

## RESEARCH ARTICLE

# Concept–Effect Relationship Weighting Based on Frequency of Concept’s Co-Occurrence for Developing Personalized Remedial Learning Path

YULIA WAHYUNINGSIH<sup>1</sup>, ARIF DJUNAIDY<sup>2</sup>, (Member, IEEE),  
AND DANIEL SIAHAAN<sup>1</sup>, (Member, IEEE)

<sup>1</sup>Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia

<sup>2</sup>Department of Information Systems, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia

Corresponding author: Arif Djunaidy (arif.djunaidy@its.ac.id)

This work was supported by the Ministry of Education, Culture, Research, and Technology of Indonesia, Doctoral Dissertation Research under Grant 1382/PKS/ITS/2022.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Lembaga Pendidikan dan Pengajaran Widyathama Parahita (LPP Widyathama Parahita).

**ABSTRACT** Notably, a personalized learning path is advantageous for students who need remedial assistance due to their struggle to master concepts and poor exam performance. Personalized learning paths produced from conceptual maps have proven highly effective. Nevertheless, in establishing personalized learning paths using conceptual maps, the manual assignment of weights that depict the degree of relationship between test items and concepts is time-consuming and ineffective. To get around the problems with giving weights by hand, we suggest an auto-generated concept relationship weighting method that is based on how often primary concepts and supporting concepts appear together in test items in this study. Moreover, we build a system that generates a personalized remedial learning path for students using the proposed weighting method and a conceptual map. To evaluate the effectiveness of the proposed approach, we carried out an experiment on high school students. As shown by the experimental results, the personalized remedial learning path generated can significantly enhance the learning achievement of underachieving students. In addition, the new auto-generated weighting method eliminated the need for experts to assign weights and resolved conflicting weight values.

**INDEX TERMS** Concept–item relationship weighting, conceptual effect relationship, frequency of concept’s co-occurrence, personalized learning path, remedial learning.

## I. INTRODUCTION

With the advancement of computer technology, many schools have adopted e-learning, which offers a practical approach to enhancing learning outcomes. E-learning has proven successful in teaching various subjects, such as science, mathematics, social sciences, the humanities, and language [1], [2], [3]. Personalized learning support in an e-learning

The associate editor coordinating the review of this manuscript and approving it for publication was Dominik Strzalka<sup>1</sup>.

environment increases students’ engagement and learning efficacy [4], [5]. Furthermore, integrating personalized learning support features and adopting the mastery learning approach in an e-learning environment has been shown to effectively improve students’ learning performance [6].

Several researchers have developed systems to diagnose learning difficulties, determine the appropriate learning path, and produce relevant additional learning materials tailored to each student’s needs to enhance personalized learning support within an e-learning environment. Initially, Hwang

et al. [7] proposed an approach to providing appropriate learning material for students by individually examining and analyzing their affective and cognitive statuses using fuzzy inference. The fuzzy inference approach was employed to establish an adaptive learning system, which was found to enhance students’ learning achievements and support under-achieving students in completing their assignments. Later, Gao et al. [8] proposed a method for generating a learning path based on modeling domain knowledge and learning resources using a knowledge graph. Gao et al. [8] applied the TextRank algorithm to extract keywords from courseware and construct the knowledge graph and employed the Ant Colony Optimization algorithm to generate the learning path. The results showed a high degree of similarity between the generated learning paths and those produced by experts. Subsequently, Subiyantoro et al. [9] proposed a method for establishing a learning path by considering a learning object ontology and accounting for students’ cognitive abilities. The Hybrid Particle Swarm Optimization (HPSO) algorithm generated the learning path. Experimental results using the HPSO algorithm depicted the effectiveness of the proposed method in recommending appropriate learning paths for several groups of students with varying cognitive abilities.

Besides the studies mentioned earlier, other research has investigated different approaches for establishing appropriate learning paths for students. One such approach involves the utilization of conceptual maps that showcase relationships between subject concepts. The Concept–Effect Relationship (CER) model, initially introduced by Hwang [10], is a specific type of conceptual map. In the CER approach, students may need prior knowledge of related concepts and an understanding of the relationships between them when encountering a new concept. In a CER model, all related subject concepts are hierarchically organized and interconnected, signifying the learning sequence. If a student has difficulty mastering a concept during the learning process, it may be because of a lack of understanding of the prerequisite concepts. The CER model effectively identifies students’ learning difficulties [11]. Furthermore, the CER model enables the provision of personalized feedback to students and assists teachers in identifying misconceptions [12]. Consequently, the utilization of the CER model in establishing learning paths has gained significant research [2], [13]. When generating a learning path using the CER model, it is important to identify and define the concepts and their relationships beforehand. In addition, we assign weightage values to specify the degree of relationship between concepts within a test item. In addition, these weight values calculate the error ratio of students when they answer test items related to each concept. This approach has several disadvantages. First, conflicts could arise among experts regarding the weight values assigned to concepts in test items. Furthermore, when dealing with a large question bank containing numerous test items, the process would become time-consuming, and

experts could experience fatigue, leading to lower-quality weight values. Furthermore, the introduction of new test items into the question bank required experts to allocate weight values to the relationships between the new items and concepts, resulting in additional time consumption. Hence, an alternative approach, such as an autogenerated method, is required to ensure the assignment of high-quality weight values between test items and concepts. This auto-generated approach would allow for the calculation of students’ error ratios for each concept, enabling the identification of concepts that necessitate further learning.

This research began with the development of a framework called the Personalized Scaffolding Adaptive Learning Management System [14]. In subsequent preliminary research, the study investigated the presence of hidden concepts within test items, which can provide insights into the relationships between various concepts [15]. We employed a classification method to categorize the multiple concepts associated with each test item [16]. Lastly, the frequency of co-occurrence among concepts was used to develop a novel weighting approach for generating personalized remedial learning paths (PRLPs).

Instead of solely relying on experts, an alternative approach can be explored to determine the weight value of items based on the occurrence of concepts within a group of test items in a question bank. Computation methods can be applied to assign weights to each test item by focusing on the characteristics of a question that involves multiple concepts in its solution, considering the simultaneous occurrence of concepts. This study suggests a better way to use conceptual maps. It uses computations to figure out weights and the level of connection between test items and concepts by looking at how often concepts appear together. Notably, the co-occurrence of concepts serves as the differentiating factor, which designates them as primary or supporting concepts within the test items. This research aims to address the following research questions:

- 1) What processes are involved in using the frequency of concept co-occurrences to generate a PRLP?
- 2) What are the benefits of implementing the frequency of concept co-occurrence in the development of a PRLP?

The experimental results of our research have proven that the proposed methodology would provide meaningful contributions to the development of science and computer technology, particularly in employing Natural Language Processing (NLP) solutions in processing and understanding text data. Our research addresses the typical problem in the prevailing process by replacing the role of an expert or multi-expert and their unending story with a machine learning algorithm for identifying weights in generating CER as a reference for designing student learning paths, particularly the learning paths for remedial students. One issue we have found in previous research studies related to CER generation methodology is that the analysis conducted by domain experts tends to be

rigid and subjective, leading to debates when determining weights. Another issue is related to the expert’s endurance and time consumption when they have to analyze a large number of test items in the question bank, which is expandable at any time, that is, more than 500 items, as exercised in our research. The consequence is that there are inconsistencies in concept weighting, which can cause difficulty in generating CER as a reference for designing learning paths for remedial students.

The rest of the paper is organized as follows: Section II provides an overview of the literature on personalized learning paths using the CER. Section III provides a concise explanation of the weighted CER approach using the frequency of concept co-occurrences to establish a PRLP. Section IV presents the experimental results obtained and analyzes the findings. Section V highlights several research discussions, and finally, Section VI concludes by summarizing the main conclusions drawn from the study and offering suggestions for future research directions.

## II. RELATED WORKS

Personalized learning path recommendations are an interesting area of research. This learning path recommendation aims to help students attain their learning objectives by recommending learning paths in the form of a sequence of learning objects that are appropriate to them. Through technological advances, personalized and adaptive learning path research is able to overcome the drawbacks of a “one-size-fits-all” approach [17]. In the learning path recommendations, the sequence of learning objects is essential for students. Several studies have introduced a graphical system in which graphs are denoted by nodes that represent the learning content and directed connections that depict the relationships between nodes. Research starts with a theory-based graph that can produce a one-way learning path by connecting all the necessary learning objects through its relationship [18].

In comparison, other studies have introduced a learning path recommendation method utilizing predetermined learning scenarios because students require various learning paths in different scenarios [19]. Research on learning path recommendations develops and builds on knowledge graph construction by recommending a learning path generation method that necessitates the determination of the start and end nodes through knowledge maps. This research recommends that a learning path be established by considering the domain knowledge and cognitive structure of students [20]. In research, the use of knowledge graphs to construct learning paths is due to the fact that they can avoid ambiguity in describing learning content. Shi et al. [21] developed a learning path recommendation model that uses a multidimensional knowledge graph framework to generate and recommend customized learning paths based on students’ target learning objectives. Furthermore, this study employs the basis of the graph to generate a path that meets the characteristics and requirements of the user. Another adaptive recommendation

system proposes predicting suitable educational paths for college preparatory years.

Researchers further developed graph-based learning path research by considering test items to find relationships between concepts, known as CER. In this study, experts will give weight to each test item based on the closeness of the test item’s relationship with the existing concept for building CER [11]. The CER weighting process is still to be further developed by other researchers [12], [22], offering a new procedure in which the integration of the opinion of the relationship of concept-test items based on the majority density of several experts to enhance the quality of the CER is proposed and used as personalized feedback. This technique is practical for reducing inconsistencies in the weighting criteria of multiple experts and enhances the overall learning diagnostic procedure for developing learning path systems. The weighting process is fundamental and a severe problem because the results of this process are decisive in learning path systems for providing learning recommendations and analyzing learning problems tailored to student needs. Considering that the weighted values act as inputs into learning path systems, poor-quality weighted values result in poor-quality feedback. Therefore, researchers are currently developing further studies on CER weighting. By leveraging the decisions made by several domain experts, Wanichsan et al. [12] have managed to provide a solution to the weaknesses of previous research using a rule-based approach.

In some cases, if there is a conflict in the weight of a test item, experts may need to be asked to reconsider their assessment, even if the majority decision has been used. Of course, this affects the quality of weighting, which consequently affects the process of building a learning path. Therefore, we must establish a method to determine the weight of test items that can solve the problem of weighting conflicts. A way of weighting, as explained in the previous paragraph, must be established because it is related to the function of the weight on each test item. We need quality weighting to produce a quality learning path. Quality weighting is a challenge, especially when building a learning path for remedial students. Remedial students are those who need special assistance [23], [24]. Learning paths designed to prepare them for retesting are required, whereas the built learning paths should be able to diagnose what concepts they require.

In previous research on CER, one or several experts were needed to determine the weight value between the concept and test items [11], [12], [22], with several potential problems including inconsistency of weight value, time-consuming conflict resolution among experts, heavy reliance on domain experts, and difficulty scaling up the number of concepts and test items. In this case, the opportunity to develop and deploy systems that can learn, recognize patterns, and make decisions with minimal human intervention is exciting [25]. Researchers have conducted extensive studies on CER, indicating the need to establish a method of weighing the interrelationships between concepts in a test item that avoids potential conflicts and produces quality

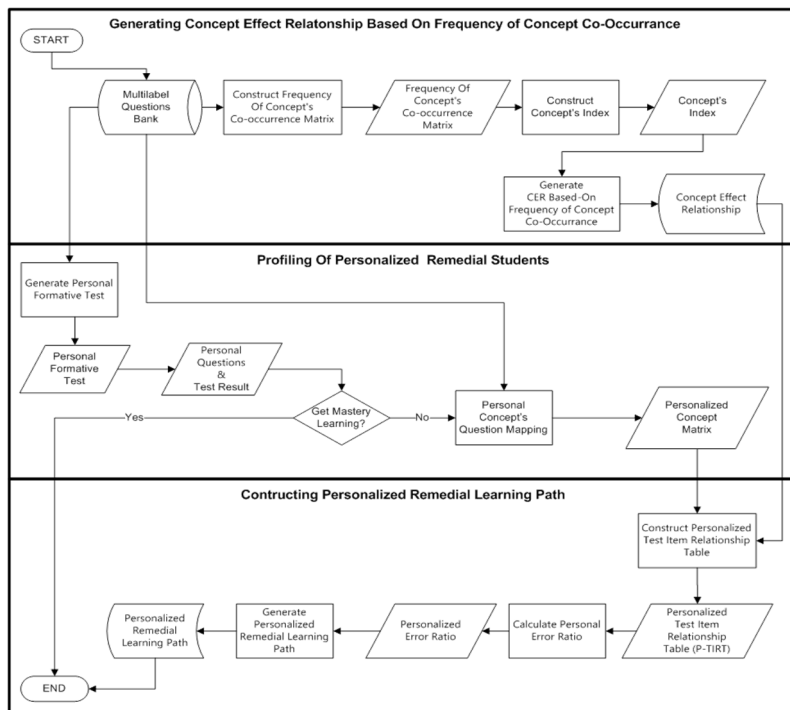


FIGURE 1. Process of Developing PRLP based on Frequency of Concept's Co-occurrence.

weighting. This method will help build a learning path for remedial students to achieve learning mastery. Consequently, by considering the frequency of the concept's appearance on each test item in the question bank, it can be assumed that the more often a concept appears, the more influential it is for students to master. This idea produced a new method, assigning a numerical weight to each test item in the question bank to measure the significance of the relationship between concepts according to the frequency of co-occurrence of these concepts in a set of test items.

### III. DEVELOPMENT OF PERSONALIZED REMEDIAL LEARNING PATHS METHOD

The process required to establish a PRLP varies from the process of building a learning path in general. Such a typical learning path is dedicated to assisting remedial students to acquire mastery. The use of CER in building a personalized learning path is advantageous in the sense that it can explicitly show the relationships between concepts. Meanwhile, the characteristics of CER in this latest research are significant in developing a mechanism that can trace back any concepts that remedial students have not mastered yet. The preliminary related study on the learning path for remedial students reported that the test items most likely contain some "hidden" concepts. The identification of those hidden concepts on the test items may facilitate the establishment of a remedial learning path that can use the co-occurrence of such concepts.

This present research identified the "hidden" concepts in the test item as supporting concepts, whereas a straightforward concept in the test item was called the

primary concept. Both primary and supporting concepts in the test item are the main ideas of the co-occurrence approach to autogenerate concepts' weightage. Utilization of the concept's co-occurrence in the test items to build a remedial learning path is described in the following paragraphs.

To obtain an established remedial learning path, one must have the weighted-directed graph of a concept and a remedial student's test results, which have a concept mapping of each test item. In obtaining the weightage of each node in that kind of graph, the concept-rank algorithm will be used based on the frequency of occurrences of the concept obtained from the labeled test items. These test items with labels taken from the question bank consist of the question data and the test item's primary and supporting concepts, which are organized afterward into a concept's co-occurrence matrix and stored as a database.

Later, a set of formative test items for each student is produced from the multilabel question bank. For each student who fails the test, personal profiling is performed on the sets of test items to generate the personalized concept matrix. Mapping the personal profile on a concept's weighted-directed graph combined with mapped fail-to-answer test items is then assumed. A personalized test item relationship table (P-TIRT) will establish a reference for calculating the personal error ratio. Then, when mapping the error ratio by the concept index, which is produced using the concept-rank algorithm, a PRLP that helps the remedial student learn the concepts according to their needs will be generated. A detailed explanation of the previously mentioned processes is within each subsection below. As an outline,



the exercised method above is shown in the flowchart in Figure 1.

**A. GENERATING CONCEPT EFFECT RELATIONSHIP BASED ON THE FREQUENCY OF THE CONCEPT’S CO-OCCURRENCE**

Based on the frequency of the concept’s co-occurrence, a CER is generated through several stages. The output produced at this stage is a CER in the form of a weighted-directed graph. The weighting of nodes and edges is based on the frequency of the number of concepts that are found in all the test items in the questions’ database. The following information helps us understand the series of processes.

**1) CONSTRUCTING FREQUENCY OF CONCEPT’S CO-OCCURRENCE MATRIX**

During a face-to-face learning session in the class, teachers usually employ a concept map to assist them in delivering the materials based on the concept map order. When students have to work on test items in a formative test, they often encounter difficulties because the concepts appear to be separated from one another. Understanding more than one concept is often required to complete a test item. In science education, the concepts contained in test items can present relationships between them, and it is an exciting subject to explore. In this research, we develop ideas according to the appearance of concepts in a test item to build a PRLP. The node’s weightage, which represents the concept’s weight, can be determined by calculating the concept’s appearance frequency or the concept’s co-occurrence frequency on all test items in the question bank. The edge’s weightage, which represents the weight of the concept’s relationship among each other, is calculated using the normalization of the concept’s appearance, which has been adjusted to the index node generated from each node so that a concept’s weighted-directed graph can be formed.

To generate a concept’s weighted-directed graph that refers to the concept’s co-occurrence, the concepts in a test item must be considered. In a test item, there is always one primary concept, and the problem can contain supporting concepts that must be mastered by students who solve the test item—exploring the co-occurrence of the concept later used as a concept’s co-occurrence matrix. Adapting the Transformer BERT architecture, the concept’s co-occurrence matrix is autogenerated from the test items and then used for weighting at nodes and edges. When test items are given to the students to solve a primary concept, they also need knowledge of other concepts called supporting concepts.

The following examples of test items, along with their primary concepts and supporting concepts, can illustrate the discussion in the previous paragraph:

- 1) Two blocks are tied together using a rope and are hung vertically from a frictionless and massless pulley. The gravitational acceleration is  $10 \text{ m/s}^2$ . If the mass of block A = 1 kg and block B = 4 kg, how much is the acceleration of that system (in  $\text{m/s}^2$ )?

(primary concept: acceleration; supporting concepts: Newton’s second law of motion, action-reaction, tension, and weight)

Blocks A and B possess a mass of 2 kg and 6 kg, respectively. Block A is on top of the frictionless Block B, and Block B is on the frictionless horizontal surface. What is the tension in the rope for each block (N) when Block B is slowly pulled by a force of 18 N?

(primary concept: tension; supporting concepts: Newton’s second law of motion, net force, acceleration, and weight)

This research has been conducted for a formative test in Physics subject with Dynamics as the main topic consisting of 14 concepts, namely, force (C1), equilibrium (C2), net force (C3), Newton’s first law of motion (C4), mass and inertia (C5), acceleration (C6), gravity (C7), weight (C8), Newton’s second law of motion (C9), normal force (C10), friction (C11), tension force (C12), action-reaction (C13), and Newton’s third law of motion (C14). The examples of test items indicate that a primary concept in each test item may become a supporting concept in another test item. Table 1 shows the co-occurrence matrix of all 14 concepts within the test items.

**TABLE 1. Co-occurrence matrix between supporting concepts (rows) and primary concepts (columns).**

Matrix Concept Co-occurrence	Supporting Concept													
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14
Primary Concept	C1	0	1	1	1	1	0	0	1	0	0	0	0	0
	C2	4	0	6	6	3	0	0	1	0	0	0	0	1
	C3	15	4	0	4	0	4	0	0	4	0	0	0	0
	C4	1	0	2	0	2	0	0	0	1	0	0	0	0
	C5	8	3	2	11	0	4	0	8	4	1	0	2	5
	C6	50	0	29	4	48	0	0	14	50	1	3	9	8
	C7	4	0	1	0	3	0	0	4	2	0	0	0	3
	C8	8	2	3	1	5	7	5	0	2	0	0	0	0
	C9	19	0	11	1	4	20	0	11	0	4	8	11	1
	C10	8	0	22	0	0	1	0	18	13	0	2	0	10
	C11	1	0	18	14	0	6	0	5	14	4	0	1	0
	C12	3	1	22	2	1	21	0	20	21	4	3	0	1
	C13	0	0	0	0	0	0	0	3	0	0	0	3	0
	C14	1	0	2	0	0	2	0	1	2	3	0	0	4

Table 1 shows that on the labeled test items in the question bank database regarding C4 as the primary concept, there is one test item that requires C1, two test items containing C3, two test items containing C5, and one test item containing C9 as the supporting concepts. Figure 2 displays the relationship between nodes in the concept’s co-occurrence matrix.

Using a simple illustration in Table 1, the concept “Newton’s first law of motion” (C4) is the primary concept, which has four supporting concepts, and “force” (C1) is the supporting concept of one test item. Conversely, the concept “force” (C1) is the supporting concept of 15 test items, of which the primary concept is “net force” (C3), which is the supporting concept of the primary concept “mass and inertia” (C5). The concept’s co-occurrence frequency as well as relationships between the concepts can be presented as a graph. Figure 3 displays an example of a graph that depicts the relationships between the concepts “Newton’s first law of motion” (C4), “mass and inertia” (C5), “net

force” (C3), “force” (C1), and “Newton’s second law of motion” (C9).

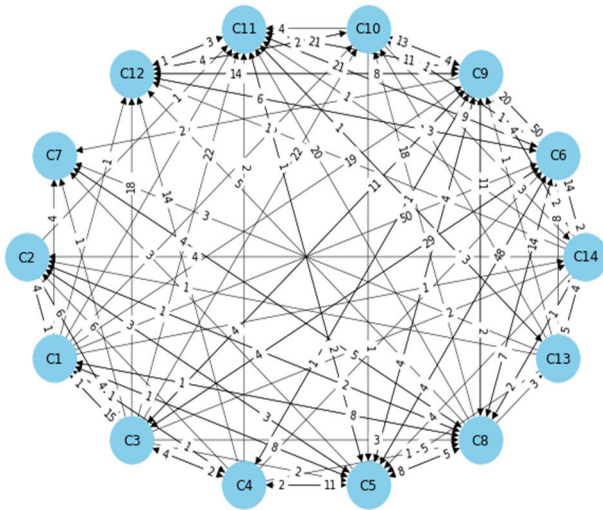


FIGURE 2. Concept’s co-occurrence relationships as a matrix representation.

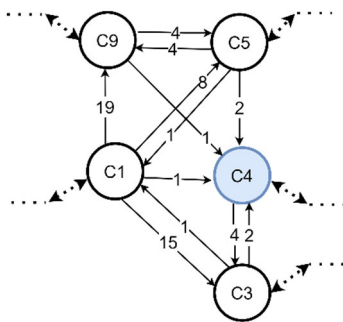


FIGURE 3. Illustration of co-occurrence relationships between “Newton’s first law of motion” (C4), “mass and inertia” (C5), “net force” (C3), “force” (C1), and “Newton’s second law of motion” (C9).

The concept emergence matrix counts the concept’s co-occurrence frequency of labeled test items in a question bank. This matrix is required to calculate the weights of both nodes and edges as the substantial components of CER.

### 2) CONSTRUCTING A CONCEPT’S INDEX

The learning process involves a specific order of obtaining concepts, whereas advanced concepts necessitate a prior understanding of fundamental concepts. This establishes a concept’s learning order, with higher-ranked concepts being more foundational and the relationships between concepts strictly one-way. A weighted-directed graph is constructed using the method introduced by Siahaan et al. [26] to identify the learning order and one-way relationships. Consider a weighted-directed graph  $G = (V, E)$  with vertices  $V$  and edges  $E$ , where each vertex  $V_i$  represents a concept  $i$ . Let  $c_{ij}$  denote the frequency of occurrence of concept  $j$  in test items with the primary concept  $i$ , and let  $N$  represent the total number of concepts. Equation (1) calculates the weight of a directed

edge  $w_{ij}$  from vertex  $V_i$  to vertex  $V_j$ .

$$w_{ij} = \frac{c_{ij}}{\sum_{k=0}^N c_{ik}} \tag{1}$$

Subsequently, the score  $S(V_i)$  was needed to rank the concepts. This method is based on the PageRank algorithm proposed, and we set the value of  $d$  to 0.85 as per the recommendation of Brin and Page [27]. The calculation of the score associated with the vertex  $i$  at iteration  $t$  is performed using equation (2).

$$S(V_i)_t = (1 - d) + d * \sum_{V_j \in in(V_i)} \frac{w_{ij}}{\sum_{V_k \in out(V_j)} w_{jk}} S(V_j)_{t-1} \tag{2}$$

In equation (2),  $in(V_i)$  is the set of predecessor vertices of the vertex  $V_i$ ,  $out(V_i)$  is the set of successor vertices of the vertex  $V_i$ , and  $d$  is a damping factor that represents the probability of jumping to a random vertex in the graph from a given vertex. The calculation continues until the error rate  $\varepsilon$  is smaller than 0.001. At iteration  $t$ , we calculate the error rate using Equation (3). Table 2 presents the results of the score calculation for each concept, in which the score index indicates the concept’s order based on its frequency of appearance in test items.

$$\varepsilon_t = \min_{i=0}^N (S(V_i)_t - S(V_i)_{t-1}) \tag{3}$$

TABLE 2. Results of the score index calculation.

No.	Concept ID	Concept Name	Score Index
1	C2	Equilibrium of a rigid body	1.500875
2	C14	Newton’s third law of motion	1.275427
3	C13	Action–reaction	1.248609
4	C4	Newton’s first law of motion	1.210815
5	C5	Mass and inertia	1.156035
6	C3	Net force	1.093273
7	C10	Normal force	0.994297
8	C8	Weight	0.971905
9	C1	Force	0.938537
10	C11	Tension	0.849631
11	C9	Newton’s second law of motion	0.789663
12	C6	Acceleration	0.729431
13	C7	Gravitational acceleration	0.681974
14	C12	Friction	0.536550

### 3) GENERATING CER BASED ON THE FREQUENCY OF THE CONCEPT’S CO-OCCURRENCE

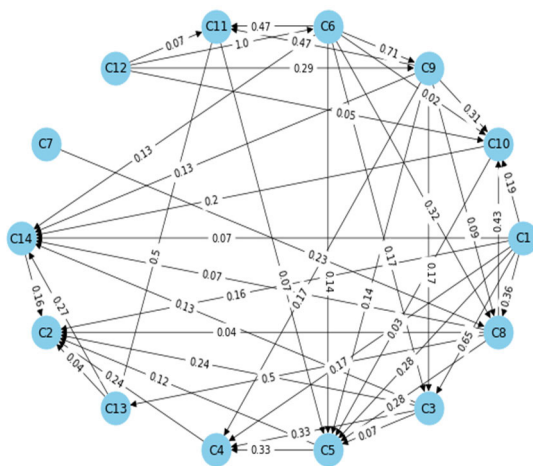
When constructing a concept’s index, the co-occurrence relationship from lower-ranked supporting concepts to

higher-ranked primary concepts is disregarded, and the co-occurrence values in the matrix are set to 0. Next, we normalize the co-occurrence matrix values between 0 and 1 by dividing each value by the sum of its respective row. As shown in Table 3, this normalization process establishes a one-way relationship between the concepts, and the resulting normalized values indicate the degree of relationship between them. The concept’s hierarchy changes based on the concept’s score index, which results in an adjusted hierarchy.

**TABLE 3.** New relationship matrix between supporting concepts and primary concepts.

	Supporting Concepts													
	C2	C14	C13	C4	C5	C3	C10	C8	C1	C11	C9	C6	C7	C12
C2	0	0.16	0.04	0.24	0.12	0.24	0	0.04	0.16	0	0	0	0	0
C14	0	0	0.267	0	0	0.133	0.2	0.067	0.067	0	0.133	0.133	0	0
C13	0	0	0	0	0	0	0	0.5	0	0.5	0	0	0	0
C4	0	0	0	0	0.333	0.333	0	0	0.167	0	0.167	0	0	0
C5	0	0	0	0	0	0.069	0.0345	0.276	0.276	0.069	0.138	0.138	0	0
C3	0	0	0	0	0	0	0	0.652	0	0.174	0.174	0	0	0
C10	0	0	0	0	0	0	0	0.429	0.190	0	0.3095	0.024	0	0.048
C8	0	0	0	0	0	0	0	0	0.364	0	0.091	0.318	0.227	0
C1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C11	0	0	0	0	0	0	0	0	0	0	0.467	0.467	0	0.067
C9	0	0	0	0	0	0	0	0	0	0	0	0.714	0	0.286
C6	0	0	0	0	0	0	0	0	0	0	0	0	0	1
C7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C12	0	0	0	0	0	0	0	0	0	0	0	0	0	0

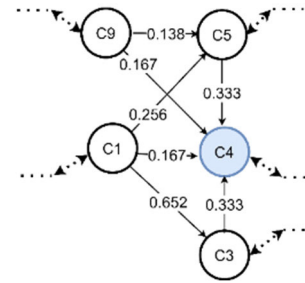
Figure 4 shows the graph representation of the changes observed in the concept’s co-occurrence matrix. The arrows in the graph depict the relationship where a concept serves as a supporting concept for the corresponding intended concept (node).



**FIGURE 4.** Concept’s co-occurrence-based concept–effect relationship.

Figure 5 presents a revised representation of the relationships between the concepts “Newton’s first law of motion” (C4), “mass and inertia” (C5), “net force” (C3), “force” (C1), and “Newton’s second law of motion” (C9)—initially

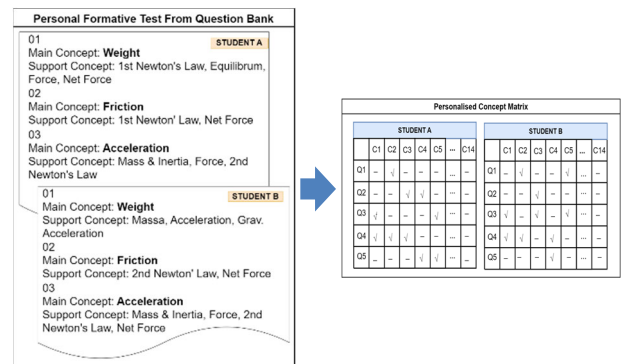
shown in Figure 3—displaying the updated concept relationships and their corresponding degrees of relationship.



**FIGURE 5.** Illustration of relationships between the concepts “Newton’s first law of motion” (C4), “mass and inertia” (C5), “net force” (C3), “force” (C1), and “Newton’s second law of motion” (C9) with their respective degree of relationship.

**B. PERSONALIZED PROFILING OF REMEDIAL STUDENTS**

In an online formative test, students may receive different test item sets with different supporting concepts. Producing the test items individually allows for personalized sets. Therefore, to identify the specific concepts that students struggle with in achieving mastery, a personalized concept profiling process is required. Figure 6 illustrates that even if students achieve the same score on a set of test items with the same primary concept, the supporting concepts vary.



**FIGURE 6.** Illustration of personalized concept matrix profiling.

In Figure 6, as an example, Student A and Student B both obtained a score of 40, which indicates that they did not achieve mastery learning. Nevertheless, their concept distribution matrices vary, which highlights the unique areas of difficulty for each student. This personalized concept profiling process is important for developing a customized learning path that caters to the diverse needs of each individual student.

**C. CONSTRUCTING PERSONALIZED REMEDIAL LEARNING PATH**

The final step involves using the CER, the weighted-directed graph, and the results of profiling remedial students to construct a personal test item relationship table. This table serves as a reference for calculating the personal error ratio. The section below discusses the process of constructing a PRLP.

1) CONSTRUCTING PERSONALIZED TEST ITEM RELATIONSHIP TABLE AND ERROR RATIO

The following steps are taken to calculate the weight values that indicate the degree of relationships between test items and concepts: (1) Identify the primary concept and supporting concepts in each test item. (2) Define the one-way relations and degree of relationships between the concepts. (3) Calculate the weight  $W(V_i)$  of a concept  $i$  to a test item  $a$  using equations (4)–(8). In these equations,  $W(V_i)_{main_a}$  is the weight of the primary concept of test item  $a$ ,  $W(V_i)_{supp_a}$  is the weight of a supporting concept of test item  $a$ ,  $n_a$  is the total number of concepts in test item  $a$ , and  $n_{max}$  is the highest number of concepts in a test item within the question bank.

$$R = \frac{1}{(n_{max} + 1)} - \frac{(1/n_{max} + 1)}{(n_{max} - 1)} \tag{4}$$

$$w_{pre} = \sum_{V_j \in in(V_i)} \left( \frac{1}{n_a} - \frac{(1/(n_{max} + 1))}{(n_a - 1)} \right) \times \frac{w_{ji}}{\sum_{V_k \in out(V_j)} w_{jk}} \times R \tag{5}$$

$$w_{suc} = \sum_{V_j \in out(V_i)} \left( \frac{1}{n_a} + \frac{(1/(n_{max} + 1))}{(n_a - 1)} \right) \times \frac{w_{ij}}{\sum_{V_k \in in(V_j)} w_{kj}} \times R \tag{6}$$

$$W(V_i)_{main_a} = \left( \frac{1}{n_a} + \frac{1}{(c_{max} + 1)} \right) + w_{pre} - w_{suc} \tag{7}$$

$$W(V_i)_{supp_a} = \left( \frac{1}{n_a} - \frac{1}{(c_{max} + 1)} \right) + w_{pre} - w_{suc} \tag{8}$$

Figure 7 illustrates the results obtained by applying equations (4) – (8). These equations help map the test items that students have failed to master in the formative test to the established weighted-directed graph, resulting in a P-TIRT. The P-TIRT serves as a reference for calculating the personalized error ratio, which provides insights into the specific areas where students are struggling based on their performance in the test items.

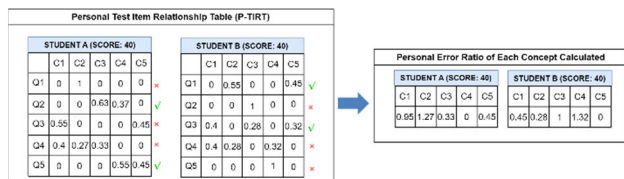


FIGURE 7. Illustration of mapping personalized test item relationship table to acquire personal error ratio.

2) GENERATING PERSONALIZED REMEDIAL LEARNING PATH

Once the weight values that indicate the degree of relationship between test items and concepts have been identified using the proposed autogenerated weighting method, these values

are stored in a Test Item–Concept Relationship Table (TIRT). In the calculation of a student’s error ratio for a concept  $C_i$ , the TIRT ( $Q_a, C_i$ ) values that represent the degree of relationship between the test item  $Q_a$  and the concept  $C_i$  are used. The error ratio (ER) of a student for a concept  $C_i$  is obtained by dividing the sum of TIRT ( $Q_a, C_i$ ) values for all incorrectly answered test items by the sum of TIRT ( $Q_a, C_i$ ) values of all the test items. If the resulting error rate is  $ER(C_i) \leq \theta$ , then the student is considered to have mastered the concept  $C_i$ , where  $\theta$  is a predetermined threshold. Conversely, if  $ER(C_i) > \theta$ , then the student is considered to have failed to master the concept  $C_i$ . Based on the concepts that the student failed to master and the established weighted-directed graph model, a suitable learning path is generated for the student, in which the sequence of concepts to be learned is recommended, as shown in Figure 8.

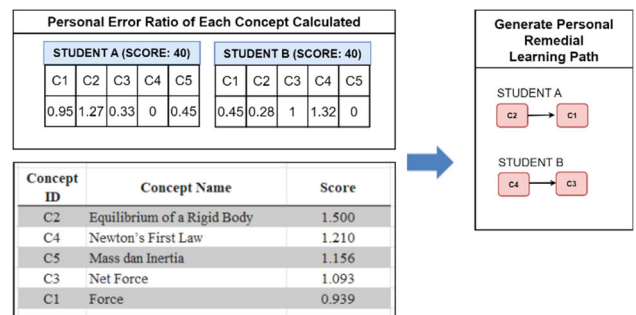


FIGURE 8. Illustration of generating personal remedial learning path.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In the experiment to validate the proposed approach, we collaborated with three experienced high school physics teachers who have over 15 years of teaching experience. The experiment focused on the topic of the dynamics of translational motion based on the Indonesian high school curriculum. This experiment aimed to gather empirical data that demonstrated how the concept’s co-occurrence approach resolves conflicting cases and the level of importance assigned by experts to concept weights while also achieving significant improvements in the learning outcomes of remedial students.

The experimental results show that students were able to improve their learning outcomes by focusing on the concepts they had not fully mastered, using the PRLP as their guide. Despite the variation in the test items used in the remedial tests in comparison to the formative test, the experimental group showed a higher percentage of remedial students who achieved passing grades when compared to the control group. This indicates that the implementation of the PRLP positively impacted the students’ performance and assisted them in achieving higher scores on their assessments.

One interesting finding is that even the lowest score among the remedial students in the experimental group was above 20, and there were even some students who achieved remarkably high scores of 95. Subsection D of this section provides the details and comprehensive data. The multistage process relies



on the frequency of co-occurrence of concepts to generate a CER. This process generates a weighted-directed graph that represents the relationships between concepts. We determine the weights assigned to nodes and edges in the graph based on the frequency of concept occurrences in the test items from the entire question database.

**A. EXPERIMENT SCENARIO**

The participants of this experiment were tenth-grade students from a high school in Indonesia who were studying the topic of the dynamics of translational motion in their physics class. We divided them into two groups, namely, the experimental group and the control group. The average age of the students was 15 years, and there were a total of 121 participants. The only difference between the groups was that the experimental group received personalized learning recommendations in the form of personalized learning paths, whereas the control group did not receive any recommendations.

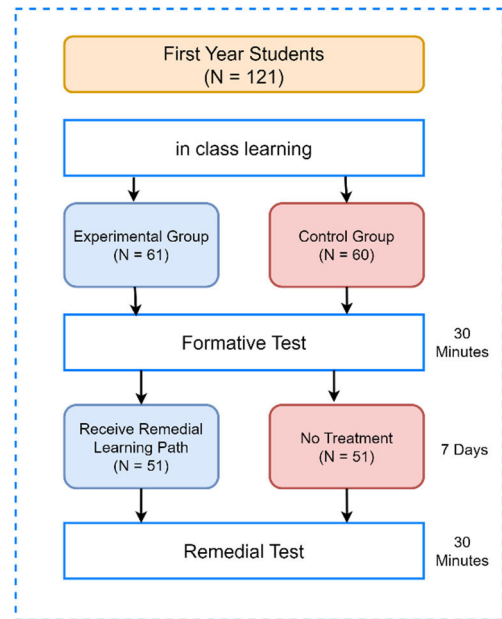
All participants were required to complete an online formative test within a time limit of 30 minutes. The test consisted of 20 items, including short-answer and multiple-choice questions, covering the 14 concepts related to the topic of the dynamics of translational motion in physics. The objectives of the formative test were to assess the students’ mastery level of the taught concepts and to determine whether they passed the physics subject. Additionally, the test results ensured that the experimental and control groups had similar levels of prerequisite knowledge and identified any learning challenges they may have encountered. We set the minimum passing grade for both tests at 70.

Students who scored below the minimum passing grade in both the experimental and control groups were required to take an online remedial test. The remedial test consisted of 20 newly generated test items in a format similar to the original formative test. This remedial test aimed to evaluate the effectiveness of the personalized learning paths generated using the proposed method. Students had 7 days to independently relearn and prepare for the test. During this preparation period, remedial students had unrestricted access to learning resources and practice test items provided to them. The domain experts played a role in ensuring the reliability and validity of the test items. Figure 9 illustrates the experimental scenario of this research.

**B. FORMATIVE TEST RESULTS ANALYSIS**

The objective of the formative test was to assess the students’ mastery level after completing the physics class. Table 4 presents the score statistics of the formative test for both the experimental and control groups. In the control group, out of 60 students, only nine passed the test. The average formative test score for the control group was 46.08, with a standard deviation of 18.69. The average score of the students who did not pass the formative test in the control group was 40.88, with a standard deviation of 14.75. In the experimental group, out of 61 students, only 10 passed the test. The average formative test score for the experimental group was

46.32, with a standard deviation of 17.60. Among the students who did not pass the formative test in the experimental group, the average score was 40.98, with a standard deviation of 14.07.



**FIGURE 9. Experimental scenario.**

Before carrying out the formative test, the students were randomly assigned to the experimental and control groups. To ensure that there were no significant statistical differences in the students’ prerequisite knowledge between the two groups, a *T*-test and an *F*-test were carried out to compare the population means and variances, respectively, using the formative test results. The results of the *F*-test showed that the variance in test scores between the experimental and control groups was nearly identical, with a *p*-value of 0.32 (>0.05). Moreover, the results of the *T*-test indicated that there was no significant difference in the mean scores between the two groups, with a *p*-value of 0.48 (>0.05). These findings suggest that students in both groups had similar levels of prerequisite knowledge, which allows for a fair evaluation of the proposed method.

**TABLE 4. Formative test score statistics.**

Formative Test				
	Mean	SD	<i>t</i> Stat	<i>t</i> Critical
Control Group ( <i>N</i> = 60)	46.08	18.69	0.04	1.66
Experiment Group ( <i>N</i> = 61)	46.23	17.60		
Underachieving Students Control Group ( <i>N</i> = 51)	40.88	14.75	0.03	0.49
Underachieving Students Experiment Group ( <i>N</i> = 51)	40.98	14.07		

**C. REMEDIAL TEST RESULTS ANALYSIS**

After the formative test, the students from both the experimental and control groups who did not pass were required to take the remedial test. Table 5 provides the remedial test score statistics for both groups. A total of 51 students from the control group and 51 students from the experimental group participated in the remedial test. The mean score for the control group on the remedial test was 46.08, with a standard deviation of 13.84. Contrarily, the experimental group achieved a higher mean score of 65.27, with a standard deviation of 11.59. The significant difference in mean scores between the experimental and control groups suggests that the proposed method had a positive impact on the students’ performance, as reflected in their remedial test scores. The results shown in Table 6 highlight the impact of the PRLP on the students’ learning achievement. In the experimental group, the mean test score significantly increased from 40.98 in the formative test to 65.27 in the remedial test. Conversely, the control group experienced a smaller increase, with the mean test score rising from 40.88 in the formative test to 44.29 in the remedial test. These findings show that the personalized learning paths produced through the proposed concept’s co-occurrence approach were highly beneficial for the students in enhancing their learning outcomes. The significant increase in the mean test score for the experimental group shows that the PRLP recommendations helped the students focus on and master the concepts they had previously struggled with, which led to higher achievement levels.

**TABLE 5. Remedial test score statistics.**

	Remedial Test			
	Mean	SD	t Stat	t Critical
Control Group (N = 51)	44.29	13.84	8.30	1.66
Experiment Group (N = 51)	65.27	11.59		

**TABLE 6. Statistical comparison of the underachieving students’ formative and remedial test scores.**

Group	Number of Students	Formative Test		Remedial Test		t
		Mean	SD	Mean	SD	
		Underachieving Students Control Group	51	40.88	14.75	
Underachieving Students Experimental Group	51	40.98	14.07	65.27	11.59	12.81

**D. LEARNING ACHIEVEMENT**

The remedial students in each group were divided into two clusters based on their knowledge level to further analyze the learning achievement of the underachieving students who did not pass the formative test. This division allowed for a more focused analysis of the students’ performance and the

effectiveness of the intervention. The first cluster consists of students with test scores of <55, whereas the second cluster consists of students with test scores of ≥55. Table 7 presents a comparison of the statistics for the remedial test scores between the experimental and control groups within these two knowledge clusters. The data in Table 7 show that among students with a remedial score of <55, the experimental group exhibited a higher mean score (46.44) and the lowest score (33) compared to the control group, which had a mean score of 35.77 and the lowest score of 10. Furthermore, when examining students with a remedial score ≥ 55, it is evident from Table 7 that the experimental group had a higher mean score (69.31) and the highest score (95) compared to the control group, which had a mean score of 57.92 and the highest score of 70. These findings propose that the implementation of the PRLP assisted the enhancement of the learning outcomes of underachieving students.

**TABLE 7. Statistical comparison of the remedial scores between the two groups in two different knowledge levels.**

Knowledge Level	Group	N	Min Score	Max Score	Mean	SD	t Stat	t Cri
<55	Control	36	10	53	35.77	11.78	2.63	1.68
	Experimental	9	33	50	46.44	5.59		
≥55	Control	15	55	70	57.92	5.48	4.81	1.67
	Experimental	42	55	95	69.31	7.96		

Figure 10 presents the formative and remedial test scores of each underachieving student in both the control and experimental groups. In Figure 10, several students in the control group obtained a lower score on the remedial test. This observation highlights the importance of implementing a PRLP to facilitate optimal preparation for the remedial test, underscoring the necessity of such a guidance system. Moreover, Figure 10 reveals that all students in the experimental group were able to improve their scores on the remedial test. The results further support the effectiveness of the PRLP, which was developed using the proposed autogenerated weighting by concept co-occurrence method, in providing students with learning guidance and improving their academic performance.

Additionally, we analyzed the learning achievement of underachieving students from the perspective of five distinct knowledge levels. Table 8 shows a comparison of the mean scores for the formative and remedial tests among underachieving students in both the control and experimental groups across these five knowledge levels. In Table 8, the number of students in the experimental group scoring <40 decreased from 26 (50.98%) in the formative test to 1 (1.96%) in the remedial test. Conversely, the number of students in the control group scoring <40 decreased from 25 (49.01%) in the formative test to 19 (37.25%) in the remedial test. The percentage of students with a score of <40 in the experimental group significantly decreased by 49.02%, whereas the control group only decreased by 11.76%. Table 8 also

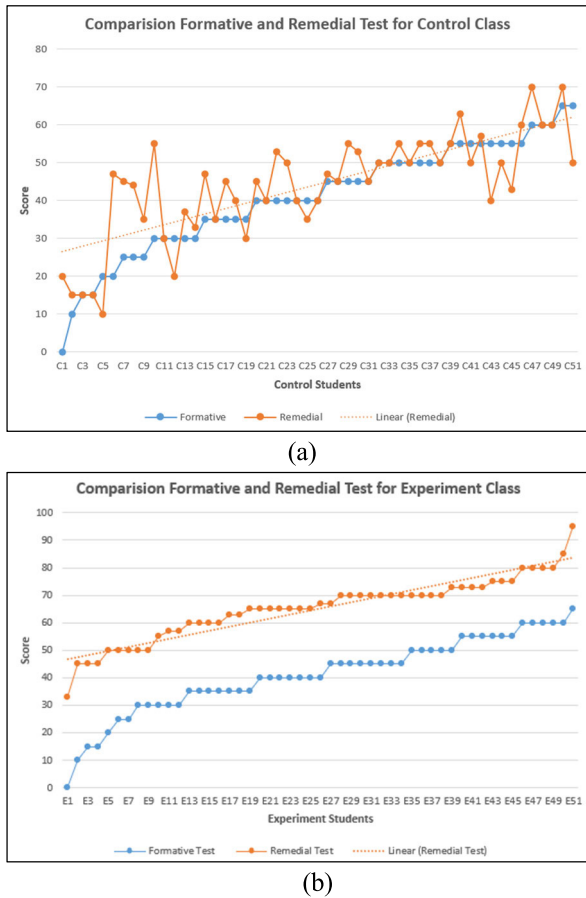


FIGURE 10. Comparison between formative (a) and remedial tests (b) of the control and experimental groups.

provides an interesting observation in terms of the performance of students in the experimental and control groups with scores of  $\geq 60.5$ . In the experimental group, the number of students achieving a score of  $\geq 60.5$  increased from 3 (5.88%) in the formative test to 35 (68.62%) in the remedial test, which indicates a significant increase of 62.74%. Alternatively, in the control group, the number of students with a score of  $\geq 60.5$  increased from 2 (3.92%) in the formative test to 3 (5.88%) in the remedial test, representing a slight increase of 1.96%. Moreover, notably, in the remedial test, two students in the experimental group were able to achieve exceptionally high scores of  $>80.5$ , whereas none of the control group reached such a level. This finding further reinforces the efficacy of the PRLP developed using the proposed autogenerated weighting method in substantially enhancing students’ learning achievement.

Based on the experiment results, it can be concluded that implementing a PRLP as a learning guide has proven to be highly beneficial in significantly improving students’ learning achievement. Moreover, the proposed autogenerated weighting method efficiently addresses the limitations of manually assigning weight values to indicate the relationship between test items and concepts. The use of the autogenerated

TABLE 8. Comparison of formative and remedial test scores of the control and experimental group in five different knowledge levels.

Group	Test	Mean Score of Knowledge Level					Score Range	
		<20.5	20.5–40.5	40.5–60.5	60.5–80.5	>80.5	Min	Max
Control	Formative Test	13.33	34.00	51.96	65	–	10	65
	Number of Students	6	19	24	2	–		
	Remedial Test	15.83	36.25	50.87	67.67	–	15	70
	Number of Students	4	14	30	3	–		
Experimental	Formative Test	12	34.52	51.67	65	–	0	65
	Number of Students	5	21	24	1			
	Remedial Test	–	33	52.93	70.36	90	33	95
	Number of Students	0	1	15	33	2		

weighting method significantly reduces the time required for establishing weight values, eliminating the requirement to resolve weight conflicts among experts. Additionally, the experiment results provide evidence of the effectiveness of the proposed concept’s co-occurrence approach in substantially enhancing the learning achievement of underachieving students.

V. DISCUSSION

A. COMPARISON WITH PREVIOUS STUDIES

Remedial students must study several concepts, and it is important to discover the weight of each concept. Hence, it is possible to determine which concepts are more important to learn as a basis for mastering other concepts in a learning process so that our research emphasizes the concept’s weight and their effect on the relationships between concepts. Then, we started to collect the results of various research related to CER methodology to compare which method was most appropriate, which included the research by Hwang et al. (2003), Panjaburee et al. (2013), and Wanichsan et al. (2021), as described thoroughly in Section II. Unfortunately, we found a typical problem in the previous research, that is, the process of identifying each related concept in the test items and determining the concept’s weights was conducted manually by domain experts, who are mostly rigid and tend to be subjective, so debates often occur between experts in determining weights. The impact is that there are inconsistencies in concept weighting, the effect of which is difficulty in building CER as a reference for designing learning paths for remedial students. Our research replaces, or at least dramatically reduces, the role of the expert by employing a deep learning architecture to analyze the concept co-occurrence in the test item in the question bank. Hence, in comparison to the high dependency on experts’ existence, our proposed method is automated, less time-consuming, but still consistent and reliable; it is even possible to apply it to other subjects that use natural language in learning and emphasize the importance of

understanding the relationships between concepts to achieve mastery.

### B. IMPLICATIONS TO ASSESSMENT-BASED LEARNING METHODOLOGY

Generally, the objectives of assessment-based learning methodologies are to ensure that students acquire knowledge and skills and demonstrate their understanding and competence through various assessment formats. One of the assessment formats for learning a subject is a formative test, which contains a set of test items taken from a database of test items or a question bank. A student is encouraged to analyze and evaluate information and to make connections by exploring the relationships between concepts within a single subject or between different subjects, since students with a conceptual understanding of the subject learned would be more knowledgeable than those who only know theoretical facts and procedures. Hence, the concept is essential in education, as in this research, when constructing both individual and adaptive PRLPs by mapping the test items that students have failed to master in the formative test to the established CER or weighted-directed graph, resulting in a P-TIRT. The P-TIRT serves as a reference for calculating the personalized error ratio, which provides insights into the specific areas where an individual student struggles based on the student’s performance in the formative test. By doing so, the most suitable remedial learning path for students to fulfill their learning needs would be established, and recommendations for the sequence of concepts to learn would improve learning outcomes.

### C. LIMITATIONS AND CHALLENGES

Our research has proven that the proposed method is more adaptive and practical in generating the learning path for remedial students. However, we acknowledge that there is still room for improvement due to the limitations we encountered during our research, including:

- The proposed method works best in subjects that require a relationship between concepts, for example, science and math.
- The ability of the applied NLP solutions to process text data that employs a local native language might require continuous adjustment due to the impact on the result of the multilabel classification of test items, considering the vocabulary used in composing a test item.
- Manually assigning a new set of random test items from the question bank in the exercise sessions was still being performed. It would be more efficient in an assessment-based learning environment to apply automation.

Several challenges that encourage us to elaborate further in the future are as follows:

- When a student has difficulty mastering a concept during the learning process, it may not merely be because of a lack of understanding of the prerequisite concepts.

However, there are external factors that may be considered physical conditions (e.g., learning environment, resources and materials, social factors, health and well-being, assessment and feedback, teaching methods, and student characteristics), which undeniably make learning a nonlinear process in terms of learning durability and assessment results. Therefore, utilizing such factors to predict probabilities of success or failure in enhancing learning outcomes as a complement to the provided personalized learning path would be beneficial. Unfortunately, it is difficult for methods based on physical or data-driven models to fully characterize this nonlinear process, and existing methods that hybridize physical and data-driven models suffer from ambiguous hybridization, which results in the vast majority of existing methods for predicting learning durability and assessment results suffering from a lack of accuracy and robustness [28]. Those conditions should have triggered the ideas to employ deep learning models, such as convolutional neural networks (CNNs), graph neural networks (GNN), long and short-term memory networks (LSTM), bidirectional LSTM, and hybrid neural networks, which can automatically learn complex feature representations of the system from large amounts of data to better capture the state evolution patterns of the system at different time steps [29].

- Currently, establishing the concepts still depends on the teaching activities in the classroom. It would be beneficial if the concepts were generated automatically from the learning material in the student’s textbook, so they would follow the textbook improvement adaptively.
- The application of the proposed method should be expanded to other subjects in physics and to other subjects from cross-disciplines as supporting concepts, such as math, chemistry, electronics, and others.

### D. FUTURE RESEARCH DIRECTIONS

Due to the limitations and challenges following the accomplishment of our research, the following are some suggestions that will serve as our contributions to the ongoing discourse in the field and to paving the avenues for future research:

- In the context of implementation to other related topics in physics or even solely different subjects, it is more than applicable if the other subject also emphasizes the importance of understanding concepts and the relationship between concepts to achieve mastery, such as concepts in biology, chemistry, and math, of all students within a course and generates a personalized remedial learning path for remedy students. The module would also be equipped with an automatic quiz generation module that selects questions from the question bank based on the personal remedial learning path. Contrariwise, in the context of scalability to other subjects, it would be more valuable if our proposed method could



build CER from multi-concepts across subjects. For example, we noticed that the concept of trigonometry in math is closely related to comprehending the concepts of dynamics in physics, so it would be more effective if our proposed method were able to analyze the test items in the same question bank and then generate the CER upon the concepts both in math and in physics, either separately or integrated. Another prior knowledge from other disciplines that will be useful to explore to mitigate the knowledge barriers is the semantic skills, either grammatical or lexical, in order to grasp the relationships between words, understand context, and recognize the subtle nuances in meanings of the concept in the textbook or a test item, particularly understanding a word from a local language that has several different definitions. Regarding the customization to different subjects, we believe that it depends on how significantly it will impact the implementation of this proposed method on the subject. Similar to our research, when we need to measure the effectiveness of an established PRLP, we can customize the passing grade in the phase of the exercise sessions and remedial test.

- Generating a weighted-directed graph as the relationships between concepts integrate expert knowledge and system data. However, expert knowledge is difficult to use directly [30], and the fusion method of expert knowledge and data is not appropriate and affects the control effect [31]. As previously mentioned, one of the desired results achieved in our research is eliminating or at least minimizing the expert intervention in generating CER for remedial students and replacing it with a deep learning algorithm to autogenerate CER based on the frequency of the concept co-occurrence. Through these automation processes, we also aim to eliminate the consensus problem that involves obtaining all nodes, representing concepts, in the system to agree on a single value or decision. In this case, there is an opportunity to explore the implementation of the Fault Tolerance Consensus (FTC), a designed mechanism in distributed systems that ensures that the system can continue to function correctly and reach consensus among nodes, even in the presence of faulty nodes, which include node crashes, so the system can continue to function correctly despite any faults.

## VI. CONCLUSION AND SUGGESTIONS FOR FUTURE WORK

This study proposes an autogenerated weighting method that employs the frequency of concept co-occurrence in a test item bank to determine the degree of relationship between test items and subject concepts. The frequency matrix is constructed according to the occurrence of concepts in the test items, and a concept ranking algorithm is applied to determine the most frequently appearing concepts. By obtaining the order of concept indices, the weights of concept relationships are calculated through normalization based on the

frequency of occurrence. This process results in the creation of a CER, which is a weighted and directed graph used to map personalized profiles for remedial students. The CER serves as a reference to determine a personal error ratio, which is then applied to define the sequence of the PRLP.

The proposed approach successfully addresses the issue of conflicts arising from assigning different weight values by experts, which results in reliable weight values. By employing the concept’s co-occurrence-based CER method in the development of a PRLP, a CER model is obtained. This model incorporates adaptive weight values on both the nodes and the relationships between nodes, which accommodates changes in the composition of the test items within the question bank. The learning path sequence is no longer fixed based on an initial learning sequence but rather adaptive, prioritizing concepts that occur most frequently in the test items. Each remedial student’s learning path is customized based on their profiling results. This personalized approach effectively helps students enhance their learning achievement in the remedial test. The learning paths are designed by determining concepts that students have yet to master and providing them with the necessary foundational knowledge. As a result, students acquire the necessary knowledge to achieve better results, even when faced with different sets of test items.

During the research process, we manually conducted the compilation of the frequency concept co-occurrence matrix, which involved main concepts and supporting concepts in multilabel classification. However, we recommend automating this process using machine learning modeling in future work to enhance its efficiency and accuracy. Despite the fact that our research has applied BERT (Bidirectional Encoder Representations), a machine learning model that, more specifically, falls under the category of natural language processing (NLP) models, the result of BERT is still processed manually to construct the co-occurrence matrix. Therefore, we are optimistic that applying a simple modification to its algorithm or hardcode, such as inserting sum-up instructions, can directly convert all the keywords, as the result of pre-trained BERT on large amounts of textual data, to become the co-occurrence matrix in an automated sequential process, so it can further optimize the feasibility and reliability of the method.

The application of PRLPs can also be extended to develop assessment-based learning methods by integrating them with the test item database to make the proposed method more widely applicable and not limited to remedial learning scenarios since the characteristic of the proposed method is produced by analyzing the relationships between concepts and presented as a weighted-directed graph. To build a concept map to design learning paths for students in the learning process, it is possible to switch the approach of CER utilization to become antemortem, as it assists students to comprehend prior knowledge in a learning session and prepare themselves prior to the formative test. The critical success factor of this altered approach is the existence of a question bank,

as we also prepared for our research, which contains a large number of labeled test items as the database and is ready to be integrated with the proposed method to achieve the objectives of the assessment-based learning methodology that students acquire knowledge and skills and demonstrate their understanding and competence through various assessment formats. Furthermore, future research endeavors must aim to determine correlations between physics concepts and prerequisite knowledge in other subjects that may hinder the acquisition of mastery. For example, understanding mathematical concepts such as logarithms and graphs of sine, cosine, and tangent is crucial for comprehending certain physics concepts. Therefore, obtaining the prior knowledge's concept relationships between different disciplines will make the proposed method remain relevant or applicable over an extended period, and it will help address knowledge barriers between disciplines, hence providing a better guide to designing PRLPs.

## REFERENCES

- [1] S. Chookaew, P. Panjaburee, D. Wanichsan, and P. Laosinchai, "A personalized e-learning environment to promote student's conceptual learning on basic computer programming," *Proc. Social Behav. Sci.*, vol. 116, pp. 815–819, Feb. 2014, doi: [10.1016/j.sbspro.2014.01.303](https://doi.org/10.1016/j.sbspro.2014.01.303).
- [2] C. Wongwatkit, N. Srisawasdi, G.-J. Hwang, and P. Panjaburee, "Influence of an integrated learning diagnosis and formative assessment-based personalized web learning approach on students learning performances and perceptions," *Interact. Learn. Environments*, vol. 25, no. 7, pp. 889–903, Oct. 2017, doi: [10.1080/10494820.2016.1224255](https://doi.org/10.1080/10494820.2016.1224255).
- [3] H. Xie, H.-C. Chu, G.-J. Hwang, and C.-C. Wang, "Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2017," *Comput. Educ.*, vol. 140, Oct. 2019, Art. no. 103599, doi: [10.1016/j.compedu.2019.103599](https://doi.org/10.1016/j.compedu.2019.103599).
- [4] G.-J. Hwang, F.-R. Kuo, P.-Y. Yin, and K.-H. Chuang, "A heuristic algorithm for planning personalized learning paths for context-aware ubiquitous learning," *Comput. Educ.*, vol. 54, no. 2, pp. 404–415, Feb. 2010, doi: [10.1016/j.compedu.2009.08.024](https://doi.org/10.1016/j.compedu.2009.08.024).
- [5] M. M. Rahman and N. A. Abdullah, "A personalized group-based recommendation approach for web search in e-learning," *IEEE Access*, vol. 6, pp. 34166–34178, 2018, doi: [10.1109/ACCESS.2018.2850376](https://doi.org/10.1109/ACCESS.2018.2850376).
- [6] C. Wongwatkit, P. Panjaburee, N. Srisawasdi, and P. Seprum, "Moderating effects of gender differences on the relationships between perceived learning support, intention to use, and learning performance in a personalized e-learning," *J. Comput. Educ.*, vol. 7, no. 2, pp. 229–255, Jun. 2020, doi: [10.1007/s40692-020-00154-9](https://doi.org/10.1007/s40692-020-00154-9).
- [7] G.-J. Hwang, H.-Y. Sung, S.-C. Chang, and X.-C. Huang, "A fuzzy expert system-based adaptive learning approach to improving students' learning performances by considering affective and cognitive factors," *Comput. Educ., Artif. Intell.*, vol. 1, Aug. 2020, Art. no. 100003, doi: [10.1016/j.caeai.2020.100003](https://doi.org/10.1016/j.caeai.2020.100003).
- [8] J. Gao, Q. Liu, and W.-B. Huang, "Learning path generator based on knowledge graph," in *Proc. 12th Int. Conf. E-Education, E-Bus., E-Manag., E-Learn*. New York, NY, USA: ACM, Jan. 2021, pp. 27–33, doi: [10.1145/3450148.3450155](https://doi.org/10.1145/3450148.3450155).
- [9] E. Subiyantoro, A. Ashari, and S. Suprpto, "Learning path recommendation using hybrid particle swarm optimization," *Adv. Sci., Technol. Eng. Syst. J.*, vol. 6, no. 1, pp. 570–576, Jan. 2021, doi: [10.25046/aj060161](https://doi.org/10.25046/aj060161).
- [10] G.-J. Hwang, "A conceptual map model for developing intelligent tutoring systems," *Comput. Educ.*, vol. 40, no. 3, pp. 217–235, Apr. 2003, doi: [10.1016/S0360-1315\(02\)00121-5](https://doi.org/10.1016/S0360-1315(02)00121-5).
- [11] P. Panjaburee, G.-J. Hwang, W. Triampo, and B.-Y. Shih, "A multi-expert approach for developing testing and diagnostic systems based on the concept-effect model," *Comput. Educ.*, vol. 55, no. 2, pp. 527–540, Sep. 2010, doi: [10.1016/j.compedu.2010.02.015](https://doi.org/10.1016/j.compedu.2010.02.015).
- [12] D. Wanichsan, P. Panjaburee, and S. Chookaew, "Enhancing knowledge integration from multiple experts to guiding personalized learning paths for testing and diagnostic systems," *Comput. Educ., Artif. Intell.*, vol. 2, Jan. 2021, Art. no. 100013, doi: [10.1016/j.caeai.2021.100013](https://doi.org/10.1016/j.caeai.2021.100013).
- [13] N. Komalawardhana, P. Panjaburee, and N. Srisawasdi, "A mobile game-based learning system with personalised conceptual level and mastery learning approach to promoting students' learning perceptions and achievements," *Int. J. Mobile Learn. Organisation*, vol. 15, no. 1, p. 29, 2021, doi: [10.1504/IJMLO.2021.111596](https://doi.org/10.1504/IJMLO.2021.111596).
- [14] Y. Wahyuningsih and A. Djunaidy, "Framework blended learning personalisasi scaffolding adaptif LMS (PSALMS) pada pelajaran matematika," *Seminar Nasional Teknologi Informasi Komunikasi*, vol. 1, no. 1, pp. 298–309, Oct. 2018.
- [15] Y. Wahyuningsih, A. Djunaidy, and D. O. Siahaan, "Remedial directed topic map on personalized scaffolding adaptive learning management system remedial directed topic map on personalized scaffolding adaptive learning management system," in *Proc. 1st Int. Conf. Comput. Mach. Intell. (ICMI)*, vol. 1, no. 1, Istanbul, Turkey, Feb. 2021, pp. 277–281. [Online]. Available: [https://icmi.aiplustech.org/icmi21\\_home.html](https://icmi.aiplustech.org/icmi21_home.html)
- [16] Y. Wahyuningsih, A. Djunaidy, and D. O. Siahaan, "A method comparison on multi-label questions classification for assessment-based personalised scaffolding adaptive learning path," in *Proc. 3rd Int. Conf. Electr. Eng. Informat. (Icon EEI)*, Oct. 2022, pp. 162–167, doi: [10.1109/IconEEI55709.2022.9972269](https://doi.org/10.1109/IconEEI55709.2022.9972269).
- [17] M. Niknam and P. Thulasiraman, "LPR: A bio-inspired intelligent learning path recommendation system based on meaningful learning theory," *Educ. Inf. Technol.*, vol. 25, no. 5, pp. 3797–3819, Sep. 2020, doi: [10.1007/s10639-020-10133-3](https://doi.org/10.1007/s10639-020-10133-3).
- [18] G. Durand, N. Belacel, and F. LaPlante, "Graph theory based model for learning path recommendation," *Inf. Sci.*, vol. 251, pp. 10–21, Dec. 2013, doi: [10.1016/j.ins.2013.04.017](https://doi.org/10.1016/j.ins.2013.04.017).
- [19] H. Zhu, F. Tian, K. Wu, N. Shah, Y. Chen, Y. Ni, X. Zhang, K.-M. Chao, and Q. Zheng, "A multi-constraint learning path recommendation algorithm based on knowledge map," *Knowl.-Based Syst.*, vol. 143, pp. 102–114, Mar. 2018, doi: [10.1016/j.knsys.2017.12.011](https://doi.org/10.1016/j.knsys.2017.12.011).
- [20] C.-L. Bian, D.-L. Wang, S.-Y. Liu, W.-G. Lu, and J.-Y. Dong, "Adaptive learning path recommendation based on graph theory and an improved immune algorithm," *KSH Trans. Internet Inf. Syst.*, vol. 13, no. 5, pp. 2277–2298, May 2019, doi: [10.3837/tiis.2019.05.003](https://doi.org/10.3837/tiis.2019.05.003).
- [21] D. Shi, T. Wang, H. Xing, and H. Xu, "A learning path recommendation model based on a multidimensional knowledge graph framework for e-learning," *Knowl.-Based Syst.*, vol. 195, May 2020, Art. no. 105618, doi: [10.1016/j.knsys.2020.105618](https://doi.org/10.1016/j.knsys.2020.105618).
- [22] D. Wanichsan, P. Panjaburee, P. Laosinchai, W. Triampo, and S. Chookaew, "A majority-density approach to developing testing and diagnostic systems with the cooperation of multiple experts based on an enhanced concept-effect relationship model," *Expert Syst. Appl.*, vol. 39, no. 9, pp. 8380–8388, Jul. 2012, doi: [10.1016/j.eswa.2012.01.182](https://doi.org/10.1016/j.eswa.2012.01.182).
- [23] L.-H. Chen, "Enhancement of student learning performance using personalized diagnosis and remedial learning system," *Comput. Educ.*, vol. 56, no. 1, pp. 289–299, Jan. 2011, doi: [10.1016/j.compedu.2010.07.015](https://doi.org/10.1016/j.compedu.2010.07.015).
- [24] G. Musumba and R. Wario, "Towards a personalized adaptive remedial e-learning model," in *Proc. IST-Africa Week Conf. (IST-Africa)*, May 2019, pp. 1–11, doi: [10.23919/ISTAFRICA.2019.8764848](https://doi.org/10.23919/ISTAFRICA.2019.8764848).
- [25] Y. Keshun, W. Chengyu, and L. Huizhong, "Research on intelligent implementation of the beneficiation process of shaking table," *Minerals Eng.*, vol. 199, Aug. 2023, Art. no. 108108, doi: [10.1016/j.mineng.2023.108108](https://doi.org/10.1016/j.mineng.2023.108108).
- [26] D. Siahaan, A. Maududie, Slamim, A. Subekti, Prihandoko, and H. Siringoringo, "A new approach for modeling researcher community based on scientific article metadata using natural language processing," in *Proc. 7th Int. Conf. Inf. Sci. Technol. (ICIST)*, Apr. 2017, pp. 386–390, doi: [10.1109/ICIST.2017.7926790](https://doi.org/10.1109/ICIST.2017.7926790).
- [27] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Comput. Netw. ISDN Syst.*, vol. 30, nos. 1–7, pp. 107–117, Apr. 1998, doi: [10.1016/S0169-7552\(98\)00110-X](https://doi.org/10.1016/S0169-7552(98)00110-X).
- [28] Y. Keshun, Q. Guangqi, and G. Yingkui, "Remaining useful life prediction of lithium-ion batteries using EM-PF-SSA-SVR with gamma stochastic process," *Meas. Sci. Technol.*, vol. 35, no. 1, Jan. 2024, Art. no. 015015, doi: [10.1088/1361-6501/acfbef](https://doi.org/10.1088/1361-6501/acfbef).
- [29] Y. Keshun, Q. Guangqi, and G. Yingkui, "Optimizing prior distribution parameters for probabilistic prediction of remaining useful life using deep learning," *Rel. Eng. Syst. Saf.*, vol. 242, Feb. 2024, Art. no. 109793, doi: [10.1016/j.res.2023.109793](https://doi.org/10.1016/j.res.2023.109793).

- [30] Z. Feng, Z. Zhou, C. Hu, X. Ban, and Y. Zhang, “A fault diagnosis and tolerant control method for aerospace relay with environmental disturbance,” *Scientia Sinica Informationis*, vol. 51, no. 4, p. 648, Apr. 2021, doi: [10.1360/SSI-2020-0189](https://doi.org/10.1360/SSI-2020-0189).
- [31] Q. Hou and J. Dong, “Robust adaptive event-triggered fault-tolerant consensus control of multiagent systems with a positive minimum interevent time,” *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 53, no. 7, pp. 4003–4014, Jul. 2023, doi: [10.1109/TSMC.2023.3238709](https://doi.org/10.1109/TSMC.2023.3238709).



**YULIA WAHYUNINGSIH** received the master’s degree in informatics from Institut Teknologi Sepuluh Nopember, Indonesia, in 2005, where she is currently pursuing the Ph.D. degree in computer science with the Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology. Her research interests include educational data mining, NLP, and artificial intelligence.



**ARIF DJUNAIDY** (Member, IEEE) received the B.Eng. degree in electrical engineering from Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia, and the M.Sc. and Ph.D. degrees in computer science from The University of Manchester, U.K. He is currently a University Professor with the Department of Information Systems, Faculty of Intelligent Electrical and Informatics Technology, ITS. He has published some articles in refereed journals and contributed to some international conferences. His research interests include data management, data and knowledge engineering, machine, and deep learning. He has been a member of the IEEE Computer Society, since 1992, and the Association for Computing Machinery (ACM), since 2002.



**DANIEL SIAHAAN** (Member, IEEE) received the master’s degree in software engineering from Technische Universiteit Delft, in 2002, and the Ph.D. degree in software engineering from Technische Universiteit Eindhoven, in 2004. He is currently a Professor with the Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. He has published more than 90 journal articles and conference papers related to software engineering. His research interests include software engineering, requirements engineering, and natural language processing.

...