Lin, Y.-C., & Huang, Y.-M. (2013). A Fuzzy-based Prior Knowledge Diagnostic Model with Multiple Attribute Evaluation. *Educational Technology & Society*, *16* (2), 119–136.

A Fuzzy-based Prior Knowledge Diagnostic Model with Multiple Attribute Evaluation

Yi-Chun Lin[1] and Yueh-Min Huang[1,2*]
[1]Department of Engineering Science, National Cheng Kung University, No. 1, Ta-Hsueh Road, Tainan 701, Taiwan, R.O.C. // [2]Department of Applied Geoinformatics, Chia Nan University of Pharmacy and Science, 60, Erh-Jen RD., Sec. 1, Jen-Te, Tainan 710, Taiwan, R.O.C. // jellyplum@gmail.com // huang@mail.ncku.edu.tw
[*]Corresponding author

(Submitted Novermber 03, 2011; Revised Februery 25, 2012; Accepted June 04, 2012)

ABSTRACT

Prior knowledge is a very important part of teaching and learning, as it affects how instructors and students interact with the learning materials. In general, tests are used to assess students' prior knowledge. Nevertheless, conventional testing approaches usually assign only an overall score to each student, and this may mean that students are unable to understand their own specific weaknesses. To address this problem, previous work has presented a prior knowledge diagnosis model with a single attribute to assist instructors and students in diagnosing and strengthening prior knowledge. However, this model neglects the fact that a diagnostic decision might involve multiple attributes. In order to provide more a precise diagnosis to instructors and students, this study thus proposes a fuzzy prior knowledge diagnosis model with a multiple attribute decision making technique for diagnosing and strengthening students' prior knowledge. The experimental results from an interdisciplinary bioinformatics course have demonstrated the utility and effectiveness of this innovative approach.

Keywords

Fuzzy multiple attribute decision making, Prior knowledge diagnosis, Interdisciplinary course, Computer-assisted testing

Background and objectives

Evaluating and strengthening the prior knowledge of individual students is an important task before teaching and learning new knowledge or skills, since prior knowledge affects how instructors and students interact with the learning materials they encounter (Chieu, 2007; Moos & Azevedo, 2008; Ozuru, Dempsey, & McNamara, 2009). From the perspective of instructors, gaps in the students' prior knowledge often confound their best efforts to deliver effective instructions (Roschelle, 1995). Moreover, they can also affect how instructors plan their teaching strategies for new material in order to enhance students' learning motivation and performance (Biswas, 2007; Tseng, Chu, Hwang, & Tsai, 2008).

If the students do not have the necessary prior knowledge, then there is a strong risk that they may build new knowledge on faulty foundations (Dochy, Moerkerke, & Marten, 1996). It can thus be seen that inadequate or fragmented prior knowledge is an important issue, and if the instructors' expectations of the students' knowledge are very different from their actual knowledge, then both teaching and learning are likely to adversely affected (Hailikari, Katajavouri, & Lindblom-Ylanne, 2008).

To avoid this risk, tests are usually adopted to assess how well students understand a concept or piece of knowledge (Panjaburee, Hwang, Triampo, & Shih, 2010; Tao, Wu, & Chang, 2008; Treagust, 1988). Nevertheless, conventional testing systems usually assign only an overall score or grade to students, and thus instructors and students may be unable to identify which specific concepts or pieces of knowledge are misunderstood, making it difficult to improve the learning performance of students (Gerber, Grund, & Grote, 2008; Gogoulou, Gouli, Grigoriadou, Samarakou, & Chinou, 2007; Hwang, Tseng, & Hwang, 2008). To work around this issue, instructors can further analyze the testing results to determine the students' learning deficiencies. However, this is a time-consuming task that presents a heavy workload for instructors, since there are often many students on a course, especially in higher education or e-learning contexts. Hence, previous work has led to the development of a prior knowledge diagnosis (PKD) model to assist instructors and students in diagnosing and strengthening prior knowledge before new instruction is undertaken (Lin, Lin, Huang, 2011).

Nevertheless, one of the major problems when applying the PKD model is that it only uses correctness rates answered by students to determine their level of understanding with regard to particular concepts, and diagnoses based on a single attribute lead to inaccurate results (Hwang, Tseng, & Hwang, 2008). Therefore, it is necessary to develop a more effective approach to assist instructors in identifying the specific learning problems of individual students in the context of multiple attributes, and this issue actually matches a traditional computer science problem, called Multiple Attribute Decision Making (MADM) (Chen & Hwang, 1992). The MAMD problem is to select the best choice among the previously specified finite number of alternatives (Seel, & Dinter, 1995), with the alternatives evaluated based on their attributes.

Therefore, this study proposes a Fuzzy Prior Knowledge Diagnostic (FPKD) model by applying the Efficient Fuzzy Weighted Average (EFWA) technique (Lee & Park, 1997) to assist instructors in diagnosing the level of students' understanding of prior knowledge, and to provide appropriate feedback to individual students. Based on this model, a testing and diagnostic system has been implemented, and an experiment on an interdisciplinary bioinformatics course was conducted to demonstrate the efficacy of the proposed approach.

Fuzzy Prior Knowledge Diagnostic Model

The aim of the FPKD model is to assist instructors in diagnosing students' prior knowledge with multiple attributes. Figure 1 shows the hierarchical structure of the decision making problem and its criteria. To realize and diagnose the knowledge strength of students, instructors usually apply tests and consider the difficulty of the test items, the relevance of the concept, and the students' answers (Saleh & Kim, 2009).

As shown in Figure1, each criterion has its rating, r_i, which is associated with the measured value of the attribute. Furthermore, each criterion has been assigned a relative weight, w_i, which is used to adjust the weight of each criterion in relation to the decision goal.

In this study, the relative weight values of the three criteria are adjusted by the instructors according to different education contexts. Therefore, in order to develop the FPKD model, the ratings of each criterion first have to be retrieved and measured.

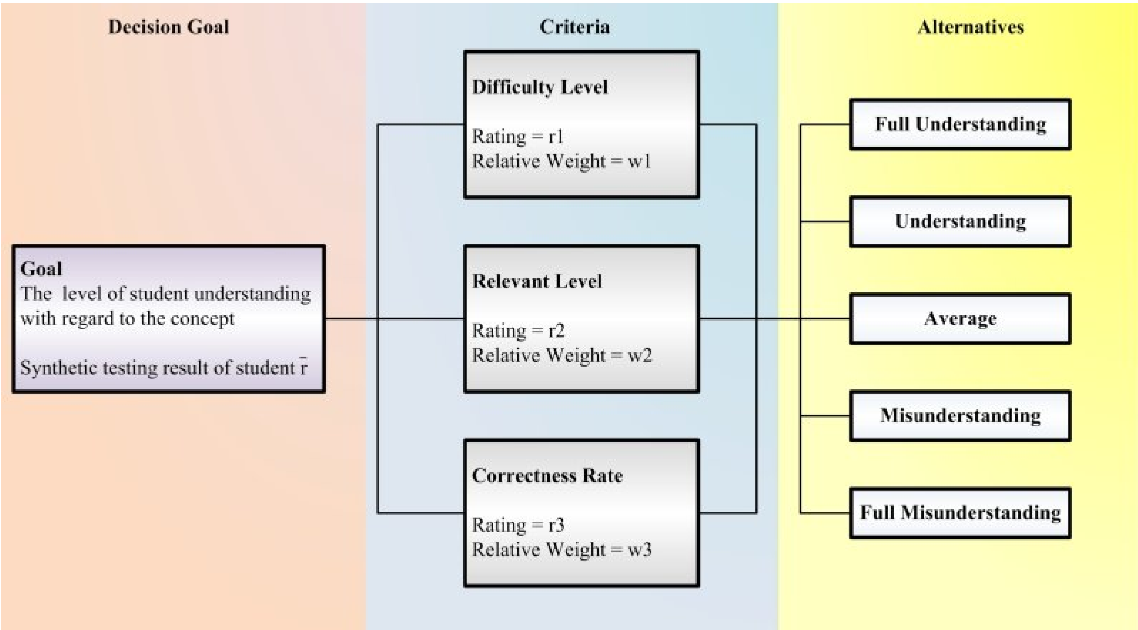

Figure 1. The hierarchical structure of the decision problem.

Analysis of decision attributes

The ratings of each criterion are retrieved from two data sources, the first is the testing information assigned by teachers, representing an association between each concept and test item, and the relationships among the concepts. The second is derived from students, which represents an association between their answers and the test items.

Assume that an instructor aims to teach a subject of a course, and the instructor specifies n concepts, C_1, C_2, C_3,…, C_i,…, C_n that are the requisite prior knowledge of the objective subject for r participating students, S_1, S_2, S_3,…, S_l,…, S_r. Before teaching the subject, the instructor selects k test items, I_1, I_2, I_3,…, I_j,…, I_k from a test item bank to form a pre-test, with k test items that possibly have different degree of difficulty , D_1, D_2, D_3,…, D_j,…, D_k.

In addition, each test item is relevant to n concepts, and each concept is possibly related to the others. The k test items and n concepts in the pre-test can be associated with each other, and the relationships among each concept can also be associated. After the initial setting of the test items, the instructor then conducts the pre-test to assess the r students' levels of understanding with regard to the n concepts using the proposed approach to measure their prior knowledge.

The test items are coded with a number ranging from one to k and each test item is relevant to from one to n concepts in the pre-test. To represent the degree of relevance between each concept and test item, an X-value is used. X_{ij} indicates the relevance between the i^{th} concept and the j^{th} test item. If the j^{th} test item is relevant to the i^{th} concept, X_{ij} is 1; otherwise X_{ij} is 0.

The concepts are coded with numbers ranging from one to n and each concept is relevant to from one to n concepts. To represent the relationships among the concepts, a Z-value is adopted that also ranges from 0 to 1. Z_{im} indicates the relationship between the i^{th} and the m^{th} concepts, $i, m \in n$.

After the r students have taken the test, their results for each item can be recorded. To represent the relationship between the students' answers and test items, an R-value is adopted using a binary coding scheme. R_{lj} indicates the answer of the l^{th} student for the j^{th} test item. If the student answers the test item correctly, then R_{lj} is 1; otherwise R_{lj} is 0.

Therefore, three assessment functions can be developed to measure the three decision attributes by using the above testing information. Firstly, based on the D, R, and X values, the highest difficulty level of concept C_i answered by student S_l correctly can be measured as:

$$HDL(S_{l,}C_i) = \max_{1 \le j \le k} \left\{ R_{lj} D_j X_{ij} \right\} \quad (1)$$

where $HDL(S_{l,}C_i)$ represents the highest difficulty level of i^{th} concept answered by l^{th} student correctly, $0 \le HDL(C_i) \le 1$; R_{lj} indicates the answer of the l^{th} student for the j^{th} test item, $R_{lj} \in \{0,1\}$; D_j represents the difficulty degree of j^{th} test item, $0 \le \mathrm{D}_j \le 1$; and X_{ij} indicates the relevance between the i^{th} concept and the j^{th} test item, $X_{ij} \in \{0,1\}$.

Furthermore, based on the R, X and Z values, the relevant level of concept C_i answered by student S_l correctly can be measured as:

$$RL(S_{l,}C_i) = \sum_{j=1}^{k} \left\{ R_{lj} \left[\frac{\sum_{w=1}^{n} x_{wj} z_{iw}}{\sum_{w=1}^{n} x_{wj} \left(\sum_{v=1}^{n} z_{wv} \right)} \right] \right\} / \sum_{j=1}^{k} R_{lj} \quad (2)$$

where $RL(S_{l,}C_i)$ represents the relevance level of i^{th} concept answered by l^{th} student correctly, $0 \le RL(S_{l,}C_i) \le 1$;

R_{lj} indicates the answer of the l^{th} student on the j^{th} test item, $R_{lj} \in \{0,1\}$; X_{ij} indicates the relevance between the i^{th} concept and the j^{th} test item, $X_{lj} \in \{0,1\}$; and Z_{iw} indicates the relationship between the i^{th} and the w^{th} concepts.

In addition, based on the R and X values, the correctness rate of student S_l with regard to concept C_i can be inferred as:

$$CR(S_{l,}C_i) = \frac{\sum_{j=1}^{k} R_{lj} X_{ij}}{\sum_{j=1}^{k} X_{ij}} \quad (3)$$

where $CR(S_{l,}C_i)$ represents the correctness rate of the l^{th} student with regard to i^{th} concept, $0 \le CR(C_i) \le 1$; X_{ij} indicates the relevance between the i^{th} concept and the j^{th} test item, $X_{ij} \in \{0,1\}$; and R_{lj} indicates the answer of the l^{th} student for the j^{th} test item, $R_{lj} \in \{0,1\}$.

Multiple attribute decision making algorithm of the FPKD model

After the rating measurements of the three criteria, the Efficient Fuzzy Weighted Average (EFWA) technique is used to produce the FPKD model. The criteria can be synthesized by *Equation* (4), and the fuzzy average $\bar{r}$ is produced based on the input criteria.

$$\bar{r} = \frac{\sum_{i=1}^{n} w_i r_i}{\sum_{i=1}^{n} w_i} \quad (4)$$

In the FPKD model, the $\bar{r}$ presents the synthetic testing result of the students with regard to a concept, and it is used to judge which alternative is appropriate to represent the students' level of understanding of the concept. Furthermore, to present the rating and relative weight of each criterion, two fuzzy membership functions are shown in Figure 2 and Figure 3, respectively. Note that a fuzzy membership function can be used to represent the extent to which a value from a domain is included in a fuzzy concept, such as "low relevance", ''high performance", and so on. In addition, each fuzzy concept can be represented in a formula form. For instance, the fuzzy concept ''Average" of Rating in Figure 2 (a triangular curve) can be mapped to the membership function of ''Average" of Rating in Table 1. In this study, the transformation between the triangular curve and the formula can be by solving the linear equations (Huang, Kuo, Lin, & Cheng, 2008).

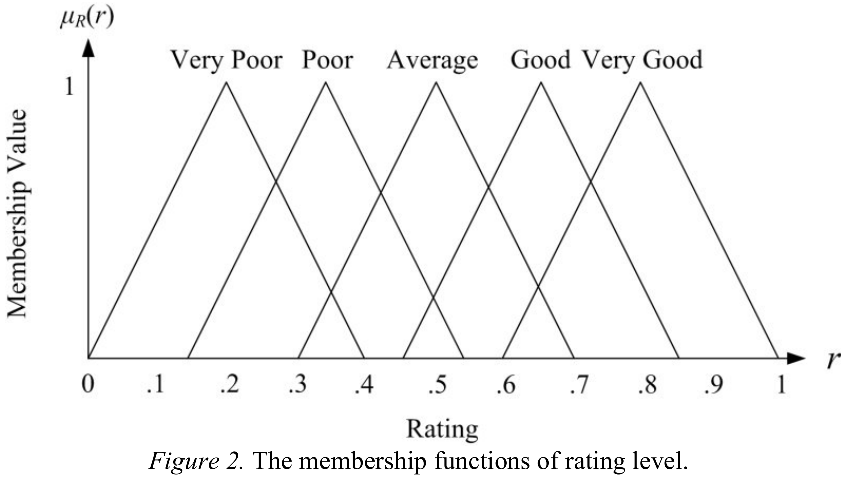

Figure 2. The membership functions of rating level.

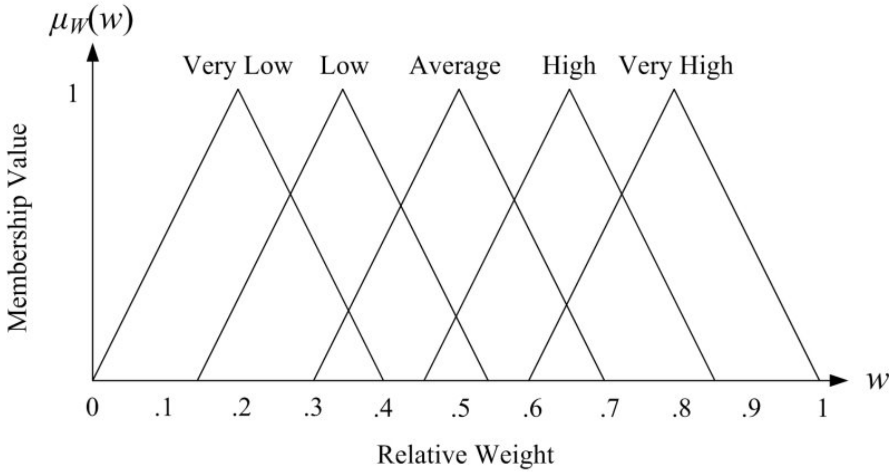

Figure 3. The membership function of relative weight.

Table 1. The definitions of the membership functions

Category	Class	Membership function
Alternative	Full Understanding	$=\begin{cases} 5a, a\in[0,0.2] \\ 2-5a, a\in[0.2,0.4] \end{cases}$
	Understanding	$=\begin{cases} 5a-0.75, a\in[0.15,0.35] \\ 2.75-5a, a\in[0.35,0.55] \end{cases}$
	Average	$=\begin{cases} 5a-1.5, a\in[0.3,0.5] \\ 3.5-5a, a\in[0.5,0.7] \end{cases}$
	Misunderstanding	$=\begin{cases} 5a-2.25a, a\in[0.45,0.65] \\ 4.26-5a, a\in[0.65,0.85] \end{cases}$
	Full Misunderstanding	$=\begin{cases} 5a-3, a\in[0.6,0.8] \\ 5-5a, a\in[0.8,1] \end{cases}$
Rating	Very poor	$=\begin{cases} 5r, r\in[0,0.2] \\ 2-5r, r\in[0.2,0.4] \end{cases}$
	Poor	$=\begin{cases} 5r-0.75, r\in[0.15,0.35] \\ 2.75-5r, r\in[0.35,0.55] \end{cases}$
	Average	$=\begin{cases} 5r-1.5, r\in[0.3,0.5] \\ 3.5-5r, r\in[0.5,0.7] \end{cases}$
	Good	$=\begin{cases} 5r-2.25, r\in[0.45,0.65] \\ 4.26-5r, r\in[0.65,0.85] \end{cases}$
	Very good	$=\begin{cases} 5r-3, r\in[0.6,0.8] \\ 5-5r, r\in[0.8,1] \end{cases}$

Relative weight	Very low	$=\begin{cases} 5w, w\in[0,0.2] \\ 2-5w, w\in[0.2,0.4] \end{cases}$
	Low	$=\begin{cases} 5w-0.75, w\in[0.15,0.35] \\ 2.75-5w, w\in[0.35,0.55] \end{cases}$
	Average	$=\begin{cases} 5w-1.5, w\in[0.3,0.5] \\ 3.5-5w, w\in[0.5,0.7] \end{cases}$
	High	$=\begin{cases} 5w-2.25, w\in[0.45,0.65] \\ 4.26-5w, w\in[0.65,0.85] \end{cases}$
	Very high	$=\begin{cases} 5w-3, w\in[0.6,0.8] \\ 5-5w, w\in[0.8,1] \end{cases}$

In the FPKD model, the alternatives are the concept understanding levels, which consist of five levels. The alternative requiring the highest synthetic testing result with regard to a concept is full understanding, and the lowest synthetic testing result is full misunderstanding. All the alternatives' membership functions are shown in Figure 4.

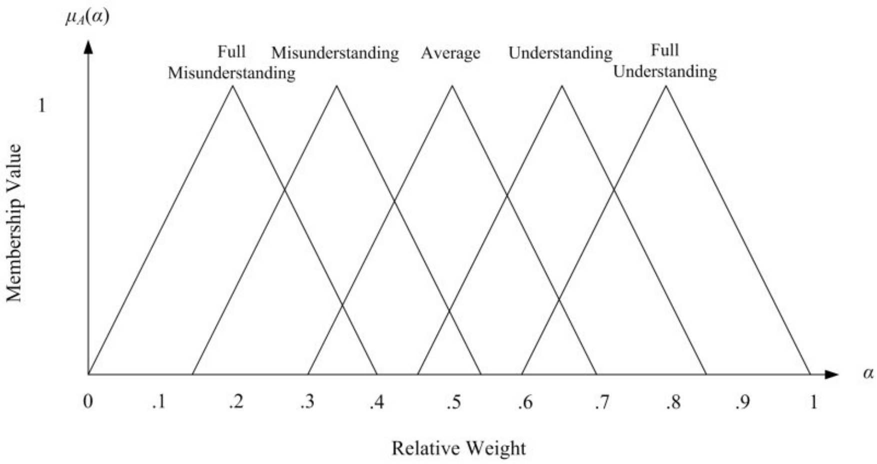

Figure 4. The membership functions of the alternatives.

Based on Figures 2, 3, and 4, Table 1 arranges the membership functions of rating levels, relative weights, and alternatives. The interval of each membership function in Table 1 is used by the EFWA for interval analysis to compute the result, $\bar{r}$. After obtaining the fuzzy weighted average, it then compares the distance between the weighted average and alternatives. The approximate Euclidean distance (Dobois & Prade, 1980; Ross, Sorensen, Savage, & Carson, 1990), as in *Equation* (5), is adopted as the measurement to determine the distance. In *Equation* (5), the parameter X represents the resulting fuzzy membership function ($\bar{r}$), the parameter A represents the pre-defined fuzzy membership function (alternatives), and the function d is the Euclidean distance, which presents the distance between X and A.

$$d(X,A)=\sqrt{(X^{\alpha=0}_{lower-bound}-A^{\alpha=0}_{lower-bound})^2+(X^{\alpha=1}-A^{\alpha=1})^2+(X^{\alpha=0}_{upper-bound}-A^{\alpha=0}_{upper-bound})^2} \quad (5)$$

Therefore, according to the decision goal, the appropriate solution is the alternative, that minimizes the Euclidean distance d. The calculation process is presented in Appendix, which includes an illustrative example to explain the entire decision making process in more detail.

Development of an FPKD-based testing system

Based on the FPKD model, a computer-assisted testing and diagnostic system is implemented in this work. A user-friendly interface is provided for instructors on the teacher side, in which they can select a specific course, subject, and concept to develop a diagnostic assessment, as shown in Figure 5. The system can then pick relevant test items from the test item bank according to the specific criteria. The instructors can thus use the interface to select the necessary test items to make a test-sheet based on their expertise. Figure 6 shows that the instructors can consult the assessment results and learning status of all students, and then use this information to improve their teaching plan before teaching a new course.

From the student side, students can log into the system and then use the student interface to take a diagnostic assessment, as shown in Figure 7. In addition, as shown in Figure 8, the system applies the FPKD model to diagnose the students' test results, and provides diagnostic results to each participant through a diagnostic interface that can clearly show the level of understanding of the students with regard to specific concepts, so that they can know what they need to pay more attention to.

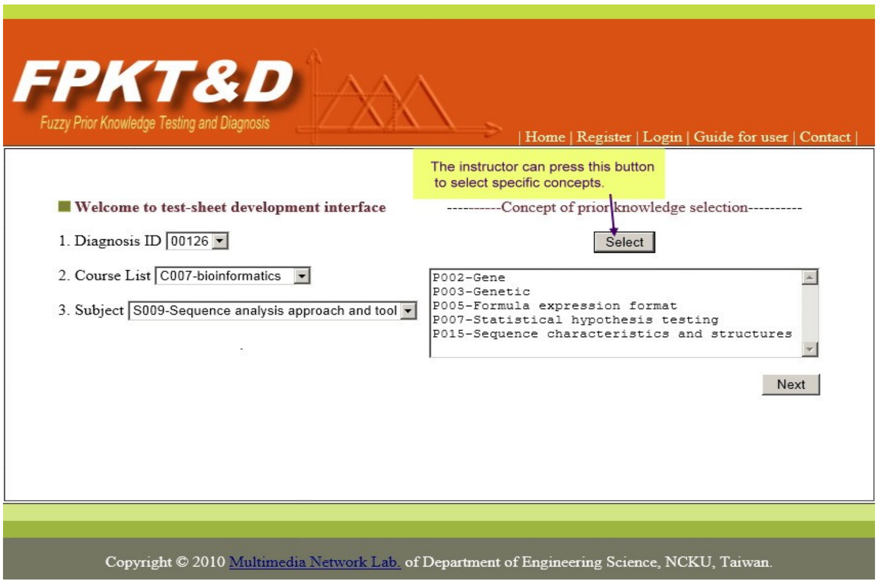

Figure 5. Screenshot of the test-sheet development interface.

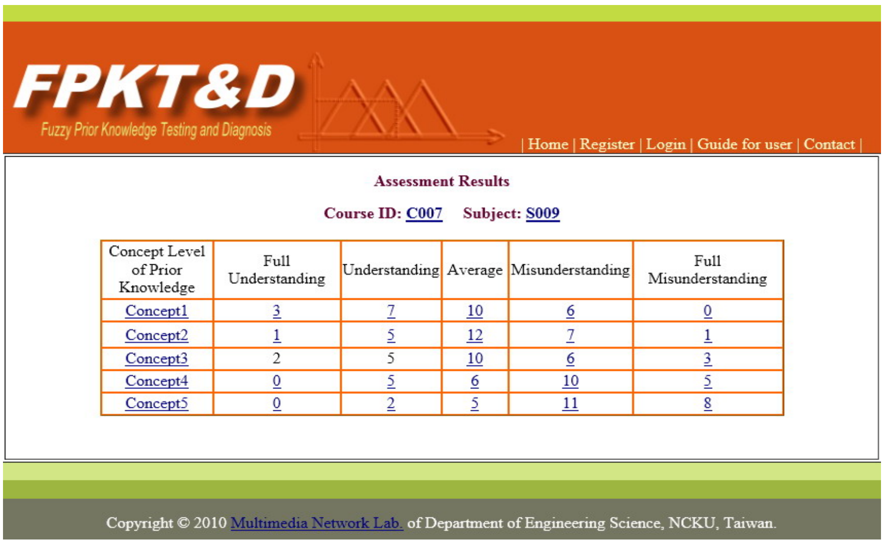

Figure 6. Screenshot of the assessment results for the instructor.

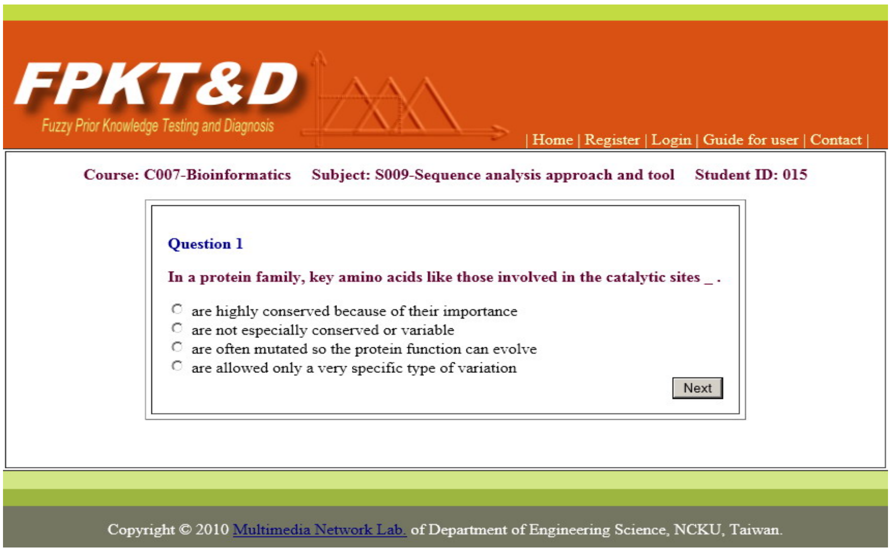

Figure 7. Screenshot of the testing interface.

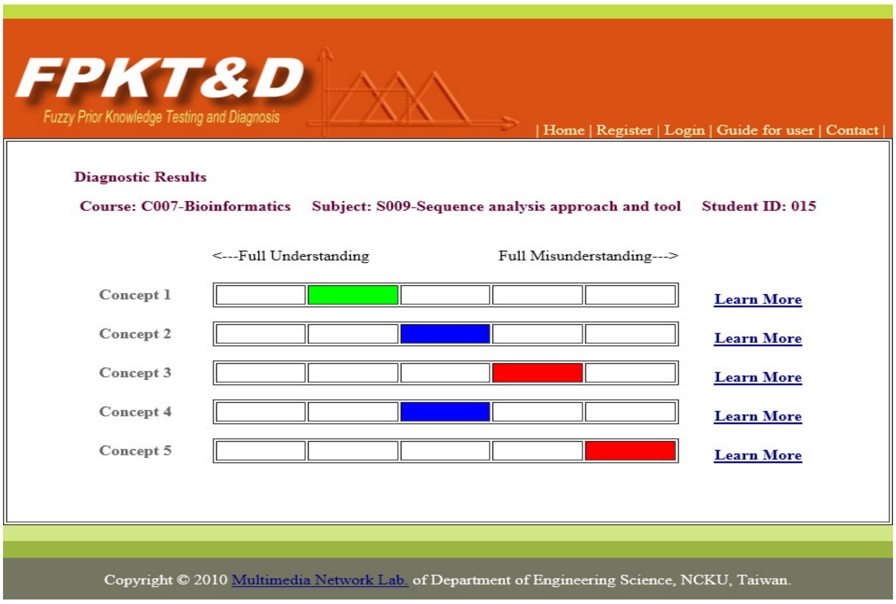

Figure 8. Screenshot of the diagnostic results interface for students.

Experiment and evaluation

Experimental design, participants, and procedure

To investigate the effectiveness of the innovative approach, a quasi-experimental study research was conducted on an interdisciplinary bioinformatics course at a university in Taiwan. The participants in the experiment were a course instructor and 86 university students. The average age of the students was 22. These students were divided into three groups. One group of 26 students served as the experiment group 1 (EG1), which used the FPKD model to diagnose and strengthen their prior knowledge before taking the bioinformatics course. Another group of 28 students served as the experiment group 2 (EG2), which used the PKD model before taking the bioinformatics course. The other group of 32 students served as the control group (CG), and did not use either of the models.

The experiment was conducted on the subject, “sequence analysis approaches and tools”. This subject was taught in the fourth week of the syllabus of the bioinformatics course, and it had a total of 180 minutes of learning activities, including both instruction and practice. The time distribution of each learning activity was planned by the course instructor, and these came in six stages, as shown in Table 2. Prior to learning the subject, the students need to have knowledge of the following five concepts that they were taught in the first three weeks of the course: gene, sequence characteristics and structures, genetics, statistical hypotheses testing, and formula expression format.

Before and after taking part in the learning activities, all the students received pre- and post-tests. The pre-test/post-test were designed to assess the students' knowledge of the sequence analysis techniques presented in the bioinformatics course, including questions about the operation of the BLAST (Basic Local Alignment Search Tool) programs, its application to various problems, and the meaning of the analytical results. Finally, a diagnostic evaluation was conducted to examine the accuracy of the diagnoses derived from the FPKD model.

Table 2. Major teaching and learning activities in the bioinformatics course

Subject: sequence analysis approaches and tools			
Concepts in prior knowledge: Gene, sequence characteristics and structures, genetics, statistical hypotheses testing, and formula expression format			
Unit		Instruction activities	Time (min)
Understanding the importance of similarity	1. 2. 3.	A series of guided questions (5) Slide presentation (15) Discussions (10)	30
Introduction to the most popular data-mining tool: BLAST	1. 2. 3.	A series of guided questions (5) Slide presentation (15) Practice (10)	30
BLASTing protein sequences	1. 2. 3.	A series of guided questions (5) Slide presentation (10) Practice (15)	30
Understanding BLAST output	1. 2.	Slide presentation (15) Discussions (15)	30
BLASTing DNA sequences	1. 2. 3.	A series of guided questions (5) Slide presentation (10) Practice (15)	30
The BLAST way of doing things	1. 2.	Slide presentation (15) Practice (15)	30

Pre-test/Post-test evaluation

Based on a previous investigation, at least two items should be used to measure an objective in order to obtain highly accurate results (Tuckman & Monetti, 2010). Therefore, 20 multiple-choice test items were used in both the pre-test and post-test. Moreover, the two tests were identical, and the maximum score that could be obtained in either of them was 100. The KR-20 reliabilities of the pre-test and post-test were .757 and .778, respectively. The item difficulty index value ranged between 0.35 - 0.85, and the mean difficulty index of items was 0.56. The item discrimination index of most items was greater than 0.35, implying that the items had good discriminative validity (Doran, 1980).

The pre-test results show that the mean and standard deviation of the EG1 (50.38 and 15.61) were similar to those of the EG2 (51.42 and 15.80) and CG (51.87 and 16.15). After preliminary analysis, an ANOVA test was used to determine whether the knowledge level of the three groups was the same with regard to learning bioinformatics. Prior to the ANOVA test, Levene's test of homogeneity of variances was applied to examine whether the variances across samples were equal. The result of this test was not significant ($p = .869 > .05$), which suggests that the difference between the variances for all groups was also not significant. Therefore, ANOVA was performed. As shown in Table 3, the results show that there were no significant differences between the experiment and control groups prior to the experiment ($F(2,83) = 0.065$, $p > .05$). That is, the students in all groups had statistically equivalent abilities before taking the bioinformatics course.

After the bioinformatics course, the course instructor administered a post-test, the results of which show that the mean and standard deviation of the EG1 (75.00 and 8.60) were slightly better than those of the EG2 (66.78 and 11.88) and CG (59.06 and 13.52). The results imply that the students who worked with the FPKD model achieved better learning performance than the others. Moreover, a Pearson's correlation coefficient was used to measure the strength of the association between the accuracy of the diagnosis of prior knowledge and the learning performance of individual students($r = 0.576$, $p < .01$), and the result reveals a significant correlation between them.

A paired *t*-test was then used to analyze the learning improvement of the three groups, as shown in Table 4 , and the results indicate that the teaching strategy could help the students in all groups to learn about bioinformatics (EG1: $t(25) = -9.631$, $p < .05$; EG2: $t(27) = -9.222$, $p < .05$; CG: $t(31) = -6.411$, $p < .05$). In addition, an ANOVA test was used to examine whether the experimental treatment could really enhance the students' learning performance. The result of Levene's test for equality of variances was not significant ($p = .142 > .05$), which indicates that the variances for all groups were assumed to be equal. A one-way ANOVA was then conducted. As shown in Table 5, the results reveal that there was a significant difference in students' post-test achievements between the three groups ($F(2, 83) = 13.36$, $p < .05$). The Scheffe test was used to make post hoc comparisons to identify statistically significant differences among the three groups with regard to their knowledge of bioinformatics, with the results shown in Table 5, and the significance level for the mean difference was $p < .05$. The results thus indicate that the FPKD model can benefit the students in terms of knowledge acquisition.

Table 3. Pre-test ANOVA on knowledge of bioinformatics of the three groups

Variable	Pre-test			$F(2, 83)$
	N	Mean	Std. dev.	
EG1	26	50.38	15.61	
EG2	28	51.42	15.80	0.065
CG	32	51.87	16.15	

Note. $^*p < .05$.

Table 4. The paired *t-test* results of the learning improvement of the three groups

Group	Tests	N	Mean	Std. dev.	*t*
EG1	Pre-test	26	50.38	15.61	-9.631*
	Post-test	26	75.00	8.60	
EG2	Pre-test	28	51.42	15.80	-9.222*
	Post-test	28	66.78	11.88	
CG	Pre-test	32	51.87	16.15	-6.411*
	Post-test	32	59.06	13.52	

Note. $^*p < .05$.

Table 5. Post-test ANOVA on the three groups' knowledge of bioinformatics

Variable	Post-test			$F(2, 83)$	Post hoc test (Scheffe)
	N	Mean	Std. dev.		
EG1	26	75.00	8.60		EG1>EG2*
EG2	28	66.78	11.88	13.36*	EG2>CG*
CG	32	59.06	13.52		EG1>CG*

Note. $^*p < .05$.

Diagnosis evaluation

To assess whether the diagnoses given by the FPKD model are consistent with expert opinions, an evaluation was conducted using the following evaluation function:

$$CR = \frac{n-(n-m)}{n} \tag{6}$$

where *CR* represents the correctness rate of the diagnoses, $0 \leq CR \leq 1$, *n* represents the number of concepts; and *m* indicates the number of matching diagnoses.

A comparison was also conducted to evaluate whether the correctness rates of the diagnoses derived from the FPKD model were superior to those obtained by the PKD model. As noted in the earlier experiment section, the 26 students in experiment group 1 and 28 students in experiment group 2 were asked to use the FPKD and PKD models to diagnose their understanding of five concepts, respectively. In this evaluation, three experts diagnosed the understanding of the 54 students with regard to the five concepts based on the students' test results.

The correctness rates of the diagnoses derived from the FPKD and PKD models were then measured using *Equation* (6). Table 6 shows the correctness rate for each student's diagnosis. It can be seen that the average correctness rates of the results diagnosed by the FPKD model were higher (i.e., 93.26%, 90.38%, and 92.30% for the students) than those diagnosed by the PKD model (i.e., 86.60%, 87.50%, and 88.39%). The results demonstrate that the diagnosis mechanism of the FPKD model is valid, since the diagnoses of the FPKD model were very similar to those from the experts. Moreover, the results of this comparison also revealed that the diagnosis mechanism of the FPKD model is superior to that of the PKD model with regard to diagnosing the learning problem of individual students.

Table 6. Evaluation of the correctness rate results for the five concepts

Student ID		Model	001	002	003	004	005	006	007	008
Expert 1	Correctness rate of diagnoses	EG1	100%	100%	100%	50%	100%	100%	75%	100%
		EG2	100%	75%	50%	100%	75%	100%	100%	100%
Expert 2		EG1	100%	75%	75%	75%	100%	100%	75%	100%
		EG2	100%	75%	100%	75%	100%	75%	100%	100%
Expert 3		EG1	100%	100%	75%	75%	100%	100%	75%	100%
		EG2	75%	100%	75%	100%	100%	100%	75%	100%
Student ID			009	010	011	012	013	014	015	016
Expert 1	Correctness rate of diagnoses	EG1	75%	75%	100%	100%	100%	100%	100%	100%
		EG2	100%	75%	100%	50%	100%	50%	100%	100%
Expert 2		EG1	100%	100%	100%	50%	100%	75%	100%	100%
		EG2	100%	100%	50%	75%	100%	75%	100%	100%
Expert 3		EG1	100%	100%	100%	75%	100%	100%	100%	100%
		EG2	100%	100%	75%	100%	50%	100%	100%	100%
Student ID			017	018	019	020	021	022	023	024
Expert 1	Correctness rate of diagnoses	EG1	100%	100%	100%	100%	75%	100%	100%	100%
		EG2	100%	75%	75%	50%	100%	100%	100%	100%
Expert 2		EG1	100%	100%	75%	100%	50%	100%	100%	100%
		EG2	50%	100%	50%	100%	75%	100%	75%	100%
Expert 3		EG1	100%	100%	75%	100%	50%	100%	100%	100%
		EG2	100%	50%	75%	100%	75%	100%	100%	100%
Student ID			025	026	027	028	029	030	031	032

Expert 1	Correctness rate of diagnoses	EG1	75%	100%	-	-	-	-	-	-
		EG2	50%	100%	100%	100%				
Expert 2		EG1	100%	100%	-	-	-	-	-	-
		EG2	100%	100%	75%	100%				
Expert 3		EG1	100%	75%	-	-	-	-	-	-
		EG2	100%	75%	50%	100%				

Conclusions and discussions

Prior knowledge diagnosis is important for both students and instructors before new instruction is undertaken. Nevertheless, conventional testing systems usually assign only an overall score or grade to instructors and students, giving no adequate way to diagnose any specific problems that they face. Although previous work has developed a prior knowledge testing and diagnosis (PKT&D) system to assist instructors and students in diagnosing and strengthening prior knowledge before new instruction is undertaken (Lin, Lin, & Huang, 2011), it only consider a single attribute to diagnose the learning problems of individual students, and this may lead to inaccurate results .

Therefore, this study applied a multiple attribute decision making technique to develop an innovative prior knowledge diagnosis model, called the FPKD model. The results of the experiment and evaluation show that the proposed model can effectively assist instructors and students in diagnosing students' understanding of prior knowledge in an interdisciplinary bioinformatics course. From a pedagogical perspective, this study applied the FPKD model to a bioinformatics course. However, based on various pedagogical objectives, instructors can use the proposed model in different educational contexts.

Although the innovative approach presented in this work seems to promising, it has some limitations with regard to its practical application. In this study, we applied triangular curves to be the fuzzy membership functions of the rating levels, relative weights, and alternatives. The results of the diagnostic evaluations reveal that this kind of assignment is suitable for our educational context, as the diagnoses that the system produced were very similar to those produced by experts. Nevertheless, in practice, the three functions may have to be adjusted based on the instructors' expertise in different educational contexts.

To enable instructors to more effectively adjust the membership functions, we are currently developing a mechanism to dynamically tune these based on different educational contexts. Moreover, other relevant models or techniques will be taken into account to further improve the diagnosis mechanism of the FPKD model, such as the analytic hierarchy process (AHP) and repertory grid. Finally, to enable instructors to use the FPKT&D system more conveniently, the number of test items in the item bank should be continually increased to address various subject objectives and the different needs of instructors.

References

Biswas, N. B. (2007). Knowledge and Pedagogy: An Essential Proposition in Response to Teacher Preparation. *US-China Education Review*, *4*(7), 1–14.

Chen, S. J., & Hwang, C. L. (1992). *Fuzzy multiple attribute decision making - Methods and applications*. Secaucus, NJ: Springer-Verlag.

Chieu, V. M. (2007). An operational approach for building learning environments supporting cognitive flexibility. *Journal of Educational Technology & Society*, *10*(3), 32–46.

Dobois, D. & Prade, H. (1980). Fuzzy *sets and systems*. New York, NY: Academic Press.

Dochy, F. J. R. C., Moerkerke, G., & Marten, R. (1996). Integrating assessment, learning and instruction: assessment of domain-specific and domain-transcending prior knowledge and program. *Studies in Educational Evaluation*, *22*(4), 309–339.

Dong, W. M., & Wong, F. S. (1987). Fuzzy weighted averages and implementation of the extension principle. *Fuzzy Sets and Systems*, *21*(2), 183–199.

Doran, R. (1980). *Basic measurement and evaluation of science instruction*. Washington D.C.: National Science Teachers Association.

Gerber, M., Grund, S., & Grote, G. (2008). Distributed collaboration activities in a blended learning scenario and the effects on learning performance. *Journal of Computer Assisted Learning*, *24*(3), 232–244.

Gogoulou, A., Gouli, E., Grigoriadou, M., Samarakou, M., & Chinou, D. (2007). A web-based educational setting supporting individualized learning, collaborative learning and assessment. *Journal of Educational Technology & Society*, *10*(4), 242–256.

Hailikari, T., Katajavouri, N., & Lindblom-Ylanne, S. (2008). The relevance of prior knowledge in learning and instructional design. *American Journal of Pharmaceutical Education*, *72*(5), 1–8.

Huang, Y. M., Kuo, Y. H., Lin, Y. T., & Cheng, S. C. (2008). Toward Interactive Mobile Synchronous Learning Environment with Context-awareness Service. *Computers & Education*, *51*(3), 1205–1226.

Hwang, G. J., Tseng, Judy. C. R., & Hwang, G. H. (2008). Diagnosing student learning problems based on historical assessment records. *Innovations in Education and Teaching International*, *45*(1), 77–89.

Lee, D. H., & Park, D. (1997). An efficient algorithm for fuzzy weighted average. *Fuzzy Sets and Systems*, 87(1), 39–45.

Lin, Y. C., Lin Y. T., Huang, Y. M. (2011). Development of a diagnostic system using a testing-based approach for strengthening student prior knowledge. *Computers & Education*, *57*(2), 1557–1570.

Moos, D. C., & Azevedo, R. (2008). Self-regulated learning with hypermedia: The role of prior domain knowledge. *Contemporary Educational Psychology*, *33*(2), 270–298.

Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction*, *19*(3), 228–242.

Panjaburee, P., Hwang, G. J., Triampo, W., & Shih, B. Y. (2010). A Multi-Expert Approach for Developing Testing and Diagnostic Systems Based on the Concept Effect Model. *Computers & Education*, *55*(2), 527–540.

Roschelle, J. (1995). *Learning in Interactive Environments: Prior knowledge and new experience*. Washington, D.C.: American Association of Museums.

Ross, T. J., Sorensen, H. C., Savage, S. J., & Carson, J. M. (1990). DAPS: Expert system for structural damage assessment. *Journal of Computer and Civil Engineering*, *4*(4), 327–348.

Saleh, I., & Kim, S. I. (2009). A fuzzy system for evaluating students' learning achievement. *Expert Systems with Applications*, *36*(3), 6236–6243.

Seel, N. M., & Dinter, F. R. (1995). Instruction and mental model progression: learner-dependent effects of teaching strategies on knowledge acquisition and analogical transfer. *Educational Research and Evaluation*, (1), 4–35.

Tao, Y. H., Wu, Y. L., & Chang, H. Y. (2008). A practical computer adaptive testing model for small-scale scenarios. *Journal of Educational Technology & Society*, *11*(3), 259–247.

Tseng, C. R., Chu, H. C., Hwang, G. J., & Tsai, C. C. (2008). Development of an adaptive learning system with two sources of personalization information. *Computers and Education, 51*(2), 776–786.

Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, *10*(2), 159–169.

Tuckman, B. W., & Monetti, D. M. (2010). *Educational Psychology*. Wadsworth, OH: Cengage Learning.

Appendix

The EFWA algorithm

- Definition: the input *a*, *b*, *c*, and *d* are the intervals of fuzzy membership functions, and the outputs are the intervals of the result fuzzy membership function. Additionally, the δ_{s_i} and the ζ_{s_i} can be calculated by *Equations* (7) and (8) respectively.

$$\delta_{s_i} = \frac{(a_1 - a_i)e_1 + (a_2 - a_i)e_2 + ... + (a_{n-} a_i)e_n}{e_1 + e_2 + ... e_n} \quad (7)$$

$$\zeta_{s_i} = \frac{(b_1 - b_i)e_1 + (b_2 - b_i)e_2 + ... + (b_{n-} b_i)e_n}{e_1 + e_2 + ... e_n} \quad (8)$$

- Description of the EFWA algorithm (Lee and Park, 1997)

(1) Sort *a*'s in non-decreasing order. Let (a_1, a_2, …, a_n) be the resulting sequence. Let *first* := 1 and *last* := *n*.

(2) Sort *a*'s in non-decreasing order. Let (a_1, a_2, …, a_n) be the resulting sequence. Let *first* := 1 and *last* := *n*.

(3) Let δ-threshold := $\lfloor (first + last)/2 \rfloor$. For each *i* = 1, 2, …, δ-threshold, let e_i := d_i and for each *i* = δ-threshold + 1, …, *n*, let e_i := c_i. For an *n*-tuple *S* = (e_1, e_2, …, e_n), evaluate $\delta_{s_{\delta-threshold}}$ and $\delta_{s_{\delta-threshold+1}}$.

(4) If $\delta_{s_{\delta-threshold}} > 0$ and $\delta_{s_{\delta-threshold+1}} \leq 0$ then *L* = f_L(e_1, e_2, …, e_n) and go to Step 4; otherwise execute the following step.

(a) If $\delta_{s_{\delta-threshold}} > 0$, then *first* := δ-threshold + 1; otherwise *last* := δ-threshold, and go to Step 2.

(5) Sort *b*'s in non-decreasing order. Let (b_1, b_2, …, b_n) be the resulting sequence. Let *first* := 1 and *last* := *n*.

(6) Let ζ-threshold := $\lfloor (first + last)/2 \rfloor$. For each *i* = 1, 2, …, ζ-threshold, let e_i := c_i and for each *i* = ζ-threshold + 1, …, *n*, let e_i := d_i. For an *n*-tuple *S* = (e_1, e_2, …, e_n), evaluate $\zeta_{s_{\zeta-threshold}}$ and $\zeta_{s_{(\zeta-threshold+1)}}$.

(7) If $\zeta_{s_{\zeta-threshold}} > 0$ and $\zeta_{s_{\zeta-threshold+1}} \leq 0$ then *U* = f_U(e_1, e_2, …, e_n) and stop; otherwise execute the following step:

(b) If $\zeta_{s_{\zeta-threshold}} > 0$, then first := ζ-threshold + 1; otherwise last := ζ-threshold, and go to step 5.

Illustrative example

Assume that an instructor aims to assess five students (S_1, S_2, S_3, S_4, S_5) to identify their level of understanding with regard to five concepts (C_1, C_2, C_3, C_4, C_5,). The instructor selects five test items (I_1, I_2, I_3, I_4, I_5) from a test item bank to form a test-sheet, that are relevant to concepts one to five, and each test item has its difficulty degree, D_1, D_2, D_3, D_4, D_5. In this test-sheet, each concept is possibly related to the others. The instructor then conducts the test to assess the five students to identify the level of understanding of the individual students with regard to the five concepts. The degree of difficulty of each test item is shown in Table 7. In addition, the relationships among the test items and concepts are shown in Table 8, and the relationships among the concepts are shown in Table 9. After the participating students have taken the test, their test results for each test item are given, as shown in Table 10.

Table 7. Illustrative example of degree of difficulty each test item

Degree of Difficulty of Test Item	Test Item				
	I_1	I_2	I_3	I_4	I_5
Difficulty Degree	0.2	0.4	0.8	0.2	0.4

Table 8. Illustrative example of the relationships among test items and concepts

Test Item	Concept				
	C_1	C_2	C_3	C_4	C_5
I_1	1	0	0	0	0
I_2	1	0	1	0	1
I_3	0	1	0	0	0
I_4	0	1	0	0	1
I_5	0	0	0	1	0

Table 9. Illustrative example of relationship among concepts

Concept	Concept				
	C_1	C_2	C_3	C_4	C_5
C_1	1.0	0.0	0.4	0.0	0.0
C_2	0.0	1.0	0.6	0.0	0.2
C_3	0.4	0.6	1.0	0.2	0.6
C_4	0.0	0.0	0.2	1.0	0.4
C_5	0.0	0.2	0.6	0.4	1.0

Table 10. Illustrative example of the relationship among students' answers and test items

Test Item	Student				
	S_1	S_2	S_3	S_4	S_5
I_1	1	1	0	1	1
I_2	1	0	1	1	1
I_3	0	0	0	0	1
I_4	0	1	0	1	1
I_5	0	0	1	0	0

Based on the association among Tables 8, 9, and 11, the highest difficulty level of the five concepts answered by the five students can be measured with *Equation* (5). In order to illustrate this clearly, Table 11 shows the association among the five test items and five concepts related to the fourth student (S_4).

Table 11. Illustrative example of relationship of test items, concepts, and the fourth student's answers

Test Item	Concept				
	C_1	C_2	C_3	C_4	C_5
I_1	1	0	0	0	0
I_2	1	0	1	0	1
I_3	0	1	0	0	0
I_4	0	1	0	0	1
I_5	0	0	0	1	0

The gray rows means that the fourth student answered the first, second, and fourth test items (I_1, I_2, and I_4) correctly, as seen in Table 10. Based on Table 11, therefore, the highest difficulty level of the five concepts answered by the five students can thus be measured. For instance, the highest difficulty level of second concept answered by the fourth student is:

$$HDL(S_{4,}C_2) = 0.2$$

Furthermore, based on Tables 9,10, and11 the relevant level of five concepts answered by the five students correctly can be measured using *Equation* (2). Similarly, as shown in Table 12, the relevant level of second concept answered by the fourth student correctly is:

$$RL(S_{4,}C_2) = 0.142$$

In addition, based on Tables 9 and 11, the correctness rate of the five students with regard to five concepts can be inferred with *Equation* (3). For instance, the correctness rate of the fourth student with regard to second concept can be measured as follows:

$$\mathrm{CR}(S_{4,}C_2) = 0.5$$

Therefore, the ratings of the fourth student's attributes are shown in Table 13. With regard to the relative weight of each criterion, this example assumes that it is very low, very low, and average for the highest difficulty level, relevant level and correctness rate respectively, as shown in Table 12. Notice that, in Table 13, the values of rating are related to Figure 2 and the values of relative weight are related to Figure 3, and the parameter settings of r_i and w_i are the triangle values of membership functions (Figures 3 and 4) with respect to α = 0 and 1 (the α-cuts is the interval analysis technique (Dong & Wong, 1987).

Table 12. The input values related to fourth student's answer with regard to the second concept

Criteria	Rating	Relative Weight
highest difficulty level	Very poor	Very low
relevant level	Very poor	Very low
correctness rate	Average	Average

Before starting the EFWA algorithm, it is necessary to choose two values for α, namely 0 and 1, as the initial input values. For α = 0, the intervals of $r_{i=1\text{-}3}$ are $[a_1 = 0, b_1 = 0.4]$, $[a_2 = 0, b_2 = 0.4]$, and $[a_3 = 0.3, b_3 = 0.7]$, and the intervals of $w_{i=1\text{-}3}$ are $[c_1 = 0, d_1 = 0.4]$, $[c_2 = 0, d_2 = 0.4]$, and $[c_3 = 0.3, d_3 = 0.7]$. Notice that the r_i shown here have not been sorted yet. The computational procedure is as follow:

Step 1: Sort a's into non-decreasing order, and the resulting sequence is $[a_1 = 0, b_1 = 0.4]$, $[a_2 = 0, b_2 = 0.4]$, and $[a_3 = 0.3, b_3 = 0.7]$. So $(a_1, a_2, a_3) = (0, 0, 0.3)$, *first* := 1, *last* := 3.

Step 2: δ-threshold := $\lfloor (1+3)/2 \rfloor = 2$, S = (d_1, d_2, c_3) = (0.4, 0.4, 0.3), then evaluating δ_{s_2} and δ_{s_3}, the evaluation results are as shown in *Equations* (9) and (10).

$$\delta_{s_2} = \frac{(0-0)\times 0.4 + (0-0)\times 0.4 + (0.3-0)\times 0.3}{0.4+0.4+0.3} = 0.01818 \tag{9}$$

$$\delta_{s_3} = \frac{(0-0.3)\times 0.4 + (0-0.3)\times 0.4 + (0.3-0.3)\times 0.3}{0.4+0.4+0.3} = -0.2181 \tag{10}$$

Step 3: Since $\delta_{s_2} > 0$ and $\delta_{s_3} \le 0$, L = $f_L(d_1, d_2, c_3) = a_1 + \delta_{s_2}$ = 0 + 0.0818 = 0.0818. Hence, the min f_L is 0.0818 and go to Step 4.

Step 4: Sort b's in to non-decreasing order, the resulting sequence is $[a_1 = 0, b_1 = 0.4]$, $[a_2 = 0, b_2 = 0.4]$, and $[a_3 = 0.3, b_3 = 0.7]$. So $(b_1, b_2, b_3) = (0.4, 0.4, 0.7)$, *first* := 1, *last* := 3.

Step 5: ζ-threshold := $\lfloor (1+3)/2 \rfloor = 2$, S = (c_1, c_2, d_3) = (0, 0, 0.7), then evaluating ζ_{s_2} and ζ_{s_3}, the evaluating results as shown in *Equation* (11) and *Equation* (12).

$$\zeta_{s_2} = \frac{(0.4-0.4)\times 0 + (0.4-0.4)\times 0 + (0.7-0.4)\times 0.7}{0+0+0.7} = 0.3 \tag{11}$$

$$\zeta_{s_3} = \frac{(0.4-0.7)\times 0 + (0.4-0.7)\times 0 + (0.7-0.7)\times 0.7}{0+0+0.7} = 0 \quad (12)$$

Step 6: Since $\zeta_{s_2} > 0$ and $\zeta_{s_3} \le 0$, $U = f_U(c_1, c_2, d_3) = b_2 + \zeta_{s_2} = 0.4 + 0.3 = 0.7$.

Hence, the max f_U is 0.7 and stop. The interval for α = 0 is [0.0818, 0.7], in which each point is corresponding to the end points of the triangle representing the membership function.

The above process finds the upper and lower bounds of the synthetic membership function, and the following process will obtain conduct the triangle value with respect to α = 1. For α = 1, the intervals of $r_{i=1\text{-}3}$ are [a_1 = 0.2, b_1 = 0.2], [a_2 = 0.2, b_2 = 0.2], and [a_3 = 0.35, b_3 = 0.35], and the intervals of $w_{i=1\text{-}3}$ are [c_1 = 0.2, d_1 = 0.2], [c_2 = 0.2, d_2 = 0.2], and [c_3 = 0.5, d_3 = 0.5]. Notice that the r_i shown here have not been sort yet.

Step 1: Sort *a*'s into non-decreasing order, and the resulting sequence is [a_1 = 0.2, b_1 = 0.2], [a_2 = 0.2, b_2 = 0.2], and [a_3 = 0.35, b_3 = 0.35]. So (a_1, a_2, a_3) = (0.2, 0.2, 0.35), *first* := 1, *last* := 3.

Step 2: δ-threshold := $\lfloor (1+3)/2 \rfloor = 2$, S = ($d_1$, d_2, c_3) = (0.2, 0.2, 0.5), then evaluating δ_{s_2} and δ_{s_3}, the evaluation results as shown in *Equations* (13) and(14).

$$\delta_{s_2} = \frac{(0.2-0.2)\times 0.2 + (0.2-0.2)\times 0.2 + (0.35-0.2)\times 0.5}{0.2+0.2+0.5} = 0.0833 \quad (13)$$

$$\delta_{s_3} = \frac{(0.2-0.5)\times 0.2 + (0.2-0.5)\times 0.2 + (0.35-0.5)\times 0.5}{0.2+0.2+0.5} = -0.2167 \quad (14)$$

Step 3: Since $\delta_{s_2} > 0$ and $\delta_{s_3} \le 0$, $L = f_L(d_1, d_2, c_3) = a_2 + \delta_{s_2}$ = 0.2 + 0.0833 = 0.2833. Hence, the min f_L is 0. 3884 and according to the $a_i = b_i$ (where i = 1-3, when α = 1), it can conclude that the min $f_L = f_U$ =0.2833. For α = 1, the obtained interval result is [0.2833, 0.2833] which corresponds to the center of the triangle. Consequently, with the intervals for α = 0 and 1, the resulting membership function is determined and is plotted in Figure 9.

As the result shown in Figure 9 is fuzzy membership function, Euclidean distance (as shown in *Equation* (5)) is used to determine the closest membership function (from Figure 4 for performance the decision goal. The following calculation shows how to use the Euclidean distance to determine the appropriate understanding level.

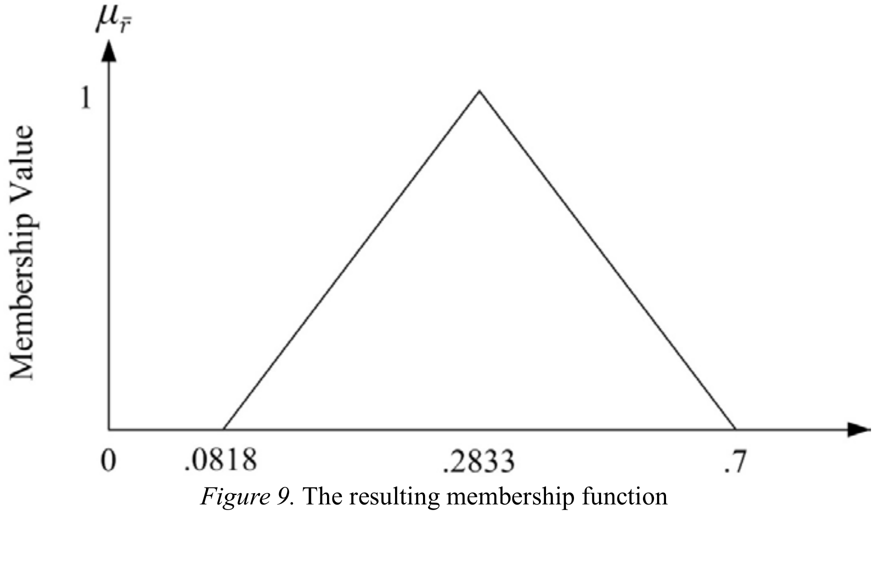

Figure 9. The resulting membership function

$$d(\bar{r}, A_{Misunderstanding}) = \sqrt{(0.0818 - 0.15)^2 + (0.2833 - 0.35)^2 + (0.7 - 0.55)^2} = 0.1777$$

Based on the measured results, the other Euclidean distances are $d(\bar{r}, A_{Misunderstanding}) = 0.3219$, $d(\bar{r}, A_{Average}) = 0.3075$, $d(\bar{r}, A_{Understanding}) = 0.5409$, and $d(\bar{r}, A_{Full\ Understanding}) = 0.7909$. After measuring all of the Euclidean distances between the resulting membership function and all alternatives, the decision model determines that Misunderstanding is the closest alternative, and then presents the corresponding feedback to the students.