Evaluation of Students' Academic Performance with Artificial Intelligence Applications

Hayrünnisa Ergin
Dokuz Eylul University, Dept. of Computer Science,
35390 Izmir, Turkiye
hayrunnisa.ergin@ogr.deu.edu.tr
0000-0003-2625-0043

Efendi Nasibov
Dokuz Eylul University, Dept. of Computer Science,
35390 Izmir, Turkiye
efendi.nasibov@deu.edu.tr
0000-0002-1889-6410

***Abstract*— The purpose of this study: First of all, to create a dataset consisting of programming questions and to determine the level of these questions according to Bloom Taxonomy and the weight of each concept they contain by taking expert opinion. Afterward, it will be determined to what extent the student has learned a concept, taking into account the score the student received and the difficulty level of the question. After analyzing the evaluation results, the numerical values obtained were expressed as linguistic variables using the fuzzy triangle-type membership function. This research was conducted with 14 10th-grade students studying at a Vocational and Technical Anatolian High School in Izmir, Turkey, within the scope of the Object Oriented Programming course. In order to measure the effectiveness of the research, the Programming Self-Efficacy Scale, the validity and reliability study, was applied to the students as a single group pre-test and post-test. In the analysis of the research data, the Wilcoxon Signed Rank Test was used, and it was seen that the method used in the study had a positive effect on the students' self-efficacy regarding programming.**

Keywords—Programming language learning, Academic performance evaluation, Fuzzy rules.

I. INTRODUCTION

Feedback is provided by informing the student how correct or incorrect, faulty or flawless the learning is, facilitating learning and making it permanent. A student who does not know exactly how far he/she has progressed in the learning and teaching process will make slow progress, may learn mistakes as if they were right, and his/her learning may be interrupted. Success in learning forms the basis of new learning. Feedback that informs the student of his/her mistakes and deficiencies is a forerunner of the student's future success. Sünbül (2011, cited in Yeşilyurt, 2020) stated that sharing the reasons for the evaluation results with students is important in terms of this principle [1].

Similar studies using a new fuzzy exam system developed using Chen and Lee's evaluation method were examined. There are questions with alternatives from easy to difficult that every student can answer. In this way, students at every level will be able to answer questions at every level, which will enable them to be better motivated for subsequent exams [2].

In a study based on the mathematics course, the number of correct answers given by students to unit questions, the number of correct answers given by other students taking the exam, and the course notes taken by the student in previous years were given as input to the fuzzy logic system and the student's learning outcomes were evaluated. According to the determined learning rates, exams consisting of questions appropriate to the student's level were produced [3].

In this study, a fuzzy logic-based smart exam system was developed. The instructor can choose the number of questions, topics for the weeks, and the difficulty level of the questions on the exam page. In this system, the difficulty levels of the questions are calculated fuzzily, and after each exam, the difficulty levels of the questions are updated using the fuzzy logic method according to the rate of correct answers. The difficulty level of a question to be asked to students for the first time is determined as the medium. Thus, a smart exam system was created, with difficulty levels of questions changing depending on the answer rate of the questions [4].

In our study, first, a database consisting of programming questions was created. Students at a vocational high school are asked questions selected from this database as the first exam questions of the first semester and are asked to answer them in the C Sharp programming language. According to the method, the students' rates of learning the concepts are given as feedback.

In 1956, University of Chicago professor Benjamin Bloom classified the development of mental skills from low-level thinking to higher-level thinking skills to explain how the learning process develops. Bloom's taxonomy shows that learning occurs in stages [5]. In this study, first of all, a database consisting of programming questions was created, and the learning stage of each question, according to Bloom Taxonomy, was determined by taking the opinions of the researcher and experts. We do this to assign a difficulty level according to the learning stage in which the questions are located. These difficulty levels are shown in Table I below.

TABLE I. BLOOM'S TAXONOMY DIFFUCULTY LEVEL

Bloom's Taxonomy	**Difficulty Level**
Remembering	1/6
Understanding	2/6
Applying	3/6
Analyzing	4/6
Synthesis	5/6
Evaluating	6/6

II. Description of the Processing System

A. Difficulty Attribute

A. Difficulty Attribute

It is known that the rate of correct answers to a question indicates the difficulty level of the question. Item difficulty index is calculated as a percentage of the total number of correct responses to items and is calculated using the formula

P=R/T

P is the item difficulty index,

R is the number of correct answers,

T is the total number of answers, both correct and incorrect.

Based on this, the difficulty level of the question was determined using the equation below.

- Od=BTdl∗(1- Id)
- Od=Overall Difficulty
- BTdl = Bloom's Taxonomy Difficulty Level
- Id = Item difficulty index

B. Concept Weights

Afterward, the weight of the concepts contained in each question was represented by fuzzy numbers in the range [0, 1]. This way, the weight ratios of the concepts each question tries to measure will be obtained. For questions that students answer correctly or incorrectly, the ratio of the score they receive from each question to the score given to the question by the instructor will be multiplied by the concept weight to obtain information about which concept the student has learned and to what extent.

The fact that the scoring of open-ended questions is based on the opinion of the teacher or the rater is one of the important factors affecting the quality of written exams. These factors also influence the validity and reliability of the written exam [6]. In this study, in order to take precautions against all these factors, the concepts in the programming questions were scored according to their weight ratios by taking the opinions of 2 more experts in addition to the researcher, and the final weight levels were determined by taking the average of the scores.

C. Calculation Learning Degree

This study aims to give detailed feedback to the student about the degree of learning each concept, taking into account the score the student received from each exam question, the weight of the concepts included in the questions, the level they are in according to Bloom Taxonomy, and the difficulty level of the question. To do this, the following equations were used. Thus, our result will be expressed with linguistic variables according to the function obtained with triangular fuzzy numbers.

$$Question_i = \frac{Received_{Points}}{Number\_of\_Students * Question\_Point}$$

$Question_i$ is the importance of the question

$$Concept_{learning\_degree} = Question_i * Concept_{Weight} * Overall_{Diffuculty}$$

D. Giving Feedback

According to the fuzzy logic base of the system, triangle-type membership, which was explained regarding the fuzzy logic function, was used in the exam evaluation system. The learning membership levels of the concepts in the analyzed questions take values between 0 and 1. The learning rate of the concepts is labeled as 7 units and gives the output value, and the evaluation range is determined as 0-100. According to these ranges, [10,30] range is very very bad in Turkish (çok çok kötü, ÇÇK), [20,40] range is very bad in Turkish (çok kötü, ÇK), [30,50] range is bad in Turkish (kötü, K), [40,60] range is medium in Turkish (orta, O), [50,70] range is good in Turkish (iyi, İ), [60,80] range is determined as very good in Turkish (çok iyi, Çİ) and [80,100] range is determined as very very good (çok çok iyi, ÇÇİ). The output membership function showing the learning percentages is as shown in the figure.

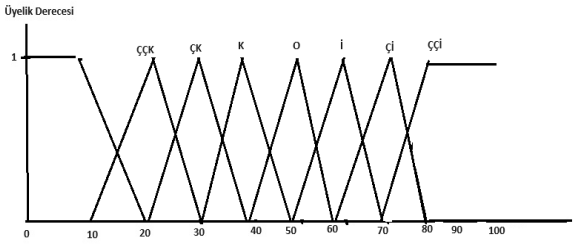

Fig. 1. The output membership function showing the learning percentages

III. Performance Improvement

The researcher conducted this research with 14 10th-grade students studying in a Vocational and Technical Anatolian High School in Izmir, Turkey, within the scope of the Object Oriented Programming course. In order to measure the effectiveness of the research, the Programming Self-Efficacy Scale, the validity and reliability study conducted by Altun and Mazman (2012), was applied to the students. Due to the validity and reliability studies, the Turkish form of the Programming Self-Efficacy Scale consists of 9 items and 2 factors. The internal consistency coefficient for the entire scale was calculated as 0.928 [7].

The results of the Wilcoxon Signed Ranks Test conducted to measure the effect of the evaluation method based on the fuzzy logic algorithm on the students' self-efficacy beliefs towards computer programming are presented in Table 2.

TABLE II. WILCOXON SIGNED-RANKS TEST RESULTS OF SPSS PRE-TEST AND POST-TEST APPLICATION SCORES

		N	Mean Rank	Sum of Ranks	z	p
Programming Self-efficacy	Negative Ranks	3	3,83	11,50	-2,378	0,017
	Positive Ranks	10	7,95	79,50		
	Ties	1				

When the Table 2 is examined, it is seen that the difference between the students' self-efficacy in programming before and after the applied evaluation method is significant (z=-2,378, p=0,017). The obtained data was also analyzed with the paired-samples t-test (t= -2,892, p= 0,013), and a significant difference was found. Accordingly, it is seen that the use of the applied fuzzy logic algorithm-based method has a positive effect on self-efficacy beliefs about programming.

IV. Conclusion

In today's education system, evaluating what students know and how they apply this knowledge is one of the important steps in the education process. Accurate evaluation of students' academic performance has significant effects on students' educational lives and futures. This study shows that individualized student evaluation can be done more accurately, and a path can be opened for them to complete their deficiencies.

References

[1] Yeşilyurt, E. (2020). Öğretmenin Pusulası: Genel Öğretim İlkeleri. *Ekev akademi dergisi*, (83), 263-288.

[2] Demirçelik, Ç. (2010). Bulanık Sınav Sistemleri. (Yüksek Lisans Tezi). (https://earsiv.gop.edu.tr/xmlui/bitstream/handle/20.500.12881/1813/T00877.pdf?sequence=1&isAllowed=y).

[3] Taşkırdı, Ö., & Ünver, H. M. (2020). Yüksek Başarım İçin Bulanık Mantık Tabanlı Kişiye Özel Sınav Üretme Sistemi Tasarımı. *International Journal of Engineering Research and Development*, *12*(2), 745-752.

[4] Ölmez, Ç. (2010). Uzaktan Eğitim Sistemlerindeki Soru Bankalarının Bulanık Mantık Yöntemi ile Analizi. (Yüksek Lisans Tezi). https://acikerisim.aku.edu.tr/xmlui/bitstream/handle/11630/6283/371802.pdf?sequence=1&isAllowed=y

[5] Bloom, B. (1956). A taxonomy of cognitive objectives. *New York: McKay*.

[6] Atılgan, H. (Ed.). 2006. Eğitimde ölçme ve değerlendirme, Ankara: Anı Yayıncılık.

[7] Altun, A. ve Mazman, S. G. (2012). Programlamaya ilişkin Öz Yeterlilik Algısı Ölçeğinin Türkçe formunun güvenirlik ve geçerlik çalışması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 3*(2), 297-308.