75	85
A13	81,52244	83,05	81	75	85
A14	83,33333	85,8	81,4	75	85
A15	84,65064	76,7	81,6	75	85

Alternative	C1	C2	C3	C4	C5
A16	83,02564	85,7	81,4	75	85
A17	84,35256	86,25	82,5	75	85
A18	83,84615	85,95	83	75	85
A19	82,19872	75,35	82,4	75	85
A20	86,12821	86,25	82,9	75	85
A21	83,69872	87,25	83,5	75	85
⋮	⋮	⋮	⋮	⋮	⋮
A115	75,82692	84,4	81,7	100	85
A116	52,80769	66,35	54,8	20	85
Nilai Max	86,29167	87,7	83,5	100	100

Handling missing values in C_4 for Class XI-A adopts the fuzzy adaptive imputation principle as formulated by Halder et al. (2022). Halder et al. emphasized that missing information in a dataset can disrupt pattern recognition and skew overall evaluation results (Halder et al., 2022). By considering the interrelationship of attributes, students with missing attendance data are not assumed to be absent without excuse, but are categorized as 'Sick/Excused with Explanation.' The data is then fuzzified using the linguistic variable 'Enough' and assigned a neutral crisp score of 75 before being normalized into the R matrix.

B. Normalization (R) with Benefit Criteria

All the data in the Decision Matrix (X) in table 4 are calculated using the benefit formula to get the normalization (R). The benefit criteria formula uses formula 1, which divides the criterion value per alternative by the maximum value per criterion. The results of the normalization (R) calculations can be seen in table 5.

Table 6. Normalization calculation results (R)

Alternative	C1	C2	C3	C4	C5
A1	0,967537	0,970924	0,978443	0,75	0,85
A2	0,971771	0,973204	0,982036	0,75	0,85
A3	0,995394	0,982326	0,997605	0,75	0,85
A4	0,960406	0,965792	0,980838	0,75	1
A5	0,990046	1	0,997605	0,75	0,85
A6	0,92486	0,792474	0,967665	0,75	0,85
A7	0,97634	0,988027	0,992814	0,75	0,85
A8	0,983471	0,982896	0,99521	0,75	0,85
A9	0,996397	0,974344	0,986826	0,75	0,85
A10	0,979869	0,965222	0,985629	0,75	0,85
A11	0,981874	0,997719	0,996407	0,75	0,85
A12	0,979943	0,972064	0,996407	0,75	0,85
A13	0,944731	0,946978	0,97006	0,75	0,85
A14	0,965717	0,978335	0,97485	0,75	0,85
A15	0,980983	0,874572	0,977246	0,75	0,85
A16	0,962151	0,977195	0,97485	0,75	0,85
A17	0,977529	0,983466	0,988024	0,75	0,85
A18	0,97166	0,980046	0,994012	0,75	0,85
A19	0,952568	0,859179	0,986826	0,75	0,85
A20	0,998106	0,983466	0,992814	0,75	0,85
A21	0,969951	0,994869	1	0,75	0,85
⋮	⋮	⋮	⋮	⋮	⋮
A115	0,878728	0,962372	0,978443	1	0,85
A116	0,611967	0,756556	0,656287	0,2	0,85

C. Calculation of criteria weights based on TFN (Triangular Fuzzy Number)

The assessment of criteria importance by experts is collected in a structured way using a digital questionnaire instrument based on linguistic variables. These are variables whose values are words or sentences in natural language (such as "Very High", "Moderate", "Low") to represent human subjective assessments that are imprecise (Tafazzoli et al., 2024; Wisnujati et al., 2023). To handle the subjectivity and uncertainty of experts' perceptions, these qualitative variables are converted into Triangular Fuzzy Number (TFN) sets with the

triplet notation $\tilde{W}_j^k = (l_{jk}, m_{jk}, u_{jk})$ (Tafazzoli et al., 2024; Yulanda & Utama, 2024). The scale mapping linguistic variables to TFN numbers is shown in Table 6.

Table 7. Linguistic variable conversion scale to TFN

Linguistic Variables	Code	TFN Scale (l, m, u)
Very Important	VI	(0.700, 0.900, 1.000)
Important	I	(0.500, 0.700, 0.900)
Quite Important	QI	(0.300, 0.500, 0.700)
Not Important	NI	(0.100, 0.300, 0.500)
Very Unimportant	VU	(0.700, 0.900, 1.000)

The TFN mapping in Table 6 is then used as the basis for weighting criteria in both the Single Expert model and the aggregation of Fuzzy Group Decision Making (FGDM).

1. Single Expert Assessment

The Single-Expert approach bases the determination of criteria weights on the perception of a single internal key informant, namely the Deputy Head of the Curriculum Division at the school under study. In this model, the expert provides qualitative assessments with an importance level preference of "Very Important" (VI) for all criteria (C_1 to C_5). Based on the TFN conversion scale, this preference is transformed into a fuzzy evaluation matrix $\tilde{W}_j = (0.700, 0.900, 1.000)$ for each $j=1,2,3,4,5$. The sum of the TFN triplet components ($\sum l_k$, $\sum m_k$, $\sum u_k$) for the five criteria is as follows:

$$\begin{aligned}\sum_{k=1}^5 l_k &= 0.700 \times 5 = 3.500 \\ \sum_{k=1}^5 m_k &= 0.900 \times 5 = 4.500 \\ \sum_{k=1}^5 u_k &= 1.000 \times 5 = 5.000\end{aligned}$$

Next, the calculation of the normalized fuzzy weights ($\tilde{w}_j$) for criteria C_1 to C_5 is done by dividing the TFN triplet components by the total accumulated value, using formula 2. The normalized TFN value is $\tilde{w}_j = (0.1400, 0.2000, 0.2857)$ for all criteria. The details of the TFN calculations for a single expert assessment are as follows:

$$\tilde{w}_j = \left(\frac{0.700}{5.000}, \frac{0.900}{4.500}, \frac{1.000}{3.500} \right) = (0.1400, 0.2000, 0.2857)$$

The final stage of this model is determining the crisp weight through defuzzification and final normalization. Using the Center of Area (COA) method, the initial crisp value (W_j^*) generated for each criterion is 0.2086. After final normalization to make the total weight equal 1.0000, the final priority weight (W_j) obtained is 0.2000 (20.00%) evenly for all five criteria (C_1 – C_5). The details of the CoA defuzzification calculation for W_j^* and the final normalization for W_j are as follows:

$$W_j^* = \frac{0.1400 + 0.2000 + 0.2857}{3} = 0.2086$$

$$W_j = \frac{0.2086}{0.2086 \times 5} = \mathbf{0.2000} \quad (20.00\%)$$

These results show the characteristics of the Single-Expert model, which produces uniform criterion weights (equal weighting). Next, the TFN calculation will be tested with combined assessments from other NU-based MA schools.

2. Fuzzy Group Decision Making (FGDM) Assessment

The assessment of criteria importance in the FGDM model involves the consensus of four experts (E1 to E4), where E1 is an internal single expert (Vice Principal of Curriculum) and E2–E4 are external experts from similar schools. The positions of these three experts are vice principal of curriculum, vice principal of student affairs, and teacher. The results of collecting linguistic variable assessments from the four experts are presented in

Table 7.

Table 8. Digital questionnaire results

Code	Criteria	Expert 1 (E1)	Expert 2 (E2)	Expert 3 (E3)	Expert 4 (E4)
C_1	Group A Score	VI	VI	VI	VI
C_2	Group B Score	VI	I	I	NI
C_3	Group C Score	VI	I	I	QI
C_4	Attendance	VI	VI	VI	VI
C_5	Extracurricular	VI	I	I	NI

Next, to combine the variations in experts' perceptions into a single group weight, fuzzy aggregation is carried out using the Fuzzy Mean Aggregation approach based on Sudakov & Sivakova's (2020) taxonomy. Each linguistic variable assessment is converted into a triangular fuzzy number (TFN), then the average of its triplet components is calculated for each criterion (C_1 – C_5) using formula 2. The aggregation formula is as follows:

$$\tilde{W}_j = \left(\frac{\sum_{i=1}^4 l_i}{4}, \frac{\sum_{i=1}^4 m_i}{4}, \frac{\sum_{i=1}^4 u_i}{4} \right)$$

The aggregation calculation per criterion uses the aggregation formula ($\tilde{W}_j$) as follows:

$$\tilde{W}_1 (C_1): \left(\frac{0.7+0.7+0.7+0.7}{4}, \frac{0.9+0.9+0.9+0.9}{4}, \frac{1.0+1.0+1.0+1.0}{4} \right) = (0.700, 0.900, 1.000)$$

$$\tilde{W}_2 (C_2): \left(\frac{0.7+0.5+0.5+0.1}{4}, \frac{0.9+0.7+0.7+0.3}{4}, \frac{1.0+0.9+0.9+0.5}{4} \right) = (0.450, 0.650, 0.825)$$

$$\tilde{W}_3 (C_3): \left(\frac{0.7+0.5+0.5+0.3}{4}, \frac{0.9+0.7+0.7+0.5}{4}, \frac{1.0+0.9+0.9+0.7}{4} \right) = (0.500, 0.700, 0.875)$$

$$\tilde{W}_4 (C_4): \left(\frac{0.7+0.7+0.7+0.7}{4}, \frac{0.9+0.9+0.9+0.9}{4}, \frac{1.0+1.0+1.0+1.0}{4} \right) = (0.700, 0.900, 1.000)$$

$$\tilde{W}_5 (C_5): \left(\frac{0.7+0.5+0.5+0.1}{4}, \frac{0.9+0.7+0.7+0.3}{4}, \frac{1.0+0.9+0.9+0.5}{4} \right) = (0.450, 0.650, 0.825)$$

The explanation of the results of this group's consensus aggregation calculation produces the aggregated fuzzy weights ($\tilde{W}_j$) as follows:

$\tilde{W}_1 (C_1): (0.700, 0.900, 1.000)$ — shows the absolute agreement of all experts (VI).

$\tilde{W}_2 (C_2): (0.450, 0.650, 0.825)$ — aggregating assessment variations from VI, I, to NI.

$\tilde{W}_3 (C_3): (0.500, 0.700, 0.875)$ — aggregating assessments from VI, I, up to QI.

$\tilde{W}_4 (C_4): (0.700, 0.900, 1.000)$ — shows the absolute consensus of all experts (VI).

$\tilde{W}_5 (C_5): (0.450, 0.650, 0.825)$ — aggregating assessment variations from VI, I, to QI.

After the aggregated TFN values are obtained, the process continues with the normalization of the fuzzy weights ($\tilde{w}_j$). Summing up the triplet components of all criteria results in totals of $\sum l = 2.800$, $\sum m = 3.800$, and $\sum u = 4.525$. Dividing the aggregated matrix by these total components gives the normalized TFN values ($\tilde{w}_j$) for C_1 through C_5 precisely, as shown in the FGDM normalization stage. The calculation process for the components of all criteria is as follows:

$$\sum l = 0.700 + 0.450 + 0.500 + 0.700 + 0.450 = 2.800$$

$$\sum m = 0.900 + 0.650 + 0.700 + 0.900 + 0.650 = 3.800$$

$$\sum u = 1.000 + 0.825 + 0.875 + 1.000 + 0.825 = 4.525$$

The process of calculating the normalization of fuzzy weights ($\tilde{w}_j$) is as follows:

$$\tilde{w}_1 (C_1): \left(\frac{0.700}{4.525}, \frac{0.900}{3.800}, \frac{1.000}{2.800} \right) = (0.1547, 0.2368, 0.3571)$$

$$\tilde{w}_2 (C_2): \left(\frac{0.450}{4.525}, \frac{0.650}{3.800}, \frac{0.825}{2.800} \right) = (0.0994, 0.1711, 0.2946)$$

$$\tilde{w}_3 (C_3): \left(\frac{0.500}{4.525}, \frac{0.700}{3.800}, \frac{0.875}{2.800} \right) = (0.1105, 0.1842, 0.3125)$$

$$\tilde{w}_4 (C_4): \left(\frac{0.700}{4.525}, \frac{0.900}{3.800}, \frac{1.000}{2.800} \right) = (0.1547, 0.2368, 0.3571)$$

$$\tilde{w}_5 (C_5): \left(\frac{0.450}{4.525}, \frac{0.650}{3.800}, \frac{0.825}{2.800} \right) = (0.0994, 0.1711, 0.2946)$$

The next step is converting it into a definite form through Center of Area (COA) defuzzification to get a crisp value (W_j^*). The process of calculating the crisp value (W_j^*) is as follows:

$$W_1^* = \frac{0.1547 + 0.2368 + 0.3571}{3} = 0.2496$$

$$W_2^* = \frac{0.0994 + 0.1711 + 0.2946}{3} = 0.1884$$

$$W_3^* = \frac{0.1105 + 0.1842 + 0.3125}{3} = 0.2024$$

$$W_4^* = \frac{0.1547 + 0.2368 + 0.3571}{3} = 0.2496$$

$$W_5^* = \frac{0.0994 + 0.1711 + 0.2946}{3} = 0.1884$$

$$\text{Total } \sum W_j^* = 0.2496 + 0.1884 + 0.2024 + 0.2496 + 0.1884 = 1.0784$$

The calculation results in values of $W_1^* = 0.2496$, $W_2^* = 0.1884$, $W_3^* = 0.2024$, $W_4^* = 0.2496$, and $W_5^* = 0.1884$, with a total accumulation of $\sum W_j^* = 1.0784$. After the final normalization ($\frac{W_j^*}{1.0784}$), the final group-based priority weights (W_{-j}) are obtained as follows:

W_1 (Group A Score): 0.2314 (23.14%)

W_2 (Group B Score): 0.1747 (17.47%)

W_3 (Group C Score): 0.1877 (18.77%)

W_4 (Attendance): 0.2314 (23.14%)

W_5 (Ekstracurricular): 0.1747 (17.47%)

Unlike the Single-Expert model, which assigns equal weighting (20% per criterion), the FGDM approach manages to reflect a more realistic hierarchy of importance. Basic Academic Criteria (C_1) and Attendance (C_4) were agreed by the expert group as the top priorities (23.14%), followed by Local Content (18.77%), and then Specialization and Extracurriculars (17.47%).

D. Vector Calculation

1. Single Expert Model

After the decision matrix is normalized (R) and the crisp weights of the criteria are obtained, the next step is to calculate the preference values (V_i) for all 116 student alternatives (A_1 to A_{116}). In the Single-Expert model, the calculation is done using the defuzzified weight matrix $W=(0.2,0.2,0.2,0.2,0.2)$ through weighted summation (Simple Additive Weighting) following the equation $V_i = \sum_{j=1}^n w_j \cdot r_{ij}$.

Since all criteria have the same weight (20%), the final preference value is largely determined by the linear distribution of the students' normalized matrix scores for each criterion. The full details of the preference vector (V_i) calculation for all students in the Single-Expert model are presented in Table 8.

Table 9. Single expert approach vector calculation

Ai	Vi	Ai	Vi	Ai	Vi	Ai	Vi	Ai	Vi
A1	0,903381	A25	0,921877	A49	0,925268	A73	0,817554	A97	0,946464
A2	0,905402	A26	0,909258	A50	0,863054	A74	0,817438	A98	0,952149
A3	0,915065	A27	0,94912	A51	0,815573	A75	0,901779	A99	0,94597
A4	0,931407	A28	0,989129	A52	0,863116	A76	0,905984	A100	0,94922
A5	0,91753	A29	0,930805	A53	0,868425	A77	0,900577	A101	0,940542
A6	0,857	A30	0,772974	A54	0,907239	A78	0,824145	A102	0,753712
A7	0,911436	A31	0,953972	A55	0,929115	A79	0,905995	A103	0,94218
A8	0,912315	A32	0,957484	A56	0,924031	A80	0,908758	A104	0,815924

Ai	Vi	Ai	Vi	Ai	Vi	Ai	Vi	Ai	Vi
A9	0,911514	A33	0,958187	A57	0,91985	A81	0,853894	A105	0,663794
A10	0,906144	A34	0,961974	A58	0,842634	A82	0,793656	A106	0,898886
A11	0,9152	A35	0,922336	A59	0,864932	A83	0,944889	A107	0,813925
A12	0,909683	A36	0,940667	A60	0,857629	A84	0,86359	A108	0,939405
A13	0,892354	A37	0,868045	A61	0,851315	A85	0,943332	A109	0,924666
A14	0,903781	A38	0,835284	A62	0,588855	A86	0,973331	A110	0,931813
A15	0,88656	A39	0,868595	A63	0,854595	A87	0,767187	A111	0,929803
A16	0,902839	A40	0,877198	A64	0,867207	A88	0,774212	A112	0,903392
A17	0,909804	A41	0,927037	A65	0,920485	A89	0,760342	A113	0,932796
A18	0,909144	A42	0,842161	A66	0,918395	A90	0,941351	A114	0,929876
A19	0,879715	A43	0,862398	A67	0,862571	A91	0,950401	A115	0,933909
A20	0,914877	A44	0,866914	A68	0,815921	A92	0,938519	A116	0,614962
A21	0,912964	A45	0,597186	A69	0,860789	A93	0,965732		
A22	0,878679	A46	0,918484	A70	0,832309	A94	0,958927		
A23	0,914293	A47	0,920685	A71	0,805292	A95	0,941832		
A24	0,901845	A48	0,896593	A72	0,887693	A96	0,950759		

2. FGDM Model

Unlike the Single-Expert model, the calculation of preference values (V_i) in the FGDM model integrates the aggregated and defuzzified group consensus weight vector, which is $W=(0.2314,0.1747,0.1877,0.2314,0.1747)$. Through the weighted sum formula $V_i = \sum_{j=1}^n w_j \cdot r_{ij}$, the weights for basic academic criteria (C_1) and attendance (C_4) contribute the most (23.14%) to determining the final value for each student alternative.

Using these non-uniform weights results in preference values that are more sensitive to student strengths on priority criteria. The overall calculation results of the preference vectors (V_i) for 116 students based on the FGDM approach are shown in Table 9.

Table 10. FGDM approach vector calculation

Ai	Vi	Ai	Vi	Ai	Vi	Ai	Vi	Ai	Vi
A1	0,899207	A25	0,920698	A49	0,931943	A73	0,807961	A97	0,950319
A2	0,90126	A26	0,906563	A50	0,854391	A74	0,804285	A98	0,954807
A3	0,911242	A27	0,952636	A51	0,799806	A75	0,89823	A99	0,950425
A4	0,923315	A28	0,988953	A52	0,854527	A76	0,902667	A100	0,953254
A5	0,913092	A29	0,93011	A53	0,854494	A77	0,897667	A101	0,944642
A6	0,856133	A30	0,767831	A54	0,905549	A78	0,808831	A102	0,730524
A7	0,90693	A31	0,95719	A55	0,927956	A79	0,902881	A103	0,946306
A8	0,908133	A32	0,961372	A56	0,923273	A80	0,905319	A104	0,801069
A9	0,908057	A33	0,962131	A57	0,91898	A81	0,842129	A105	0,649133
A10	0,902413	A34	0,965876	A58	0,831593	A82	0,783314	A106	0,899003
A11	0,910578	A35	0,920892	A59	0,856572	A83	0,94924	A107	0,796027
A12	0,905649	A36	0,944403	A60	0,848915	A84	0,854983	A108	0,941094
A13	0,888173	A37	0,859366	A61	0,841718	A85	0,947484	A109	0,926249
A14	0,899406	A38	0,834589	A62	0,605525	A86	0,973638	A110	0,933311
A15	0,885261	A39	0,872109	A63	0,845743	A87	0,755514	A111	0,93148
A16	0,898382	A40	0,868154	A64	0,85819	A73	0,807961	A112	0,900266
A17	0,905509	A41	0,92898	A65	0,918876	A74	0,804285	A113	0,934009
A18	0,904677	A42	0,830585	A66	0,917123	A88	0,752742	A114	0,931091
A19	0,877795	A43	0,853057	A67	0,854261	A89	0,738235	A115	0,935013
A20	0,911169	A44	0,858123	A68	0,806922	A90	0,945291	A116	0,59174
A21	0,907995	A45	0,617629	A69	0,851975	A91	0,954407		
A22	0,876071	A46	0,917286	A70	0,819492	A92	0,942232		
A23	0,909504	A47	0,919944	A71	0,796399	A93	0,969815		
A24	0,897403	A48	0,893557	A72	0,88636	A94	0,963136		

E. Model Validation and Evaluation (Spearman Rank Correlation and MAE)

To test the alignment level of the hierarchy and the magnitude of position shifts between assessment models, a quantitative evaluation was carried out using two main metrics, namely Spearman Rank Correlation (r_s)

and Mean Absolute Error (MAE). This test compared the recommendation results from the Single Expert (SE) model, Fuzzy Group Decision Making (FGDM), and the pure Academic Report (AC) scores across four experimental classes (XI-A to XI-D). A recap of the validation metric calculation results for all classes is presented in Table 10.

Table 11. Evaluation Results of Spearman's Rank Correlation (r_s) and Mean Absolute Error (MAE) Across Decision Models (SE, FGDM, and Academic Report)

Classes	SPEARMAN			MAE		
	SE-FGDM	AC-SE	AC-FGDM	SE-FGDM	AC-SE	AC-FGDM
XI-A	0,993913	0,846957	0,85826087	0,416667	1,8333333	1,5
XI-B	0,996154	0,768462	0,748461538	0,32	3,76	4
XI-C	0,996992	0,732286	0,752339572	0,363636	5,3939394	5,090909091
XI-D	0,998472	0,877617	0,877616501	0,294118	3,5	3,5
AVER-AGE	0,996383	0,80633	0,80916962	0,348605	3,6218182	3,522727273

*notes: Catatan: SE = Single Expert; FGDM = Fuzzy Group Decision Making; AC = Academic Report

The results of the comparative validation testing using Spearman Rank Correlation (r_s) and Mean Absolute Error (MAE) across four experimental classes (XI-A, XI-B, XI-C, and XI-D) show very high algorithm stability. Based on the evaluation between the Single Expert (SE) model and Fuzzy Group Decision Making (FGDM), very high Spearman correlation values were obtained for all classes: 0,9939 (XI-A), 0,9962 (XI-B), 0,9970 (XI-C), and 0,9985 (XI-D), with an average $r_s = 0,9964$. Correlation values close to 1.00 refer to a "Very Strong" relationship category according to Schober et al. (2018). This is further supported by very low MAE values between fuzzy models, with an average MAE = 0,35 levels (ranging from 0,29 to 0,42). The MAE value of < 0,50 proves micro-level that the aggregation of multi-expert preferences through FGDM does not randomize or undermine the hierarchy of key decisions formed by a single expert, but rather only provides very minimal position adjustment. Meanwhile, the comparison between the pure Rapot ranking and the Fuzzy SAW-based model (SE and FGDM) results in an average correlation of r_s approx 0,81 and an average shift in position of MAE approx 3,52 to 3,62 level. The existence of an average position shift of 3 to 5 levels indicates that Fuzzy SAW modeling successfully accommodates and reconstructs the weighting of qualitative criteria that is not accommodated in the accumulation of pure academic scores. Furthermore, the correlation of the FGDM model with the report is generally higher ($r_s = 0.8092$) compared to the Single Expert model ($r_s = 0.8063$), which indicates that multi-expert consensus is effective in reducing the subjectivity of a single evaluator.

A visual representation of the shifts in position between individual models can be seen in the ranking comparison chart for the experiment class sample shown in Figure X.

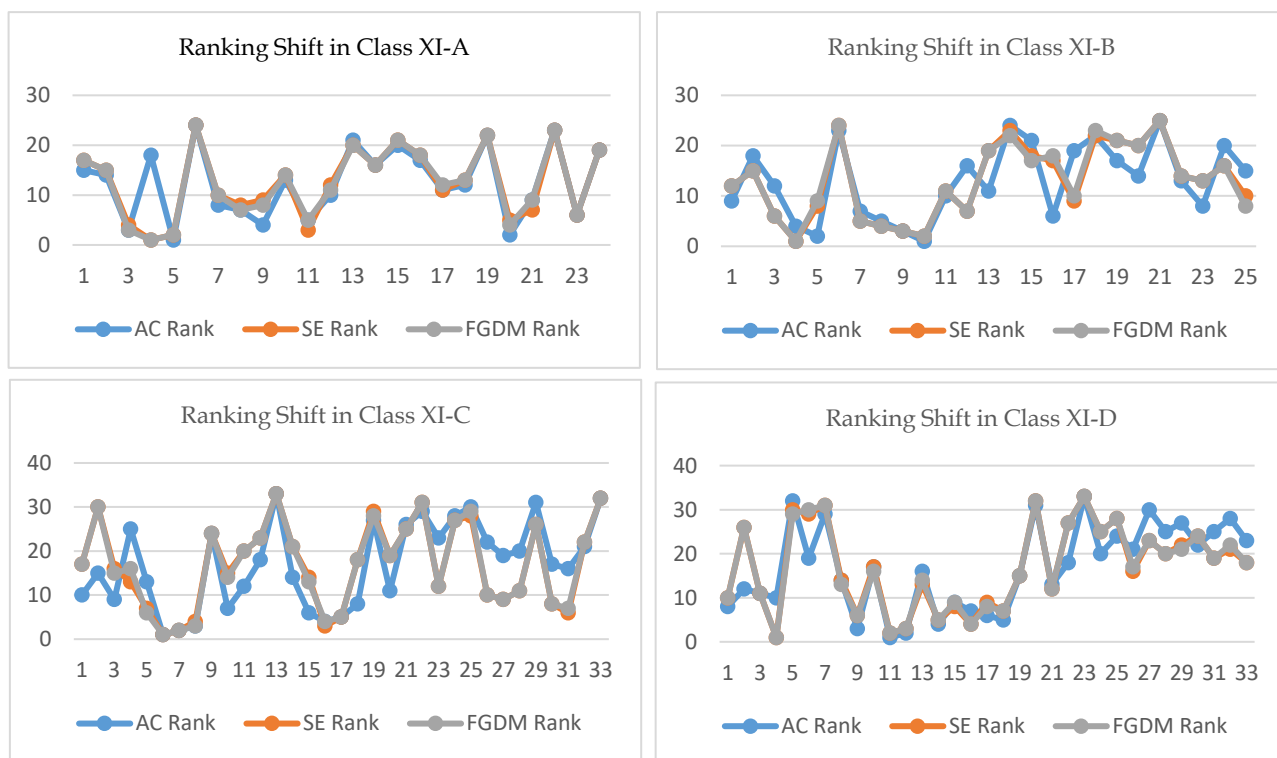


Figure 2. Ranking shift in 4 classes

As illustrated in Figure 2, the rank shift visualisations across all four experimental classes (XI-A to XI-D) strongly reinforce the quantitative metrics obtained in Table 10. The trajectory lines representing the Single Expert model (orange) and the Fuzzy Group Decision Making model (grey) exhibit near-perfect overlap across all individual student profiles. This visual alignment confirms the exceptionally high macro-correlation ($r_s = 0.9964$) and minimal micro-shift ($MAE = 0.35$), demonstrating that multi-expert aggregation via FGDM maintains the structural integrity of expert judgments while smoothing out individual evaluator bias.

Conversely, the Academic Report line (blue) displays notable vertical fluctuations and rank divergence when compared against both Fuzzy-SAW models across all classes. Sharp position shifts observed in specific student cases highlight the structural difference between pure academic score accumulation and the Fuzzy-SAW approach. These visual discrepancies illustrate how the integration of qualitative criteria (such as non-academic achievements and behavioral evaluations) successfully reconstructs the final student rankings, resulting in a more comprehensive, balanced, and objective evaluation system.

5. CONCLUSION, IMPLICATION, SUGGESTION, AND LIMITATIONS

The comparative evaluation of Single Expert (SE) and Fuzzy Group Decision Making (FGDM) within the Fuzzy Simple Additive Weighting (Fuzzy-SAW) framework revealed a very high degree of agreement between the two ranking approaches. This finding is supported by a near-perfect average Spearman's rank correlation coefficient ($r_s = 0.9964$) and a relatively small Mean Absolute Error ($MAE = 0.35$ ranking positions). These results indicate that incorporating judgments from four experts produced rankings that were highly consistent with those generated using a single expert. However, the findings should be interpreted as evidence of ranking agreement and limited positional differences, rather than definitive proof of greater accuracy, objectivity, or reduced cognitive bias.

The comparison with academic report-based rankings also revealed meaningful differences in student positions. Both SE and FGDM demonstrated relatively strong correlations with academic report rankings ($r_s \approx 0.81$), while the average positional differences ranged from approximately 3.5 to 3.6 ranking positions. These findings suggest that incorporating attendance and extracurricular participation alongside academic performance produces a different representation of student achievement compared with rankings

based solely on academic report scores. Nevertheless, these differences should not automatically be interpreted as corrections or improvements to the original academic rankings, since no independent reference standard was employed to determine which ranking arrangement more accurately represents overall student achievement.

The main methodological contribution of this study lies in providing a direct comparison of single-expert and multi-expert criterion weighting within the same Fuzzy-SAW framework. By applying both approaches to identical student records, assessment criteria, and normalization procedures, the study demonstrates how alternative expert-weighting arrangements can be evaluated through ranking agreement and positional differences. The findings indicate that group-based aggregation can incorporate multiple professional perspectives without substantially altering the overall ranking structure in the examined dataset. From a practical perspective, the proposed framework offers schools a structured and transparent procedure for combining academic and non-academic assessment criteria. The use of FGDM may be considered when educational institutions wish to incorporate judgments from several stakeholders in determining criterion importance. However, schools should recognize that the inclusion of multiple experts does not automatically guarantee fairer or more accurate decisions. The assessment criteria, scoring rules, and expert-weighting procedures must remain transparent, clearly justified, and consistent with institutional policies.

Several limitations should be acknowledged. First, the study was conducted using data from 116 students at a single Madrasah Aliyah, which limits the generalizability of the findings to other educational institutions with different curricula, student characteristics, and assessment practices. Second, the FGDM model involved only four experts, consisting of one internal expert and three experts from comparable institutions. Although this arrangement enabled a comparison between individual and collective judgments, it may not fully represent the diversity of perspectives involved in educational assessment. Third, the treatment of missing attendance records involved assigning an imputed score of 75. This assumption may influence individual ranking positions, particularly for students whose final preference scores are relatively close. Furthermore, the study did not establish an independent benchmark for evaluating the accuracy or fairness of the resulting rankings.

Future research should extend the evaluation to multiple schools and larger student populations to examine whether similar ranking patterns emerge across different educational contexts. Involving a broader and more diverse group of experts would also help assess the sensitivity of criterion weights to expert composition. Additionally, future studies should conduct sensitivity analyses using alternative attendance imputation values and scoring schemes to determine their effects on student ranking stability. Comparisons with other multi-criteria decision-making methods and the inclusion of independently established assessment benchmarks are also recommended to strengthen the evaluation of decision accuracy, fairness, and practical usefulness. Overall, the study demonstrates that Fuzzy-SAW can provide a systematic framework for multidimensional student assessment, while the comparison between SE and FGDM offers empirical insight into the effects of alternative expert-weighting arrangements. The findings support the usefulness of transparent and reproducible ranking procedures, but further validation is required before claims regarding superior decision accuracy or bias reduction can be established.

REFERENCES

- Arzian, R., Abidin, Z., Irfan, P., & Yunus, M. (2020). Penerapan Fuzzy SAW untuk rekomendasi penentuan penerima bantuan pembangunan rumah tidak layak huni (Application of Fuzzy SAW for recommendations on determining recipients of development assistance: The house is not livable). *JTIM: Jurnal Teknologi Informasi dan Multimedia*, 2(1), 36–42.
- Cahyani, L. D., Yulianti, L., & Fredricka, J. (2024). Sistem pendukung keputusan pemilihan siswa berprestasi menggunakan metode Simple Additive Weighting (SAW) pada SMK Negeri 03 Kota Bengkulu. *Jurnal Media Infotama*, 20(1), 55–60.
- Christioko, B. V., Indriyawati, H., & Hidayati, N. (2017). Fuzzy multi-attribute decision making (Fuzzy MADM) dengan metode SAW untuk pemilihan mahasiswa berprestasi. *Transformatika*, 14(2), 82–85.
- Eti, S., & Yüksel, S. (2024). Integrating Pythagorean fuzzy SAW and entropy in decision-making for legal effectiveness in renewable energy projects. *Computer and Decision Making: An International Journal*, 1, 13–22.
- Halder, B., Ahmed, M. M., Amagasa, T., Isa, N. A. M., Faisal, R. H., & Rahman, M. M. (2022). Missing information in imbalanced data stream: Fuzzy adaptive imputation approach. *Applied Intelligence*, 52(5),

- 5561–5583. <https://doi.org/10.1007/s10489-021-02741-4>
- Hermansyah, M., Fauziah, D. A., & Muliawan, A. (2023). Decision support system for identifying best students in the class with Simple Additive Weighting (SAW) method. *Inside Jurnal Sistem Informatika Cerdas*, 1(1), 27–35.
- Karunasingha, D. S. K. (2022). Root mean square error or mean absolute error? Use their ratio as well. *Information Sciences*, 585, 609–629. <https://doi.org/10.1016/j.ins.2021.11.036>
- Khazaei, A., & Naseri, M. (2025). Risk assessment of water structure projects using fuzzy multi-attribute decision-making methods: Fuzzy OWA and fuzzy SAW (Case study: S1 wellhead platform in the Salman oil field). *Water Harvesting Research*, 7(2), 334–348.
- Lestari, P. F. I., Prabowo, T. T., & Utomo, W. M. (2020). The effectiveness of Fuzzy-SAW method for the selection of new student admissions in vocational high school. *Letters in Information Technology Education (LITE)*, 3(1), 18–22.
- Mahbub, A. R., Khaerudin, M., & Kharoh, I. (2022). Penerapan metode SAW (Simple Additive Weighting) untuk menentukan siswa berprestasi (Studi kasus pada SMP Negeri 24 Jakarta). *JSI (Jurnal Sistem Informatika)*, 9(1), 193–202. <https://doi.org/10.35968/jsi.v9i1.854>
- Melati, S., & Triyono, G. (2020). Siswa terbaik menggunakan metode Simple Addictive. *Jurnal Idealis*, 3(2), 574–580.
- Putri, D. A. (2018). Penerapan metode Fuzzy SAW sebagai pendukung keputusan pengangkatan karyawan tetap perusahaan. *Jurnal Techno Nusa Mandiri*, 15(1), 31–36.
- Rachmad, A., Syakur, M. A., Widjaya, E., Pramudita, Y. D., Anamisa, D. R., Putro, S. S., Rochman, E. M. S., & Purwanti, E. (2018). The selection of new students RSBI using Fuzzy SAW based application. *The 1st International Conference on Computer Science and Engineering Technology Universitas Muria Kudus*, 30–36. <https://doi.org/10.4108/eai.24-10-2018.2280568>
- Rahman, I. A., Daniati, E., & Wardani, A. S. (2023). Penerapan metode Simple Additive Weighting (SAW) dalam penentuan siswa kelas unggulan (Studi kasus MTsN 1 Nganjuk). *Jurnal Borneo Informatika & Teknik Komputer*, 3(1), 12–24.
- Ramadhan, R. F. (2024). Achievement student selection system using the SAW method and WSM method. *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, 10(1), 108–115.
- Robeson, S. M., & Willmott, C. J. (2023). Decomposition of the mean absolute error (MAE) into systematic and unsystematic components. *PLOS ONE*, 18(2), 1–8. <https://doi.org/10.1371/journal.pone.0279774>
- Sagar, M. K., Jayaswal, P., & Kushwah, K. (2013). Exploring fuzzy SAW method for maintenance strategy selection problem of material handling equipment. *International Journal of Current Engineering and Technology*, 3(2), 600–605.
- Salabum, W., & Urbaniak, K. (2020). A new coefficient of rankings similarity in decision-making problems. *International Conference on Computational Science*, 632–645. <https://doi.org/10.1007/978-3-030-50417-5>
- Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia*, 126(5), 1763–1768. <https://doi.org/10.1213/ANE.0000000000002864>
- Sudakov, V., & Sivakova, T. (2020). Features of aggregation of fuzzy expert judgments. *IOP Conference Series: Materials Science and Engineering*. <https://doi.org/10.1088/1757-899X/927/1/012034>
- Syafi'i, M., Samsudin, M., Abidin, Z., & Basarrudin, M. (2025). Evaluasi pendidikan sebagai dasar pengembangan instrumen penilaian berbasis kompetensi. *Jurnal Akuntansi, Manajemen, dan Ilmu Pendidikan (JAMED)*, 1(4), 1–12.
- Tafazzoli, M., Hazrati, A., Shrestha, K., & Kisi, K. (2024). Enhancing contractor selection through fuzzy TOPSIS and fuzzy SAW techniques. *Buildings*, 14(6), 1–24.
- Wardani, I. K., Utomo, P., & Prabowo, D. W. S. (2023). Sistem pendukung keputusan pemilihan peserta musyawarah dusun menggunakan metode Fuzzy SAW. *Journal of Information Technology Ampera*, 4(2), 169–183.
- Wijaya, A. H., & Ipriadi. (2025). Implementasi metode Simple Additive Weighting (SAW) dalam penentuan siswa berprestasi di SDN 36 Lubuk Batu. *Jurnal Ilmu Komputer dan Informatika*, 1(3), 89–95.
- Wisnujati, A., Widodo, A. M., Mudjijana, Tambunan, K., Rahaman, M., & Chen, H.-C. (2023). Implementation of fuzzy Simple Additive Weighting (Fuzzy-SAW) as decision support system for leave scheduling. *AIP Conference Proceedings*, 1–11.
- Witanto, M. A., Santoso, E., & Suprpto. (2020). Sistem pendukung keputusan menentukan siswa berprestasi

menggunakan metode Weighted Product dan Simple Additive Weighting (Studi kasus: SMPN 2 Bululawang Kabupaten Malang). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 4(10), 3770–3776.

Yulanda, R., & Utama, D. N. (2024). Fuzzy SAW based decision model for determining the priority scale of ICT handling in public sector organizations. *Journal of Applied Data Sciences*, 5(4), 1963–1976.

Yusnira, D. R., & Saputri, T. A. (2023). Penerapan metode Simple Additive Weighting (SAW) pemilihan mahasiswa terbaik pada STMIK Dharma Wacana. *Cyberspace: Jurnal Pendidikan Teknologi Informasi*, 7(2), 93–108.