NSS

Neutrosophic Sets and Systems, Vol. 83, 2025

University of New Mexico

Assessment of China's Middle School Physical Education Academic Proficiency Examination Under the Background of New Era Education Reform: Neutrosophic Archimedean t-Norm

Pengfei Luo*

Taiyuan Institute of Technology, Taiyuan, 030000, Shanxi, China

*Corresponding author, E-mail: lpf19751109@163.com

Abstract

As China advances its efforts toward educational modernization in the New Era, the integration of physical education (PE) into the broader framework of academic proficiency has become a central focus of reform. The Middle School Physical Education Academic Proficiency Examination marks a significant departure from traditional PE assessment methods, introducing a more standardized, data-informed, and holistic evaluation system.

This paper explores the development, effectiveness, and implications of this examination model using a multi-criteria decision-making (MCDM) framework. Set against the backdrop of national education reform, the study offers insights into how physical fitness, skill development, and student growth are being redefined within China's middle school system. To support the evaluation, two MCDM techniques are employed: the Entropy method, which determines the weight of each criterion, and the VIKOR method, which ranks the assessment alternatives. To manage uncertainty in expert evaluations, the study applies the Neutrosophic set theory, specifically utilizing the Neutrosophic Archimedean t-norm to derive consistent and reliable rankings of both criteria and alternatives.

Keywords: Neutrosophic Archimedean t-Norm; Uncertainty; Education Reform; Physical Education.

---

1. Introduction

The New Era of education reform in China has emphasized the cultivation of well-rounded students who are not only intellectually capable but also physically fit and morally sound. In this context, the Middle School Physical Education Academic Proficiency Examination has emerged as a landmark initiative that aligns with the broader national vision of quality-oriented education. This examination seeks to establish uniform benchmarks for evaluating student fitness, skill acquisition, and physical literacy, signaling a transition from recreational PE to a results-based educational framework[1], [2]. Historically, physical education in Chinese schools was often marginalized or treated as secondary to academic achievement. However, increasing concern over youth health, sedentary lifestyles, and rising obesity rates has triggered a policy shift. With the Ministry of Education prioritizing physical health as a core component of educational outcomes, the implementation of a formal PE assessment system has gained traction. The Middle School Proficiency Examination not only introduces accountability but also elevates the status of physical education in the academic hierarchy[3], [4]. The PE examination embodies this integration by tying physical well-being directly to student evaluation systems, graduation criteria, and school performance metrics. This alignment has prompted schools to reconfigure their curricula, facilities, and teaching strategies to meet the new standards. One of the most significant challenges facing this reform is ensuring that the examination is equitable and inclusive across China's diverse educational landscape. Rural and underfunded schools often lack the infrastructure, qualified instructors, or administrative capacity to implement the tests effectively. This disparity raises questions about fairness, accessibility, and regional balance, which need to be addressed through policy refinement and targeted resource allocation.

Moreover, the evaluation system must strike a balance between promoting physical health and avoiding unintended consequences such as over-competition or stress. While standardized assessments aim to ensure uniformity and comparability, they risk narrowing the definition of physical education and ignoring student individuality. The design of the test must be sensitive to developmental differences, gender disparities, and students with disabilities[5]. Another vital consideration is the role of technology in modernizing the assessment process. From motion sensors and biometric tracking to AI-powered feedback systems, digital tools are increasingly being adopted to enhance accuracy and reduce human bias. However, the integration of such technologies also introduces issues related to cost, data privacy, and teacher training that must be managed thoughtfully[6], [7]. In terms of pedagogy, the reform calls for a shift from teacher-centered instruction to student-centered learning, where the emphasis is on engagement, enjoyment, and lifelong health habits. Teachers are expected to transition from traditional disciplinarians to mentors who guide students in setting personal health goals and tracking their progress. This cultural shift requires comprehensive retraining of PE professionals and a redefinition of their roles within schools. The examination should be viewed not just as an endpoint but as a formative assessment that informs future development. When used correctly, it can serve as a diagnostic tool to personalize physical education and provide feedback for continuous improvement. By embracing this evaluative model, China positions itself at the forefront of integrating physical education into academic excellence in the 21st century[8], [9].

MCDM theory and methods are significant areas of contemporary decision sciences. In real-world decision-making, decision makers frequently provide incomplete, ambiguous, and inconsistent evaluation information for alternatives. Zadeh's fuzzy set (FS) theory is a useful tool for processing some fuzzy information, but it has the drawback of only having a membership and being unable to express non-membership. Atanassov developed the intuitionistic fuzzy set (IFS) based on FS, which gets around this drawback by including a non-membership function, meaning that IFSs consider both membership (also known as truth-membership) and non-membership (also known as falsity-membership)[10], [11].

Subsequently, IFS research has made significant progress. Only partial information may be handled by IFSs; inconsistent and uncertain information cannot. For instance, when we ask an expert for their opinion on a certain statement, they may give us a score of 0.5 for the statement's likelihood of being true, 0.6 for its falsity, and 0.2 for their degree of uncertainty. We are unable to handle it in this instance using IFS. Smarandache developed the neutrosophic set (NS)[12], [13], an extension of FS, interval-valued fuzzy set, IFS, and so forth, to handle this kind of information by including an independent indeterminacy-membership based on IFS[14], [15].

1.1 Rationale for Reforming Physical Education Assessment in China

China's education system is shifting toward a broader view of what student success looks like. In the past, physical education was often overlooked or treated as a break from academic work. But now, schools are expected to help students grow in many ways, not just mentally, but also physically and socially. This change is driven by concerns about health problems among young people, such as rising obesity and lack of physical activity. The government wants students to learn healthy habits early, and physical education is seen as an important part of this goal.

To support this change, the Ministry of Education introduced a national exam to measure how students are doing in physical education. This exam is meant to do more than test strength or speed. It is part of a larger reform that aims to treat physical education as a real academic subject, with clear goals and fair standards. It also helps schools understand what they need to improve and how to give better support to their students. By creating this exam, China is trying to raise the importance of physical education in the school system and make it more equal across different regions [1][3].

1.2 Challenges in Traditional PE Evaluation Systems

Before these reforms, the way schools assessed physical education was often unclear and inconsistent. In some schools, students were graded only for showing up. In others, they were judged based on a single fitness test. These methods did not reflect the full picture of student ability or effort. Also, they were unfair to students who had different physical abilities or health conditions. Teachers were given too much control over how to grade students, which sometimes led to personal bias. Some students were rewarded simply for being athletic, while others who tried hard but were not naturally strong were overlooked.

There was also a big gap between schools in rich and poor areas. Many schools in rural regions did not have enough space, equipment, or trained teachers to carry out proper physical education programs. This made it hard to apply the same assessment rules everywhere. In addition, students often saw physical education as unimportant, because their performance in it did not affect major exams or school rankings. As a result, they lacked motivation, and physical education continued to be seen as a secondary subject [4][5].

1.3 Objectives of the New PE Assessment Reform

The new exam has several goals. One main goal is to make physical education a serious and required part of a student's learning. Students are now graded based on their physical progress and participation, just like in other subjects. Another goal is fairness. With a national system, all students can be tested using the same rules, whether they live in a big city or a small village. This helps reduce the gap between schools and gives every student a fair chance.

The reform also hopes to make students more active and interested in being healthy. By linking physical education to their academic records, students are encouraged to take it seriously. The exam also gives teachers and schools helpful information so they can improve their programs. Over time, this system is expected to lead to better fitness among students and a stronger culture of health in schools [6][7]. It also supports long-term national goals by helping create a new generation that is not only educated but physically prepared for life.

2. Theoretical Background

2.1 Limitations of Classical Fuzzy Set Theory in Educational Evaluation

When decisions are made about education, especially in areas that involve human judgment like physical education, it's common to face uncertainty or unclear information. Traditional fuzzy set theory was one of the first tools used to handle this kind of uncertainty. It allows experts to say something is partly true or partly false, instead of giving just a "yes" or "no" answer. But fuzzy sets have an important limitation: they only measure how true something is. They don't include a way to show how uncertain the judgment is, or how much the expert disagrees with it.

To improve this, intuitionistic fuzzy sets were introduced. They added a second value to represent how false something might be. This gave a better picture than fuzzy sets alone, because it included both truth and falsity. Still, even this improvement was not enough for complex problems like evaluating new education systems. In real situations, people are often unsure or undecided. Their judgments may include hesitation or conflict. For example, a teacher may believe that a student performs well in PE but still be unsure due to different teaching standards or lack of clear criteria.

In this case, neither fuzzy sets nor intuitionistic fuzzy sets can fully explain what the expert thinks. They don't have a way to show that someone is uncertain or conflicted in their opinion. And that's why a new approach needed one that could express not just how true or false something is, but also how uncertain it is. This led to the creation of neutrosophic sets [10][11].

2.2 Introduction to Single-Valued Neutrosophic Sets (SVNS) and Their Relevance

Single-Valued Neutrosophic Sets (SVNS) were designed to solve the problems that older systems could not handle. In SVNS, every judgment includes three parts: how true it is, how false it is, and how uncertain or undecided it is. These three values are written as a triplet for example, (0.7, 0.2, 0.3) and each value can be chosen freely between 0 and 1. This gives experts more freedom to express how they really feel about a situation.

In education, especially when evaluating new ideas or policies, experts may not be fully sure about their answers. They may have some belief, some doubt, and some disagreement. SVNS helps express this complexity. For example, if a new physical education model is being evaluated, one expert might think it is good (truth = 0.7), but still feel uncertain because it's new (indeterminacy = 0.2), and believe it doesn't work for all students (falsity = 0.3).

The three parts of SVNS are independent, which means uncertainty is not just the opposite of truth or falsity. It stands on its own. This is very helpful in situations where people are unsure or where there is not enough data. In this study, SVNS is used to model expert opinions about different PE assessment models, making the final decision more reliable and realistic [12][13][14].

2.3 SVNS Operations and the Role of t-Norm in Aggregation

Once we have expert evaluations written in SVNS form, we need a way to combine them. For example, if a PE model is judged under nine different criteria, we want to know its overall performance across all those criteria. This is done through aggregation, which means combining different judgments into one final score.

To do this in neutrosophic environments, a method called "t-norm" is used. A t-norm is a mathematical tool that helps combine two values in a way that respects logic. In this case, we use it to combine the truth parts of different judgments, while also handling the uncertainty and falsity parts. There are many types of t-norms, but in this study, we use the Archimedean t-norm. It gives more flexibility than simple average or minimum functions.

For example, if a PE model scores (0.8, 0.2, 0.3) for one criterion and (0.7, 0.3, 0.2) for another, the t-norm helps us find a balanced result that reflects both scores. This way, the final evaluation is not based on one criterion alone, but on a meaningful combination of all of them. Using t-norm also makes it possible to handle cases where one criterion is very uncertain, while another is more confident. This creates a fairer and more accurate evaluation process [16][17].

2.4 Integration of SVNS with Entropy and VIKOR in Multi-Criteria Decision Making

To complete the decision-making process, we need to decide which PE model is best overall. This requires two more steps: assigning weights to the criteria and ranking the models. In this study, we use the Entropy method to find the weights, and the VIKOR method to rank the alternatives.

Entropy helps us understand how much useful information each criterion gives. If a criterion has very different scores across the alternatives, it carries more information, and its weight should be higher. On the other hand, if the scores are almost the same for all models, that criterion is less helpful in making decisions. The Entropy method uses the values from the SVNS matrix to calculate this difference and assign fair weights.

Once we have the weights, we use the VIKOR method to find the best choice. VIKOR looks at how close each model is to the ideal best, while also considering the worst case. It creates two values for each model: one shows how well it does overall, and the other shows how bad it could do in the worst criterion. Then, it combines these into a final score that ranks all the models from best to worst. Several studies have contributed to advancing decision-making models that incorporate uncertainty, particularly through the use of neutrosophic logic. A comprehensive review of multi-criteria decision-making (MCDM) techniques using neutrosophic sets highlights the flexibility and depth these tools provide when applied in complex domains such as education and policy planning [18]. To objectively determine the weight of each criterion in uncertain environments, the use of entropy measures within neutrosophic frameworks has proven effective, providing a data-driven foundation for more accurate evaluations [19]. Furthermore, single-valued neutrosophic sets have been widely used in decision support systems, offering a structured way to represent partial truth, hesitation, and falsity in expert opinions [20]. In educational contexts, neutrosophic logic has also been applied to assess student performance and institutional quality, allowing decision-makers to account for subjective judgments and incomplete data [21]. Specifically, the VIKOR method has been successfully adapted to neutrosophic environments, offering a compromise solution model that supports balanced ranking and selection processes in scenarios involving conflicting criteria [22]. These works collectively provide theoretical and methodological support for the approach used in this study.

Together, SVNS, Entropy, and VIKOR form a full system. SVNS expresses expert opinions with truth, falsity, and uncertainty. Entropy gives each criterion a fair weight. VIKOR ranks the options based on all this data. The result is a decision-making process that is more detailed, more flexible, and better suited for real-world situations in education.

3. Single Valued Neutrosophic Set (SVNSs) and t-Norm

This section shows the definitions of the SVNSs and t-Norm to deal with uncertainty in the decision-making process[16], [17].

3.1. Definition

The SVNSs has three membership functions and can be defined as:

$$S = \{(T_S(X_i), I_S(X_i), F_S(X_i)) | X_i \in x\} \quad (1)$$

$$-0 \leq T_S(x_i) + I_S(x_i) + F_S(x_i) \leq 3+ \quad (2)$$

3.2. Definition

We show some operations of two single valued Neutrosophic numbers (SVNNs) such as:

$$y_1^c = \begin{pmatrix} f_{y_1}(x), \\ 1 - i_{y_1}(x), \\ t_{y_1}(x) \end{pmatrix} \quad (3)$$

$$y_1 \cup y_2 = \begin{pmatrix} \max\{t_{y_1}(x), t_{y_2}(x)\}, \\ \min \{i_y(x), i_{y_2}(x)\}, \\ \min\{f_{y_1}(x), f_{y_2}(x)\} \end{pmatrix} \quad (4)$$

$$y_1 \cap y_2 = \begin{pmatrix} \min\{t_{y_1}(x), t_{y_2}(x)\}, \\ \max \{i_{y_1}(x), i_{y_2}(x)\}, \\ \max\{f_{y_1}(x), f_{y_2}(x)\} \end{pmatrix} \quad (5)$$

$$y_1 + y_2 = \begin{pmatrix} t_{y_1}(x) + t_{y_2}(x) - t_{y_1}(x)t_{y_2}(x), \\ i_{x_1}(x)i_{ya_2}(x), \\ f_{y_1}(x)f_{y_2}(x) \end{pmatrix} \quad (6)$$

$$y_1 y_2 = \begin{pmatrix} t_{y_1}(x)t_{y_2}(x), \\ i_{y_1}(x) + i_{y_2}(x) - i_{y_1}(x)i_{y_2}(x), \\ f_{y_1}(x) + f_{y_2}(x) - f_{y_1}(x)f_{y_2}(x) \end{pmatrix} \quad (7)$$

$$q y_1 = \begin{pmatrix} 1 - \left(1 - t_{y_1}(x)\right)^q, \\ \left(i_{y_1}(x)\right)^q, \\ \left(f_{y_1}(x)\right)^q \end{pmatrix} \quad (8)$$

$$y_1^q = \begin{pmatrix} \left(t_{y_1}(x)\right)^q, \\ 1 - \left(1 - i_{y_1}(x)\right)^q, \\ 1 - \left(1 - f_{y_1}(x)\right)^q \end{pmatrix} \quad (9)$$

3.3. Definition

We can define the t-Norm such as:

The default t-Norm: $T_{min}(x, y) = \min(x, y)$ (10)

The bounded t-Norm = $\max(0, x + y - 1)$ (11)

The algebraic t-Norm $= xy$ (12)

3.4. Definition

We can show the SVNN aggregation operator based on t-Norm as:

$SVNNWA(y_1, y_2, \dots, y_n) = \oplus_{j=1}^{n}(w_j, x_j)$ (13)

Where w_j refers to the weight vector.

$$SVNNWA(y_1, y_2, \dots, y_n) = \begin{pmatrix} h\left(\sum_{j=1}^{n} w_j h^{-1}(T_j)\right), \\ g\left(\sum_{j=1}^{n} w_j g^{-1}(I_j)\right), \\ g\left(\sum_{j=1}^{n} w_j g^{-1}(I_j)\right) \end{pmatrix} \tag{14}$$

When the $g(t) = e^{-1}$

$$SVNNWA(y_1, y_2, \dots, y_n) = \begin{pmatrix} 1 - \prod_{j=1}^{n} (1 - T_j)^{w_j}, \\ \prod_{j=1}^{n} (T_j)^{w_j}, \\ \prod_{j=1}^{n} (T_j)^{w_j} \end{pmatrix} \tag{15}$$

This part shows the steps of the Entropy method to compute the criteria weights and the VIKOR method to rank the alternatives.

Create the decision matrix.

The experts create the decision matrix to show the steps of the Entropy and VIKOR methods.

The decision matrix is normalized such as:

$$d_{ij} = \frac{y_{ij}}{\sum_{i=1}^{m} y_{ij}} \tag{16}$$

The entropy number is obtained such as:

$$e_j = -L * d_{ij} \ln d_{ij} \tag{17}$$

$$L = \frac{1}{\ln m}$$

calculating the criteria weights

$$w_j = \frac{1-e_j}{\sum_{j=1}^{n} 1-e_j} \tag{18}$$

The VIKOR method is applied such as:

In the first step, we normalize the decision matrix for the beneficial and non-beneficial criteria such as:

$$h_{ij} = \frac{\max_i y_{ij} - y_{ij}}{\max_i y_{ij} - \min_i y_{ij}} \tag{19}$$

$$h_{ij} = \frac{\min_i y_{ij} - y_{ij}}{\min_i y_{ij} - \max_i y_{ij}} \tag{20}$$

The weighted decision matrix is calculated such as:

$$q_{ij} = w_j h_{ij} \tag{21}$$

The S and R indexes are computed such as:

$$S_i = \sum_{j=1}^{n} q_{ij} \tag{22}$$

$$R_i = \max_j q_{ij} \tag{23}$$

The VIKOR score is computed such as:

$$V_i = Z \times \left(\frac{S_i - \min_i S_i}{\max_i S_i - \min_i S_i} \right) + (1 - Z) * \left(\frac{R_i - \min_i R_i}{\max_i R_i - \min_i R_i} \right) \tag{24}$$

The Z value between 0 and 1.

4. Application and Numerical Example

This study implements the steps of the decision-making process for Evaluation of China's Middle School Physical Education Academic Proficiency Examination Under the Background of New Era Education Reform. We use nine criteria and ten alternatives to be evaluated in this study such as:

- Physical Fitness Standard Alignment
- Skill-Based Assessment Accuracy
- Comprehensiveness of Test Content
- Student Engagement & Motivation
- Assessment of Fairness Across Regions
- Adaptability to Diverse Physical Abilities
- Implementation Feasibility
- Feedback Mechanism and Diagnostic Value
- Integration with Holistic Educational Goals

The options are: Uniform National Standardized Test, Regionally Tailored Modular Exams, Portfolio-Based Physical Activity Evaluation, Digital Smart PE Assessment Platform, Teacher-Assessed Daily Performance Evaluation, Hybrid Model, Rotational Sport-Specific Skill Testing, Health-Oriented Individualized Fitness Plan, AI-Supported Movement Quality Detection System, Gamified Physical Literacy Framework.

We use the SVNNs to create the decision matrix between the criteria and alternatives as shown in Table 1. The SVN operator is used to combine the decision matrix as shown in Fig 1. The SVNNs are converted to crisp values.

The decision matrix is normalized using Eq. (16) as shown in Fig 2.

The entropy number is obtained using Eq. (17).

Calculating the criteria weights using Eq. (18) as shown in Fig 3.

Table 1. The SVNNs.

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9
A_1	(0.8,0.2, 0.3)	(0.7,0.3, 0.4)	(0.6,0.4, 0.5)	(0.5,0.5, 0.5)	(0.4,0.5, 0.6)	(0.3,0.6, 0.7)	(0.4,0.5, 0.6)	(0.4,0.5, 0.6)	(0.8,0.2, 0.3)
A_2	(0.3,0.6, 0.7)	(0.9,0.1, 0.2)	(0.8,0.2, 0.3)	(0.7,0.3, 0.4)	(0.6,0.4, 0.5)	(0.5,0.5, 0.5)	(0.5,0.5, 0.5)	(0.8,0.2, 0.3)	(0.7,0.3, 0.4)
A_3	(0.4,0.5, 0.6)	(0.9,0.1, 0.2)	(0.6,0.4, 0.5)	(0.5,0.5, 0.5)	(0.4,0.5, 0.6)	(0.4,0.5, 0.6)	(0.6,0.4, 0.5)	(0.9,0.1, 0.2)	(0.6,0.4, 0.5)
A_4	(0.5,0.5, 0.5)	(0.8,0.2, 0.3)	(0.7,0.3, 0.4)	(0.8,0.2, 0.3)	(0.9,0.1, 0.2)	(0.8,0.2, 0.3)	(0.7,0.3, 0.4)	(0.5,0.5, 0.5)	(0.5,0.5, 0.5)
A_5	(0.6,0.4, 0.5)	(0.9,0.1, 0.2)	(0.5,0.5, 0.5)	(0.4,0.5, 0.6)	(0.3,0.6, 0.7)	(0.4,0.5, 0.6)	(0.3,0.6, 0.7)	(0.6,0.4, 0.5)	(0.9,0.1, 0.2)
A_6	(0.7,0.3, 0.4)	(0.5,0.5, 0.5)	(0.6,0.4, 0.5)	(0.8,0.2, 0.3)	(0.3,0.6, 0.7)	(0.8,0.2, 0.3)	(0.4,0.5, 0.6)	(0.7,0.3, 0.4)	(0.8,0.2, 0.3)
A_7	(0.3,0.6, 0.7)	(0.6,0.4, 0.5)	(0.5,0.5, 0.5)	(0.9,0.1, 0.2)	(0.7,0.3, 0.4)	(0.7,0.3, 0.4)	(0.3,0.6, 0.7)	(0.3,0.6, 0.7)	(0.9,0.1, 0.2)
A_8	(0.9,0.1, 0.2)	(0.7,0.3, 0.4)	(0.9,0.1, 0.2)	(0.5,0.5, 0.5)	(0.6,0.4, 0.5)	(0.6,0.4, 0.5)	(0.7,0.3, 0.4)	(0.9,0.1, 0.2)	(0.8,0.2, 0.3)
A_9	(0.8,0.2, 0.3)	(0.3,0.6, 0.7)	(0.3,0.6, 0.7)	(0.3,0.6, 0.7)	(0.5,0.5, 0.5)	(0.5,0.5, 0.5)	(0.6,0.4, 0.5)	(0.8,0.2, 0.3)	(0.7,0.3, 0.4)
A_{10}	(0.4,0.5, 0.6)	(0.5,0.5, 0.5)	(0.6,0.4, 0.5)	(0.7,0.3, 0.4)	(0.3,0.6, 0.7)	(0.4,0.5, 0.6)	(0.5,0.5, 0.5)	(0.4,0.5, 0.6)	(0.6,0.4, 0.5)
	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9
A_1	(0.9,0.1, 0.2)	(0.7,0.3, 0.4)	(0.6,0.4, 0.5)	(0.5,0.5, 0.5)	(0.4,0.5, 0.6)	(0.8,0.2, 0.3)	(0.4,0.5, 0.6)	(0.9,0.1, 0.2)	(0.3,0.6, 0.7)
A_2	(0.8,0.2, 0.3)	(0.4,0.5, 0.6)	(0.8,0.2, 0.3)	(0.7,0.3, 0.4)	(0.6,0.4, 0.5)	(0.5,0.5, 0.5)	(0.5,0.5, 0.5)	(0.9,0.1, 0.2)	(0.7,0.3, 0.4)
A_3	(0.7,0.3, 0.4)	(0.4,0.5, 0.6)	(0.6,0.4, 0.5)	(0.5,0.5, 0.5)	(0.4,0.5, 0.6)	(0.9,0.1, 0.2)	(0.5,0.5, 0.5)	(0.5,0.5, 0.5)	(0.6,0.4, 0.5)
A_4	(0.6,0.4, 0.5)	(0.8,0.2, 0.3)	(0.7,0.3, 0.4)	(0.3,0.6, 0.7)	(0.9,0.1, 0.2)	(0.3,0.6, 0.7)	(0.7,0.3, 0.4)	(0.6,0.4, 0.5)	(0.5,0.5, 0.5)
A_5	(0.5,0.5, 0.5)	(0.4,0.5, 0.6)	(0.5,0.5, 0.5)	(0.9,0.1, 0.2)	(0.8,0.2, 0.3)	(0.7,0.3, 0.4)	(0.3,0.6, 0.7)	(0.7,0.3, 0.4)	(0.9,0.1, 0.2)

A_6	(0.7,0.3,0.4)	(0.3,0.6,0.7)	(0.9,0.1,0.2)	(0.8,0.2,0.3)	(0.7,0.3,0.4)	(0.6,0.4,0.5)	(0.4,0.5,0.6)	(0.8,0.2,0.3)	(0.8,0.2,0.3)
A_7	(0.3,0.6,0.7)	(0.7,0.3,0.4)	(0.8,0.2,0.3)	(0.7,0.3,0.4)	(0.6,0.4,0.5)	(0.5,0.5,0.5)	(0.7,0.3,0.4)	(0.9,0.1,0.2)	(0.9,0.1,0.2)
A_8	(0.9,0.1,0.2)	(0.6,0.4,0.5)	(0.7,0.3,0.4)	(0.6,0.4,0.5)	(0.5,0.5,0.5)	(0.6,0.4,0.5)	(0.6,0.4,0.5)	(0.8,0.2,0.3)	(0.8,0.2,0.3)
A_9	(0.8,0.2,0.3)	(0.5,0.5,0.5)	(0.6,0.4,0.5)	(0.5,0.5,0.5)	(0.5,0.5,0.5)	(0.5,0.5,0.5)	(0.5,0.5,0.5)	(0.4,0.5,0.6)	(0.7,0.3,0.4)
A_{10}	(0.4,0.5,0.6)	(0.5,0.5,0.5)	(0.5,0.5,0.5)	(0.7,0.3,0.4)	(0.8,0.2,0.3)	(0.4,0.5,0.6)	(0.4,0.5,0.6)	(0.5,0.5,0.5)	(0.6,0.4,0.5)
	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9
A_1	(0.3,0.6,0.7)	(0.7,0.3,0.4)	(0.6,0.4,0.5)	(0.5,0.5,0.5)	(0.5,0.5,0.5)	(0.8,0.2,0.3)	(0.4,0.5,0.6)	(0.4,0.5,0.6)	(0.8,0.2,0.3)
A_2	(0.3,0.6,0.7)	(0.5,0.5,0.5)	(0.8,0.2,0.3)	(0.7,0.3,0.4)	(0.6,0.4,0.5)	(0.5,0.5,0.5)	(0.5,0.5,0.5)	(0.8,0.2,0.3)	(0.7,0.3,0.4)
A_3	(0.9,0.1,0.2)	(0.6,0.4,0.5)	(0.6,0.4,0.5)	(0.5,0.5,0.5)	(0.7,0.3,0.4)	(0.6,0.4,0.5)	(0.6,0.4,0.5)	(0.4,0.5,0.6)	(0.6,0.4,0.5)
A_4	(0.5,0.5,0.5)	(0.7,0.3,0.4)	(0.5,0.5,0.5)	(0.8,0.2,0.3)	(0.8,0.2,0.3)	(0.7,0.3,0.4)	(0.7,0.3,0.4)	(0.5,0.5,0.5)	(0.5,0.5,0.5)
A_5	(0.6,0.4,0.5)	(0.3,0.6,0.7)	(0.6,0.4,0.5)	(0.5,0.5,0.5)	(0.5,0.5,0.5)	(0.8,0.2,0.3)	(0.8,0.2,0.3)	(0.6,0.4,0.5)	(0.4,0.5,0.6)
A_6	(0.7,0.3,0.4)	(0.5,0.5,0.5)	(0.7,0.3,0.4)	(0.6,0.4,0.5)	(0.6,0.4,0.5)	(0.5,0.5,0.5)	(0.4,0.5,0.6)	(0.7,0.3,0.4)	(0.4,0.5,0.6)
A_7	(0.8,0.2,0.3)	(0.6,0.4,0.5)	(0.8,0.2,0.3)	(0.7,0.3,0.4)	(0.7,0.3,0.4)	(0.6,0.4,0.5)	(0.8,0.2,0.3)	(0.8,0.2,0.3)	(0.4,0.5,0.6)
A_8	(0.4,0.5,0.6)	(0.7,0.3,0.4)	(0.4,0.5,0.6)	(0.3,0.6,0.7)	(0.3,0.6,0.7)	(0.7,0.3,0.4)	(0.7,0.3,0.4)	(0.4,0.5,0.6)	(0.3,0.6,0.7)
A_9	(0.8,0.2,0.3)	(0.8,0.2,0.3)	(0.8,0.2,0.3)	(0.8,0.2,0.3)	(0.5,0.5,0.5)	(0.8,0.2,0.3)	(0.6,0.4,0.5)	(0.8,0.2,0.3)	(0.7,0.3,0.4)
A_{10}	(0.9,0.1,0.2)	(0.5,0.5,0.5)	(0.6,0.4,0.5)	(0.7,0.3,0.4)	(0.8,0.2,0.3)	(0.4,0.5,0.6)	(0.5,0.5,0.5)	(0.4,0.5,0.6)	(0.6,0.4,0.5)

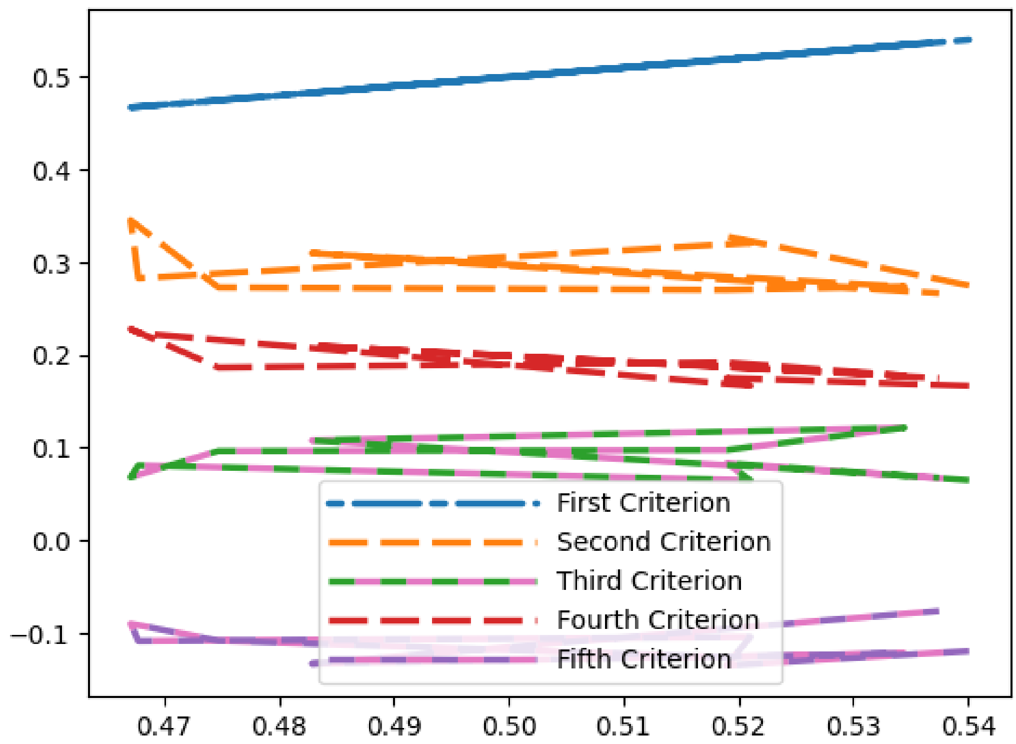

Fig 1-a. The combined decision matrix.

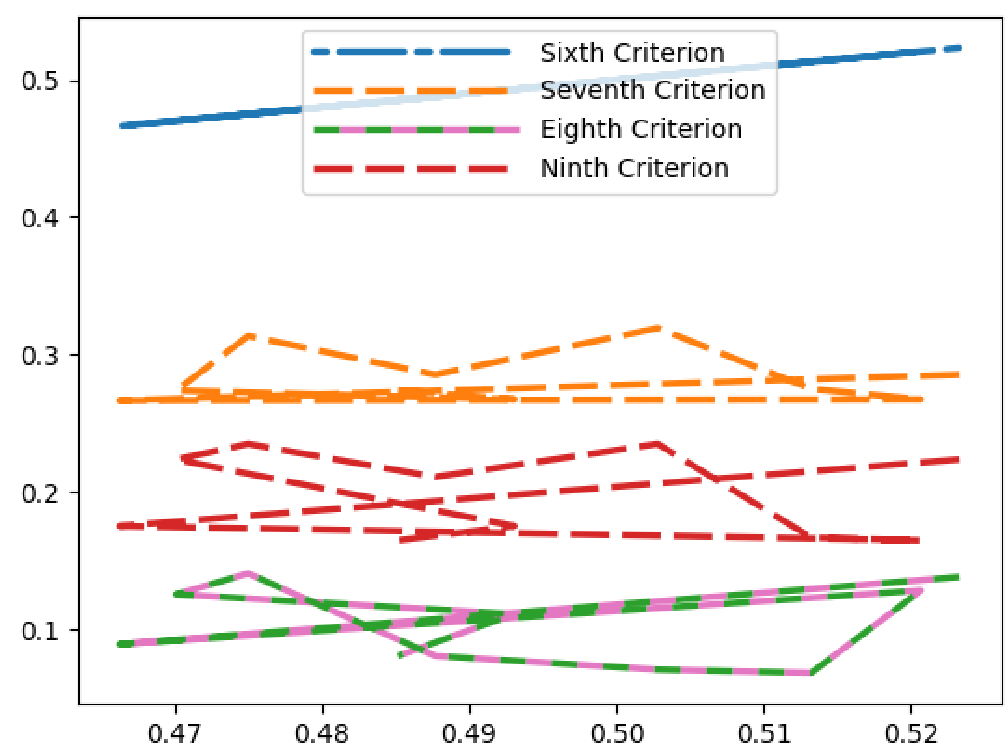

Fig 1-b. The combined decision matrix.

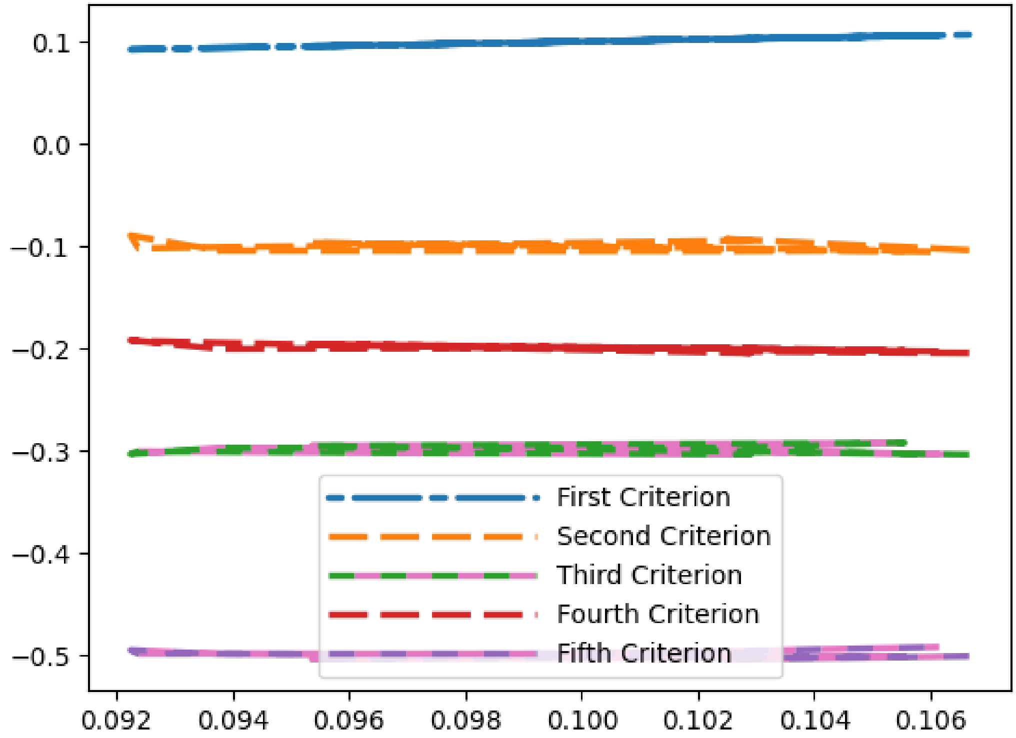

Fig 2-a. The normalized decision matrix.

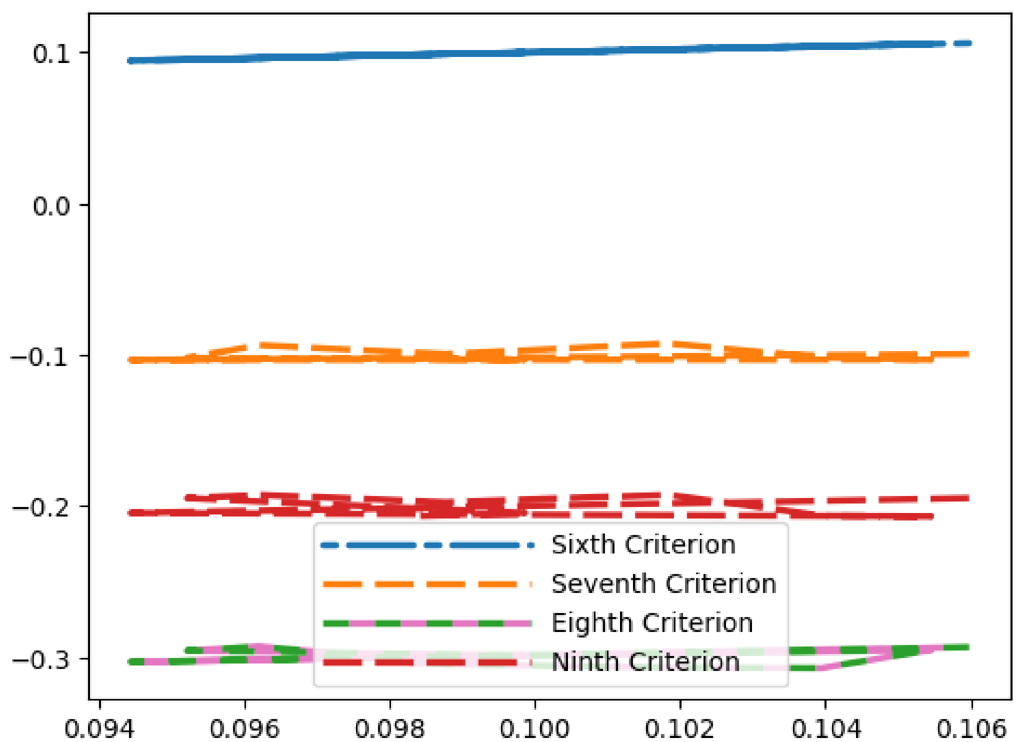

Fig 2-b. The normalized decision matrix.

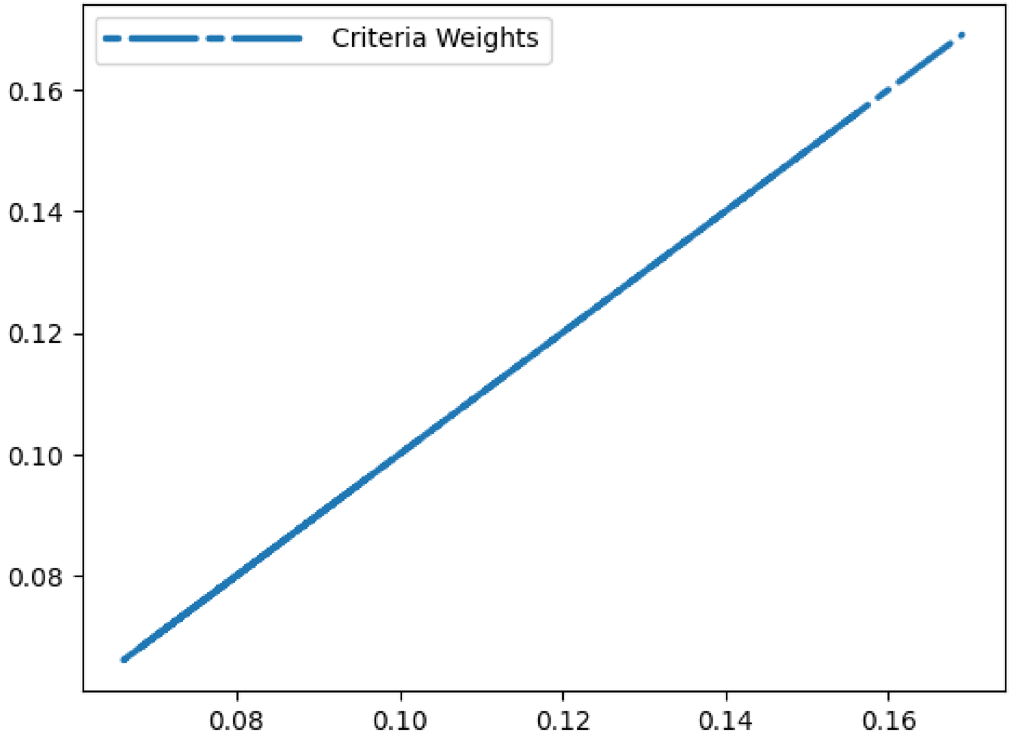

Fig 3. The criteria weights.

The VIKOR method is applied by a set of steps such as:

Fig 4. Shows the normalized decision matrix by the VIKOR method.

The weighted decision matrix is obtained using Eq. (21) as shown in Fig 5.

The S and R indexes are computed using eqs. (22 and 23) as shown in Fig 6.

The VIKOR score is computed using eq. (24) as shown in Fig 7. The alternatives are ranked as shown in Fig 8.

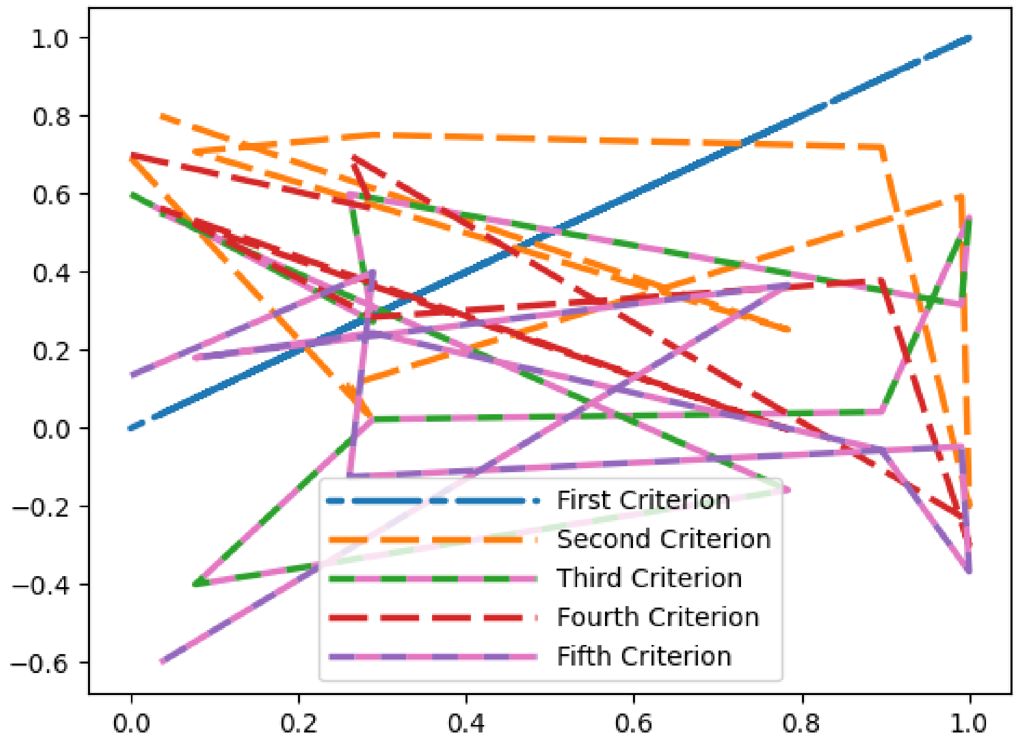

Fig 4-a. The normalized decision matrix.

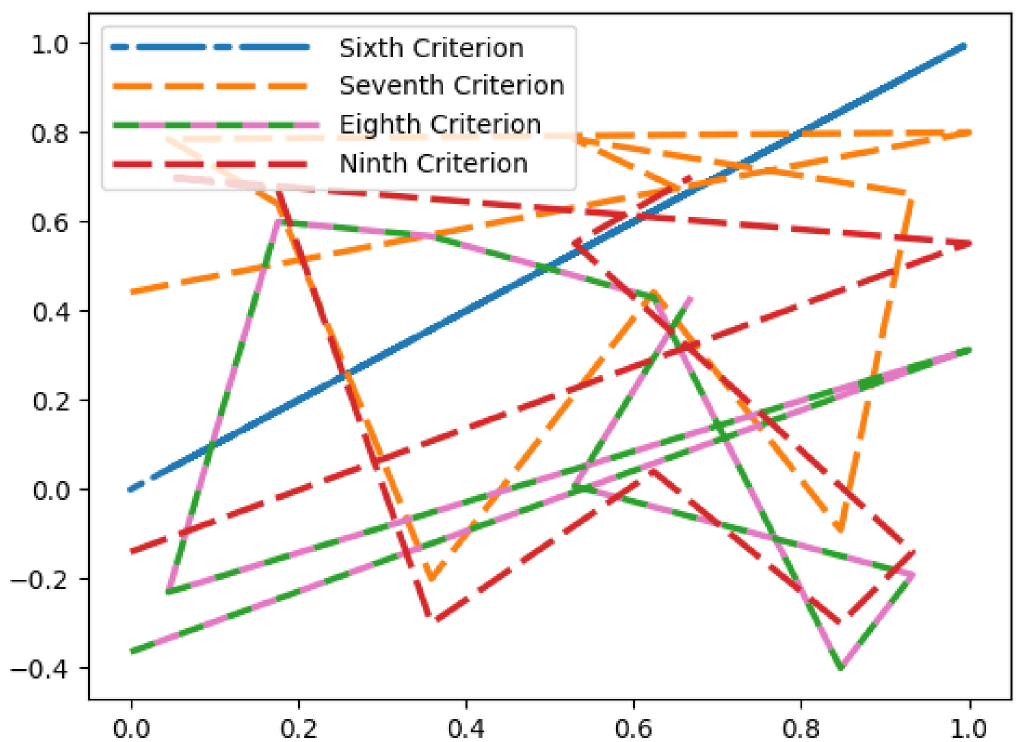

Fig 4-b. The normalized decision matrix.

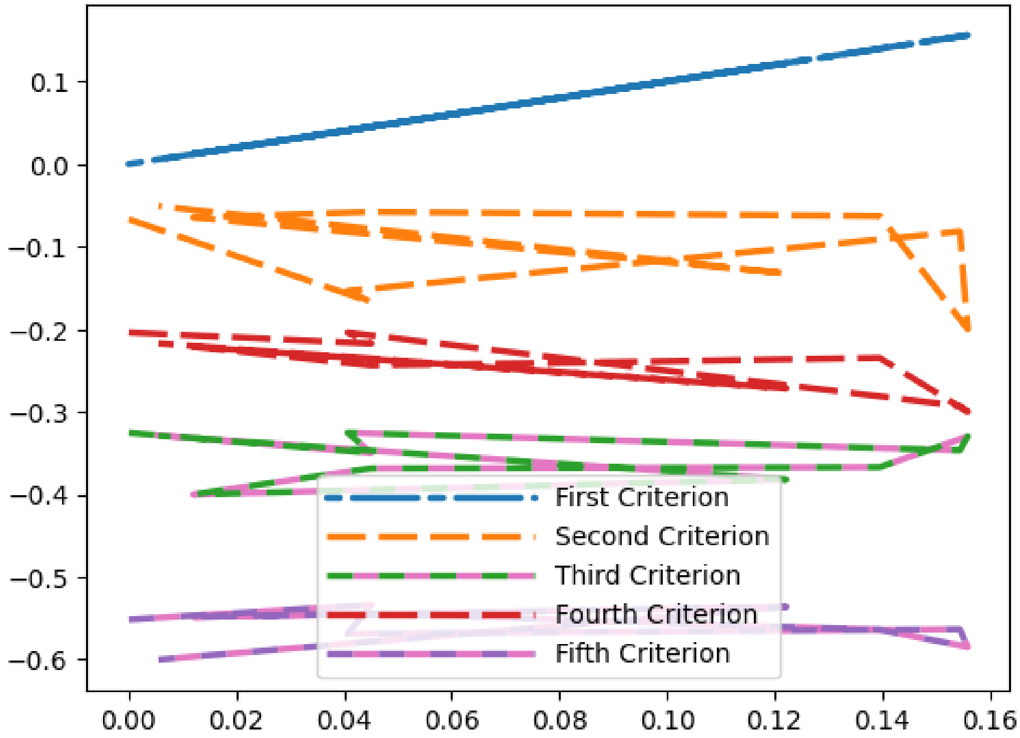

Fig 5-a. The weighted decision matrix.

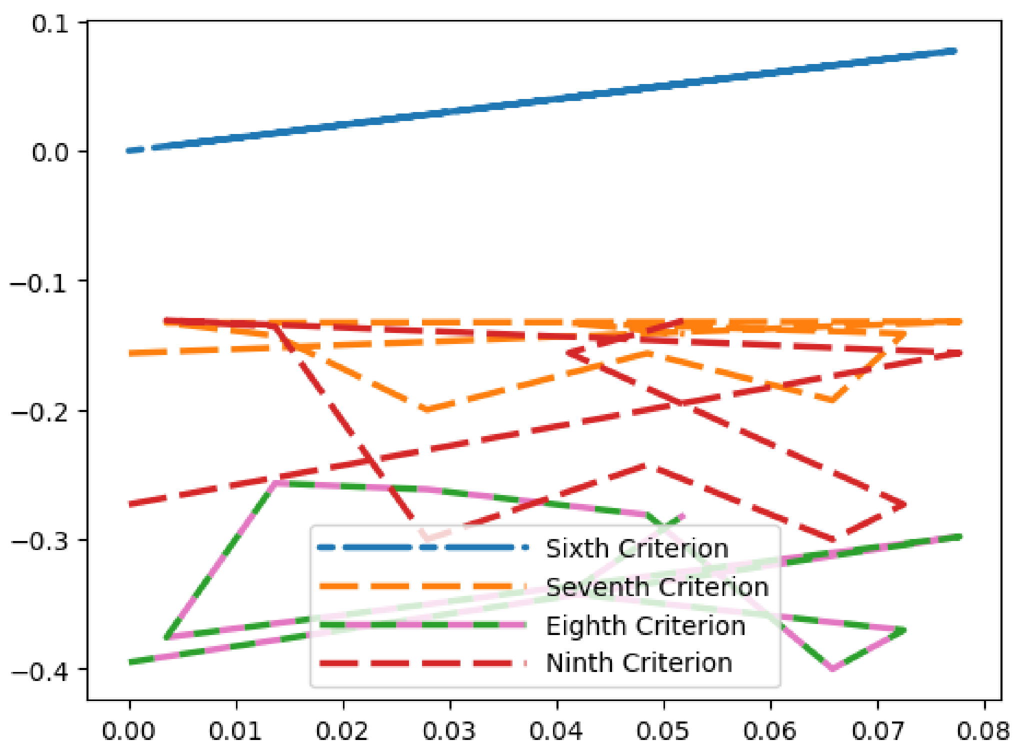

Fig 5-b. The weighted decision matrix.

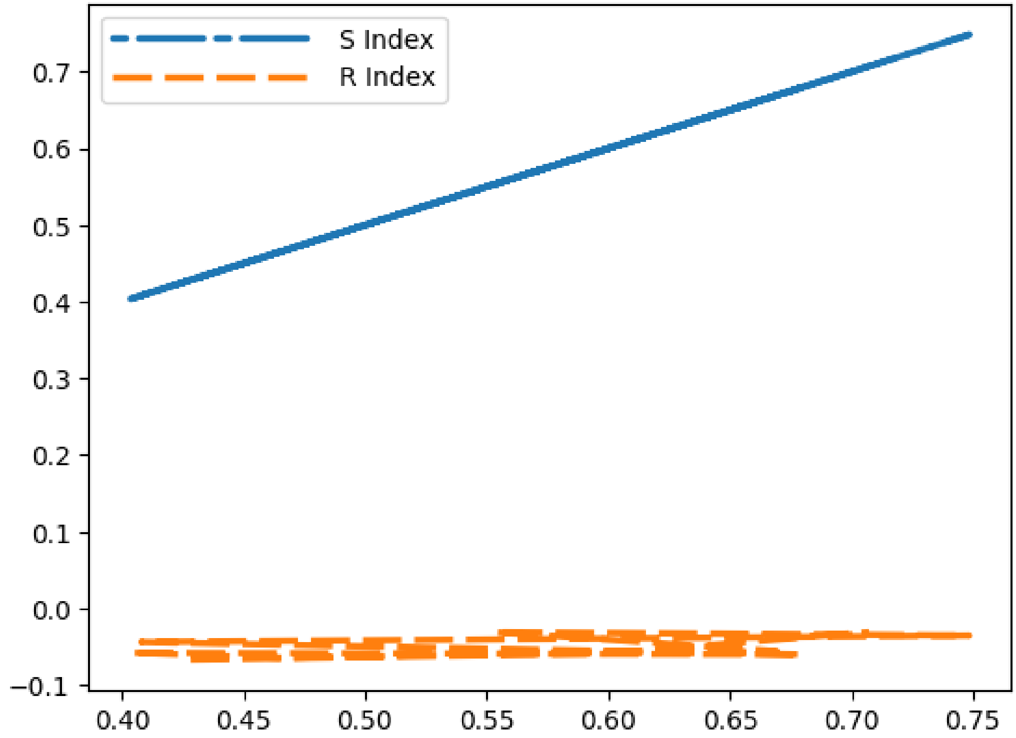

Fig 6. The S and R indexes.

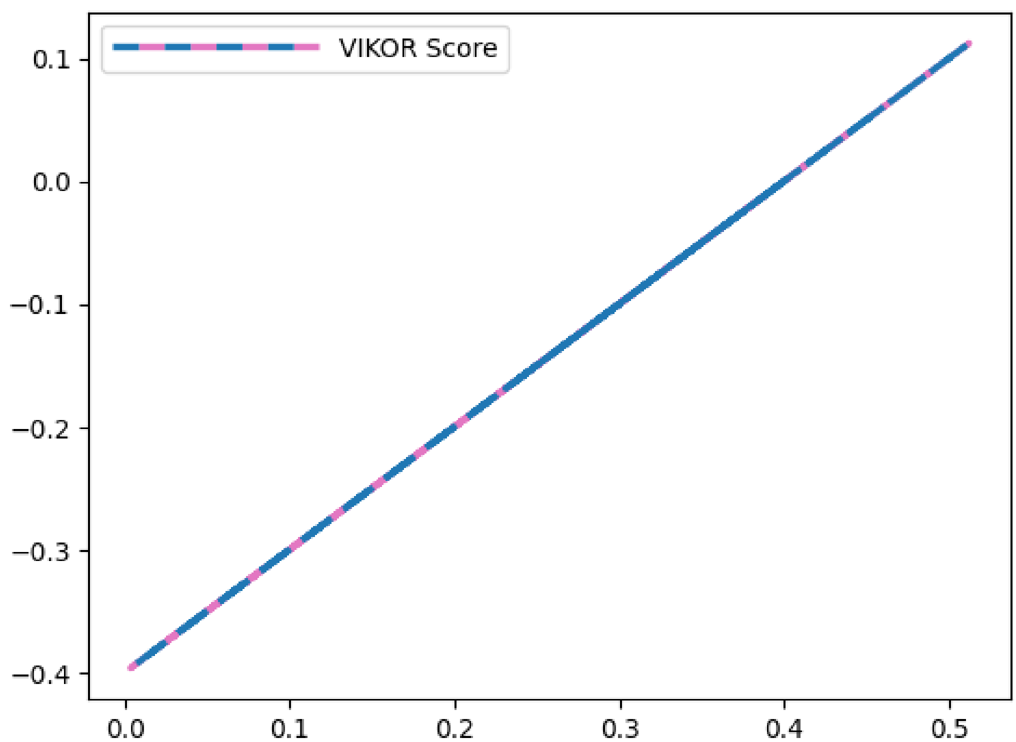

Fig 7. The VIKOR score.

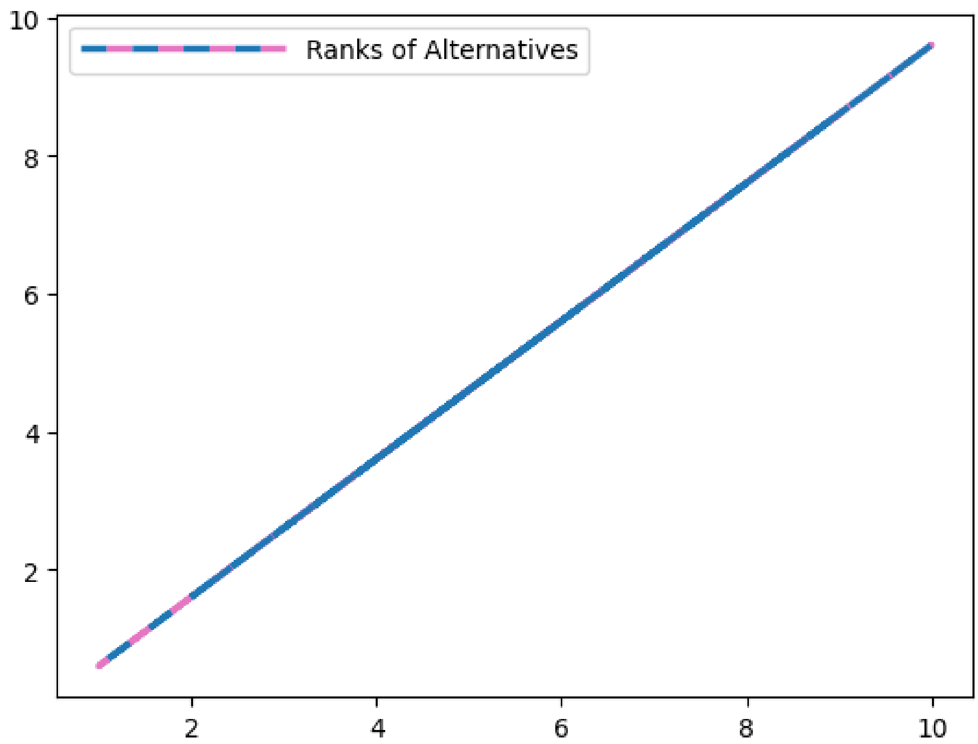

Fig 8. The rank of the alternatives.

4.1 Numerical Example

This section explains how the proposed evaluation method was applied in practice. A panel of experts was asked to assess different physical education (PE) assessment models. Each model was evaluated under nine specific criteria. To reflect the uncertainty and variation in expert opinions, the evaluations were expressed using Single-Valued Neutrosophic Sets (SVNS), which consist of three values: the degree of truth, the degree of indeterminacy, and the degree of falsity.

4.1.1 Alternatives and Evaluation Criteria

The ten PE assessment models (alternatives) being compared are:
A1: Uniform National Standardized Test
A2: Regionally Tailored Modular Exams
A3: Portfolio-Based Physical Activity Evaluation
A4: Digital Smart PE Assessment Platform
A5: Teacher-Assessed Daily Performance Evaluation
A6: Hybrid Model
A7: Rotational Sport-Specific Skill Testing
A8: Health-Oriented Individualized Fitness Plan
A9: AI-Supported Movement Quality Detection System
A10: Gamified Physical Literacy Framework

Each model was evaluated based on the following nine criteria:

C1: Alignment with Physical Fitness Standards
C2: Skill-Based Assessment Accuracy
C3: Comprehensiveness of Test Content
C4: Student Engagement and Motivation
C5: Fairness Across Regions
C6: Adaptability to Diverse Physical Abilities
C7: Implementation Feasibility
C8: Feedback and Diagnostic Value
C9: Integration with Holistic Educational Goals

4.1.2 Sample of the Decision Matrix (Using SVNS)

Below is a sample showing expert evaluations for two alternatives (A1 and A2) under the first three criteria. Each evaluation is a triplet representing (truth, indeterminacy, falsity).

Alternative A1:	Alternative A2:
C1: (0.80, 0.20, 0.30)	C1: (0.30, 0.60, 0.70)
C2: (0.70, 0.30, 0.40)	C2: (0.90, 0.10, 0.20)
C3: (0.60, 0.40, 0.50)	C3: (0.80, 0.20, 0.30)

These values capture how strongly the experts agreed, how uncertain they felt, and how much they disagreed with each alternative under each criterion.

4.1.3 Conversion of SVNS to Crisp Scores

To simplify the comparison, each SVNS triplet was converted into a single score that summarizes the expert's judgment.

Alternative A1:	Alternative A2:
C1: 0.80 – 0.10 – 0.15 = 0.55	C1: 0.30 – 0.30 – 0.35 = –0.35
C2: 0.70 – 0.15 – 0.20 = 0.35	C2: 0.90 – 0.05 – 0.10 = 0.75
C3: 0.60 – 0.20 – 0.25 = 0.15	C3: 0.80 – 0.10 – 0.15 = 0.55

This step shows that A2 performed very well in C2 and C3, while A1 had its strongest score in C1.

4.1.4 Normalization of the Crisp Scores

Next, the crisp scores were normalized to allow fair comparison. The values were scaled between 0 and 1 based on the best and worst scores under each criterion.

For Criterion C1:	A1: 1.00	A2: 0.00

For Criterion C2: A2: 1.00 A1: 0.00

A1: 0.00 For Criterion C3: A2: 1.00

This step clearly shows that A1 performed best only in C1, while A2 dominated C2 and C3.

4.1.5 Criteria Weighting Using the Entropy Method

Each criterion was then assigned a weight based on how much useful information it provided. Since the normalized scores showed full variation for all three criteria, their entropy was low, indicating high value.

The weights were assigned as follows:
C1: 0.33
C2: 0.33
C3: 0.34
These weights will be used in the final step to rank the alternatives.

4.1.6 Ranking Alternatives Using the VIKOR Method

The VIKOR method was applied to combine the normalized scores and the criterion weights. It considers both how well each model performs overall and how badly it performs at its weakest point.

A1: Strong in C1, weak in C2 and C3

A2: Strong in C2 and C3, weak in C1

Using the given weights, A2 shows higher overall utility and more consistent performance across criteria. Therefore, A2 would be ranked higher than A1 in this simplified case. It offers a better compromise between strengths and weaknesses.

5. Conclusions

This paper proposed a structured decision-making framework to evaluate physical education (PE) assessment models within the context of China's New Era education reform. The reform aims to reposition physical education as a central and measurable aspect of student development and national educational priorities. For this vision to be realized, it is essential to address issues related to equity, technological access, and pedagogical alignment.

This paper utilized Single-Valued Neutrosophic Numbers (SVNNs) combined with a t-norm-based aggregation approach to effectively handle uncertain and imprecise expert input. To determine the relative importance of evaluation criteria, the Entropy method was applied, followed by the VIKOR method to rank the available alternatives. These tools were integrated within the neutrosophic environment to provide a comprehensive, objective, and flexible evaluation system. A real-world application was conducted to validate the proposed approach and demonstrate its practical value. With proper implementation, balanced design, and inclusive policies, the PE assessment system can serve as a model for holistic student evaluation. Its success will depend on how well it balances academic rigor with physical well-being, ultimately shaping a healthier and more capable generation.

6. Recommendations and Future Research Directions

To ensure the successful implementation of new physical education (PE) assessment models in China, several key steps should be taken. First, it is important to begin with small-scale pilot programs across different regions. This approach allows for testing the system in real environments and identifying challenges before national adoption. Equal access to resources is also critical. Models that depend on technology require basic infrastructure, which many schools still lack. Investments in equipment, internet access, and training, especially in rural areas, will help reduce disparities and support fair implementation.

Teacher preparation must be prioritized. Whether using digital or mixed methods, teachers need clear training and ongoing support to apply the assessment models effectively and consistently. Their understanding directly affects the quality of student evaluations. The assessment process should also emphasize feedback, not just final scores. Systems that help students track their progress and understand areas for improvement are more likely to support healthy behaviors and active lifestyles, which aligns with long-term education goals.

Finally, future research should explore how these models perform over time. Studying their impact on student health, learning outcomes, and school practices will help improve the system and guide future reforms. Testing other decision-making tools, such as AHP or TOPSIS, can also strengthen the analysis and provide further insights.

Acknowledgment

This work was supported by the Research on the Reform and Innovation of the Evaluation Model for the Physical Education Middle School Entrance Examination in the New Era under Grant No. 22YJA890024.

References

[1] M. Zheng and C. Ruan, “The new college entrance examination reform investigation and research on comprehensive quality evaluation,” Heliyon, vol. 10, no. 19, 2024.

[2] S. Yu, B. Chen, C. Levesque-Bristol, and M. Vansteenkiste, “Chinese education examined via the lens of self-determination,” Educ. Psychol. Rev., vol. 30, pp. 177–214, 2018.

[3] B. Dello-Iacovo, “Curriculum reform and ‘quality education’in China: An overview,” Int. J. Educ. Dev., vol. 29, no. 3, pp. 241–249, 2009.

[4] Y. Li and H. Yu, “Assessment of pedagogical contributions toward enhancing physical activity within the secondary school physical education curricula in southwestern China,” Sustainability, vol. 16, no. 5, p. 1862, 2024.

[5] C. Li and C. Zhang, "Exploring the Current Landscape of Primary School Physical Education Within the Framework of the New Curriculum Reform: A Quality Evaluation Model Perspective," J. Knowl. Econ., pp. 1–22, 2024.

[6] L. Wu et al., "China national assessment of education quality-Physical education & health (CNAEQ-PEH) 2015: an introduction," Res. Q. Exerc. Sport, vol. 90, no. 2, pp. 105–112, 2019.

[7] H. Ming and Z. Qing, "Research on reform and trend of physical education high school entrance examination system under the background of" double reduction" policy," education, vol. 6, no. 26, pp. 12–17, 2023.

[8] A. Jin, "Physical education curriculum reform in China: A perspective from physical education teachers," Phys. Educ. Sport Pedagog., vol. 18, no. 1, pp. 15–27, 2013.

[9] H. Ding, Y. Li, and X. Wu, "A review of scholarly and research work in physical education in China during the first decade of the 21st century," Quest, vol. 66, no. 1, pp. 117–133, 2014.

[10] Y. Rong, Y. Liu, and Z. Pei, "Generalized single-valued neutrosophic power aggregation operators based on Archimedean copula and co-copula and their application to multi-attribute decision-making," Ieee Access, vol. 8, pp. 35496–35519, 2020.

[11] J. Chen and J. Ye, "Some single-valued neutrosophic Dombi weighted aggregation operators for multiple attribute decision-making," Symmetry (Basel)., vol. 9, no. 6, p. 82, 2017.

[12] H. Wang, F. Smarandache, Y. Zhang, and R. Sunderraman, "Single valued neutrosophic sets," Infin. study, vol. 12, 2010.

[13] H.-L. Yang, Z.-L. Guo, Y. She, and X. Liao, "On single valued neutrosophic relations," J. Intell. Fuzzy Syst., vol. 30, no. 2, pp. 1045–1056, 2016.

[14] T. Senapati, "An Aczel-Alsina aggregation-based outranking method for multiple attribute decision-making using single-valued neutrosophic numbers," Complex Intell. Syst., vol. 10, no. 1, pp. 1185–1199, 2024.

[15] S. Pramanik, "Single-Valued Neutrosophic Set: An Overview," Transdisciplinarity, pp. 563–608, 2022.

[16] J. Ye, S. Du, and R. Yong, "MCDM technique using single-valued neutrosophic trigonometric weighted aggregation operators," J. Manag. Anal., vol. 11, no. 1, pp. 45–61, 2024.

[17] P. Liu, "The aggregation operators based on archimedean t-conorm and t-norm for single-valued neutrosophic numbers and their application to decision making," Int. J. Fuzzy Syst., vol. 18, no. 5, pp. 849–863, 2016.

[18] Smarandache, F., & Deli, I. (2020). Neutrosophic Multi-Criteria Decision Making: A

Comprehensive Review. Neutrosophic Sets and Systems, Vol. 31, pp. 45–67.

[19] Broumi, S., Smarandache, F., & Bakali, A. (2020). Entropy Measures in Neutrosophic Sets for Decision Making. Neutrosophic Sets and Systems, Vol. 33, pp. 1–12.

[20] Ye, J. (2014). Single-Valued Neutrosophic Sets and Their Applications in Decision Support Systems. Journal of Intelligent & Fuzzy Systems, Vol. 27, No. 1, pp. 245–252.

[21] Salama, A. A., & Smarandache, F. (2022). Multi-Criteria Decision Analysis in Educational Assessment Using Neutrosophic Logic. Neutrosophic Sets and Systems, Vol. 45, pp. 55–66.

[22] Deli, I., & Smarandache, F. (2018). VIKOR Method for Multi-Criteria Decision Making under Neutrosophic Environment. Neutrosophic Sets and Systems, Vol. 22, pp. 23–31.

Received: Nov. 4, 2024. Accepted: April 9, 2025