# A Fuzzy Logic-Based Scoring Model for Comprehensive Physical Fitness Assessment in Senior High School

**Puji Ratno[1], Usman Nasution[2], Ichwanul Muslim Karo Karo[3], Zulfan Heri[4], Haris Kurniawan[5]**

[1,2,4,5]Faculty of Sports Science and Health, Universitas Negeri Medan
[3]Computer Science, Universitas Negeri Medan

**HOW TO CITE:**
Puji Ratno, Usman Nasution, Ichwanul Muslim Karo Karo, Zulfan Heri, Haris Kurniawan (2026). A Fuzzy Logic-Based Scoring Model for Comprehensive Physical Fitness Assessment in Senior High School. International Journal of Special Education, 41(17s), 1090-1098

**COPYRIGHT STATEMENT:**
Copyright: © 2026 Authors. Open access publication underthe terms and conditions of the Creative CommonsAttribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

**ABSTRACT:**
Physical fitness assessments were implemented in educational institutions, military, or police for admission selection and generally still used a simple average method. The approach was fully compensatory, where a low score in a particular test component can be compensated by a high score in another test, even though the objectives of the tests are different and cannot substitute for each other. This condition has the potential to create hidden failure, which is a significant weakness hidden behind a composite score. This study aims to develop and validate an assessment model based on the Mamdani Fuzzy Inference System (FIS) as an alternative to conventional methods. Data were obtained from 136 candidates (61 males, 75 females aged 15–17 years) on a high school in Medan, Indonesia, with having complete six test components: push-up, sit-up, pull-up/chinning, shuttle run, 12-minute run, and swimming. The fuzzy membership functions were designed separately by gender based on the institution's official conversion table, supported by statistical analysis that confirmed significant performance differences between genders in five out of six components. Three rule base scenarios were implemented. The results showed that the conventional method passed 56 participants (41.2%), while the non-compensatory fuzzy rule only passed 5 participants (3.7%) and confirmed a significant scale of hidden failure. The main contribution of this study is the design of membership functions and the application of FIS in the context of institutional selection/certification (not pedagogical), with a rule base that can be adjusted according to selection needs. The study concludes that the fuzzy approach, explicitly and measurably, is capable of revealing hidden weaknesses behind conventional average scores..

Keywords: Fuzzy inference system, Physical fitness assessment, Hidden failure, Membership function, Admission.

## 1. Introduction

Physical fitness assessments are instruments used widely in the military, police, paramilitary institutions, even educational institutions in Indonesia to assess physical readiness and fitness (Orr et al., 2022; Sepahvand et al., 2025). This assessment consists of several tests: pull-

up/chin-up, push-up, sit-up, shuttle run, swimming, and a timed long-distance run (12 minutes) (Zhang et al., 2025). The raw scores (number of repetitions, completion time, or distance) for each test are converted into standardized scores (0–100) using an institutional conversion table. The final test score is calculated by averaging all test scores.

Although the test score with the average is practically administrative, the simple average is a fully compensatory aggregation rule (Blancas & Lozano-Oyola, 2022; Yadlowsky et al., 2025). A low score on an easier test can be offset by a high score on another test. The underlying weakness has long been recognized in the literature on composite index measurement. The existing formulas implicitly assume full substitution between tests, whereas each test measures distinct elements of physical readiness and fitness and is not equivalent to another test. An illustration, a candidate is unable to swim adequately, but is strong in chin-ups. She still has the potential to pass with the existing calculation method. The condition is crucial and high-risk in the workplace. This situation gives rise to hidden failures and is serious enough to jeopardize operational readiness.

This study proposes a Mamdani-type fuzzy inference system (FIS) as an alternative aggregation method for fitness assessment. FIS aggregation is based on the discretization of the six test scores using an institutional conversion table, membership functions, and a rule base derived from expert judgment. The results of this fuzzy aggregation are then empirically compared with existing methods. This comparative validation is intended to provide empirical evidence, rather than merely theoretical arguments. In addition, it aims to determine whether the compensatory nature of the simple average poses a practically significant risk in fitness tests, and whether the non-compensatory fuzzy-based alternative actually alters the pass/fail results.

Two clusters of literature underpin this study. First, the application of fuzzy logic to the assessment of physical fitness and athletic performance (Li & Jawis, 2024; Tüű-Szabó & Kóczy, 2025). This application has consistently produced results superior to those of linear scoring. The fuzzy logic approach is also capable of capturing nonlinear relationships and threshold sensitivity between subcategories: Mamdani-based classification of student fitness (Bora et al., 2025), the FIS framework for ranking college students' fitness levels (Xie, Bai, & Li, 2026), and a comprehensive AHP-entropy-based fuzzy evaluation of the effects of sprint training (Song, 2026). However, their application remains within pedagogical contexts or elite athletic performance, not professional pass/fail certification. The second approach focuses on additive aggregation or weighted averaging. This approach allows for full compensation across dimensions that are, in reality, not interchangeable, and proposes non-compensatory alternatives such as geometric aggregation (Munda, 2012; Mazziotta & Pareto, 2013; OECD, 2008). However, this criticism of compensability has never been linked to fuzzy inference systems nor applied to institutional physical readiness tests such as physical fitness assessments.

The novelty of this study lies in three aspects: (1) The FIS is generated based on an institutionally standardized score conversion table, so it is not a new assessment instrument; (2) this study is conducted within the context of operational selection/certification, not a pedagogical context; (3) the criticism regarding compensability from the composite index literature is operationalized into concrete and measurable comparisons.

To empirically validate the above argument, the following section outlines the methodological steps taken in this study. It begins with the sources and structure of the physical fitness test data used, followed by the process of fuzzifying component scores based on the standardized conversion table. The formulation of the rule base for the Mamdani fuzzy inference system is then described. Finally, a comparison scheme is presented, specifically designed to identify and quantify cases of hidden failure between the simple-average method and the proposed fuzzy method.

## 2. Methodology

### 2.1. Research Design

This study employs a quantitative applied computational modeling design. The study develops and validates a fuzzy logic-based decision support model as an alternative for determining physical fitness test results. The proposed instrument is based on the Mamdani Fuzzy Inference System (FIS), and its performance is compared with that of conventional methods. This design combines two stages: (1) a descriptive/exploratory analysis of empirical test data to characterize score distributions, gender-based disparities, and intercomponent correlations, and (2) the development and validation of the fuzzy inference system, in which the empirical findings from the first stage serve as the basis for formulating membership functions and the rule base. Figure 2.1 shows the research process. It begins with the collection of exam data from entrance exam participants, which is then processed in parallel using two approaches: the baseline method, which calculates the simple average of component scores, and the Mamdani Fuzzy Inference System (FIS) method, which involves fuzzifying each input variable into linguistic terms (Low, Medium, High), rule-based inference, and defuzzification into a single comprehensive score. The final results of both approaches are then quantitatively compared to identify and measure cases of hidden failure.

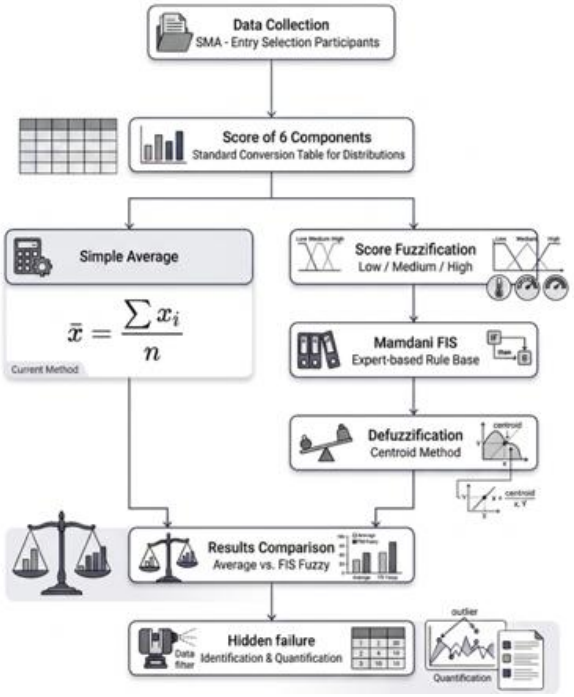


Figure 2.1: Research Workflow

### 2.2. Participants

This dataset consists of 195 prospective students aged 15–17 who applied to a high school in Medan, Indonesia. They were required to take a physical fitness test as part of the admissions process. A total of 136 candidates (68.0%) consisted of 61 male candidates (M) and 75 female candidates (F) who completed the full series of six fitness tests and were included in the quantitative analysis; the remainder were excluded.

### 2.3. Instruments

Physical fitness tests are evaluated using a series of six test components (push-ups, sit-ups, pull-ups/chin-ups, shuttle run, 12-minute run, and swimming).

**Existing Instrument**. The conventional approach converts the achievement scores for each test into a 0–100 scale using a reference table developed by the school, then calculates a comprehensive fitness score using the arithmetic mean. The resulting average is then mapped to

linguistic categories (Very Good/Good/Fair/Poor/Very Poor). This method treats all six components as equally important and assumes that their relationship to overall fitness is linear and compensatory**.**

**Proposed Instrument.** The proposed method involves replacing the linear average with fuzzy logic based on the Mamdani Fuzzy Inference System (FIS). Each of the six component scores is treated as an input variable and fuzzified using membership functions. The Mamdani-type FIS is implemented in five sequential stages (shown in Figure 2.2). **Stage 1** — Fuzzification. Each score of the six tests (push-ups, sit-ups, pull-ups/chin-ups, shuttle run, 12-minute run, and swimming) is converted into a membership degree across five categories (very poor, poor, fair, good, and very good). The membership functions were formulated, and their boundary parameters were derived from the applicable institutional regulations. **Stage 2** — Generated Rule Base. A set of IF–THEN rules encodes expert/empirical knowledge that links the six fuzzified inputs to a single output variable. These rules are formulated separately for male and female participants. There are three scenario rules in this study : (1) IF Score >= 50 THEN Pass ELSE Fail. (2) IF one of tests ∈ {Very Poor} THEN Fail; ELSE IF Score >= 50 THEN Pass ELSE Fail. (3) IF one of tests ∈ {Very Poor, Poor} THEN Fail; ELSE IF Score >= 50 THEN Pass; ELSE Fail. Number of 50 was guidance from institution. Based on these rules, which rules can prevent and eliminate hidden failure issues. **Step 3** — Rule Evaluation (Inference). For each prospective student, every rule in the base is evaluated based on the fuzzified input values. **Step 4** — Aggregation. The fuzzy output sets generated by all triggered rules are combined into a single aggregate fuzzy set for the output variable. **Step 5** — Defuzzification. The aggregated fuzzy output is converted back into a single, crisp comprehensive suitability score using the centroid (center of gravity) method, shown in equation below, where $\mu(z)$ is the aggregated membership function and $z^*$is the resulting crisp score.

$$z^* = \frac{\int \mu(z)\, z\, dz}{\int \mu(z)\; dz}$$

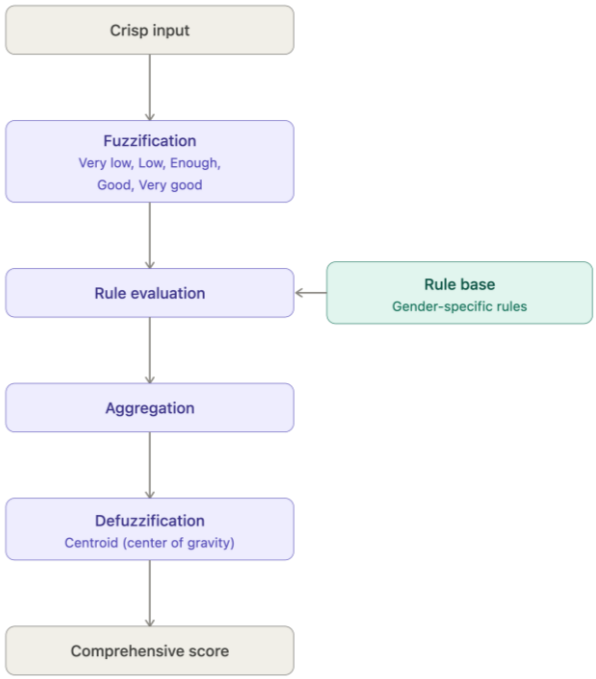


Figure 2.2: FIS Workflow

### 2.4. Procedure

**Data cleaning and standardization.** Prior to processing, the raw dataset underwent a pre-processing stage to address inconsistencies in units. In addition, prospective students who did not take a series of tests were excluded from the

analysis. This section also split the dataset by gender.

**Score conversion.** Raw scores are converted to a scale of 0–100 using the institution's specific conversion table based on gender. This is the data used for the existing instrument. Conversely, the proposed instrument employs raw scores**. Exploratory statistical analysis.** This study presents descriptive statistical information by gender to empirically describe the behaviour of each component. These findings directly serve as the basis for the generation of FIS-based fuzzy logic membership functions.

**Parallel scoring with both instruments.** The dataset was processed in parallel using two tools. Both tools were implemented in Python**.** The results of the physical fitness test from each instrument were compared. Cases of inconsistency were identified, in which the two instruments assigned different categories to the same participant. The characteristics of these cases provide quantitative evidence for the phenomenon of hidden failure discussed in this study.

### 2.5. Data Analysis

Data analysis in this study was conducted in three stages: (1) descriptive statistical analysis of the empirical data, (2) development and application of the Mamdani fuzzy inference system, and (3) a quantitative comparison between the proposed instrument and existing instruments.

## 3. Results

### 3.1. Data Characteristic

This study obtained test data from 195 prospective students (PS). Table 1 shows raw score of prospective student. Only 136 participants (61 males, 75 females) had complete data for all six tests, while the rest did not take the tests. The raw scores (number of repetitions, completion time, or distance).

**Table 1. Prospective Student Raw Score**

| ID PS | M/ F | Push-Up | Sit-Up | Pull-Up/ Chinning | Shuttle Run (s) | 12-minute run (km) | Swimming (m) |
|---|---|---|---|---|---|---|---|
| 001 | M | 24 | 50 | 0 | 16.91 | 1.694 | 32.58 |
| 002 | M | 55 | 46 | 7 | 17.01 | 2.509 | 21.48 |
| 003 | M | 45 | 35 | 8 | 17.89 | 1.839 | 34.55 |
| . . . | . . . | . . . | . . . | . . . | . . . | . . . | . . . |
| 133 | F | 19 | 8 | 38 | 21.42 | 1.267 | 0 |
| 134 | F | 29 | 40 | 49 | 19.9 | 1.2 | 0 |
| 135 | F | 24 | 39 | 45 | 18.85 | 1.97 | 0 |
| 136 | F | 24 | 46 | 44 | 19.57 | 2.06 | 0 |

Additionally, the statistical data are presented in Table 2. There was a performance disparity between male and female participants, although it was not absolute. Male participants outperformed female participants in five of the six components (push-ups, sit-ups, shuttle run, 12-minute run, and swimming), but not in the pull-up/chin-up test. This discrepancy is due to differences in the test protocols and scoring standards applied to the modified pull-up (chin-up) for female participants. Additionally, a clear disparity was also in performance across the tests: sit-ups and the shuttle run scored relatively high (mean >70), while pull-ups, the 12-minute run, and swimming scored lower (mean <30). The available statistical data directly supports the argument of "hidden failure." The pull-up, 12-minute run, and swimming tests are weak components and are at risk of being "masked" by tests with high scores.

**Table 2. Descriptive Statistics of Test Components**

| Component Tests | Mean (M) | Mean (F) | SD (M) | SD (F) | *t* | *p* | Cohen's *d* |
|---|---|---|---|---|---|---|---|
| Push-up | 78.2 | 37.2 | 22.4 | 24.7 | 10.12 | <0.001 | 1.73 |
| Sit-up | 94.8 | 61.7 | 10.8 | 24.6 | 10.44 | <0.001 | 1.68 |
| Pull-up/chinning | 11.6 | 29.9 | 17.4 | 18.8 | −5.89 | <0.001 | −1.01 |
| Shuttle run | 76.6 | 72.5 | 18.2 | 10.9 | 1.54 | 0.126 | 0.28 |
| Run for 12 minutes | 34.5 | 25.7 | 14.9 | 11.7 | 3.74 | <0.001 | 0.66 |
| Swimming | 43.4 | 15.7 | 41.1 | 29.7 | 4.38 | <0.001 | 0.78 |

In addition, the independent t-test confirmed significant differences between genders in five out of six test components ($p < 0.001$). The effect sizes varied. Push-Ups and Sit-Ups show very large effects ($d = 1.73$ and 1.68). Pull-Ups, on the other hand, show a reverse effect ($d = -1.01$), with females outperforming males. Only the Shuttle Run did not show a significant difference ($p = 0.126$).

### 3.2. Design of Fuzzy Membership Functions

Each membership function is derived directly from the institution's official raw score conversion table. Since, there are differences in performance between men and women, the test protocols are also differentiated by gender. Full overlap between adjacent groups allows a single score to have partial membership in two adjacent categories, enabling this model to capture performance at the cut-off point more realistically than conventional threshold-based methods.

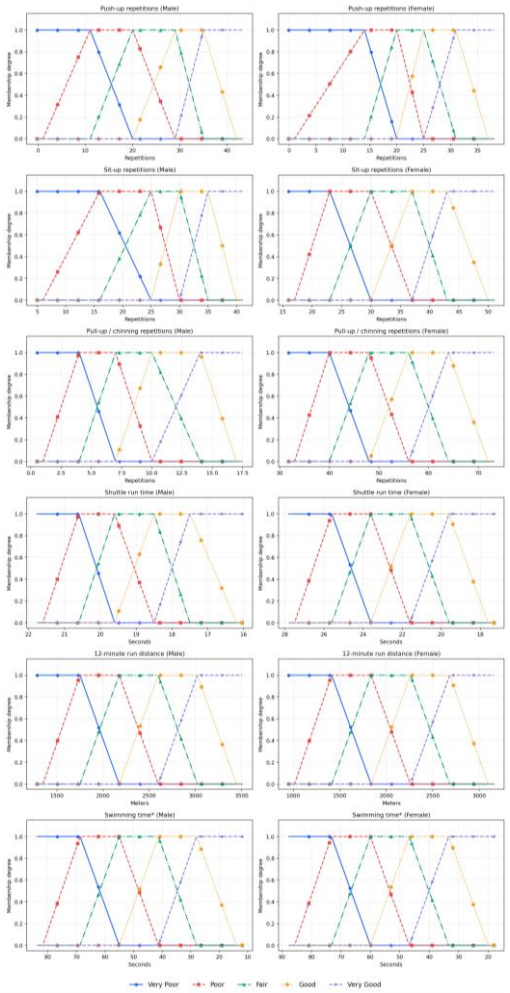


Figure 3.1: Membership Functions of Physical Fitness Tests Based on Gender

### 3.3. Fuzzification

Table 3 presents the fuzzification output for all participants. Each column of the test components displays the 1090-1098 highest membership degrees (μ). Meanwhile, the z* column is the comprehensive score that has been defuzzified through centroid aggregation. Each raw test data was successfully converted using the designed Fuzzy Membership Functions. For example, a participant with ID PS 1 (male) shows a diverse fitness profile: full membership in the Very Good (VG) category for Sit-Ups and a tendency toward Fair (F) for Push-Ups (with partial membership of 0.56 in Poor (P)), strong performance in Shuttle Run and Swimming (both leaning toward VG), but a clear weakness in Pull-Ups/Chinning (Very Poor (VP), μ = 1.00) and the 12-Minute Run (dominated by Very Poor (VP) with an additional membership of 0.86 in Poor (P)), and this combination results in a final comprehensive score z* = 54.01 after defuzzification.

**Table 3. Fuzzification of Prospective Student Raw Score**

| ID PS | M/F | Push-Up | Sit-Up | Pull-Up/ Chinning | Shuttle Run (s) | Run for 12 Minutes (m) | Swimming (m) | $z^*$ |
|---|---|---|---|---|---|---|---|---|
| 1 | M | F(1.00) P(0.56) | VG(1.0) | VP(1.00) | VG(1.0) G(0.55) | VP(1.00) P(0.86) | G(1.00) VG(0.66) | 54.01 |
| 2 | M | VG(1.00) | VG(1.0) | P(0.5) F(0.5) | VG(1.0) G(0.62) | F(1.00) G(0.79) | VG(1.00) G(0.53) | 69.09 |
| 3 | M | VG(1.00) | G(0.5) VG(0.5) | F(1.00) P(0.67) | G(1.00) VG(0.6) | P(1.00) VP(0.79) | G(1.00) VG(0.51) | 61.49 |
| . . . | . . . | . . . | . . . | . . . | . . . | . . . | . . . | |
| 133 | F | P(1.00) F(0.83) | VP(1.00) | VP(1.00) P(0.71) | G(1.00) F(0.91) | VP(1.00) P(0.64) | VP(1.00) | 29.67 |
| 134 | F | G(1.00) VG(0.67) | G(1.00) F(0.50) | F(1.00) P(0.88) | G(1.00) VG(0.8) | VP(1.00) P(0.47) | VP(1.00) | 53.47 |
| 135 | F | F(1.00) G(0.80) | G(1.00) F(0.67) | P(1.00) F(0.62) | VG(1.00 G(0.62) | F(1.00) P(0.68) | VP(1.00) | 51.46 |
| 136 | F | F(1.00) G(0.80) | VG(1.0) G(0.57) | P(1.00) VP(0.50) | VG(1.00 G(0.98) | F(1.00) G(0.53) | VP(1.00) | 54.2 |

### 3.4. Comprehensive Decision Outcomes

Table 4 presents an overview of the decision results between the conventional instrument and FIS for all 136 participants. The conventional instrument uses the unweighted arithmetic mean of the six scaled test component scores (0–100). With a threshold of 50, the conventional method decided that 56 out of 136 participants (41.2%) passed the fitness test. There are three rule-based scenarios under FIS. Rule-based (1) only considers the score (IF Score >= 50 THEN Pass ELSE Fail). This rule decided that 68 participants (50.0%) passed. Rule-based (2) evaluates the performance of each test and score z* (IF one of the tests ∈ {Very Poor} THEN Fail; ELSE IF Score >= 50 THEN Pass ELSE Fail). This rule set a stricter passing criterion than before, with only 18 participants (13.2%) passing. Meanwhile, rule-based (3) has the strictest criteria (IF one of

the tests ∈ {Very Poor, Poor} THEN Fail; ELSE IF Score >= 50 THEN Pass; ELSE Fail), allowing only 5 participants (3.7%) to pass.

**Table 4. Fuzzification of Prospective Student Raw Score**

| ID PS | Conventional Method | Decision by Rule-Based (1) | Decision by Rule-Based (2) | Decision by Rule-Based (3) |
|---|---|---|---|---|
| 1 | Pass | Pass | Fail | Fail |
| 2 | Pass | Pass | Pass | Fail |
| 3 | Pass | Pass | Pass | Fail |
| . . . | . . . | . . . | . . . | . . . |
| 133 | Fail | Fail | Fail | Fail |
| 134 | Fail | Pass | Fail | Fail |
| 135 | Fail | Pass | Fail | Fail |
| 136 | Pass | Pass | Fail | Fail |

The rows in Table 6 illustrate the mechanisms behind each scenario. a prospective student with ID PS 1 ($z^* = 54.01$) passed conventionally and under Rule-Based (1), but failed under Rule-Based (2) because Pull-Up was dominated by the VP category (Table 4). Two prospective students with IDs PS 2 and 3 ($z^* = 69.09$; 61.49) were decided to pass under Rule-Based (2) because no component reached VP, but failed under Rule-Based (3) because there was a test that scored Poor (P). Cases like this are further discussed in the hidden failure discussion section.

## 4. Discussion

Statistical evidence in Section 3.1 supports the decision to separate membership functions based on gender. The effect size is very large for Push-Ups, Sit-Ups, and Pull-Ups ($|d| > 1.0$), indicating that the distributions of both genders are nearly non-overlapping. The combined membership function is definitely biased toward one gender. The reversed effect direction of the Pull-Up is the most significant. Women are actually superior to men. One membership function is impossible to represent this pattern without distorting one of the genders.

### 4.1. Interpretation of Hidden Failure

The results in Section 3.4 confirm that the conventional compensatory assessment obscures the weaknesses of PS in several tests. The FIS evaluator with rule-based (1) passed 50% of PS, which turned out to reveal a compensatory issue. The passing status of PS is due to several strong tests compensating for at least one weak test. For example, PS with Id 134 can still pass, even though she has two tests (Run for 12 Minutes and Swimming) rated as Very Poor. This study notes that rule-based (1) still results in 100% compensatory issues. Therefore, it is not recommended to replace the conventional method. The FIS evaluator with rule-based (2) passed 18 PS (13.2%). However, 13 of 18 PS did not fully pass each test. Some tests performed were still Poor (P), for example, PS with ID 003 was declared passed even though the pull-up and running for 12 minutes tests were still poor. This study refers to the aforementioned cases as partially compensatory, where PS can still be declared passed : if there is progress in the test components, regardless of whether the scores are poor. Of course, FIS with rule-based (2) approach is not yet perfect for addressing the

compensatory issue, but considering the need for student admissions, it can still be considered. FIS assessment with rule-based (3) passed 5 participants (3.7%). This study identified that the five participants passed with a minimum score of Fair on each test and $z^* > 50$. So it is ensured that there is no compensatory issue. Of course, those selection rules are very strict for the selection of high school students.

### 4.2. Practical Implication for Admission Policy

Physical fitness tests are not exclusively used for school admission selection; they are also commonly used for military, police, and paramilitary institution admission selection. The distinction in objectives serves as the basis for establishing graduation criteria. Military and police recruitment demand highly stringent standards and minimal compensation, because personnel must be ready to face real field conditions. For example, a police officer on duty requires swimming skills; no matter how fast someone can run, it cannot replace that. On the other hand, school admissions are developmental, not operational: the goal is to select students with adequate physical abilities and basic knowledge, to be further developed. Therefore, it is reasonable that the criteria for school student admissions are partially compensatory.

The three rule-based FIS configurations should not be viewed as competing candidates, but rather as demonstrations of the FIS's ability to adapt to different acceptance needs. Our findings show that the FIS, which adapts to the conditions of need, rule-based (1) passed 50.0%, rule-based (2) is partially compensatory passing 13.2%, and rule-based (3) is non-compensatory. That fact actually serves as evidence that the designed FIS membership architecture can be configured for stricter selection contexts (such as military/police) simply by tightening the rule base. This flexibility is the main practical advantage of the fuzzy approach compared to the conventional average. Instead of hiding a single compensation assumption within the average formula, the rule base becomes an explicit policy instrument that can be tuned according to needs. For example, if an institution needs workers who rely more on hand strength, then the rule-based tests for pull-ups/chin-ups, push-ups, or swimming can be adjusted accordingly.

### 4.3. Limitation of Study

This research designs limited fuzzy membership functions based on the assessment guidelines of physical fitness from a single educational institution. The extension of fuzzy membership functions to military or police admission rules requires further adjustments, as the guidelines are classified/limited. Thus, although the developed FIS architecture is structurally adaptive across admission contexts, the boundaries of membership functions, rule base thresholds, and graduation rates reported in this study cannot be assumed to directly apply to military or police selection.

## 5. Conclusion

- Summarize key findings
- Suggest recommendations or future research directions

## 6. References

- Follow APA 7th edition format
- Example:
  - Journal article: Smith, J. (2022). Inclusive education practices. *Journal of Special Education*, 34(2), 45–60.
  - Book: Brown, L. (2021). *Disability and learning*. New York, NY: Academic Press.

## 7. Acknowledgements

The authors fully acknowledged Universitas Negeri Medan, Indonesia for the approved fund which makes this important research viable and effective.

## 8. Author Declaration

Authors declare that there is no conflict of interest regarding the publication of this paper