



Evaluating the user's experience, adaptivity and learning outcomes of a fuzzy-based intelligent tutoring system for computer programming for academic students in Greece

Konstantina Chrysafiadi¹  · Maria Virvou¹ · George A. Tsihrintzis¹ · Ioannis Hatzilygeroudis²

Received: 28 July 2022 / Accepted: 27 October 2022 / Published online: 17 November 2022
© The Author(s) 2022

Abstract

Nowadays, the improvement of digital learning with Artificial Intelligence has attracted a lot of research, as it provides solutions for individualized education styles which are independent of place and time. This is particularly the case for computer science, as a tutoring domain, which is rapidly growing and changing and as such, learners need frequent update courses. In this paper, we present a thorough evaluation of a fuzzy-based intelligent tutoring system (ITS), that teaches computer programming. The evaluation concerns multiple aspects of the ITS. The evaluation criteria are: (i) context, (ii) effectiveness, (iii) efficiency, (iv) accuracy, (v) usability and satisfaction, and (vi) engagement and motivation. In the evaluation process students of an undergraduate program in Informatics of the University of Piraeus in Greece participated. The evaluation method that was used included questionnaires, analysis of log files and experiments. Also, t-tests were conducted to certify the validity of the evaluation results. Indeed, the evaluation results are very positive and show that the incorporated fuzzy mechanism to the presented ITS enhances the system with Artificial Intelligence and through this, it increases the learners' satisfaction and new knowledge learning and mastering, improves the recommendation accuracy of the system, the efficacy of interactions, and contributes positively to the learners' engagement in the learning process.

Keywords Artificial Intelligence · Intelligent Tutoring System · e-learning · Evaluation · Fuzzy logic

✉ Konstantina Chrysafiadi
kchrysafiadi@unipi.gr

Extended author information available on the last page of the article

1 Introduction

Nowadays, education has benefitted from advances in computer technology. Particularly, individual students can participate in a lesson from wherever they are and whenever they can, receiving learning material tailored to their needs. This has been achieved to a large extent due to the development of advanced software for computer-assisted learning, such as Intelligent Tutoring Systems (ITS) (Sáiz-Manzanares et al., 2021; Cho & Kim, 2021; Urdaneta-Ponte et al., 2021; Alonso-Secades et al., 2022). Indeed, ITSs constitute a special kind of educational software programs that aim to model the cognitive state and the learning needs of the individual students and provide a personalized learning experience (Akyuz, 2020; Chrysafiadi et al., 2022). They incorporate Artificial Intelligence, which enhances the learning process making it attached to each individual learner's needs (Sotiropoulos et al., 2019; Tsihrintzis et al., 2019, 2021; Virvou et al., 2020). They model the students' characteristics and needs and imitate the way that a human tutor thinks and reacts during the teaching process (Chrysafiadi & Virvou, 2013a; Clancey & Hoffman, 2021; Khazanchi & Khazanchi, 2021). This is particularly significant in the case of computer science education, in which the learners have heterogeneous background, characteristics and needs. Moreover, according to (Nesbit et al., 2014), there is a significant advantage of ITS over teacher-led classroom instruction and computer-based instruction that are not based on intelligent techniques.

The main aim of an ITS is to provide a student-oriented learning process that helps learners acquire knowledge and accomplish the learning goal (Polson & Richardson, 2013; Erümit & Çetin, 2020; Paladines & Ramírez, 2020). To achieve this, it has to be able to (i) recognize the learner's knowledge level, misconceptions and learning needs, (ii) provide lessons and feedback that are tailored to each individual learner's needs, (iii) create positive feelings to the student and motivate her/him to participate in the learning process (Graesser et al., 2018). Therefore, the success of an ITS depends on several factors (Kulik & Fletcher, 2016; Mousavinasab et al., 2021; Feng et al., 2021). Consequently, the evaluation of an ITS has to include usability evaluation (Chughtai et al., 2015; Chrysafiadi & Virvou 2021a; Wang et al., 2021), learning outcomes evaluation (Hosseini et al., 2020; Rebolledo-Mendez et al., 2022; Binh & Trung 2021; Chrysafiadi & Virvou, 2021b), student modeling and recommendation validity evaluation of the system (Chrysafiadi & Virvou, 2013b; Sosnovsky & Brusilovsky, 2015; Effenberger & Pelánek, 2021).

In view of the previous, in this paper we present a thorough evaluation of a fuzzy-based ITS that teaches computer programming. The aim is to examine how useful and effective the system is in terms of the learning process and how the educational process benefits from it. Therefore, the following questions are seeking answers in this research:

- How helpful the system is in the learning process?
- Does the system contribute to the acquisition of new knowledge?
- How efficient the system is concerning the number of interactions needed to achieve the learning goal?

- How accurate are the system's recommendations?
- How usable and pleasant the system is?
- How the system affects the students' engagement in the learning process?

For answering the above questions, a thorough evaluation of the system was conducted. For the evaluation, we combined two evaluation frameworks, the CIAO! framework (Jones et al., 1999) and the evaluation framework that was proposed by Lynch and Ghergulescu (2016), that were developed for evaluating educational software. In this way, we accomplish to assess multiple aspects of the tutoring system that include the intelligent features as well as the necessary educational aspects. The evaluation process was based on the participation of 140 learners who attended an undergraduate program in Informatics at the University of Piraeus, Greece. For the evaluation questionnaires and experiments were used.

The remainder of this paper is organized as follows. In Section 2, we present background knowledge about ITS evaluation. In Section 3, we present the theoretical framework and the methodology of research. In Section 4, we present the fuzzy-based ITS which was evaluated. In Section 5, we describe the evaluation method, testbed and results. In Section 6, we discussed the evaluation results and present the research's impication. Finally, in Section 7, we draw conclusions from this work.

2 Related work

The evaluation of an ITS is crucial to its acceptance and contribution to the learning process. The evaluation criteria of most ITSs include usability, learners' performance and learning outputs. However, a thorough evaluation should include additional criteria, like accuracy, precision, sensitivity, adaptivity, reliability, recognition rate, usability, and mean square error (MSE) (Lampropoulou et al., 2010; Mousavinasab et al., 2021). Furthermore, the most common techniques for an ITS evaluation are observations, questionnaires, and experiments. According to (Greer & Mark, 2016) experiments are ideal for ITS evaluation because they enable researchers to examine relationships between teaching interventions and student-related teaching outcomes, and to obtain quantitative measures of the significance of such relationships.

In recent literature review there is a variety of ITSs that have been evaluated trough experiment and questionnaires. The authors in Wambsganss et al. (2020) used a questionnaire with 38 items to evaluate the usability, usefulness, adaptivity and effectiveness of an adaptive dialog-based tutoring system for augmentation skills. Similarly, the authors in Wang et al. (2021) used questionnaires to evaluate the usability of an affective emotional mobile tutoring system and the user satisfaction. On the other hand, an experimental evaluation, which includes the performing of a pre-test and a post-test and the comparing of their results, was used for the evaluation of a tutoring system that teaches Algebraic concerning its contribution to the students' performance (VanLehn et al., 2020). A similar experimental evaluation was used in Singh et al. (2022) to evaluate a custom-tailored tutoring system that was called SeisTutor. Particularly, pre-test and post-test method and questionnaires

were used in order to evaluate the system, according to the four phases of the Kirkpatrick model (Kirkpatrick, 1994), which are: (i) evaluation of reaction, (ii) evaluation of learning, (iii) evaluation of behaviour, (iv) evaluation of results. However, this model was created to evaluate traditional tutoring systems and programs. It does not evaluate specific characteristics of an adaptive e-learning tutoring system, like accuracy of recommendations, usability, usefulness, interactions etc. Furthermore, the authors in Eryilmaz and Adabashi (2020) present an experimental study to evaluate the effectiveness of an intelligent tutoring system, which embeds artificial intelligence methods to support the higher student academic performance. They compared the developed tutoring system with other versions of it and used t-test to compare the different academic performance of the students, who used the systems. Moreover, a pilot study was conducted to evaluate the ability of an intelligent team tutoring system to provide feedback to positively influence team behaviour and improve team task performance (Ostrander et al., 2020). Two groups of 16 humans participated in the study, which included performance measuring and comparing through statistical t-test method, and a self-assessment survey through questionnaire. Also, in Kochmar et al. (2020) an experiment was conducted to measure the student's learning gain and check if it is improved by a tutoring system, which uses machine learning, to provide automated personalised feedback. Another experiment, which concerned the use of an intelligent tutoring system, that is called WinITS, by students of Hanoi National University, was described in Binh and Trung (2021). The aim of the experiment was to evaluate the learning effectiveness of a proposed student model that is based on learning styles. The participants completed a final test, after the use of the tutoring system, to evaluate their performance and the time they need to finish the test. The results were compared with the corresponding results of a group of students, who did not use WinITS. In addition, students, who used WinITS, completed a questionnaire to evaluate the effect of adaptation of the system to students.

Taking into account the above, we come up with the conclusion that the most common-used evaluation methods of an ITS are: questionnaires and experiments. Performance is the most frequent evaluated metric in the experiments. Other common-evaluated metrics are users' satisfaction and system's usability. Furthermore, experiments, usually, include measuring performance through pre-test and post-test and using statistical t-test method for the comparison of measurements' results of groups that used different versions of the evaluated system. However, in the literature review, there is not a widely approved evaluation framework and technique for the assessment of an ITS, especially since ITSs need to be evaluated concerning their intelligent features as well as their educational effectiveness and usability aspects. Therefore, after a thorough investigation in the literature review, we decided to perform the evaluation of the fuzzy-based ITS following well-known and accepted evaluation methodologies: the CIAO! framework (Jones et al., 1999) and the evaluation framework that was proposed by Lynch and Ghergulescu (2016). We chose to use these frameworks because the CIAO! framework was developed especially for the evaluation of general educational aspects of computer assisted learning systems and the Lynch and Chergulescu framework concerns the evaluation of adaptive and intelligent learning systems. Therefore, the combination of these two

evaluation frameworks is ideal for performing a thorough evaluation, which include the assessment of multiple aspect of the ITS.

3 Theoretical framework and methodology

The fuzzy-based tutoring systems embeds intelligent techniques for supporting the learning process. Therefore, its evaluation has to include both aspects that concern in general an educational software and aspects that concern its intelligent operation. To succeed it we combined two evaluation frameworks: the CIAO! framework (Jones et al., 1999), which evaluates in general aspects of a computer assisted learning (CAL) system, and the evaluation framework that was proposed by Lynch and Ghergulescu (2016), which evaluates aspects of adaptive and intelligent learning systems.

According to the CIAO! framework, the following three dimensions of a CAL system have to be evaluated:

1. The CAL aim and its context of use. This dimension is assessed through questionnaires, interviews and analyzing policy documents.
2. Interactions: Data that concern the learners' interaction with the CAL system. These data are gathered, measured, and analyzed through observations, audio and/or video recording, interactions recording and log files.
3. Attitudes and outcomes: Learning outcomes, students' performance and changes in students' perceptions and attitudes. For the evaluation of this dimension, questionnaires, interviews, and tests are used.

According to the evaluation framework of Lynch and Chergulescu, the following four criteria have to be assessed:

1. Learning and training: It concerns factors, such as learning outcomes, knowledge acquisition and learning improvements, that are related to the effectiveness and factors, such as number and duration of interactions needed to achieve the learning goal, which are related to the efficiency.
2. System: It concerns factors, such how accurate is the system grading in comparison to grading by physical teachers, how accurate are the predictive errors and the feedback, that are related to the accuracy of the student model and system recommendations.
3. User experience: It concerns the system usability and the learners' satisfaction.
4. Affective: It concerns learners' motivation and engagement in the learning process.

The combination of these two frameworks span more generic aspects that should be evaluated in an educational software that has the features of an Intelligent Tutoring Systems. From the combination of these evaluation frameworks, six evaluation criteria have arisen, namely (i) context, (ii) effectiveness, (iii) efficiency, (iv)

accuracy, (v) usability and satisfaction, and (vi) engagement and motivation. In this way, we accomplish to assess multiple aspects of the tutoring system that include the intelligent features as well as the necessary educational aspects. Table 1 presents the criteria of our evaluation model, how they are mapped to CIAO! and Lynch and Chergulescu evaluation frameworks, their metrics and the method that was chosen to evaluate them.

To apply the evaluation process we selected the participants and defined the experiment's conditions. Then, we delivered the tutoring system to participants and end of its usage we asked them to completed the questionnaires. Subsequently, we collected and analyzed the data either from the questionnaires or from the log files. The research methodology for the presented evaluation process is depicted in Fig. 1.

4 An overview of the fuzzy-based ITS

The fuzzy-based ITS that is evaluated is a web-based educational environment for personalized tutoring of computer programming (Chrysafiadi & Virvou, 2013c). The ITS dynamically adapts the lesson flow to the learner's learning needs. The system adaptation is based on the learner's current knowledge level and the knowledge dependencies that exist among the knowledge concepts of the learning material. The adaptation is realized through a fuzzy rule-based mechanism. This mechanism takes as input the learner's knowledge level and the knowledge dependencies among the domain concepts of the learning material and returns as output the learner's estimated knowledge level for each domain concept. Then, the lesson sequence definer module of the ITS takes into account the output of the fuzzy inference system and decides about the domain concepts that s/he has to study. The architecture of the fuzzy-based ITS is outlined in Fig. 2.

The domain knowledge of the system is separated into 31 chapters that concern the following concepts: declarations of variables and constants; expressions and operators; input and output expressions; the sequential execution of a program; the if-else statement; the iteration statements; sorting and searching algorithms; arrays and subprogramming. For the representation of the learner's knowledge level of each domain concept of the learning material, we use a quartet $(\mu_{Un}, \mu_{InK}, \mu_K, \mu_L)$, where μ_x is the value of the membership function of the fuzzy set x . Particularly, we use four fuzzy sets: (i) Unknown (Un), (ii) Insufficiently Known (InK), (iii) Known (K), and (iv) Learned (L). The membership function of each fuzzy set is trapezoid as in Fig. 3; Table 2. The input to the membership function is the learner's degree of success in the test of the corresponding knowledge domain concept. Therefore, for example, if a learner achieves 73/100 in the test of knowledge concept C5, then her/his knowledge level for C5 is described by the quartet $(0, 0.4, 0.6, 0)$, which means that her/his knowledge level belongs to 'Insufficiently Known' with 0.4 degree of membership and, simultaneously, it belongs to 'Known' with 0.6 degree of membership. Also, if a learner achieves 65/100 in the test of knowledge concept C2, then her/his knowledge level for C3 is described by the quartet $(0, 1, 0, 0)$, which means that her/his knowledge level belongs entirely to 'Insufficiently Known' fuzzy set.

Table 1 Fuzzy sets: linguistic values and trapezoidal membership functions

Criteria of our model	Criteria of Lynch and Ghergulescu (2016)	Criteria of CIAO!	Metrics	Method
Context	-	Context	The aims of educational software and the context of use	questionnaire
Effectiveness	learning & training	Outcomes	New knowledge learning and managing	t-test
Efficiency	learning & training	Outcomes-interactions	The mean number of interactions until completing a knowledge domain chapter & the mean number of interactions until achieving the learning goal	t-test
Accuracy	System	interactions	Accuracy of recommendations	Questionnaire
usability and Satisfaction	user experience	interactions	Easiness of use, satisfaction	Questionnaire
Engagement, motivation	affective direction	interactions	Number of learners that drop out, learners' interest and willingness to participate in the learning process	Questionnaire and t-test

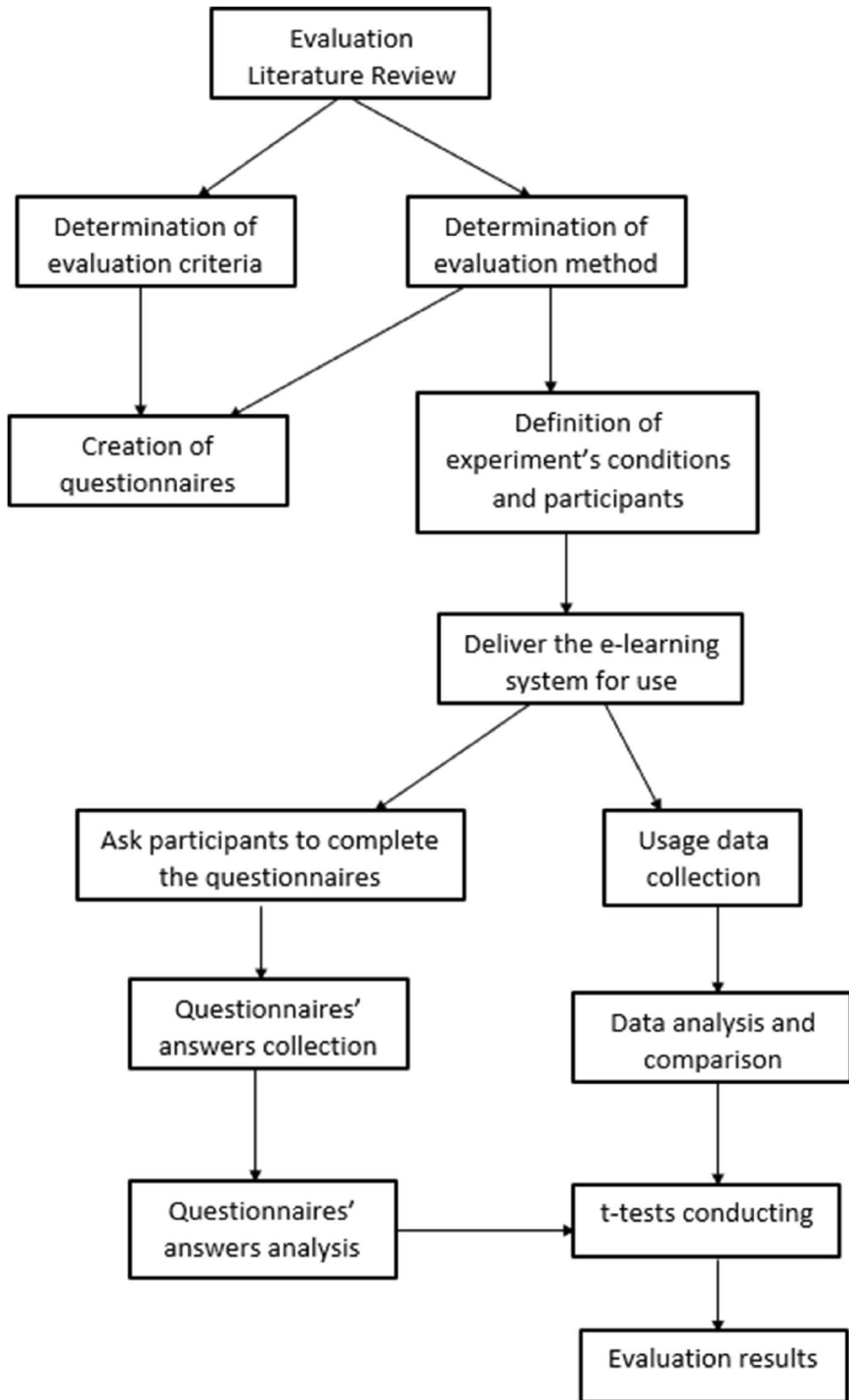


Fig. 1 The evaluation research methodology

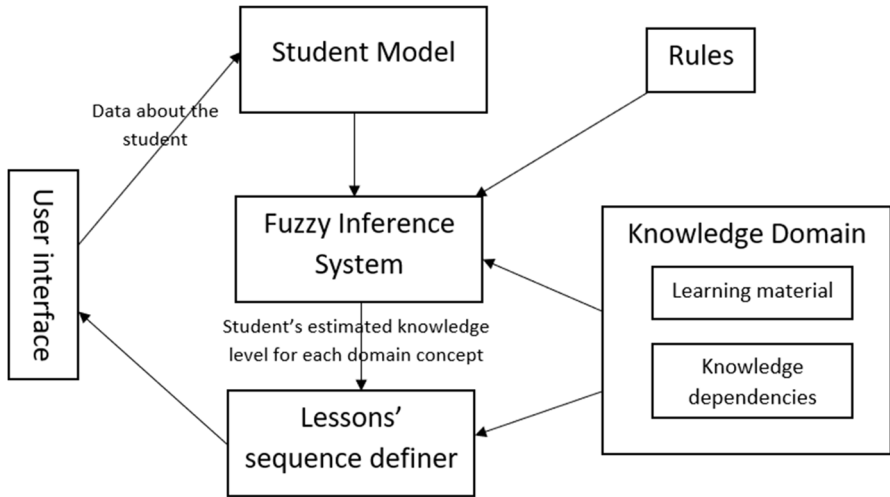


Fig. 2 The architecture of the fuzzy-based ITS

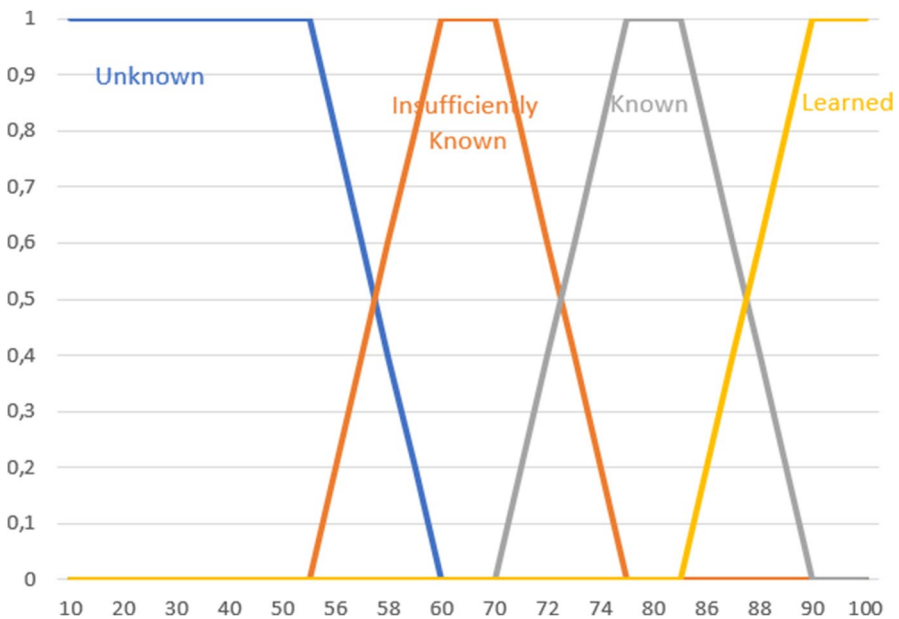


Fig. 3 Fuzzy sets partition

At the first interaction of the learner with the system, the ITS considers that s/he is ‘novice’ for all the knowledge domain concepts of the learning material and delivers to her/him the basic concepts of the learning material to study, which include variables, constants and operators. Then, s/he completes a test to assess the knowledge that s/he acquired. The learner’s grade in the test is used to identify the fuzzy

Table 2 Fuzzy sets: linguistic values and trapezoidal membership functions

Fuzzy set - linguistic value	Abbreviation	Membership function
Unknown	Un	(0, 0, 55, 60)
Insufficiently Known	InK	(55, 60, 70, 75)
Known	K	(70, 75, 85, 90)
Learned	L	(85, 90, 100, 100)

set (or sets) to which her/his knowledge level belongs concerning the knowledge domain C_i and to calculate the corresponding degree of membership. Then, the system takes into account the knowledge dependencies that exist among the domain concept C_i and the other domain concepts of the learning material, and applying a mechanism of fuzzy rules, updates the learner's knowledge level for all the related knowledge domain concepts of the learning material. The description of the system's mechanism of fuzzy rules is out of the scope of this paper and has been presented in a previous work of the first two authors (Chrysafiadi & Virvou, 2014). Thus, the system detects:

- the domain concepts of the learning material that are completely known to the learner and do not need study.
- the domain concepts of the learning material that are partially known to the learner and do need little additional study.
- the domain concepts of the learning material that are unknown to the learner and need careful additional study.
- the domain concepts of the learning material that have been forgotten and need significant revision.

Consequently, the presented fuzzy-based ITS detects changes in the learner's knowledge level at each interaction s/he has with the system, recognizes if her/his knowledge level increases or decreases, and decides about the most appropriate knowledge concepts of the learning material that have to be delivered to the learner. This is illustrated in Figs. 4 and 5, which present screenshots from the interaction with the system. On the other hand, Fig. 6 provides an overview of the major system components and their relationships.

5 Evaluation

5.1 Implementation

For the evaluation implementation, we followed the next phases:

- 1: Define the evaluation's goal.
- 2: Define the criteria.
- 3: Define the evaluation method.

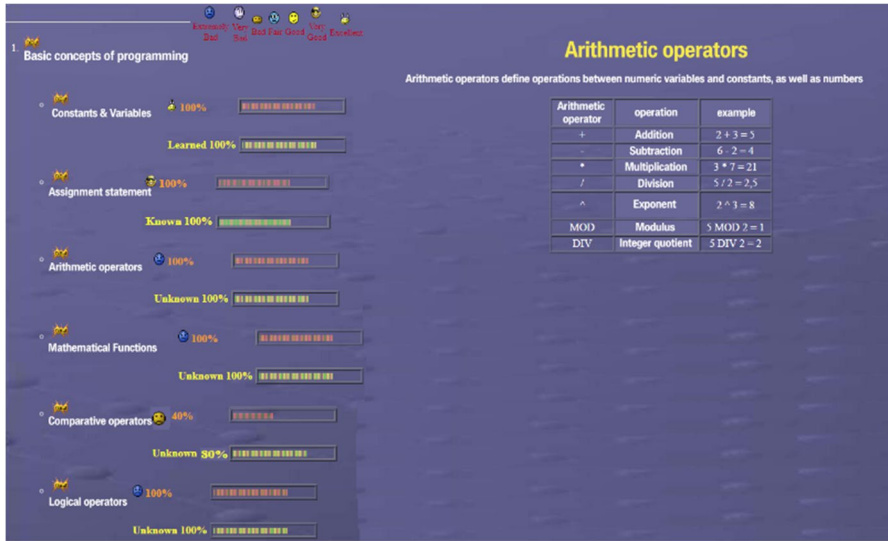
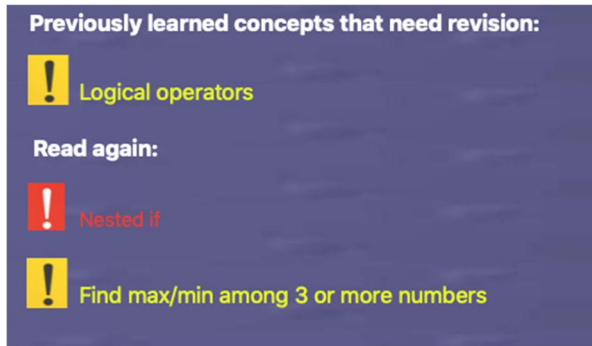


Fig. 4 A sample of a learner’s screen. The system uses different icons and words to inform her/him about her/his knowledge level

Fig. 5 System recommendation for concepts to study



- 3.1: Define the data collections methods.
- 3.2: Define the data analysis methods.

- 4: Define the experiment’s conditions.
- 5: Select the evaluators and identify roles and responsibilities.
- 6: Select the participants and record their characteristics.
- 7: Conduct the experiment.
- 8: Collect data.
- 9: Analyze data.
- 10: Export results.
- 11: Draw conclusions.

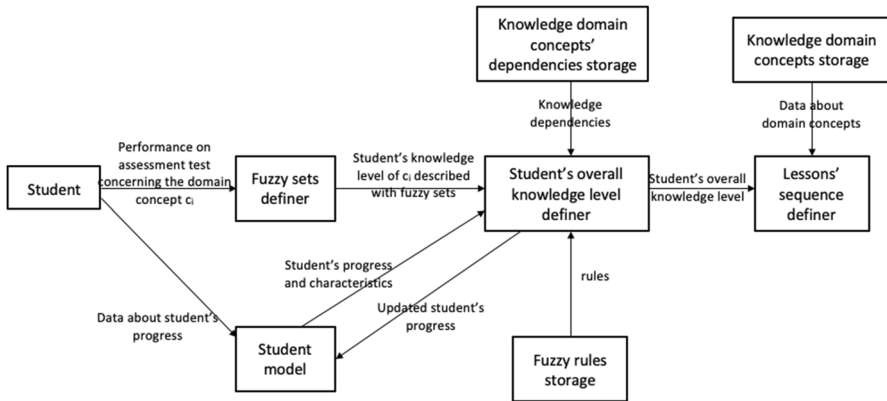


Fig. 6 The components of the system

5.2 The method

For the evaluation of the system, both questionnaires and experiments were used. In more detail, 70 learners of an undergraduate program in Informatics of the University of Piraeus in Greece (Group A) used the presented fuzzy-based ITS for a period of 6 weeks. After that period, they were asked to complete questionnaires and take a test which examined the knowledge on computer programming that they had acquired. Also, data were gathered via the system log files and records. Next, the results were compared with the corresponding answers and measures of Group B, which included 70 other learners in the same undergraduate program that used a system similar to the presented ITS, but in which the fuzzy mechanism was absent. For the comparison of the results, t-tests (Pallant, 2020) were used. In the following, the methods that were used to evaluate each criterion are presented in more details.

1. **Context:** We assessed the aims of educational software and the context of its use through the questionnaire of Table 3. It consists of close-end questions based on the Likert scale (Schrum et al., 2020) with five responses ranging from “not at all” (1) to “very much” (5).
2. **Effectiveness:** We evaluated new knowledge learning and managing through an experiment. Particularly, the participants and users of the presented tutoring system were asked to complete a test after a period of 6 weeks of system usage. Then, this mean was compared with the corresponding mean learner performance of a group of students who had not used the presented tutoring system. For the comparison, a t-test was employed.
3. **Efficiency:** We assessed how efficiency of the presented ITS in the learning process. In more detail, we calculated the mean number of interactions until completing a knowledge domain chapter and the mean number of interactions until achieving the learning goal for Group A and Group B and we compared these means via a t-test.

Table 3 Questionnaire for ‘context’ criterion evaluation

	Question	Answers			
1	How clear is the aim of the tutoring system?	1 – not at all	2	3	4
2	Does the tutoring system help you to learn computer programming?	1 – not at all	2	3	4
3	Does the tutoring system meet your expectations?	1 – not at all	2	3	4
4	Is the content of the tutoring system associated with the purpose of the learning process?	1 – not at all	2	3	4
5	Is the system environment helpful for learning computer programming?	1 – not at all	2	3	4
6	Evaluate your experience from using the tutoring system for learning computer programming.	1 – not at all	2	3	4

4. Accuracy: We assessed, through the questionnaire of Table 4, the accuracy of system recommendations made to each individual learner. The questionnaire includes close-end questions based on the Likert scale (Schrum et al., 2020) with five responses ranging from “not at all” (1) to “very much” (5).
5. Usability and satisfaction: We evaluated the learners’ satisfaction of the use of the presented fuzzy-based ITS, as well as the easiness of its use, through the questionnaire of Table 5. The questionnaire consists of close-end questions based on the Likert scale (Schrum et al., 2020) with five responses ranging from “not at all” (1) to “very much” (5).
6. Engagement and motivation: We assessed the learners’ interest and willingness to participate in the learning process, through the questionnaire of Table 6, which includes close-end questions based on the Likert scale (Schrum et al., 2020) with five responses ranging from “not at all” (1) to “very much” (5). Furthermore, we calculated the number of the learners that dropped out from usage of the system. Then, we compared this number with the corresponding number of Group B via a t-test.

5.3 The testbed

In the evaluation process, 140 learners participated who attended an undergraduate program in Informatics of the University of Piraeus, Greece. Specifically, they attended a computer programming class for a period of three months. All lectures in the class were presented in a classroom, with physical presence. After the completion of the lectures, the learners were divided into two, equal-size groups, namely: (i) Group A, which consisted of 70 learners who were asked to use the presented fuzzy-based ITS for a period of six weeks as a complementary tool for their education in computer programming, and (ii) Group B, which consisted of 70 learners who were asked to use, for the same period, a system similar to the presented ITS from which the fuzzy mechanism was absent. Before, the assignment of the systems to the learners, both systems were fully demonstrated to each group. Furthermore, learners of both groups were provided with detailed user manuals concerning the tutoring systems. Also, during the period of system usage, instructors were available to provide help to the learners. The learners’ characteristics and their distribution in the two groups are depicted in Tables 7 and 8.

5.4 Results and discussion

In this section the results of the evaluation process are presented per criterion and discussed.

5.4.1 Context

The mean answers to the questionnaire that concerns context assessment (Table 3) for both groups are presented in Table 9. We notice that all answers are similar for

Table 4 Questionnaire for ‘accuracy’ criterion evaluation

Question		Answers				
1	Was the system aware of your knowledge level at each interaction?	1 – not at all	2	3	4	5- very much
2	Was the system aware of your learning weaknesses and needs at each interaction?	1 – not at all	2	3	4	5- very much
3	Were the system reactions adjusted to your learning needs and misconceptions?	1 – not at all	2	3	4	5- very much
4	Did the system feedback help you realize your misconceptions and needs?	1 – not at all	2	3	4	5- very much
5	How helpful was the system feedback to the learning process?	1 – not at all	2	3	4	5- very much
6	Was the delivered learning material tailored to your knowledge level at each interaction?	1 – not at all	2	3	4	5- very much
7	How tailored to your needs were the system pieces of advice?	1 – not at all	2	3	4	5- very much
8	Did the system succeed to recognize your misconceptions and needs?	1 – not at all	2	3	4	5- very much
9	Was the error diagnosis valid?	1 – not at all	2	3	4	5- very much
10	Rate the adaptability of the system to your learning needs.	1 – not at all	2	3	4	5- very much

Table 5 Questionnaire for ‘usability and satisfaction’ criterion evaluation

	Question	Answers		
1	How much did you like the tutoring system?	1 – not at all	2	3
2	How positive your overall experience with the tutoring system was?	1 – not at all	2	3
3	Did the tutoring system cause your frustration?	1 – not at all	2	3
4	How usable is the tutoring system environment?	1 – not at all	2	3
5	How easy is the navigation in the tutoring system?	1 – not at all	2	3
6	How often did you seek help during your navigation in the tutoring system?	1 – not at all	2	3
7	How friendly is the tutoring system interface?	1 – not at all	2	3
8	How satisfied are you about the presentation of the learning material?	1 – not at all	2	3
9	Is the presented learning content of the system clear and straightforward?	1 – not at all	2	3
10	How satisfied are you about the feedback by the system?	1 – not at all	2	3
11	Is the system feedback clear and straightforward?	1 – not at all	2	3
12	Would you use the tutoring system again?	1 – not at all	2	3
13	Would you recommend the tutoring system to a friend?	1 – not at all	2	3

Table 6 Questionnaire for ‘engagement and motivation’ criterion evaluation

Question		Answers				
1	How interesting was the tutoring system?	1 – not at all	2	3	4	5- very much
2	How willing were you to use the tutoring system for learning computer programming?	1 – not at all	2	3	4	5- very much
3	Did you like to use the tutoring system for learning computer programming?	1 – not at all	2	3	4	5- very much
4	Will you use a similar tutoring system for learning any other knowledge domain?	1 – not at all	2	3	4	5- very much
5	Did the tutoring system motivate you to participate in the learning process?	1 – not at all	2	3	4	5- very much
6	Did the tutoring system contribute to the increase of your engagement in the learning process?	1 – not at all	2	3	4	5- very much
7	Was the time you spent using the tutoring system a waste of time?	1 – not at all	2	3	4	5- very much
8	Did the tutoring system help you to absorb the learning process?	1 – not at all	2	3	4	5- very much

Table 7 Participants' age and gender

	Age	18–20	21–22	22+
	Gender			
Group A	Female	25	5	2
	Male	26	9	3
Group B	Female	24	3	1
	Male	29	7	6

Table 8 Participants' experience in using computers, educational software and their background in computer programming

	Experience in using computers				
	none	low	medium	high	excellent
Group A	0	0	8	25	37
Group B	0	0	10	19	41
	Experience in using educational software				
	none	low	medium	high	excellent
Group A	14	30	18	7	1
Group B	12	27	23	8	0
	Previous Knowledge on computer programming				
	none	low	medium	high	excellent
Group A	18	32	19	1	0
Group B	15	38	14	3	0

Table 9 Learners' mean answers to the questionnaire concerning 'context'

No of question	Group A	Group B
1	4.37	4.27
2	3.96	3.83
3	3.87	3.71
4	4.53	4.5
5	3.97	3.79
6	3.93	3.8

the learners of both groups. Therefore, the integration of the fuzzy mechanism to the tutoring system does not affect its context.

5.4.2 Effectiveness

For the evaluation of the system effectiveness, a test was delivered to the participants, which included quizzes and exercises that concern computer programming. The test was given to the participants after they had used the corresponding tutoring system. The learners' grades on the test vary from 0 (lowest) to 100 (excellent). The results are presented in Table 10 We notice that the mean grade of Group A is higher than the corresponding mean grade of Group B. To ensure that the difference in the

Table 10 t-test results concerning the learners' performance

	Group A	Group B
Mean	8.314286	7.15
Variance	2.537474	2.531522
Observations	70	70
Pooled Variance	2.534498	
Hypothesized Mean Difference	0	
Df	138	
t Stat	4.326609	
P(T ≤ t) one-tail	1.44E-05	
t Critical one-tail	1.65597	
P(T ≤ t) two-tail	2.89E-05	
t Critical two-tail	1.977304	

learners' per-formance of the two groups was not caused by chance or due to differences in the characteristics of the participants, we compared the results through a t-test. As it is depicted in Table 10, the value of “P(T ≤ t) two-tail” is lower than 0.05. Therefore, the difference in means of performance between the two groups is statistically significant. As a consequence, the incorporated fuzzy mechanism makes the tutoring system more effective and improves the learners' performance. Therefore, the presented fuzzy-based ITS contributes significantly to new knowledge learning and managing.

5.4.3 Efficiency

The results concerning the t-tests that were conducted to calculate and compare (i) the mean number of interactions until completion a knowledge domain chapter and (ii) the mean number of interactions until achieving the learning goal, for Group A and Group B, are presented in Tables 11 and 12, respectively. We notice that both mean values for Group A are lower than the corresponding mean values for Group B. Furthermore, the value of “P(T ≤ t) two-tail” is lower than 0.05 for the comparison of both mean values, which means that the difference of both means is statistically significant. Therefore, the fuzzy-based ITS allows the learner to complete a knowledge domain chapter in a reduced interaction time. This is due to the fact that the fuzzy-based ITS succeeds in identifying the learning needs of each learner. Specifically, it identifies the chapters the learners known fully or partially, the chapters s/he does not know and the chapters s/he has forgotten. Thus, it provides the learner with the most suitable learning material to her/him and adapts the lesson flow accordingly. Consequently, learners succeed to complete all of the lessons and reach their target knowledge within a smaller number of interactions.

To analyze further the results concerning the number of interactions and time savings, we separated the participants in the experiment learners into three categories based on their background. Therefore, three categories of learners are derived: (i) learners with computer-related background, (ii) learners with background in hard

Table 11 t-test results concerning the mean number of interactions until completing a knowledge domain chapter

	Group A	Group B
Mean	3.128571429	4.571428571
Variance	2.548447205	2.915113872
Observations	70	70
Pooled Variance	2.731780538	
Hypothesized Mean Difference	0	
Df	138	
t Stat	-5.164573382	
P(T ≤ t) one-tail	4.12482E-07	
t Critical one-tail	1.655970382	
P(T ≤ t) two-tail	8.,24964E-07	
t Critical two-tail	1.977303542	

Table 12 t-test results concerning the mean number of interactions until reaching the target knowledge

	Group A	Group B
Mean	66.17142857	75.9
Variance	630.9556936	872.2072
Observations	70	70
Pooled Variance	751.58147	
Hypothesized Mean Difference	0	
Df	138	
t Stat	-2.099399367	
P(T ≤ t) one-tail	0.018801303	
t Critical one-tail	1.655970382	
P(T ≤ t) two-tail	0.037602607	
t Critical two-tail	1.977303542	

sciences (like math, physics, chemistry etc.), and (iii) learners with background in soft sciences (like pedagogy, philosophy, psychology etc.). Then we recorded the maximum number, the minimum number and the mean number of the following type of interactions with the system for all the three categories of learners (Tables 13 and 14). The types of interactions are:

- The interactions in total, which include reading of a concept, revision of a concept, completing the practice test of a concept, complete the assessment test of the concept.
- The total “local” revisions, which include the revisions of a concept that belongs to the learner’s current knowledge level.
- The total revisions of prerequisites, which include the revisions of a concept that the learner has previously learned in a precedent interaction and the system considered later that s/he has forgotten it.

Table 13 Number of interactions for participants of group A

	Interactions in total			Total “local” revisions			Total revisions of prerequisites		
	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean
Learners with computer-related background	66	35	48.92	15	0	5.78	6	0	2.08
Learners wit background in hard sciences	87	41	62.85	28	4	12.31	8	1	3.22
Learner’s wit background in soft sciences	124	52	86,74	48	9	24.78	12	2	5.46

Table 14 Number of interactions for participants of group B

	Interactions in total			Total “local” revisions			Total revisions of prerequisites		
	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean
Learners with computer-related background	79	48	61.5	17	0	6.91	3	0	0.86
Learners wit background in hard sciences	98	52	73.48	39	6	19.38	6	0	2.36
Learner’s wit background in soft sciences	123	65	92.72	64	17	38.42	8	1	3.27

Comparing the numbers in Tables 13 and 14, we concluded that the fuzzy mechanism leads to decrease in the total number of interactions, although increases the total revisions of prerequisites. This happens because the fuzzy-based system detects the concepts that the learner knows, and it does not deliver them to her/him for reading. Also, we noticed that the decrease in interactions’ numbers are greater for the learners with computer-based background. Therefore, the sequence of lessons is more tailored to the students’ learning needs and contributes to provide effective learning results in less time.

5.4.4 Accuracy

The mean answers to the questionnaire that concerns the assessment of the system accuracy for Group A and Group B are presented in Table 15. We notice that the mean answers of the learners of Group A are higher than the corresponding mean answers of Group B for all the questions. To ensure that the differences in the answers of the two groups are statistically valid, t-tests were conducted. The results of the t-tests are presented in Table 16. The “P(T <= t) two-tail” value, which reveals if the difference between the means is statistically significant, is lower than 0.05 for all the questions. Therefore, the fuzzy mechanism contributes to more accurate system recommendations.

5.4.5 Usability and satisfaction

The mean answers to the questionnaire that concerns the evaluation of the systems usability and the users’ satisfaction are presented in Table 17. We notice that the

Table 15 Learners' mean answers to the questionnaire concerning 'accuracy'

No of question	Group A	Group B
1	4.36	3.56
2	4.31	3.5
3	4.33	3.51
4	4.34	3.51
5	4.37	3.54
6	4.36	3.56
7	4.26	3.46
8	4.3	3.51
9	4.27	3.44
10	4.29	3.49

mean answers to questions 1, 2, 4, 5, 6, 7, 8, 9 and 11 do not differ significantly between the two groups. However, differences in the mean answers of the two groups are observed in questions 3, 10, 12 and 13. For these questions, the mean answers of Group A are better than the corresponding mean answers of Group B. To ensure the statistical validity of differences in these questions, t-tests were conducted. According to the t-test results (Table 18) the “ $P(T \leq t)$ two-tail” value is lower than 0.05 for all four questions (i.e., 3, 10, 12, 13). This indicates that the differences are statistically significant. Consequently, the ability is higher of the presented fuzzy-based ITS to recognize a learner's knowledge level and learning needs and to adapt the lesson flow on the fly to better satisfy the learners.

5.4.6 Engagement and motivation

The mean answers to the questionnaire that concern the evaluation of the learners' engagement and motivation are presented in Table 19. We notice that the mean answers to all of the questions, except of questions 4 and 7, do not differ between the two groups. To ensure the statistical validity of differences in questions 4 and 7, t-tests were conducted. The “ $P(T \leq t)$ two-tail” value of t-test has to be lower than 0.05 to indicate that the difference in the means is statistically valid. We noticed that the “ $P(T \leq t)$ two-tail” value is, indeed, lower than 0.05 for both questions 4 and 7, as shown in Table 20. Therefore, the higher mean answers of the learners of Group A to question 4 indicates that the adaptation of the tutoring system, which is based on a fuzzy logic mechanism, creates more positive feelings to the learners and contributes to a greater acceptance of the ITS. Furthermore, the lower mean answers of the learners of Group A to the question when compared with that of Group B indicates that the presented fuzzy-based ITS allows learners to complete the lessons in less time as it appears to have the ability to identify at each interaction which chapters the learner knows fully or partially, which chapters s/he does not know and which chapters s/he has forgotten, and to adapt the lesson flow accordingly. As a result, the learners remain more engaged to the learning process.

Table 16 t-test results of the learners' mean answers concerning the 'accuracy' criterion

	Q1		Q2		Q3		Q4		Q5	
	group A	group B	group A	group B	group A	group B	group A	group B	group A	group B
Mean	4.357	3.557	4.31	3.5	4.329	3.514	4.343	3.514	4.371	3.543
Variance	0.494	0.511	0.509	0.457	0.514	0.456	0.489	0.456	0.469	0.426
Observations	70	70	70	70	70	70	70	70	70	70
Pooled Variance	0.503		0.483		0.485		0.473		0.447	
Hypothesized Mean Difference	0		0		0		0		0	
Df	138	138	138	138	138	138	138	138	138	138
t Stat	6.677	6.677	6.935	6.935	6.917427	6.917427	7.128	7.128	7.330	7.330
P(T <= t) one-tail	2.75E-10	2.75E-10	7.16329E-11	7.16329E-11	7.86E-11	7.86E-11	2.59E-11	2.59E-11	8.81E-12	8.81E-12
t Critical one-tail	1.656	1.656	1.656	1.656	1.656	1.656	1.656	1.656	1.656	1.656
P(T <= t) two-tail	5.49E-10	5.49E-10	1.43266E-10	1.43266E-10	1.57E-10	1.57E-10	5.17E-11	5.17E-11	1.76E-11	1.76E-11
t Critical two-tail	1.977	1.977	1.977	1.977	1.977	1.977	1.977	1.977	1.977	1.977
	Q6	Q7	Q8	Q9	Q10					
Mean	group A 4.357	group B 3.557	group A 4.257	group B 3.457	group A 4.3	group B 3.514	group A 4.271	group B 3.443	group A 4.286	group B 3.486
Variance	0.494	0.511	0.513	0.397	0.474	0.427	0.490	0.395	0.526	0.427
Observations	70	70	70	70	70	70	70	70	70	70
Pooled Variance	0.502		0.455		0.451		0.442		0.477	
Hypothesized Mean Difference	0		0		0		0		0	
Df	138	138	138	138	138	138	138	138	138	138
t Stat	6.677	6.677	7.019	7.019	6.924	6.924	7.366	7.366	6.856	6.856
P(T <= t) one-tail	2.75E-10	2.75E-10	4.61E-11	4.61E-11	7.57E-11	7.57E-11	7.26E-12	7.26E-12	1.09E-10	1.09E-10
t Critical one-tail	1.656	1.656	1.656	1.656	1.656	1.656	1.656	1.656	1.656	1.656
P(T <= t) two-tail	5.49E-10	5.49E-10	9.22E-11	9.22E-11	1.51E-10	1.51E-10	1.45E-11	1.45E-11	2.17E-10	2.17E-10
t Critical two-tail	1.977	1.977	1.977	1.977	1.977	1.977	1.977	1.977	1.977	1.977

In addition, the number of the learners that dropped out from the usage of the system was calculated for both groups. The percentage of the dropout learners of Group A is 14.29%. The corresponding percentage for Group B is 32.86%. Then, a t-test was conducted to compare the two different means and ensure the statistical validity of the difference between the two means. The results of the t-test are presented in Table 21. We notice that the “P(T ≤ t) two-tail” value is lower than 0.05. Therefore, the difference in the number of learners that dropout from the usage of the tutoring system is statistically significant. Therefore, the presented fuzzy-based ITS increases the learners’ engagement.

6 Discussion and implication

Evaluation results showed that the fuzzy-based tutoring system impacts positively the learning outcomes and the educational process. The intelligent features of the system, which are supported by a fuzzy-based mechanism, makes it effective and helpful. More specifically, the system is able to recognize the following parts in relation to the students’ knowledge and needs:

- (i) the chapters, in which the learners has misconceptions,
- (ii) the chapters, which needs revision,
- (iii) the chapters, which the learner has forgotten.
- (iv) the chapters that the learner already knows and do not need read,

As such, the fuzzy-based system was found to have successfully made more accurate and content-relevant recommendations about the flow of the lessons than the educational system without fuzzy reasoning. Therefore, in comparison of the two

Table 17 Learners’ mean answers to the questionnaire concerning ‘usability and satisfaction’

No of question	Group A	Group B
1	4.13	3.91
2	4.1	3.86
3	2.13	2.74
4	4.46	4.43
5	4.47	4.5
6	1.36	1.34
7	4.19	4.2
8	4.17	4.14
9	4.2	4.17
10	4.3	3.64
11	4.26	4.21
12	4.09	3.79
13	4.11	3.84

The mean answers that are obviously different between the two groups and the corresponding number of the questions are presented in bold

Table 18 t-test results of the learners' mean answers concerning the 'usability and satisfaction' criterion

	Q3		Q10		Q12		Q13	
	group A	group B	group A	group B	group A	group B	group A	group B
Mean	2.129	2.743	4.3	3.642	4.086	3.786	4.114	3.843
Variance	0.577	1.092	0.503	0.784	0.630	0.751	0.624	0.656
Observations	70	70	70	70	70	70	70	70
Pooled Variance	0.835		0.643		0.690		0.640	
Hypothesized Mean Difference	0		0		0		0	
Df	138		138		138		138	
t Stat	-3.977		4.847		2.136		2.007	
P(T <=t) one-tail	5.60073E-05		1.66199E-06		0.017		0.023	
t Critical one-tail	1.656		1.656		1.656		1.656	
P(T <=t) two-tail	0.00011		3.32399E-06		0.034		0.047	
t Critical two-tail	1.977		1.977		1.977		1.977	

Table 19 Learners' mean answers to the questionnaire concerning 'engagement and motivation'

No of question	Group A	Group B
1	4.12	3.99
2	4.13	3.99
3	4.12	3.99
4	4.14	3.77
5	4.01	3.97
6	4.06	4.03
7	2.63	3.31
8	4.01	3.99

The mean answers that are obviously different between the two groups and the corresponding number of the questions are presented in bold

Table 20 t-test results of the learners' mean answers concerning the questions 4 and 7 of 'usability and satisfaction' criterion

	Q4		Q7	
	group A	group B	group A	group B
Mean	4.143	3.771	2.629	3.314
Variance	0.617	0.729	1.628	1.871
Observations	70	70	70	70
Pooled Variance	0.673		1.749	
Hypothesized Mean Difference	0		0	
Df	138		138	
t Stat	2,678		-3.067	
P(T<=t) one-tail	0.004		0.001	
t Critical one-tail	1.656		1.656	
P(T<=t) two-tail	0.0083		0,0026	
t Critical two-tail	1.977		1.977	

systems, the learners of the fuzzy-based system achieved their respective learning goals with a lower number of interactions and in less time. Furthermore, the adaptive sequence of lessons is found to be very efficient regarding the educational benefits of students. This efficiency creates a better user experience, which is showed by the lower number of students' dropouts. In particular, the fact that the system identifies the individual learners' needs and misconceptions, which in turn results in the presentation of sequences of lessons that are more tailored to them, leads learners to better and quicker performances. This creates positive feelings to the learners and prevents them from dropping out.

The findings of the presented study are very important for the fuzzy-based ITS. They underline that the use of fuzzy logic in managing learner's knowledge and modeling the educational process, enhances the learning outcomes. As a consequence, the findings of this research contribute to the design and development educational software and applications that provide individually tailored and more efficient educational support to learners.

Table 21 t-test results concerning the number of dropped out learners

	Group A	Group B
Mean	0.142857	0.328571
Variance	0.124224	0.22381
Observations	70	70
Pooled Variance	0.174017	
Hypothesized Mean Difference	0	
Df	138	
t Stat	-2.63381	
P(T ≤ t) one-tail	0.004704	
t Critical one-tail	1.65597	
P(T ≤ t) two-tail	0.009407	
t Critical two-tail	1.977304	

7 Conclusion

In this paper, we presented a multispect evaluation of a fuzzy-based ITS that teaches computer programming. The evaluated ITS employs a fuzzy mechanism to identify the learners' current knowledge level and learning needs and decide about and adapt the lesson flow accordingly. For the evaluation, we assessed six criteria that arose as combination of two evaluation frameworks for computer-based tutoring systems, namely: the CIAO! framework (Jones et al., 1999) and the evaluation framework that was proposed by Lynch and Ghergulescu (2016). The criteria include: (i) context, (ii) effectiveness, (iii) efficiency, (iv) accuracy, (v) usability and satisfaction, and (vi) engagement and motivation. The criteria were evaluated through questionnaires, experimental research and log file analysis. Seventy (70) students of an undergraduate program in Informatics at the University of Piraeus, Greece used the fuzzy-based ITS under real conditions and for a period of six weeks. The data of the system's usage were compared with the corresponding data of the usage of a similar ITS, from which the fuzzy mechanism was absent. This other ITS was used by a group of another seventy (70) students in the same undergraduate program. The validity of the differences in the evaluation results between the two groups of participants was certified through conducting t-tests.

The evaluation results are positive and very significant for the learning process. Particularly, the fuzzy-based ITS significantly improves the learners' performance and allows the learners to complete the lessons and reach the learning goal with fewer number of interactions with the system. Also, it improves the accuracy of identifying the students' learning needs detection and the system recommendations. Furthermore, it diminishes the number of learners that dropout of the tutoring system usage. Therefore, it encourages learners to remain more engaged in the learning process. Finally, the system enhances the learners' satisfaction. The findings of this study show that the use of fuzzy logic in the learning process modeling helps significantly the educational process. They provide important insight to the designers of educational software and applications, and to researchers, who deal with application of intelligent techniques into educational software and tutoring systems.

In the future, we will conduct further evaluations, in which additional learners participate from various programs of study. We will also apply the fuzzy-based ITS in other educational fields, besides computer programming, and compare the effectiveness of the presented fuzzy-based ITS with other ITSs. These and other related research avenues are currently being followed and the corresponding results will be announced elsewhere in the near future.

Funding Open access funding provided by HEAL-Link Greece.

Data availability All data generated or analysed during this study are included in this published article.

Declarations

Conflict of interest None.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Akyuz, Y. (2020). Effects of intelligent tutoring systems (ITS) on personalized learning (PL). *Creative Education*, 11(6), 953–978. <https://doi.org/10.4236/ce.2020.116069>
- Alonso-Secades, V., López-Rivero, A. J., Martín-Merino-Acero, M., Ruiz-García, M. J., & Arranz-García, O. (2022). Designing an intelligent virtual educational system to improve the efficiency of primary education in developing countries. *Electronics*, 11(9), 1487.
- Binh, H. T., & Trung, N. Q. (2021). Responsive student model in an intelligent tutoring system and its evaluation. *Education and Information Technologies*, 26(4), 4969–4991. <https://doi.org/10.1007/s10639-021-10485-4>
- Cho, Y., & Kim, J. (2021). Production of mobile english language teaching application based on text interface using deep learning. *Electronics*, 10(15), 1809.
- Chrysafiadi, K., & Virvou, M. (2013a). Student modeling approaches: A literature review for the last decade. *Expert Systems with Applications*, 40(11), 4715–4729. <https://doi.org/10.1016/j.eswa.2013.02.007>
- Chrysafiadi, K., & Virvou, M. (2013b). PeRSIVA: An empirical evaluation method of a student model of an intelligent e-learning environment for computer programming. *Computers & Education*, 68, 322–333. <https://doi.org/10.1016/j.compedu.2013.05.020>
- Chrysafiadi, K., & Virvou, M. (2013c). Dynamically personalized e-training in computer programming and the language C. *IEEE Transactions on Education*, 56(4), 385–392.
- Chrysafiadi, K., & Virvou, M. (2014). Fuzzy logic for adaptive instruction in an e-learning environment for computer programming. *IEEE Transactions on Fuzzy Systems*, 23(1), 164–177. <https://doi.org/10.1109/TFUZZ.2014.2310242>

- Chrysafiadi, K., & Virvou, M. (2021a). Evaluating the user experience of a fuzzy-based Intelligent Tutoring System. In *2021 12th International Conference On Information, Intelligence, Systems & Applications (IISA)* (pp. 1–7). IEEE.
- Chrysafiadi, K., & Virvou, M. (2021b). Evaluating the learning outcomes of a fuzzy-based Intelligent Tutoring System. In *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)* (pp. 1392–1397). IEEE.
- Chrysafiadi, K., Papadimitriou, S., & Virvou, M. (2022). Cognitive-based adaptive scenarios in educational games using fuzzy reasoning. *Knowledge-Based Systems*, 109111. <https://doi.org/10.1016/j.knsys.2022.109111>.
- Chughtai, R., Zhang, S., & Craig, S. D. (2015, September). Usability evaluation of intelligent tutoring system: ITS from a usability perspective. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 59, No. 1, pp. 367–371). SAGE Publications.
- Clancey, W. J., & Hoffman, R. R. (2021). Methods and standards for research on explainable artificial intelligence: Lessons from intelligent tutoring systems. *Applied AI Letters*, 2(4), e53.
- Effenberger, T., & Pelánek, R. (2021, April). Validity and reliability of student models for problem-solving activities. In *LAK21: 11th International Learning Analytics and Knowledge Conference* (pp. 1–11).
- Erümit, A. K., & Çetin, İ. (2020). Design framework of adaptive intelligent tutoring systems. *Education and Information Technologies*, 25(5), 4477–4500. <https://doi.org/10.1007/s10639-020-10182-8>
- Eryılmaz, M., & Adabashi, A. (2020). Development of an intelligent tutoring system using bayesian networks and fuzzy logic for a higher student academic performance. *Applied Sciences*, 10(19), 6638. <https://doi.org/10.3390/app10196638>
- Feng, S., Magana, A. J., & Kao, D. (2021, October). A systematic review of literature on the effectiveness of intelligent tutoring systems in STEM. In *2021 IEEE Frontiers in Education Conference (FIE)* (pp. 1–9). IEEE.
- Graesser, A. C., Hu, X., & Sottolare, R. (2018). Intelligent tutoring systems. *International handbook of the learning sciences* (pp. 246–255). Routledge.
- Greer, J., & Mark, M. (2016). Evaluation methods for intelligent tutoring systems revisited. *International Journal of Artificial Intelligence in Education*, 26(1), 387–392. <https://doi.org/10.1007/s40593-015-0043-2>
- Hosseini, R., Akhuseyinoglu, K., Brusilovsky, P., Malmi, L., Pollari-Malmi, K., Schunn, C., & Sirkiä, T. (2020). Improving engagement in program construction examples for learning Python programming. *International Journal of Artificial Intelligence in Education*, 30(2), 299–336. <https://doi.org/10.1007/s40593-020-00197-0>
- Jones, A., Scanlon, E., Tosunoglu, C., Morris, E., Ross, S., Butcher, P., & Greenberg, J. (1999). Contexts for evaluating educational software. *Interacting with Computers*, 11(5), 499–516. [https://doi.org/10.1016/S0953-5438\(98\)00064-2](https://doi.org/10.1016/S0953-5438(98)00064-2)
- Khazanchi, R., & Khazanchi, P. (2021). Artificial intelligence in education: A closer look into intelligent tutoring systems. *Handbook of research on critical issues in special education for school rehabilitation practices* (pp. 256–277). IGI Global.
- Kirkpatrick, D. (1994). *Evaluating training programs*. Berrett-Koehler.
- Kochmar, E., Vu, D. D., Belfer, R., Gupta, V., Serban, I. V., & Pineau, J. (2020, July). Automated personalized feedback improves learning gains in an intelligent tutoring system. In *International Conference on Artificial Intelligence in Education* (pp. 140–146). Springer.
- Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: a meta-analytic review. *Review of Educational Research*, 86(1), 42–78. <https://doi.org/10.3102/0034654315581420>
- Lampropoulou, P. S., Lampropoulos, A. S., & Tsihrintzis, G. A. (2010, October). A framework for evaluation of middleware systems of mobile multimedia services. In *2010 IEEE International Conference on Systems, Man and Cybernetics* (pp. 1041–1045). IEEE.
- Lynch, T., & Ghergulescu, I. (2016, November). An evaluation framework for adaptive and intelligent tutoring systems. In *E-learn: world conference on e-learning in corporate, government, healthcare, and higher education* (pp. 1385–1390). Association for the Advancement of Computing in Education (AACE).
- Mousaviniasab, E., Zarifsanaiy, N., Niakan Kalhori, R., Rakhshan, S., Keikha, M., & Ghazi Saedi, M. (2021). Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods. *Interactive Learning Environments*, 29(1), 142–163. <https://doi.org/10.1080/10494820.2018.1558257>

- Nesbit, J. C., Adesope, O. O., Liu, Q., & Ma, W. (2014, July). How effective are intelligent tutoring systems in computer science education?. In *2014 IEEE 14th international conference on advanced learning technologies* (pp. 99–103). IEEE.
- Ostrander, A., Bonner, D., Walton, J., Slavina, A., Ouverson, K., Kohl, A., & Winer, E. (2020). Evaluation of an intelligent team tutoring system for a collaborative two-person problem: Surveillance. *Computers in Human Behavior*, *104*, 105873. <https://doi.org/10.1016/j.chb.2019.01.006>
- Paladines, J., & Ramírez, J. (2020). A systematic literature review of intelligent tutoring systems with dialogue in natural language. *IEEE Access: Practical Innovations, Open Solutions*, *8*, 164246–164267. <https://doi.org/10.1109/ACCESS.2020.3021383>
- Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. Routledge.
- Polson, M. C., & Richardson, J. J. (2013). *Foundations of intelligent tutoring systems*. Psychology Press.
- Rebolledo-Mendez, G., Huerta-Pacheco, N. S., Baker, R. S., & du Boulay, B. (2022). Meta-affective behaviour within an intelligent tutoring system for mathematics. *International Journal of Artificial Intelligence in Education*, *32*(1), 174–195. <https://doi.org/10.1007/s40593-021-00247-1>
- Sáiz-Manzanares, M. C., Marticorena-Sánchez, R., & Ochoa-Orihuel, J. (2021). Using advanced learning technologies with university students: An analysis with machine learning techniques. *Electronics*, *10*(21), 2620.
- Schrum, M. L., Johnson, M., Ghuy, M., & Gombolay, M. C. (2020, March). Four years in review: Statistical practices of likert scales in human-robot interaction studies. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 43–52).
- Singh, N., Gunjan, V. K., Mishra, A. K., Mishra, R. K., & Nawaz, N. (2022). SeisTutor: A custom-tailored intelligent tutoring system and sustainable education. *Sustainability*, *14*(7), 4167. <https://doi.org/10.3390/su14074167>
- Sosnovsky, S., & Brusilovsky, P. (2015). Evaluation of topic-based adaptation and student modeling in QuizGuide. *User Modeling and User-Adapted Interaction*, *25*(4), 371–424. <https://doi.org/10.1007/s11257-015-9164-4>
- Sotiropoulos, D. N., Alepis, E., Kabassi, K., Virvou, M. K., Tsihrintzis, G. A., & Sakkopoulos, E. (2019). Artificial immune system-based learning style stereotypes. *International Journal on Artificial Intelligence Tools*, *28*(04), 1940008. <https://doi.org/10.1142/S0218213019400086>
- Tsihrintzis, G. A., Virvou, M., & Hatzilygeroudis, I. (2021). Special collection of extended selected papers on “Novel Research Results Presented in The 12th International Conference on Information, Intelligence, Systems and Applications (IISA2021), 12–14 July 2021, Chania, Crete. *Intelligent Decision Technologies*, *15*(4).
- Tsihrintzis, G. A., Virvou, M., & Phillips-Wren, G. (2019). Editorial Note to Surveys in artificial intelligence-based technologies. *Intelligent Decision Technologies*, *13*(4), 393–394.
- Urdaneta-Ponte, M. C., Mendez-Zorrilla, A., & Oleagordia-Ruiz, I. (2021). Recommendation systems for education: systematic review. *Electronics*, *10*(14), 1611.
- VanLehn, K., Banerjee, C., Milner, F., & Wetzel, J. (2020). Teaching Algebraic model construction: a tutoring system, lessons learned and an evaluation. *International Journal of Artificial Intelligence in Education*, *30*(3), 459–480. <https://doi.org/10.1007/s40593-020-00205-3>
- Virvou, M., Alepis, E., Tsihrintzis, G. A., & Jain, L. C. (2020). Machine learning paradigms. *Machine learning paradigms* (pp. 1–5). Springer.
- Wambsgans, T., Söllner, M., & Leimeister, J. M. (2020). Design and evaluation of an adaptive dialog-based tutoring system for argumentation skills. In *International Conference on Information Systems (ICIS)*.-Hyderabad, India.
- Wang, T. H., Lin, H. C. K., Chen, H. R., Huang, Y. M., Yeh, W. T., & Li, C. T. (2021). Usability of an affective emotional learning tutoring system for mobile devices. *Sustainability*, *13*(14), 7890. <https://doi.org/10.3390/su13147890>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Konstantina Chrysafiadi¹  · Maria Virvou¹ · George A. Tsihrintzis¹ · Ioannis Hatzilygeroudis²

Maria Virvou
mvirvou@unipi.gr

George A. Tsihrintzis
geoatsi@unipi.gr

Ioannis Hatzilygeroudis
ihatz@ceid.upatras.gr

¹ Department of Informatics, University of Piraeus, 80, M. Karaoli & A. Dimitriou St, 18534 Piraeus, Greece

² Department of Computer Engineering and Informatics, School of Engineering, University of Patras, B' Building, University Campus, 26504 Patras, Greece