Article

A Multi-Project Evaluation of Engineering Students' Performance for Online PBL: Taking the Sustainable Decision Analysis Course as an Example

Fan Zhang *, Hongxia Yang * and Shengbin Li

Department of Architecture and Engineering, Yan'an University, Yan'an 716000, China
* Correspondence: fanzhang@yau.edu.cn (F.Z.); yadxyhx@yau.edu.cn (H.Y.)

Abstract: In order to meet the growing demand for engineering professionals who can incorporate sustainable solutions into their work, sustainability courses have been launched in online problem-based learning (PBL) environments through various real-life projects. Nonetheless, the conventional one-off grading approach may fail to capture the intricate variations in students' performance across different projects. To address this problem, a multi-project evaluation framework utilizing the probability exceedance method (PEM) is proposed, which can fuse linguistic evaluation data presented in probability distributions without the need to obtain weights of criteria. In the case study, a comprehensive evaluation of the performance of students majoring in engineering management is conducted within a study group over an online PBL course on sustainable decision analysis. The sensitivity analysis demonstrates that consistent scores can be achieved after assigning different values of fuzzy measures to each criterion. This study enables teachers to holistically evaluate students without being bound by rigid numerical standards or strict weighting schemes, thus allowing them to focus on other educational tasks while ensuring effective and reliable results. Moreover, it contributes to educational innovation by introducing a modern and comprehensive approach for engineering student assessment in online PBL, aligning with the evolving needs of educational sustainability in higher education.

Keywords: online PBL; multi-project evaluation; sustainability courses; probability exceedance method; fuzzy measures

1. Introduction

As the engineering profession is moving towards more environmentally friendly practices, sustainability has become an integral part of the process of engineering design, construction, and operation [1,2]. To actively promote sustainability as a core value in tertiary education, sustainability courses have been integrated as mandatory parts of engineering programs in many universities. These courses aim to equip students with knowledge about sustainable principles, technologies, and practices in engineering fields. To enhance the learning outcomes in these courses, students are encouraged to approach sustainability issues through real-world projects, utilizing problem-based learning (PBL) in an online and collaborative setting [3]. By sharing ideas and working together remotely, the online PBL approach facilitates students finding innovative solutions that consider environmental, social, and economic factors, as well as seeking effective collaboration with other disciplines [4].

Nevertheless, one significant challenge impeding the implementation of online PBL for sustainability courses lies in the evaluation of engineering students' performance, particularly when multiple projects are typically involved. Traditionally, students' final scores in many universities in China are comprised of two parts, namely seventy or eighty percent from the final written exams at the end of the courses and the rest from regular performance evaluations. However, this prevalent grading system may not be suitable

for the PBL courses, as students' comprehensive abilities and skills demonstrated through the collaborative completion of multiple projects are emphasized and can hardly be fully reflected by final written exams. To foster engineering professionals who can seamlessly integrate sustainability into their practical work, it is essential to capture and assess various aspects of students' abilities and accomplishments, including knowledge proficiency, problem-solving skills, teamwork capabilities, critical thinking skills, etc. [5–8]. When multiple factors with varying importance must be considered simultaneously, teachers usually face a dilemma in presenting reasonable assessment results for students in online PBL. In this context, it is imperative to develop a standardized, computerized, and objective multi-criteria performance evaluation approach to enhance the efficiency of the scoring process in online PBL over sustainability courses.

To date, while numerous studies have examined multi-criteria evaluation models or other quantitative strategies for evaluating students' academic performance, there is a lack of studies focused on PBL scenarios. Among the previous studies, ref. [9] identified relevant criteria and utilized a performance measurement philosophy to evaluate perceived student learning after a project-based assignment was applied as an instructional tool. Ref. [10] developed a program evaluation rubric via a qualitative rating system to assist teachers in identifying key learning goals, dimensions, and principles related to the socio-scientific issue (SSI)-based science, technology, engineering, arts, and mathematics (STEAM) approach for science education. Ref. [11] presented an empirical study examining the application of machine learning (ML) to assess students' engagement in PBL and provide process-oriented feedback on their collaboration. However, despite the fact that few studies have focused on the multi-criteria evaluation of student performance in PBL, a comprehensive assessment using other teaching strategies can be found. For example, ref. [12] applied the technique for order preference by similarity to ideal solution (TOPSIS) to measure and evaluate the academic performance of students enrolled in the first year of science and engineering in a university from 2016 to 2019. Ref. [13] developed an assessment model of students' flipped classroom learning via an analytic hierarchy process (AHP) towards the subject of chemistry in Iraqi schools. Ref. [14] examined multiple evaluation indexes and utilized an artificial intelligence-based evaluation method to evaluate students' classroom performance.

However, the majority of prior studies on student performance assessment have been conducted using one-off evaluations, which can be relatively limited and, in some cases, even inappropriate for online PBL. In fact, in PBL or online PBL courses, not all projects are created equal in terms of their focus, importance, or difficulty [15]. Some may require more in-depth understanding and critical thinking, while others may be more straightforward. Meanwhile, even if a student has a consistent performance in general, they may excel in some projects and struggle in others due to differences in project characteristics, students' personal interests, and motivation over different projects. Therefore, compared with conventional one-off evaluations at the end of a course, evaluating students' efforts based on individual projects allows for a more nuanced understanding of students' capabilities and areas for improvement, leading to fairer and more accurate evaluations.

When assessing students' performance in online PBL, linguistic variables such as "excellent" and "slightly poor" are typically preferred among teachers as they align better with the way humans express themselves [16,17]. Using these descriptive terms offers more flexibility and nuance in communicating feedback compared to using precise numerical values. However, the performance of a student can vary from one project to another, thus necessitating the use of different linguistic data to represent students' performance across different projects. As a result, the overall performance of a student can be presented in the form of probability distributions. For example, for an online PBL curriculum, the performance of a student against the criterion of teamwork skill based on a linguist term set $S = \{s_0, s_1, s_2, s_3, s_4, s_5, s_6\} = \{$"extremely poor", "poor", "slightly poor", "medium", "slightly good", "good", and "extremely good"$\}$ is evaluated as s_3, s_2, s_3, and s_3 over four projects, respectively. If each project is considered equally important, the overall evaluation

value would be $\{< s_3, 0.67 >, < s_4, 0.33 >\}$. Likewise, a similar form of collective linguistic data against other criteria can also be obtained. To derive students' overall assessment data against multiple criteria, linguistic evaluation values in the probability distributions need to be further aggregated. However, the traditional fusion methods or operators, such as the simple additive method (SAM) and the Choquet integral, fail to order or aggregate this type of data as they can only deal with values in real numbers [18,19]. To this end, ref. [20] proposed a probabilistic exceedance method (PEM) based on the Choquet integral to solve this problem.

Meanwhile, the PEM eliminates the need to determine the weight distribution of criteria by assigning values to fuzzy measures of criteria to represent their relative importance. In regards the common multi-criteria evaluation methods, such as the methods of VlseKriterijumska Optimizacija I Kompromisno Resenje (VIKOR), decision-making trial and evaluation laboratory (DEMATEL), TOPSIS and multi-objective optimization on the basis of a ratio analysis plus the full multiplicative form (MULTIMOORA), weights of criteria have to be calculated before the final evaluation can be derived [21]. However, for the PEM-based algorithm, fuzzy measures of criteria representing their relative significance are determined by teachers or educators and can be directly applied to derive the evaluation results. Moreover, compared with the classic fuzzy measurement, fuzzy measures take the importance of the combination of decision attributes into account rather than only considering the importance of the attributes themselves [22]. For example, teachers may consider that the criteria of online engagement and sustained learning in student performance assessment are both important, while the combined importance of the two could be greater or less than their sum. Under this circumstance, synergy or antagonism among the factors can probably exist when the importance of the combination of decision attributes is measured [22]. In fact, fuzzy measures have been widely used in comprehensive education evaluation [23,24], risk assessment [25–27], resource allocation [28–30], and other scenarios to provide quantitative analysis support for solving complex problems.

In this study, to evaluate students majoring in engineering management for the online PBL curriculum on sustainable decision analysis, a multi-project evaluation framework based on the PEM is implemented, with each of the five projects serving as a distinct assessment unit. Firstly, a concise evaluation index system for assessing student performance is established via a literature review and questionnaire surveys. Subsequently, to better align with human expression habits, linguistic variables are employed by teachers to evaluate students' performance in each project before final scores are obtained by aggregating assessment data in the form of linguistic probability distributions using the PEM.

The rest of this paper is organized as follows. Section 2 briefly reviews the fundamental concepts of fuzzy measures and the PEM and introduces a multi-project student evaluation framework in online PBL over sustainability courses. In Section 3, a case study of students' performance evaluation is presented before sensitivity analyses are conducted in Section 4. Finally, conclusions are summarized in Section 5.

2. Materials and Methods

2.1. Fuzzy Measures

To assess the performance of students over online PBL courses emphasizing sustainability, fuzzy measures can be applied to reflect the importance or priority of the indicators as well as the subjectivity and uncertainty among the evaluators or teachers. Typically, the larger the fuzzy measure value of an indicator, the greater influence it exerts on the assessment score of students' performance. The definition of a fuzzy measure is as follows:

Definition 1 ([31]). *A fuzzy measure on a finite set* $N = \{1, 2, \ldots, n\}$ *is a set function* $\mu : P(X) \to [0, 1]$ *satisfying:*

1. $\mu(\varnothing) = 0, \mu(X) = 1;$
2. *If* $A, B \in P(X)$ *and* $A \subseteq B$, *then* $\mu(A) \leq \mu(B)$.

   where $P(X)$ *is the power set of X.*

2.2. Probabilistic Exceedance Method

In this part, the probabilistic exceedance method (PEM) will be reviewed as a data aggregation method for the multi-project evaluation of student performance in a linguistic environment. Compared with traditional data fusion methods mostly developed to fuse real numbers, the PEM can not only aggregate linguistic evaluation information presented in probability distributions but also eliminate the need to determine the weights of criteria by assigning values to the fuzzy measures of criteria to represent their relative importance. Furthermore, the PEM-based algorithm for evaluating students in online PBL has a straightforward calculation process, rendering it comprehensible to teachers of diverse academic backgrounds.

Assume that there are n criteria denoted as $C = \{C_1, C_2, \ldots, C_n\}$, and m levels of satisfaction represented as $S = \{S_1, S_2, \ldots, S_m\}$, where $S_1 > S_2 > \ldots > S_m$. Furthermore, the satisfaction of a student's performance against criterion C_i $(i = 1, 2, \ldots, n)$ is denoted in the form of the probability distribution as $P_i = \{p_{i1}, p_{i2}, \ldots, p_{im}\}$, $i = 1, 2, \ldots, n$. The main steps of the PEM are presented as follows [20]:

First, the exceedance distribution function for criterion C_i $(i = 1, 2, \ldots, n)$ is calculated based on P_i $(i = 1, 2, \ldots, n)$, which is represented as $EDF_i = [EDF_i(1), EDF_i(2), \ldots, EDF_i(j), \ldots, EDF_i(m)]$, where $EDF_i(j) = \sum_{t=1}^{j} p_{it}$, is used to express the probability that a student's satisfaction to criterion C_i $(i = 1, 2, \ldots, n)$ is at least S_j $(j = 1, 2, \ldots, m)$, and it satisfies $EDF_i(m) = 1$, $EDF_i(1) = p_{i1}$.

Afterward, by ranking $EDF_i(j)$ under a given S_j $(j = 1, 2, \ldots, m)$, the kth largest value of $EDF_i(j)$ can be obtained and denoted as $EDF_{g_j(k)}$. Here, we note that the fuzzy measure is represented as μ. According to the nature of the Choquet integral, the weight determination is shown as follows:

$$w_j(k) = \mu(H_{g_j(k)}) - \mu(H_{g_j(k-1)}) \tag{1}$$

where $H_{g_j(k)}$ is the subset of the criteria with the kth largest value of $EDF_{g_j(k)}$.

Furthermore, $EDF_{g_j(k)}$ can be fused to obtain the aggregated satisfaction values under the satisfaction level S_j $(j = 1, 2, \ldots, m)$ using the following equation:

$$B_j = \sum_{k=1}^{n} w_j(k) \cdot EDF_{g_j(k)} \tag{2}$$

where $w_j(k)$ is the weight of the criterion with the kth largest value of $EDF_{g_j(k)}$.

Thus, the satisfaction values for this student's performance under S_j $(j = 1, 2, \ldots, m)$ can be calculated using $\widetilde{p}_j = B_j - B_{j-1}$, where $B_0 = 0$. Correspondingly, the satisfaction distribution can be obtained as $\widetilde{P} = [\widetilde{p}_1, \widetilde{p}_2, \ldots, \widetilde{p}_m]$.

Finally, we can calculate the expected values for this student, which is shown as:

$$EV = \sum_{j=1}^{n} \widetilde{p}_j y_j \tag{3}$$

where y_j is the scalar value to represent the satisfaction level of S_j $(j = 1, 2, \ldots, m)$.

2.3. The Proposed Students' Performance Evaluation Framework Based on the PEM for Online PBL

To ensure a fair assessment, an evaluation index system must be established. This system should be created using criteria that accurately reflect the multi-faceted nature of engineering students' online PBL performance. Furthermore, given that a complex index system with numerous criteria can be cumbersome for teachers to manage, it is advisable to select only four to seven criteria. After key criteria are determined, students' performance will be assessed in each project based on a linguistic term set of $S = \{s_0, s_1, \ldots, s_l\}$, where l represents an even number. The five-phased framework to evaluate engineering students' online PBL performance over sustainability courses is presented in Figure 1. In the first phase, the evaluation index system of student comprehensive performance in online PBL is

constructed. Afterward, the weights of the projects, along with the fuzzy measures of the criteria, are assigned by several experienced teachers in phase 2. In phase 3, multi-project evaluation matrices are developed with evaluation values presented by linguistic variables, which are then aggregated by the approach PEM in phase 4. Finally, the comprehensive scores of students are obtained in phase 5 to represent engineering students' performance over multiple projects in the online PBL curriculum.
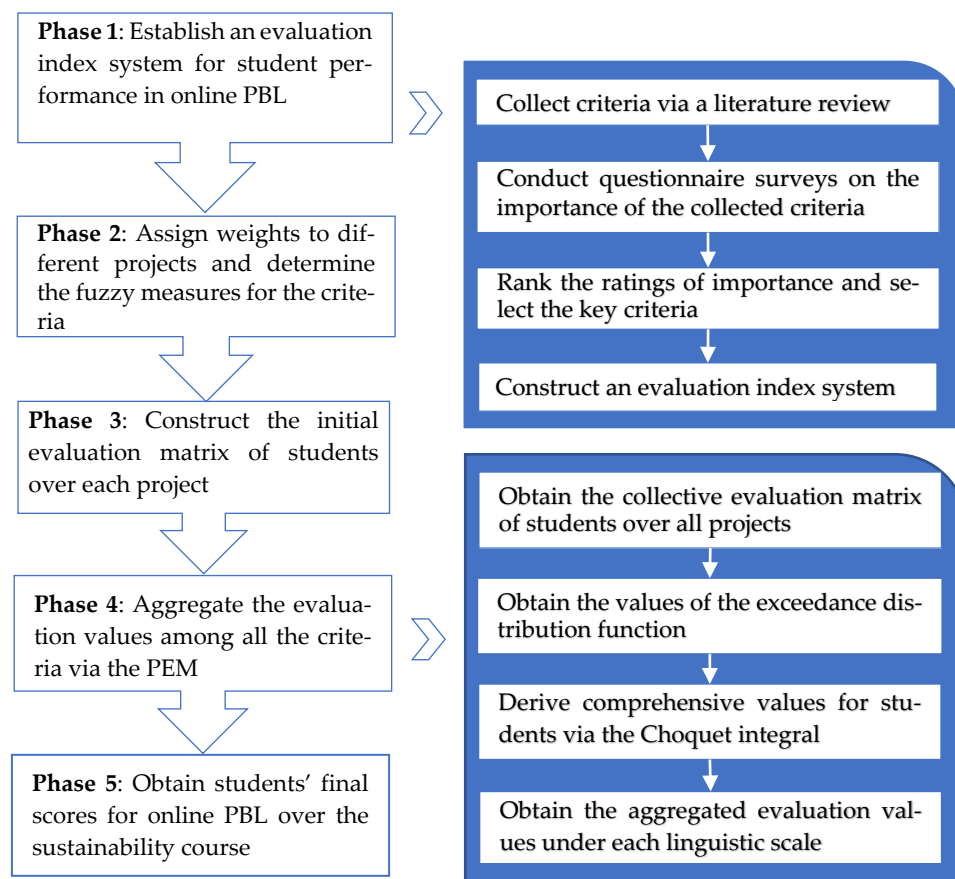
Phase 1: Establish an evaluation index system for student performance in online PBL

Collect criteria via a literature review

Conduct questionnaire surveys on the importance of the collected criteria

Rank the ratings of importance and select the key criteria

Construct an evaluation index system

Phase 2: Assign weights to different projects and determine the fuzzy measures for the criteria

Phase 3: Construct the initial evaluation matrix of students over each project

Obtain the collective evaluation matrix of students over all projects

Obtain the values of the exceedance distribution function

Derive comprehensive values for students via the Choquet integral

Obtain the aggregated evaluation values under each linguistic scale

Phase 4: Aggregate the evaluation values among all the criteria via the PEM

Phase 5: Obtain students' final scores for online PBL over the sustainability course

Figure 1. The proposed framework to evaluate engineering students' online PBL performance over the sustainability courses.

In this multi-criteria evaluation problem, supposing that there are m students forming a set of $A = \{A_1, A_2, \ldots, A_m\}$, n criteria forming a set of $C = \{C_1, C_2, \ldots, C_n\}$, and t projects forming a set of $P = \{P_1, P_2, \ldots, P_t\}$, then we can obtain the evaluation value r_{ij}^k for a student A_i $(i = 1, 2, \ldots, m)$ against criterion C_j $(j = 1, 2, \ldots, n)$ over the kth project P_k $(k = 1, 2, \ldots, t)$. Thus, an evaluation matrix of $R^k = (r_{ij}^k)_{m \times n}$, $k = 1, 2, \ldots, t$ can be developed, where the evaluation value r_{ij}^k is represented by a linguistic variable based on $S = \{s_0, s_1, \ldots, s_l\}$. In addition, the weight vector of t projects during the implementation of online PLB is assigned by the teachers and is denoted as $\omega = (\omega_1, \omega_2, \ldots, \omega_t)^T$. Meanwhile, the fuzzy measures of criteria C_j $(j = 1, 2, \ldots, n)$ are determined by teachers regarding the importance of criteria, which is denoted as $\mu(C_j)$, $j = 1, 2, \ldots, n$. The specific steps for this PEM-based evaluation algorithm are explained below:

Step 1. To obtain the assessing result of each student over t projects, the initial evaluation matrices $R^k = (r_{ij}^k)_{m \times n}$, $k = 1, 2, \ldots, t$ are fused into a collective evaluation matrix $R = (r_{ij})_{m \times n}$ using Equation (4):

$$r_{ij} = \{< s_g, \beta_{ij} > | g = 0, 1, \ldots, l\} \tag{4}$$

where r_{ij} is the collective evaluation value of student A_i $(i = 1, 2, \ldots, m)$ against criterion C_j $(j = 1, 2, \ldots, n)$, s_g is the linguistic variable based on $S = \{s_0, s_1, \ldots, s_l\}$ and $\beta_{ij} = \sum_{q \in G_{ij}} \omega_q$, $G_{ij} = \left\{ q \middle| r_{ij}^q == s_k, k = 0, 1, \ldots, l \right\}$.

Step 2. According to the collective evaluation matrix $R = (r_{ij})_{m \times n}$, the probability distribution degree p_g^{ij} of a student A_i $(i = 1, 2, \ldots, m)$ regarding the linguistic scale s_g $(g = 0, 1, \ldots, l)$ under criterion C_j $(j = 1, 2, \ldots, n)$ can be presented in the form of Table 1.

Table 1. The probability distribution degree p_g^{ij} of a student A_i $(i = 1, 2, \ldots, m)$.

Criteria	s_l	s_{l-1}	$\ldots$	s_0
C_1	p_l^{i1}	p_{l-1}^{i1}	$\ldots$	p_0^{i1}
C_2	p_l^{i2}	p_{l-1}^{i2}	$\ldots$	p_0^{i2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
C_n	p_l^{in}	p_{l-1}^{in}	$\ldots$	p_0^{in}

Step 3. The values of the exceedance distribution function for a student A_i $(i = 1, 2, \ldots, m)$ against each criterion C_j $(j = 1, 2, \ldots, n)$ are then calculated based on the probability vector $p^{ij} = (p_l^{ij}, p_{l-1}^{ij}, \ldots, p_0^{ij})$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$, and are represented as:

$$EDF_{ij} = [EDF_{ij}(l), EDF_{ij}(l - 1), \ldots, EDF_{ij}(g), \ldots, EDF_{ij}(0)] \tag{5}$$

where $EDF_{ij}(g) = \sum_{q=g}^{l} p_q^{ij}$. $EDF_{ij}(g)$ is used to express the probability that the satisfaction level of a student A_i $(i = 1, 2, \ldots, m)$ to criterion C_j $(j = 1, 2, \ldots, n)$ is at least s_g $(g = 0, 1, \ldots, l)$. For $g = 0$ and $g = l$, we can obtain $EDF_{ij}(0) = 1$ and $EDF_{ij}(l) = p_l^{ij}$ respectively.

Step 4. Based on the obtained exceedance distribution of a student A_i $(i = 1, 2, \ldots, m)$ against each criterion C_j $(j = 1, 2, \ldots, n)$, the values of probability $EDF_{ij}(g), j = 1, 2, \ldots, n$ are ranked from the largest to the smallest before we obtain the uth largest value of $EDF_{ij}(g)$ for A_i $(i = 1, 2, \ldots, m)$ under the linguistic scale s_g $(g = 0, 1, \ldots, l)$, which is denoted as $EDF_i^g(u)$. Correspondingly, the criterion with the uth largest value of $EDF_i^g(u)$ is denoted as $C_i^g(u)$. Meanwhile, the criteria set with the largest to the uth largest value of $EDF_{ij}(g)$ is denoted as $\mathbb{C}_i^g(u)$.

Step 5. Based on the criteria set $C = \{C_1, C_2, \ldots, C_n\}$, all of its subsets can be listed, and the fuzzy measures of a given subset Ψ within $C = \{C_1, C_2, \ldots, C_n\}$ are obtained using the following equation:

$$\mu(\Psi) = \max_{C_j \in \Psi} \{\lambda_j\} \tag{6}$$

where λ_j is the fuzzy measure of criterion C_j $(j = 1, 2, \ldots, n)$.

Step 6. According to the obtained fuzzy measures, the weight of $C_i^g(u)$, namely the criterion with the uth largest value of $EDF_i^g(u)$, can be calculated using the following equation:

$$v_i^g(u) = \mu(\mathbb{C}_i^g(u)) - \mu(\mathbb{C}_i^g(u - 1)), u = 1, 2, \ldots n. \tag{7}$$

where $\mu(\mathbb{C}_i^g(u))$ is the fuzzy measure of the criteria set $\mu(\mathbb{C}_i^g(u))$.

Step 7. The aggregated comprehensive values for a student A_i $(i = 1, 2, \ldots, m)$ under the linguistic scale s_g $(g = 0, 1, \ldots, l)$ can be obtained using the following equation:

$$B_i^g = \sum_{u=1}^{n} v_i^g(u) \cdot EDF_i^g(u) \tag{8}$$

where $v_i^g(u)$ is the weight of $C_i^g(u)$ and $EDF_i^g(u)$ is the uth largest value of $EDF_{ij}(g)$ for a student A_i $(i = 1, 2, \ldots, m)$ under the linguistic scale s_g $(g = 0, 1, \ldots, l)$.

Step 8. The comprehensive evaluation value for a student A_i $(i = 1, 2, \ldots, m)$ under the linguistic scale s_g $(g = 0, 1, \ldots, l)$ can be obtained using the following equation:

$$\Im_i^g = B_i^g - B_i^{g+1}, g = 0, 1, \ldots, l \tag{9}$$

where B_i^g is the aggregated comprehensive values for a student A_i $(i = 1, 2, \ldots, m)$ under the linguistic scale s_g and $B_i^{l+1} = 0$.

Step 9. The final score for a student A_i $(i = 1, 2, \ldots, m)$ in online PBL can be obtained using the following equation:

$$S(A_i) = 100 \cdot \sum_{g=0}^{l} \Im_i^g N(s_g) \tag{10}$$

where $N(s_g) = g$ and $\Im_i^g$ is the comprehensive evaluation value for student A_i $(i = 1, 2, \ldots, m)$ under the linguistic scale s_g.

3. Case Study and Results

As businesses and organizations increasingly recognize the importance of sustainability, there is a growing demand for professionals who can manage projects in a way that is environmentally responsible and socially beneficial. To provide university students majoring in engineering management with relevant knowledge and skills, the online PBL course on sustainable decision analysis was launched for students in a university in China and is implemented with collaborative learning within groups. To obtain more precise assessment outcomes for students, a phased assessment method is utilized in which each individual project is evaluated as a standalone unit before the comprehensive values are aggregated via the PEM to represent each student's final score. During the assessment, students' performance is evaluated according to a linguist term set $S = \{s_0, s_1, s_2, s_3, s_4, s_5, s_6\} = \{$"extremely poor", "poor", "slightly poor", "medium", "slightly good", "good", "extremely good"$\}$.

In this case study, the performance of four students, namely A_1, A_2, A_3, and A_4, within an online PBL learning group is assessed, and a total number of five projects P_k $(k = 1, 2, \ldots, 5)$ are involved in this course. The chosen four students exhibited different age ranges, educational levels, and genders. Their individual details are summarized in Table 2. Furthermore, to eliminate potential personal biases and preferences when assigning fuzzy measures to the criteria, a panel of experts was convened, consisting of five highly experienced teachers, namely E_1, E_2, E_3, E_4, and E_5. Each of these teachers had a minimum of four years of experience in online PBL in the field of green project management. Similarly, in order to gather a diverse range of opinions, the recruited experts were from various age groups, genders, and levels of PBL experience, which can also be observed in Table 2. Additionally, a comprehensive description of the five projects on sustainable decision analysis within the online PBL course is presented in Supplementary Table S1. In order to ensure a diverse and inclusive learning experience, these five projects vary in terms of complexity and practicality, as well as thematic focus, objective orientation, and evaluation emphasis. Based on these comprehensive factors, as well as the importance of each project towards the achievement of the teaching goal, the weights vector of five projects were assigned by the expert panel, which is given as $\omega = (0.2, 0.3, 0.1, 0.3, 0.1)^T$ in this case study.

3.1. The Establishment of an Evaluation Index System for Students' Performance in Online PBL

During the evaluation of engineering students' performance in online PBL, it is crucial to design a well-considered and robust evaluation index system that aligns with the goals and outcomes of sustainability courses to ensure effective, meaningful, and fair assessments. In regards to the sustainable decision analysis course, it emphasizes sustainability, environmental responsibility, and the integration of ecological and social considerations

into project planning and execution. Meanwhile, transversal skills, such as environmental literacy, innovation, cooperation, and leadership, are also essential.

Table 2. The individual details for the four students and five experts involved in the evaluation.

Student/Teacher	Age Range	Gender	Educational Level	Online PBL Course Experience	Major/Research Areas
Student A_1	26–27	Female	PhD	2 Semesters	Engineering management
Student A_2	22–23	Male	Undergraduate	2 Semesters	Engineering management
Student A_3	24–25	Female	Masters	1.5 Semesters	Engineering management
Student A_4	18–19	Male	Undergraduate	4 Semesters	Engineering management
Teacher E_1	35–44	Male	Masters	4 Years	Green project management
Teacher E_2	45–54	Female	PhD	4 Years	Green project management
Teacher E_3	25–34	Male	Masters	4.5 Years	Green project management
Teacher E_4	25–34	Female	PhD	4 Years	Green project management
Teacher E_5	35–44	Female	Masters	4 Years	Green project management

To establish a reasonable assessment index system, some key indicators are first gleaned through a literature review, which can be seen in Table 3. However, applying all the listed criteria by teachers would be relatively time-consuming and less efficient due to the relatively large number of criteria. Owing to this, a screening process is further conducted by distributing 100 questionnaires to students and teachers in local universities. During the selection process, it is essential to choose teachers and students who possess prior experience with online PBL, thereby guaranteeing their familiarity with the online PBL learning environment and its intricacies. Specifically, all the teachers should possess a minimum of 1.5 years of experience in online PBL, while all the students should have at least one semester's involvement. Meanwhile, well-balanced gender and age demographics of both the teachers and students are required to ensure a diverse and representative sample. After this screening process, only 20 students and 15 teachers were selected, as online PBL strategies are still not common in universities in China. The details of age, gender, and levels of experience among the recruited teachers and students are presented in Supplementary Tables S2 and S3, respectively. After obtaining their consent, each individual was invited to select a rating for each gleaned criterion among five scales of "extremely important", "very important", "important", "neutral", and "unimportant". For the convenience of the data process, the above five scales were then converted to real numbers of "4", "3", "2", "1", and "0" after the questionnaires were collected. Afterward, by averaging the scores for each criterion, the mean rating of each criterion was derived and can be conveniently observed in Figure 2. The original data can be observed in Supplementary Table S4.

Table 3. The gleaned criteria to evaluate engineering students' performance for online PBL courses.

Criteria	Explanations	References
Technical competency (B_1)	This indicator evaluates engineering students' ability to use technology effectively for learning purposes, including e-learning platform navigation, use of learning tools and applications, and technical problem-solving.	[32,33]
Knowledge application (B_2)	This involves assessing whether engineering students have applied the relevant knowledge and skills gained from their studies to actual project work.	[34–36]
Collaboration and contribution (B_3)	This indicator assesses students' ability to collaborate and interact effectively with peers. Additionally, it measures individual contributions within a group setting, such as contributing ideas and suggestions to group projects.	[37,38]
Online engagement (B_4)	This indicator measures whether engineering students can actively engage in various online PLB activities and whether students can respond to the quizzes and discussions and complete assignments on the e-learning platform.	[39–41]

Table 3. *Cont.*

| Criteria	Explanations	References
Sustained learning (B$_5$)	This indicator focuses on students' ability to retain and apply acquired knowledge over time as well as their motivation and enthusiasm for continued learning beyond the course's completion.	[42,43]
Mastery of content (B$_6$)	This indicator assesses students' comprehension and mastery of course content, such as key concepts, theories, methods, or ideas, through assignments, tests, and quizzes.	[44,45]
Critical thinking (B$_7$)	This criterion assesses students' ability to think critically about sustainable decision-making issues in the engineering field. It evaluates their analytical skills and ability to evaluate evidence and produce sound judgments based on logic and reasoning.	[46–49]
Problem-solving skills (B$_8$)	This criterion evaluates how effectively engineering students are able to address sustainability challenges they encounter. This can be measured by looking at the variety of solutions they come up with or their ability to solve technical difficulties while working on practical projects.	[50–53]
Time management (B$_9$)	This indicator evaluates whether students can ensure an adequate allocation of time to ensure the timely completion of tasks and projects in online PBL.	[54]
Reflection and learning (B$_{10}$)	This criterion evaluates students' ability to reflect on their learning and identify areas for improvement. It assesses their self-awareness, openness to feedback, and their approach to continuous learning in sustainable decision-making in construction and project management.	[55]
Sustainability integration (B$_{11}$)	This criterion measures engineering students' ability to integrate sustainability principles into sustainable decision-making. It evaluates their understanding of sustainability requirements and their implementation in project planning and execution.	[56,57]
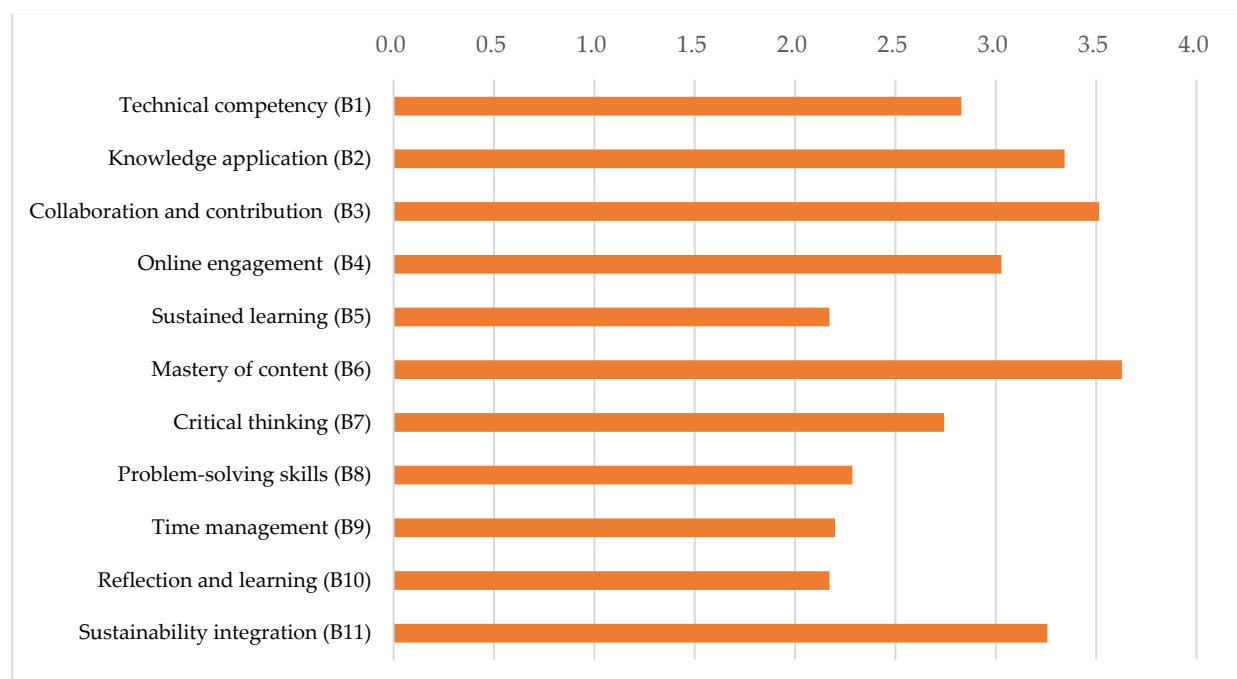
Figure 2. The average scores for the gleaned criteria based on the questionnaire surveys.

Based on the average score obtained for the gleaned criteria, five criteria with the highest importance scores were ultimately chosen to construct a comprehensive evaluation index, which are mastery of content (C$_1$), sustainability integration (C$_2$), collaboration and

contribution (C_3), knowledge application (C_4), and online engagement (C_5). The definitions for these five indicators are explained below:

- **Mastery of content (C_1)**: This indicator assesses the degree to which engineering students can understand and grasp the course content, such as the key concepts, theories, methods, or ideas. This is measured through the quality of online assignments submitted as well as the results of tests and quizzes in online PBL;
- **Sustainability integration (C_2)**: This criterion measures students' ability to integrate environmental, social, and economic sustainability into online PBL activities. For this criterion, students' performance is evaluated based on their ability to incorporate sustainability considerations into their project proposals, problem-solving approaches, and decision-making processes;
- **Collaboration and contribution (C_3)**: This indicator assesses students' ability to collaborate and interact effectively with peers, both asynchronously (through forums or message boards) and synchronously (through live chats or video conferencing). Additionally, it measures individual contributions within a group setting, such as contributing ideas and suggestions to group projects;
- **Knowledge application (C_4)**: This involves assessing whether students have applied the relevant knowledge and skills gained from the courses to project work and whether students can effectively address challenges encountered during project execution and propose sustainable solutions for project management;
- **Online engagement (C_5)**: This indicator measures the extent to which students actively engage in the various online PBL activities, including discussions, collaborations, the completion of assigned tasks, the response rates to assignments and quizzes on the e-learning platform, and so on.

3.2. The Evaluation of Students' Performance for Online PBL for the Course on Sustainable Decision Analysis

In this part, students' performance is assessed for the online PBL course on sustainable decision analysis via the PEM. Based on the established evaluation index system, the performance of four students was measured over each project, and the original evaluation matrix is shown in Table 4. Meanwhile, the fuzzy measures signifying the relative importance levels of the five criteria were provided by five teachers based on their expertise and experience, which are shown in Table 5. For the convenience of data processing, the average fuzzy measures were adopted to represent the collective fuzzy measures of five criteria, which are $\mu(C_1) = 0.9$, $\mu(C_2) = 0.6$, $\mu(C_3) = 0.8$, $\mu(C_4) = 0.7$, and $\mu(C_5) = 0.4$, respectively.

The detailed evaluation process is presented as follows: first, the collective decision-making matrix over five projects was obtained by fusing the initial evaluation matrix using Equation (4), and the results are shown in Table 6. For example, according to Table 4, the linguistic evaluation data for student A_1 against the criterion C_1 over five projects are s_5, s_4, s_5, s_3, and s_5, respectively. Given that the weight vector of five projects is $\omega = (0.2, 0.3, 0.1, 0.3, 0.1)^T$, the fused data for student A_1 against C_1 was obtained as $r_{11} = \{< s_3, 0.3 >, < s_4, 0.3 >, < s_5, 0.4 >\}$. Afterward, r_{11} was presented in the form of Table 1 by putting the probability degrees to the corresponding linguistic scale, and the results can be seen in the second line of Table 6. Similarly, all the probability distribution degrees for student A_1 regarding the linguistic scale s_g ($g = 0, 1, \ldots, 6$) under criterion C_j ($j = 1, 2, \ldots, 5$) are presented in Table 6.

Afterward, the values of the exceedance distribution function of student A_1 against each criterion under the linguistic scale s_g ($g = 0, 1, \ldots, 6$) were calculated and presented in Table 7 before the uth ($u = 1, 2, \ldots, 5$) largest value was derived along with its corresponding criterion, which is shown in the form of $EDF_1^g(u)/C_1^g(u)$ in Table 8. Meanwhile, the criteria sets $\mathbb{C}_1^g(u)$ with the largest to the uth largest value of $EDF_{1j}(g)$ are shown in Table 9. Based on the given fuzzy measures of the five criteria, the fuzzy measures of the subsets of $C = \{C_1, C_2, \ldots, C_5\}$ were derived using Equation (6). For example,

$\max\{(\mu(C_1), \mu(C_2), \mu(C_3)\} = \max\{0.9, 0.6, 0.8\} = 0.9$. Likewise, the fuzzy measures of all the subsets in Table 9 could be obtained, and the results are presented in Table 10. Based on the obtained fuzzy measures, the weights of $C_1^g(u)$ were calculated using Equation (7), and the results are shown in Table 11.

Table 4. The original evaluation matrix of four students over each project in online PBL.

Students	Projects	C_1	C_2	C_3	C_4	C_5
A_1	P_1	s_5	s_5	s_5	s_6	s_5
	P_2	s_4	s_5	s_6	s_5	s_5
	P_3	s_5	s_4	s_5	s_5	s_5
	P_4	s_3	s_5	s_5	s_6	s_5
	P_5	s_5	s_6	s_5	s_4	s_4
A_2	P_1	s_2	s_3	s_4	s_3	s_5
	P_2	s_3	s_3	s_3	s_4	s_4
	P_3	s_3	s_3	s_4	s_2	s_4
	P_4	s_4	s_4	s_3	s_2	s_4
	P_5	s_4	s_3	s_3	s_3	s_4
A_3	P_1	s_4	s_3	s_4	s_3	s_4
	P_2	s_3	s_3	s_4	s_2	s_5
	P_3	s_3	s_2	s_3	s_3	s_5
	P_4	s_3	s_2	s_5	s_3	s_5
	P_5	s_3	s_3	s_5	s_2	s_5
A_4	P_1	s_4	s_3	s_5	s_5	s_4
	P_2	s_5	s_3	s_3	s_4	s_3
	P_3	s_3	s_4	s_3	s_3	s_4
	P_4	s_3	s_4	s_3	s_4	s_4
	P_5	s_5	s_4	s_4	s_5	s_4

Table 5. The fuzzy measures of the criteria provided by five experts.

Experts	C_1	C_2	C_3	C_4	C_5
Expert 1	0.95	0.60	0.80	0.70	0.40
Expert 2	0.90	0.70	0.90	0.70	0.50
Expert 3	0.90	0.70	0.80	0.80	0.40
Expert 4	0.85	0.50	0.70	0.70	0.40
Expert 5	0.90	0.50	0.80	0.60	0.30
Average values	0.90	0.60	0.80	0.70	0.40

Table 6. The probability distribution of the collective evaluation values for student A_1.

Criteria	s_6	s_5	s_4	s_3	s_2	s_1	s_0
C_1	0.0	0.4	0.3	0.3	0.0	0.0	0.0
C_2	0.1	0.8	0.1	0.0	0.0	0.0	0.0
C_3	0.3	0.7	0.0	0.0	0.0	0.0	0.0
C_4	0.5	0.4	0.1	0.0	0.0	0.0	0.0
C_5	0.0	0.9	0.1	0.0	0.0	0.0	0.0

Table 7. The exceedance distribution function values of student *A₁* under five criteria.

EDF_{1j}	s_6	s_5	s_4	s_3	s_2	s_1	s_0
EDF_{11}	0.0	0.4	0.7	1.0	1.0	1.0	1.0
EDF_{12}	0.1	0.9	1.0	1.0	1.0	1.0	1.0
EDF_{13}	0.3	1.0	1.0	1.0	1.0	1.0	1.0
EDF_{14}	0.5	0.9	1.0	1.0	1.0	1.0	1.0
EDF_{15}	0.0	0.9	1.0	1.0	1.0	10	1.0

Table 8. The uth largest value of $EDF_1^g(u)$ and its corresponding $C_1^g(u)$.

u	s_6	s_5	s_4	s_3	s_2	s_1	s_0
$u = 1$	$0.5/C_4$	$1.0/C_3$	$1.0/C_2$	$1.0/C_1$	$1.0/C_1$	$1.0/C_1$	$1.0/C_1$
$u = 2$	$0.3/C_3$	$0.9/C_2$	$1.0/C_3$	$1.0/C_2$	$1.0/C_2$	$1.0/C_2$	$1.0/C_2$
$u = 3$	$0.1/C_2$	$0.9/C_4$	$1.0/C_4$	$1.0/C_3$	$1.0/C_3$	$1.0/C_3$	$1.0/C_3$
$u = 4$	$0.0/C_1$	$0.9/C_5$	$1.0/C_5$	$1.0/C_4$	$1.0/C_4$	$1.0/C_4$	$1.0/C_4$
$u = 5$	$0.0/C_5$	$0.4/C_1$	$0.7/C_1$	$1.0/C_5$	$1.0/C_5$	$1.0/C_5$	$1.0/C_5$

Table 9. The criteria subsets with the largest to the uth largest value of $EDF_{1j}(g)$.

u	s_6	s_5	s_4	s_3	s_2	s_1	s_0
$u = 1$	$\{C_4\}$	$\{C_3\}$	$\{C_2\}$	$\{C_1\}$	$\{C_1\}$	$\{C_1\}$	$\{C_1\}$
$u = 2$	$\{C_3, C_4\}$	$\{C_2, C_3\}$	$\{C_2, C_3\}$	$\{C_1, C_2\}$	$\{C_1, C_2\}$	$\{C_1, C_2\}$	$\{C_1, C_2\}$
$u = 3$	$\{C_2, C_3, C_4\}$	$\{C_2, C_3, C_4\}$	$\{C_2, C_3, C_4\}$	$\{C_1, C_2, C_3\}$	$\{C_1, C_2, C_3\}$	$\{C_1, C_2, C_3\}$	$\{C_1, C_2, C_3\}$
$u = 4$	$\{C_1, C_2, C_3, C_4\}$	$\{C_2, C_3, C_4, C_5\}$	$\{C_2, C_3, C_4, C_5\}$	$\{C_1, C_2, C_3, C_4\}$	$\{C_1, C_2, C_3, C_4\}$	$\{C_1, C_2, C_3, C_4\}$	$\{C_1, C_2, C_3, C_4\}$
$u = 5$	$\{C_1, C_2, C_3, C_4, C_5\}$	$\{C_1, C_2, C_3, C_4, C_5\}$	$\{C_1, C_2, C_3, C_4, C_5\}$	$\{C_1, C_2, C_3, C_4, C_5\}$	$\{C_1, C_2, C_3, C_4, C_5\}$	$\{C_1, C_2, C_3, C_4, C_5\}$	$\{C_1, C_2, C_3, C_4, C_5\}$

Table 10. The fuzzy measures of the criteria set $\mathbb{C}_{1j}^t(u)$ for student A_1.

u	s_6	s_5	s_4	s_3	s_2	s_1	s_0
$u = 1$	0.7	0.8	0.6	0.9	0.9	0.9	0.9
$u = 2$	0.8	0.8	0.8	0.9	0.9	0.9	0.9
$u = 3$	0.8	0.8	0.8	0.9	0.9	0.9	0.9
$u = 4$	0.9	0.8	0.8	0.9	0.9	0.9	0.9
$u = 5$	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Table 11. The weight of the criterion with the uth largest value of $EDF_1^g(u)$.

u	s_6	s_5	s_4	s_3	s_2	s_1	s_0
$u = 1$	0.7	0.8	0.6	0.9	0.9	0.9	0.9
$u = 2$	0.1	0.0	0.2	0.0	0.0	0.0	0.0
$u = 3$	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$u = 4$	0.1	0.0	0.0	0.0	0.0	0.0	0.0
$u = 5$	0.1	0.2	0.2	0.1	0.1	0.1	0.1

According to the ranked values of the exceedance distribution function in Table 8 and the weight of $C_1^g(u)$ in Table 11, the aggregated comprehensive values for student A_1 under each linguistic scale could be obtained, which are shown in Table 12. Likewise, the aggregated comprehensive values for the other students could also be derived and are presented in Table 12. The detailed calculation process for the four students via the PEM can be found in Supplementary Tables S6–S11. Afterward, the comprehensive values for the four students were derived using Equation (9), and the results are shown in Table 13. Finally, the final scores of the four students were computed using Equation (10), and the results are $S(A_1) = 86.6667$, $S(A_2) = 60.8333$, $S(A_3) = 70.3333$, and $S(A_4) = 69.0000$.

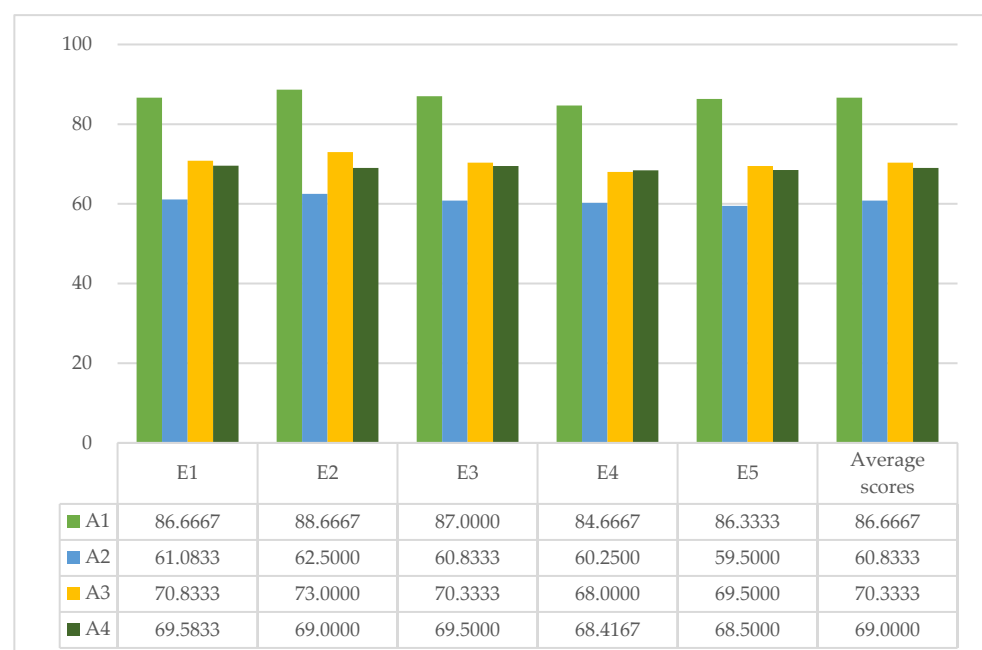**Table 12.** The aggregated comprehensive values of the students under each linguistic scale.

Students	s_6	s_5	s_4	s_3	s_2	s_1	s_0
A_1	0.38	0.88	0.94	1.00	1.00	1.00	1.00
A_2	0.00	0.08	0.63	0.94	1.00	1.00	1.00
A_3	0.00	0.48	0.85	0.96	1.00	1.00	1.00
A_4	0.00	0.36	0.78	1.00	1.00	1.00	1.00

Table 13. The comprehensive values for the students under each linguistic scale.

Students	s_6	s_5	s_4	s_3	s_2	s_1	s_0
A_1	0.38	0.50	0.06	0.06	0.00	0.00	0.00
A_2	0.00	0.08	0.55	0.31	0.06	0.00	0.00
A_3	0.00	0.48	0.37	0.11	0.04	0.00	0.00
A_4	0.00	0.36	0.42	0.22	0.00	0.00	0.00

4. Discussion

During the PEM-based evaluation of students' online PBL performance over the sustainability course, one significant feature was that it eliminated the need to determine the weight distribution of the criteria by providing the fuzzy measures to the criteria by the experts. To avoid individual preferences and bias, the ratings from an expert panel were collected instead of applying the fuzzy measure presented by a single expert. In this section, an analysis is first carried out to investigate the evaluation results of applying different fuzzy measures provided by each expert. As shown in Table 5, the fuzzy measures provided by different experts in representing the relative importance of criteria are distinctive. By applying each vector of fuzzy measures provided by individual experts, students' final scores were derived and are illustrated in Figure 3. In Figure 3, the rankings of four students based on their final scores are mostly $A_1 > A_3 > A_4 > A_2$, except for the case of expert 4, where the inverse orders are found between A_3 and A_4. Furthermore, by averaging the scores of each student under five vectors of fuzzy measures, the final scores of the students were derived as $S(A_1) = 86.6667$, $S(A_2) = 60.8333$, $S(A_3) = 70.3333$, and $S(A_4) = 69.0000$, which are exactly the same as the results obtained in the previous section when the average fuzzy measures were applied for the evaluation. In practical settings, the fuzzy measures can be collected from a larger pool of experts to ensure a more comprehensive and representative evaluation. By utilizing the mean values of the fuzzy measures provided by these experts, the PEM algorithm's results remain reliable and unbiased. Therefore, this approach significantly improves the evaluation efficiency for teachers while minimizing potential inaccuracies or biases.

	E1	E2	E3	E4	E5	Average scores
■ A1	86.6667	88.6667	87.0000	84.6667	86.3333	86.6667
■ A2	61.0833	62.5000	60.8333	60.2500	59.5000	60.8333
■ A3	70.8333	73.0000	70.3333	68.0000	69.5000	70.3333
■ A4	69.5833	69.0000	69.5000	68.4167	68.5000	69.0000

Figure 3. Students' final scores obtained based on the fuzzy measures of the criteria provided by five experts.

Furthermore, to establish the reliability of the PEM evaluation approach in assessing students' online PBL performance, sensitivity analyses were conducted by assigning variable values of the fuzzy measures to each criterion. In each scenario, the scores of four students were calculated. Initially, various values of fuzzy measures were assigned to the criterion of mastery of content (C_1), specifically 0.5, 0.6, 0.7, 0.8, 0.9, and 0.95, while maintaining the same fuzzy measures for the other criteria as in the previous case study, namely $\mu(C_2) = 0.6$, $\mu(C_3) = 0.8$, $\mu(C_4) = 0.7$, and $\mu(C_5) = 0.4$. In this scenario, the scores of the students were obtained and are displayed in Figure 4. The evaluation outcomes of the four students demonstrated minimal variations. As indicated in Figure 4, the final score of student A_1 remained unchanged, while the scores of the other three students exhibited a slight increase as the importance of C_1 increased. Despite the inverse order of rankings between A_3 and A_4 when the fuzzy measure of C_1 approached 0.70, the maximum score difference between them was only approximately 2 points when the fuzzy measure of C_1 was assigned a value of 0.50.
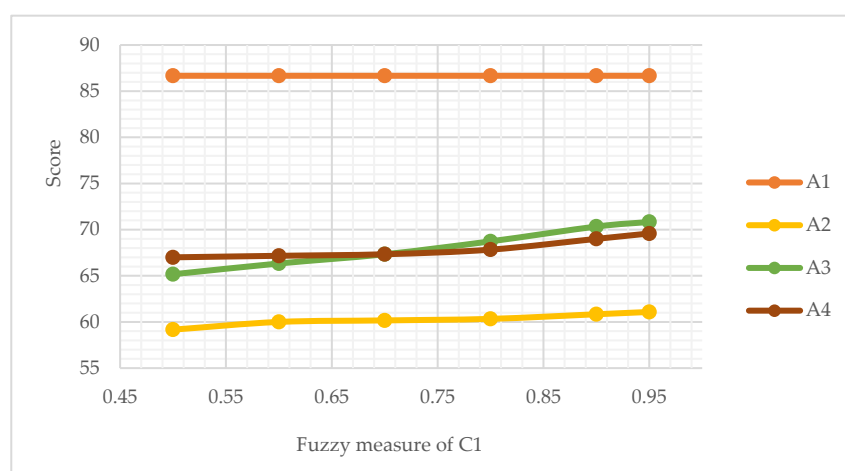
Figure 4. The scores of four students with varying values of fuzzy measures for C1.

Similarly, the impact of the varying values of fuzzy measures for C_2, C_3, C_4, and C_5 on students' scores was also investigated and is displayed in Figure 5. The sensitivity analysis data for the five criteria are detailed in Supplementary Tables S12–S16. This figure also demonstrates a relatively consistent ranking of $A_1 > A_3 > A_4 > A_2$. In Figure 5, the most obvious changes in students' scores are found for student A_1 in Figure 5b and for student A_2 in Figure 5d, when collaboration and contribution (C_3) and online engagement (C_5) are attached with higher importance, respectively. Meanwhile, it can be observed that the scores of student A_4 are the most steady under any given case, with only a minor increase occurring in Figure 5c when different fuzzy measures are assigned to knowledge application (C_4). Similarly, the tendency of inverse rankings between A_3 and A_4 can be noticed for most scenarios except in the analysis for sustainability integration (C_2) in Figure 5a, where the scores of A_3 and A_4 seem to be constant. Therefore, based on the sensitivity analysis, it can be inferred that the scores obtained by students are only slightly influenced by the variations in fuzzy measures across the criteria. In addition, the subjective nature of determining the values of fuzzy measures does not significantly impact the derivation of a reasonable assessment result.
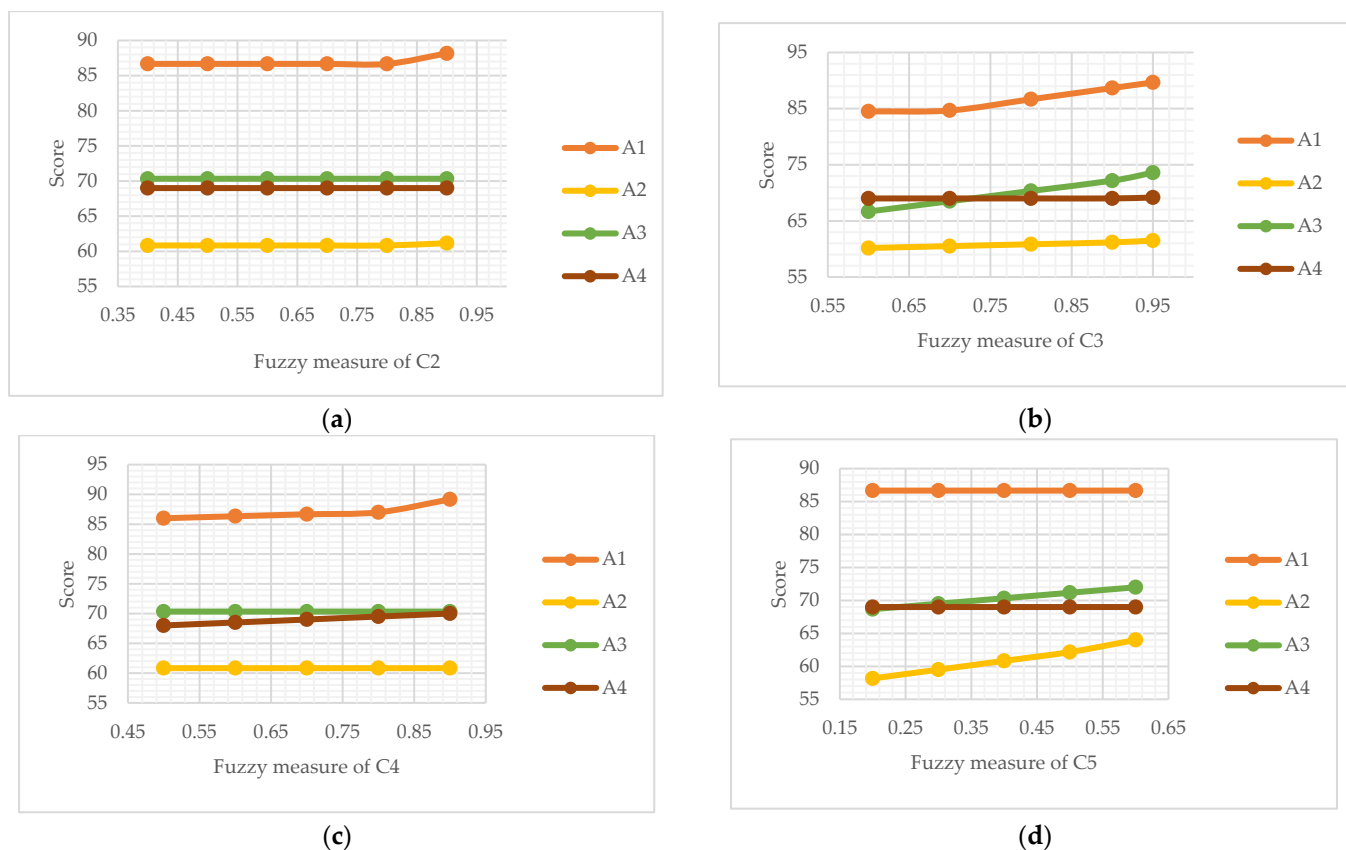
Figure 5. (**a**) The scores of four students with varying values of the fuzzy measures for C_2; (**b**) the scores of four students with varying values of the fuzzy measures for C_3; (**c**) the scores of four students with varying values of the fuzzy measures for C_4; (**d**) the scores of four students with varying values of the fuzzy measures for C_5.

In regards to the practical application of the PEM-based evaluation approach, it can be applied in a range of evaluation cases in PBL. For other curricula, a brief evaluation index system needs to be re-established by collecting the pertinent criteria representing the key evaluation dimensions towards the achievement of the teaching goal. Meanwhile, if many projects are involved in PBL, they can be bundled or divided into multiple phases to simplify the assessment process. For example, if seven projects P_k ($k = 1, 2, \ldots, 7$) are involved, the assessment can be carried out three times by grouping projects P_1 and P_2, projects P_3 and P_4, and projects P_5, P_6, and P_7 in three phases, respectively. By assigning specific weights to each phase, the PEM can be effectively utilized using Matlab R2020a software to calculate and derive students' scores in an efficient manner. Although the equations of the PEM-based algorithm for evaluating students may appear complex, it is, in fact, a relatively straightforward process as observed in the case study's calculation. Given the ongoing technological advancements, the PEM-based evaluation algorithm has the potential to be seamlessly integrated with various evaluation applications and software that are already installed on phones or computers. This integration renders it incredibly user-friendly and convenient for teachers who may not have a strong background in mathematics, thus eliminating any potential barriers to effective assessment.

In terms of the limitations of this study, the greatest subjectivity of the proposed model lies in teachers' judgement of student performance from various measuring dimensions. To represent students' performance via this PEM-based model, teachers are required to present linguistic variables for assessment, which may sometimes be unfair or too subjective, especially when only one teacher is in charge of students' performance evaluation. To address this issue, a team of teachers, comprising a minimum of three members, should provide individual evaluations prior to reaching a consensus on each assessment data point.

Moreover, to enhance the objectivity of the original evaluation data, students' performance under certain criteria, such as mastery of content (C_1), can be reflected via exams or tests. After the scores are obtained in the form of real numbers, they can be converted to linguistic variables based on the same linguistic term set. An example of a conversion from score ranges to the corresponding linguistic variables is shown in Supplementary Table S5. Afterward, all the linguistic variables can be processed using the PEM-based algorithm so as to derive students' comprehensive evaluation scores.

Another factor of subjectivity lies in the number of criteria applied for students' evaluations. While theoretically, the use of a comprehensive set of detailed criteria may lead to more accurate evaluation results, it can also increase the workload for teachers as each student needs to be evaluated against multiple criteria within a single project. On the other hand, if the number of criteria is quite limited, the selected criteria would not be able to cover the key assessment aspects of student performance in the online PBL. Therefore, it is important to balance the level of detail and the burden on teachers when selecting criteria for evaluation. To achieve this, it is advisable to base the number of criteria on the consensus of an expert panel, which typically falls within the range of four to seven criteria. Furthermore, the assignment of weights to projects involves a certain level of subjectivity. In the case study, five projects with diverse thematic focuses, objective orientations, evaluation emphases, and varying levels of complexity and practicality are described. This intricate scenario renders it challenging to determine the relative significance of the weight distributions across the five projects, which could be determined using techniques like the AHP or the best-worst method (BWM). Consequently, achieving a more reasonable weight distribution outcome is a crucial aspect that warrants further consideration in future studies regarding student evaluation in online PBL courses.

5. Conclusions

Online PBL for courses on sustainability in universities can provide an effective and engaging way to teach and learn about complex sustainability engineering issues. Given the distinct project characteristics, personal interests, and varying motivations among students in online PBL, it is crucial to assess their performance accurately and comprehensively over individual projects. To facilitate the evaluation of engineering management students participating in an online PBL course on sustainable decision analysis, a multi-project evaluation framework based on the PEM was implemented. First and foremost, a succinct evaluation index system for assessing students' performance was established through literature reviews and questionnaire surveys, which involved the criteria of mastery of content (C_1), sustainability integration (C_2), collaboration and contribution (C_3), knowledge application (C_4), and online engagement (C_5). Afterward, linguistic variables that aligned more closely with human expression habits were employed to assess students' performance, and the fuzzy measures of the criteria representing their relative importance were assigned by experts or teachers. Finally, student's scores were obtained via the PEM by aggregating the evaluation values presented in linguistic probability distributions. To better clarify the evaluation process, the performance of four students within a study group was assessed over five projects, and the final scores of the four students were 86.67, 60.83, 70.33, and 69.00, respectively.

In this PEM-based algorithm, by assigning fuzzy measures to the criteria, there is no need to determine the weight distribution of the criteria. Despite the subjective nature of determining the values of fuzzy measures, it was proven that varying values of fuzzy measures of each criterion did not significantly impact the derivation of consistent and stable assessment scores. Moreover, to ensure a more comprehensive and representative evaluation, the fuzzy measures can be collected from a broader range of experts. By utilizing the mean values of these measures, it was proven that the PEM algorithm could generate reliable and unbiased results. Another advantage of the PEM-based algorithm is that it enables the aggregation of linguistic evaluation information presented in probability distributions, which is a common data format encountered during multi-project evaluations

of student performance in linguistic environments. Lastly, the calculation of the PEM-based algorithm for evaluating students is relatively straightforward and accessible to teachers from diverse academic backgrounds.

In regard to the practical significance of this work, the proposed framework evaluates engineering students across multiple projects by providing a holistic view of their capabilities, skills, and knowledge application, thus ensuring a more comprehensive and nuanced assessment of their performance. Meanwhile, by flexibly adjusting the evaluation criteria and weights based on different projects and course objectives, it can be tailored to different educational contexts for diverse educational settings. In terms of its social justifications, by automating the evaluation process through the PEM algorithm, this study can significantly improve evaluation efficiency for teachers, allowing them to focus on other educational tasks while ensuring accurate and reliable results. It also benefits students by providing timely feedback on their performance in each project, allowing them to refine their learning strategies and approaches. Furthermore, by integrating sustainability into the evaluation framework, this research sends a powerful message to engineering students about considering the environmental, social, and economic impacts of their professional work. Doing so ensures that graduates are equipped with the necessary awareness and responsibilities required in their future careers, which aligns with broader social objectives of promoting sustainable development and environmental awareness.

References

1. Delaney, K.D. Multistage sustainability education for university engineering students: A case study from mechanical engineering in Technological University Dublin. *Int. J. Eng. Educ.* **2023**, *39*, 74–86.
2. Lozano, R.; Barreiro-Gen, M.; D'Amato, D.; Gago-Cortes, C.; Favi, C.; Martins, R.; Monus, F.; Caeiro, S.; Benayas, J.; Caldera, S.; et al. Improving sustainability teaching by grouping and interrelating pedagogical approaches and sustainability competences: Evidence from 15 worldwide higher education institutions. *Sustain. Dev.* **2023**, *31*, 349–359. [CrossRef]
3. Ota, E.; Murakami-Suzuki, R. Effects of online problem-based learning to increase global competencies for first-year undergraduate students majoring in science and engineering in Japan. *Sustainability* **2022**, *14*, 2988. [CrossRef]
4. Nicolaou, S.A.; Petrou, I. Digital redesign of problem-based learning (PBL) from face-to-face to synchronous online in biomedical sciences MSc courses and the student perspective. *Educ. Sci.* **2023**, *13*, 850. [CrossRef]
5. Saavedra, A.R.; Morgan, K.L.; Liu, Y.; Garland, M.W.; Rapaport, A.; Hu, A.; Hoepfner, D.; Haderlein, S.K. The impact of project-based learning on AP exam performance. *Educ. Eval. Policy Anal.* **2022**, *44*, 638–666. [CrossRef]
6. Li, S.H.; Li, Y.; Lin, H.H. Research on sustainable teaching models of new business-take chinese university business school as an example. *Sustainability* **2023**, *15*, 8037. [CrossRef]
7. Wang, X.; Sun, D.; Cheng, G.; Luo, H. Key factors predicting problem-based learning in online environments: Evidence from multimodal learning analytics. *Front. Psychol.* **2023**, *14*, 1080294. [CrossRef] [PubMed]
8. Baidal-Bustamante, E.; Mora, C.; Alvarez-Alvarado, M.S. STEAM project-based learning approach to enhance teaching-learning process in the topic of pascal's principle. *IEEE Trans. Educ.* **2023**, *66*, 632–641. [CrossRef]
9. Ngereja, B.; Hussein, B.; Andersen, B. Does project-based learning (PBL) promote student learning? A Performance Evaluation. *Educ. Sci.* **2020**, *10*, 330. [CrossRef]
10. Mang, H.A.; Chu, H.E.; Martin, S.N.; Kim, C.J. Developing an evaluation rubric for planning and assessing SSI-based steam programs in science classrooms. *Res. Sci. Educ.* **2023**, *53*, 1119–1144. [CrossRef]
11. Farshad, S.; Zorin, E.; Amangeldiuly, N.; Fortin, C. Engagement assessment in project-based education: A machine learning approach in team chat analysis. *Educ. Inf. Technol.* **2023**. [CrossRef]
12. Blasco-Blasco, O.; Liern-García, M.; López-García, A.; Parada-Rico, S.E. An academic performance indicator using flexible multi-criteria methods. *Mathematics* **2021**, *9*, 2396. [CrossRef]
13. Mohammed, H.J.; Daham, H.A. Analytic hierarchy process for evaluating flipped classroom learning. *CMC-Comput. Mater. Contin.* **2021**, *66*, 2229–2239.
14. Liu, Y.; Chen, H.X.; Thoff, A. Research on evaluation method of students' classroom performance based on artificial intelligence. *Int. J. Contin. Eng. Edu.* **2020**, *30*, 476–491. [CrossRef]
15. Zhang, W.J.; Yang, A.; Huang, L.S.; Leung, D.; Lau, N. Correlation between the composition of personalities and project success in project-based learning among design students. *Int. J. Technol. Des. Ed.* **2022**, *32*, 2873–2895. [CrossRef]
16. Parte, L.; Mellado, L. Academic performance in distance education: Quizzes as a moderator variable and students? Perception and expectation through linguistic analysis. *Online Learn.* **2022**, *26*, 124–147. [CrossRef]
17. Certa, A.; Enea, M.; Hopps, F. A multi-criteria approach for the group assessment of an academic course: A case study. *Stud. Educ. Eval.* **2015**, *44*, 16–22. [CrossRef]
18. Yager, R.R. OWA aggregation of probability distributions using the probabilistic exceedance method. *Fuzzy Sets Rough Sets Multisets Clust.* **2017**, *671*, 277–289.
19. Liu, Z.Y.; Xiao, F.Y. An intuitionistic linguistic MCDM model based on probabilistic exceedance method and evidence theory. *Appl. Intell.* **2020**, *50*, 1979–1995. [CrossRef]
20. Yager, R.R.; Alajlan, N. Multi-criteria formulations with uncertain satisfactions. *Eng. Appl. Artif. Intel.* **2018**, *69*, 104–111. [CrossRef]
21. Keshavarz-Ghorabaee, M.; Amiri, M.; Zavadskas, E.K.; Turskis, Z.; Antucheviciene, J. Determination of objective weights using a new method based on the removal effects of criteria (MEREC). *Symmetry* **2021**, *13*, 525. [CrossRef]
22. Torra, V.; Narukawa, Y.; Sugeno, M. On the f-divergence for discrete non-additive measures. *Inf. Sci.* **2020**, *512*, 50–63. [CrossRef]
23. Niu, P.L. An artificial intelligence method for comprehensive evaluation of preschool education quality. *Front. Psychol.* **2022**, *13*, 955870. [CrossRef] [PubMed]
24. Menon, S.; Suresh, M. Development of assessment framework for environmental sustainability in higher education institutions. *Int. J. Sustain. Higher Educ.* **2022**, *23*, 1445–1468. [CrossRef]
25. Wu, Y.N.; Jia, W.B.; Li, L.; Song, Z.X.; Xu, C.B.; Liu, F.T. Risk assessment of electric vehicle supply chain based on fuzzy synthetic evaluation. *Energy* **2019**, *182*, 397–411. [CrossRef]
26. Mao, Q.H.; Guo, M.X.; Lv, J.; Chen, J.J.; Xie, P.Z.; Li, M. A risk assessment framework of hybrid offshore wind-solar PV power plants under a probabilistic linguistic environment. *Sustainability* **2022**, *14*, 4197. [CrossRef]
27. Wang, G.P.; Liu, Y.; Hu, Z.Y.; Lyu, Y.L.; Zhang, G.M.; Liu, J.F.; Liu, Y.; Gu, Y.; Huang, X.C.; Zheng, H.; et al. Flood risk assessment based on fuzzy synthetic evaluation method in the Beijing-Tianjin-Hebei metropolitan area, China. *Sustainability* **2020**, *12*, 1451. [CrossRef]
28. Zhang, H.; He, W.; Xu, H.H.; Yang, H.; Ren, Z.X.; Yang, L.Z.; Sun, P.X.; Deng, Z.Y.; Li, M.H.; Wang, S.P.; et al. Investigating a water resource allocation model by using interval fuzzy two-stage robust planning for the Yinma River Basin, Jilin Province, China. *Water* **2021**, *13*, 2974. [CrossRef]

29. Bandara, J.; Shen, X.M.; Nurmohamed, Z. Resource allocator for non real-time traffic in wireless networks using fuzzy logic. *Wirel. Pers. Commun.* **2002**, *21*, 329–344. [CrossRef]

30. Ferreira, L.; Borenstein, D.; Righi, M.B.; de Almeida, A.T. A fuzzy hybrid integrated framework for portfolio optimization in private banking. *Expert Syst. Appl.* **2018**, *92*, 350–362. [CrossRef]

31. Sugeno, M. *Teory of Fuzzy Integral and Its Application*; Tokyo Institute of Technology: Tokyo, Japan, 1974.

32. Barrot, J.S.; Llenares, I.I.; Del Rosario, L.S. Students' online learning challenges during the pandemic and how they cope with them: The case of the Philippines. *Educ. Inf. Technol.* **2021**, *26*, 7321–7338. [CrossRef]

33. Blayone, T.; Mykhailenko, O.; VanOostveen, R.; Grebeshkov, O.; Hrebeshkova, O.; Vostryakov, O. Surveying digital competencies of university students and professors in Ukraine for fully online collaborative learning. *Technol. Pedagog. Educ.* **2018**, *27*, 279–296. [CrossRef]

34. Nicolay, B.; Krieger, F.; Stadler, M.; Vainikainen, M.P.; Lindner, M.A.; Hansen, A.; Greiff, S. Examining the development of metacognitive strategy knowledge and its link to strategy application in complex problem solving-a longitudinal analysis. *Metacogn. Learn.* **2022**, *17*, 837–854. [CrossRef]

35. McLean, S.; Attardi, S.M. Sage or guide? Student perceptions of the role of the instructor in a flipped classroom. *Act. Learn. High. Educ.* **2023**, *24*, 49–61. [CrossRef]

36. Al-Sharafi, M.A.; Al-Emran, M.; Iranmanesh, M.; Al-Qaysi, N.; Iahad, N.A.; Arpaci, I. Understanding the impact of knowledge management factors on the sustainable use of AI-based chatbots for educational purposes using a hybrid SEM-ANN approach. *Interact. Learn. Environ.* **2022**, *31*, 7491–7510. [CrossRef]

37. Verstegen, D.; Dailey-Hebert, A.; Fonteijn, H.; Clarebout, G.; Spruijt, A. How do virtual teams collaborate in online learning tasks in a MOOC? *Int. Rev. Res. Open Distrib. Learn.* **2018**, *19*, 39–55. [CrossRef]

38. Amira, T.; Lamia, M.; Hafidi, M. Learning styles in a collaborative algorithmic problem-based learning. *Rev. Socionetw. Strateg.* **2019**, *13*, 3–17. [CrossRef]

39. Salas-Pilco, S.Z.; Yang, Y.Q.; Zhang, Z. Student engagement in online learning in Latin American higher education during the COVID-19 pandemic: A systematic review. *Br. J. Educ. Technol.* **2022**, *53*, 593–619. [CrossRef]

40. Liu, Z.; Zhang, N.; Peng, X.; Liu, S.; Yang, Z.K. Students' social-cognitive engagement in online discussions: An integrated analysis perspective. *Educ. Technol. Soc.* **2023**, *26*, 1–15.

41. Wang, Y.Q.; Cao, Y.; Gong, S.Y.; Wang, Z.; Li, N.; Ai, L. Interaction and learning engagement in online learning: The mediating roles of online learning self-efficacy and academic emotions. *Learn. Individ. Differ.* **2022**, *94*, 102128. [CrossRef]

42. Terrón-López, M.J.; Velasco-Quintana, P.J.; Lavado-Anguera, S.; Espinosa-Elvira, M.D. Preparing sustainable engineers: A project-based learning experience in logistics with refugee camps. *Sustainability* **2020**, *12*, 4817. [CrossRef]

43. Yang, S.S.; Noughabi, M.A.; Jahedizadeh, S. Modelling the contribution of English language learners' academic buoyancy and self-efficacy to L2 grit: Evidence from Iran and China. *J. Multiling. Multicult. Dev.* **2022**. [CrossRef]

44. Khan, K.R.; Haque, M.M.; Alshemary, A.; AbouArkoub, A. BLDC motor-driven fluid pumping system design: An extrapolated active learning case study for electrical machines classes. *IEEE Trans. Educ.* **2020**, *63*, 173–182. [CrossRef]

45. Salimi, E.A.; Mirian, E.S.; Younesi, J. Anxiety level of mastery and performance avoid goal oriented EAP learners: The effect of teacher supportive motivational discourse. *Learn. Motiv.* **2022**, *78*, 101809. [CrossRef]

46. Tian, S.P.; Wang, L.J.; Zhang, F.; Han, T. Application of problem-based learning to promote critical thinking disposition among engineering undergraduates. *Int. J. Eng. Educ.* **2023**, *39*, 877–885.

47. Saad, A.; Zainudin, S. A review of project-based learning (PBL) and computational thinking (CT) in teaching and learning. *Learn. Motiv.* **2022**, *78*, 101802. [CrossRef]

48. Santos-Meneses, L.F.; Pashchenko, T.; Mikhailova, A. Critical thinking in the context of adult learning through PBL and e-learning: A course framework. *Think. Skills Creat.* **2023**, *49*, 101358. [CrossRef]

49. Anggraeni, D.M.; Prahani, B.K.; Suprapto, N.; Shofiyah, N.; Jatmiko, B. Systematic review of problem-based learning research in fostering critical thinking skills. *Think. Skills Creat.* **2023**, *49*, 101334. [CrossRef]

50. Wang, X.M.; Yu, X.H.; Hwang, G.J.; Hu, Q.N. An online progressive peer assessment approach to project-based learning: A constructivist perspective. *Educ. Technol. Res. Dev.* **2023**, *71*, 2073–2101. [CrossRef]

51. Higuera-Martinez, O.I.; Fernandez-Samacá, L.; Alvarado-Fajardo, A.C. PBL intervention for fostering creativity in first-year engineering students. *IEEE Trans. Educ.* **2023**, *66*, 442–449. [CrossRef]

52. Miliou, O.; Ioannou, A.; Georgiou, Y.; Vyrides, I.; Xekoukoulotakis, N.; Willert, S.; Andreou, A.; Andreou, P.; Komnitsas, K.; Zaphiris, P.; et al. The design of a postgraduate vocational training programme to enhance engineering graduates? Problem-solving skills through PBL. *Int. J. Eng. Educ.* **2022**, *38*, 1257–1273.

53. Manuaba, I.; Yi-No; Wu, C.C. The effectiveness of problem-based learning in improving critical thinking, problem-solving and self-directed learning in first-year medical students: A meta-analysis. *PLoS ONE* **2022**, *17*, e0277339. [CrossRef] [PubMed]

54. Belwal, R.; Belwal, S.; Sufian, A.B.; Al Badi, A. Project-based learning (PBL): Outcomes of students' engagement in an external consultancy project in Oman. *Educ. Train.* **2021**, *63*, 336–359. [CrossRef]

55. Du, X.Y.; Spliid, C.M.; Kolmos, A.; Lyngdorf, N.; Ruan, Y.J. Development of critical reflection for transformative learning of engineering educators in a PBL-based professional learning program. *Int. J. Eng. Educ.* **2020**, *36*, 1356–1371.

56. Gupta, C. The impact and measurement of today's learning technologies in teaching software engineering course using design-based learning and project-based learning. *IEEE Trans. Educ.* **2022**, *65*, 703–712. [CrossRef]

57. Pradhananga, P.; ElZomor, M. Developing social sustainability knowledge and cultural proficiency among the future construction workforce. *J. Civ. Eng. Educ.* **2023**, *149*, 04022011. [CrossRef]