



Article

Implementation of Fuzzy Functions Aimed at Fairer Grading of Students' Tests

Vanya Ivanova ^{1,*}  and Boyan Zlatanov ^{2,*} 

¹ Department of Education in Mathematics, Informatics and Information Technology, Faculty of Mathematics and Informatics, University of Plovdiv "Paisii Hilendarski", Plovdiv 4000, Bulgaria

² Department of Mathematical Analysis, Faculty of Mathematics and Informatics, University of Plovdiv "Paisii Hilendarski", Plovdiv 4000, Bulgaria

* Correspondence: vantod@uni-plovdiv.bg (V.I.); bobbyz@uni-plovdiv.bg (B.Z.)

Received: 27 June 2019; Accepted: 9 August 2019; Published: 12 August 2019



Abstract: The main goal of this research is to achieve fairer assessment of students' knowledge and skills in tests in view of the fact that learners with similar abilities are often assigned different test grades if their achievements are borderline cases between two grades. As an example, a student who has scored 29 points out of 60 in a test will fail, while another one with 30 or 31 points will pass, which can be viewed as unjust. This problem is addressed by using fuzzy logic and different fuzzy functions to determine whether students with borderline test scores are assigned the higher or the lower grade of the two. Furthermore, although changing the grades of individual students, the authors endeavor to preserve the overall score of all test takers statistically unchanged.

Keywords: fuzzy logic; fuzzy sets; evaluation; assessment; testing

1. Introduction

The education process at the Faculty of Mathematics and Informatics (FMI) at the University of Plovdiv is divided into three trimesters of 10 weeks each per academic year. The students from most majors study English only for two trimesters during their first year at university with a total of 100 seminars. In the course of the English language education students at FMI undergo a number of written examinations. First of all, they sit for a placement test, which determines the language level of each individual learner and, based on their results, students are redistributed from academic groups into language ones. The placement test scores are essential because they ensure the homogeneity of the language groups of learners with similar abilities. Then, students at FMI have classes once a week during which they use web-based materials and work on various projects. For homework, they do online tests in accordance with the exact teaching material covered during the seminars. These are self-study tests, created by the teacher, which are founded on criteria ranging from a lower to a higher level of difficulty [1]. Students take the self-study tests in their own place and time within the week before their next seminars in English at FMI. The grading of each test contributes towards the final grade that students are assigned at the end of the English course; for this reason, it is crucial that the students' tests are evaluated fairly and impartially. Besides the placement test and the self-study tests, students also sit for Midterm and Final exams during their course in English.

Fair and impartial assessment of the students' knowledge and skills is essential not only when high-stakes evaluations are concerned, for instance tests that determine whether a learner is eligible to enroll in or graduate from a school, but also for low-stakes ones. Assessments can help students to gain insight into their own achievements as well as to build confidence in their ability to learn and remain positive and motivated to continue studying. Since the global work culture is changing, students will

have to become life-long learners to remain competitive on the job market. Assessment plays a key role in developing the learners' confidence in their ability to learn and skills to do it throughout their lives.

The main reason to create and administer test examinations to students is to measure their understanding of specific content or the effective application of critical thinking skills. As a consequence, tests evaluate student learning and the level of their academic achievements at a specific stage of the course of studies. Online tests provide a number of benefits; select-response tests such as multiple-choice or true/false ones immediately provide students with their final scores, while open questions allow students to improve their writing by drafting, editing, and rewriting their texts, if necessary. Moreover, the data that tests provide, gathered over time, can help instructors to identify trends and improve their own teaching. In conclusion, assessment allows teachers to collect, record, score, and interpret information about learning.

Some of the advantages of self-study tests are that they support learners of different ages and different abilities; they help them to memorize the content for a longer time and increase student achievement [2,3]. Testing not only registers the student's level of knowledge but also changes it, thus improving the long-term retention of information, which is known as the testing effect [4]. Self-study tests provide feedback to both the student and the teacher and in this way, they allow for adjustments to the teaching process in order to achieve better final results.

One of the most important indicators of the quality of education is the result from it presented as a grade based on some assessment of a learner's knowledge and skills. In addition to fulfilling a specific function, the assessment grades must be objective and unbiased; they should indicate the achievement of certain significant results and be understandable for the learners, who have to be informed in advance about the evaluation method and its criteria. Moreover, the grades themselves must not be turned into a goal; they must inform about the student's progress, and the methods for calculating the grade have to be objective and statistically valid.

Nevertheless, students with similar abilities may be assigned different grades if their achievements are borderline cases, for example, a student who has scored 29 points out of 60 will probably fail, while another one with 30 or 31 points will pass the test, which can be viewed as unfair. In an attempt to find a reasonable and impartial solution to this problem, we have used fuzzy logic to review the results of a specific test taken by 78 first-year students of Informatics at FMI. The test assessed the students' reading and writing skills in English based on the learning materials and it comprised 61 items, 60 of which were closed-ended questions and one open-ended. The closed-ended items were multiple choice questions with four response options, each one awarding one point for a correct answer, while the open question required students to write a long-text answer to a question and it granted a maximum of 20 points. All the open-ended questions were evaluated by the same teacher following explicit predetermined criteria.

Fuzzy logic has a number of applications in analyzing studies from various areas of education as can be seen from Reference [5]. Research has been made previously endeavoring to incorporate fuzzy logic and education [6] or to use fuzzy logic as a tool for assessing students' knowledge and skills collectively [7]. However, instead of assessing student groups' knowledge and skills by means of fuzzy logic, we have used fuzzy functions to reconsider students' individual assessments in cases where there may be doubts of unfair grading.

2. Fuzzy Sets

Uncertainty in some test-related conclusions may appear due to vagueness or imprecision rather than to randomness, and it is frequently encountered in the evaluation of student results. The idea of fuzzy sets, which was introduced by Lofti Zadeh [8], can solve this problem to some extent. A typical example may be presented by the following two sets: set A = {if a man is taller than or equal to 1.85 m, we consider him a tall person}; set B = {if a man is less than 1.85 m tall, we consider him not to be tall}.

We can define the characteristic functions of sets A and B as listed below:

$$\mu_{tall}(x) = \begin{cases} 0 & x \leq 1.85 \\ 1 & x > 1.85 \end{cases}, \mu_{not\ tall}(x) = \begin{cases} 1 & x \leq 1.85 \\ 0 & x > 1.85 \end{cases}.$$

What can we say about a person who is 1.71 m tall? We obtain that $\mu_{tall}(1.71) = 0$ and $\mu_{not\ tall}(1.71) = 1$. Therefore, we can definitely conclude that he or she is not tall.

On the other hand, what can we say about two people who are 1.84 m and 1.86 m tall respectively? We get that $\mu_{tall}(1.84) = 0$, $\mu_{not\ tall}(1.86) = 1$, $\mu_{tall}(1.86) = 1$ and $\mu_{not\ tall}(1.86) = 0$. Therefore, the first person should be regarded as not tall and the second one as tall. However, that would be a mistake since both of them are approximately the same height. In cases like these, it is reasonable to apply fuzzy sets. Fuzzy sets are considered with respect to a nonempty base set X of elements of interest (in our case X will be the set of all people). The essential idea is that each element $x \in X$ is assigned a membership grade μ_{tall} to set A, taking values in $[0,1]$, with $\mu_{tall}(x) = 0$ corresponding to non-membership, $0 < \mu_{tall}(x) < 1$ to partial membership, and $\mu_{not\ tall}(x) = 1$ to full membership. One possible definition that provides an example of fuzzy sets of the tall and not tall individuals can be:

$$\mu_{tall} = \begin{cases} 0 & x \leq 1.8 \\ 10x - 18 & 1.8 < x \leq 1.9 \\ 1 & x > 1.9 \end{cases}, \mu_{not\ tall} = \begin{cases} 1 & x \leq 1.8 \\ -10x + 19 & 1.8 < x \leq 1.9 \\ 0 & x > 1.9 \end{cases}.$$

We can plot the function μ_{tall} with blue color and $\mu_{not\ tall}$ with red color (Figure 1):

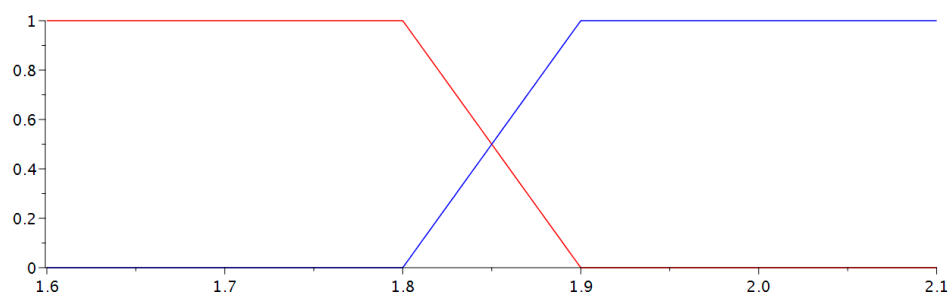


Figure 1. Plots of characteristic functions μ_A and μ_B .

Now a person who is 1.87 m tall can be viewed as partly tall and partly not tall. In fact, we get that he/she belongs to the set of tall people with a degree $\mu_{tall}(1.87) = 0.7$, and he/she belongs to the set of not tall people $\mu_{not\ tall}(1.87) = 0.3$. If we consider a 1.82 m tall person, then he/she belongs to the set of tall individuals with a degree $\mu_{tall}(1.82) = 0.2$, and to the set of not tall people $\mu_{not\ tall}(1.82) = 0.8$. If we consider a person that is 1.95 m tall, then he/she belongs to the set of tall men with a degree $\mu_{tall}(1.95) = 1$, and to the set of not tall people $\mu_{not\ tall}(1.95) = 0$, so he/she is definitely a tall individual.

There are different functions that can be used to assign a membership grade. The previous example illustrates the so-called trapezium membership function. Owing to the use of the bell-shaped function (Figure 2), when the height of a person gets closer and closer to the limit of 1.9 m, the degree of his/her membership grade to the tall people approaches 1 more rapidly, and this is evident from $\mu_{tall}(1.87) = 0.7$ (Figure 1) and $\nu_{tall}(1.87) = 0.79$ (Figure 2, blue curve). If a person's height is close to the boundary value of 1.9 for tall persons, his/her membership grade, calculated with the help of the bell-shaped function, is closer to one, than if it is calculated with a trapezium membership function. A sample for such a membership function is displayed in Figure 2.

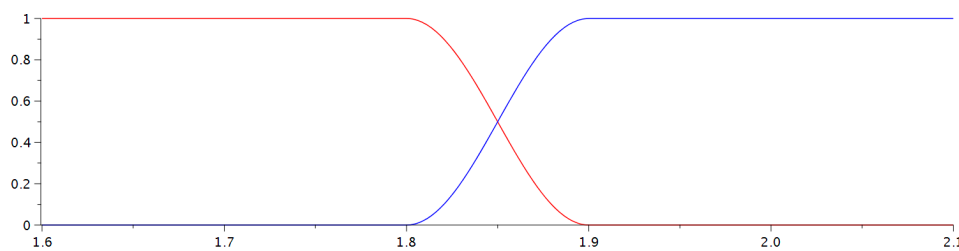


Figure 2. Bell-like membership functions of v_{tall} (blue color) and $v_{not\ tall}$ (red color).

Now a person who is 1.87 m tall can be viewed as partly tall and partly not tall. In fact, we get that he/she belongs to the set of tall people with a degree $\mu_{tall}(1.87) = 0.79$ and to the set of not tall people $\mu_{not\ tall}(1.87) = 0.21$. If we consider a person that is 1.82 m tall, then he/she belongs to the set of tall people with a degree $\mu_{tall}(1.82) = 0.09$ and to the set of not tall people $\mu_{not\ tall}(1.82) = 0.9$. If we consider a 1.95 m tall person, then he/she belongs to the set of tall people with a degree $\mu_{tall}(1.95) = 1$ and to the set of not tall individuals $\mu_{not\ tall}(1.95) = 0$ so he/she is definitely a tall person.

3. Fuzzification of Score Metrics

As previously stated, the score metric represented the number of correct answers to 60 questions with a closed answer, each one awarding one point for a correct answer and zero points for an incorrect one, and one question with an open answer, which was graded by the teacher on a scale from zero to 20.

The criteria for constructing the e-tests that we used are based on Bloom's taxonomy. The closed-ended questions were of similar difficulty and length, they were scored automatically and aim at the lower levels of cognitive thinking by testing the students' knowledge and comprehension of the learning material. For this reason, the same weight of one point for a correct answer was assigned to each of them. The open question targeted the higher levels of cognitive thinking; it required students to compose a text, in which they provided an opinion and justified it, and this question was evaluated manually by the teacher.

Besides the difference in the level of difficulty in the two types of questions, there was a chance that closed-ended ones may not reflect a student's knowledge and skills objectively for a variety of reasons, for example guessing, technical errors (automatic change of the selection of an answer while scrolling up or down the webpage), etc. On that account we assumed that the test score metrics relative to the open question were the critical ones, wherefore that item was assigned a higher weight.

The criteria for determining the weight of the open question were based on a scoring guide that we developed, which outlined what information was important in each response and how much credit was given for each part of it. These criteria reviewed the content of the responses—compliance with the topic and logical sequence of the text, adherence to the set format of the text (for example, an opinion essay, letter, etc.) and number of words, adherence to the grammatical norms and rules, correct and accurate use of the vocabulary, wealth of means of expression, and correct spelling.

When items of a test measured the same kind of ability or knowledge, they yielded a high internal consistent reliability. If a test was made up of different kinds of items assessing different types of abilities and knowledge, Cronbach's alpha coefficient tended to be low as a result of the heterogeneity of the items in terms of format and content. That was why first we decided what weight to place on the open question to ensure that the test was reliable. A wrong decision can sometimes be unfair. Cronbach's alpha coefficient can easily be calculated by the formula:

$$\alpha = \frac{k}{(k-1)} \left(1 - \frac{\text{Sum of Item Variances}}{\text{Test Variance}} \right).$$

The number of questions was 61, the sum of the variances was 50.91, and the test variance was 179.45. Applying these numbers in the formula for Cronbach's alpha coefficient, we got $\alpha = 0.72$,

which could be interpreted that the reliability was just acceptable. The greatest variance appeared in the open question, therefore we searched for a coefficient γ to multiply the results of the open question, so that α became bigger. We chose $\gamma = 0.2$, then we got: the number of questions was 61, the sum of the variances was 12.92, and the test variance was 114.99. Applying these numbers in the formula for Cronbach's alpha coefficient, we obtained $\alpha = 0.902$, which could be interpreted that the reliability was excellent.

Now the maximum number of points that a student could get was 64 ($60 + 20 \times 0.2$). We defined five categories of grades depending on the number of correct answers obtained by each student: Fail (from zero to 31 points), Satisfactory (from 32 to 39), Good (from 40 to 47), Very good (from 48 to 55), and Excellent (from 56 to 64).

As pointed out by the authors of References [6,9–11], it is difficult and unfair to assign a Good grade to a student with 47 points and to give a Very Good grade to another one with 48 points. Therefore, the use of fuzzy sets is justified. We will need to assume that someone whose test score belongs to the set (37–43) can be assigned either a grade Satisfactory or a grade Good. Thus comes the necessity to use fuzzy sets. Our goal was to apply fuzzy sets and fuzzy logic in the estimation of students' tests in order to assign them a fairer mark.

We defined five functions that represent the fuzzy membership trapezium functions to the sets of grades; their plots are given in Figure 3.

$$\begin{aligned}\mu_{Fail} &= \begin{cases} 1 & x < 29.5 \\ \frac{\cos(0.62x-18.53)+1}{2} & 29.5 \leq x < 34.5, \\ 0 & 34.5 \leq x \leq 64 \end{cases} \\ \mu_{Satis.} &= \begin{cases} 0 & 0 \leq x < 29.5 \\ \frac{\cos(0.62x-18.53)+1}{2} & 29.5 \leq x < 34.5 \\ 1 & 35 \leq x < 37.5, \\ \frac{\cos(0.62x-23.56)+1}{2} & 37.5 \leq x < 42.5 \\ 0 & 42.5 \leq x \leq 64 \end{cases} \\ \mu_{Good} &= \begin{cases} 0 & 0 \leq x < 37.5 \\ \frac{\cos(0.62x-20.42)+1}{2} & 37.5 \leq x < 42.5 \\ 1 & 42.5 \leq x < 45.5, \\ \frac{\cos(0.62x-28.58)+1}{2} & 45.5 \leq x < 50.5 \\ 0 & 50.5 \leq x \leq 64 \end{cases} \\ \mu_{VeryGood} &= \begin{cases} 0 & 0 \leq x < 45.5 \\ \frac{\cos(0.62x-25.44)+1}{2} & 45.5 \leq x < 50.5 \\ 1 & 50.5 \leq x < 53.5, \\ \frac{\cos(0.62x-33.61)+1}{2} & 53.5 \leq x < 58.5 \\ 0 & 58.5 \leq x \leq 64 \end{cases} \\ \mu_{Excellent} &= \begin{cases} 0 & 0 \leq x < 53.5 \\ \frac{\cos(0.62x-30.47)+1}{2} & 53.5 \leq x < 58.5. \\ 1 & 58.5 \leq x \leq 64 \end{cases}\end{aligned}$$

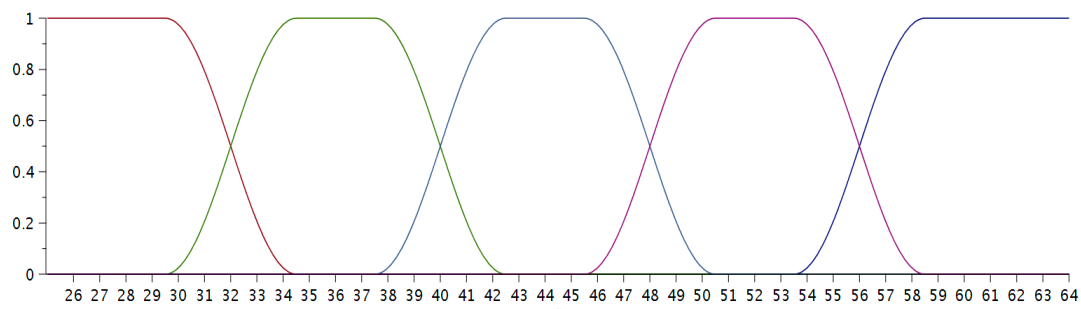


Figure 3. Plots of μ_{Fail} , $\mu_{Satis.}$, μ_{Good} , $\mu_{VeryGood}$, and $\mu_{Excellent}$.

In the case when the points that a student had obtained in a test did not belong definitely to a given set, we needed to select another criterion, dependent on the student's score, in order to take a decision what grade to assign him/her. Consequently, following the ideas from Reference [1], the second assessment that was applied to estimate a student's knowledge was their result of the open question. Furthermore, we divided the students' grades into five groups again: Fail (from zero to eight points), Satisfactory (from nine to 11), Good (from 12 to 14), Very good (from 15 to 17), and Excellent (from 17 to 20), and we also defined their membership functions (Figure 4):

$$\begin{aligned}
 v_{Fail} &= \begin{cases} 1 & x < 6.9 \\ \frac{\cos(1.42x-9.85)+1}{2} & 6.9 \leq x < 9.1, \\ 0 & 9.1 \leq x \leq 20 \end{cases} \\
 v_{Satis.} &= \begin{cases} 0 & 0 \leq x < 6.9 \\ \frac{\cos(1.42x-6.71)+1}{2} & 6.9 \leq x < 9.1 \\ 1 & 9.1 \leq x < 10.4, \\ \frac{\cos(1.42x-14.85)+1}{2} & 10.4 \leq x < 12.6 \\ 0 & 12.6 \leq x \leq 20 \end{cases} \\
 v_{Good} &= \begin{cases} 0 & 0 \leq x < 10.4 \\ \frac{\cos(1.42x-11.71)+1}{2} & 10.4 \leq x < 12.6 \\ 1 & 12.6 \leq x < 13.4, \\ \frac{\cos(1.42x-19.14)+1}{2} & 13.4 \leq x < 15.6 \\ 0 & 15.6 \leq x \leq 20 \end{cases} \\
 v_{VeryGood} &= \begin{cases} 0 & 0 \leq x < 13.4 \\ \frac{\cos(1.42x-15.99)+1}{2} & 13.4 \leq x < 15.6 \\ 1 & 15.6 \leq x < 16.4, \\ \frac{\cos(1.42x-23.41)+1}{2} & 16.4 \leq x < 18.6 \\ 0 & 18.6 \leq x \leq 20 \end{cases} \\
 v_{Excellent} &= \begin{cases} 0 & 0 < 16.4 \\ \frac{\cos(1.42x-20.28)+1}{2} & 16.4 \leq x < 18.6. \\ 1 & 18.6 \leq x \leq 20 \end{cases}
 \end{aligned}$$

The fuzzy associative matrix (Table 1) provided a convenient way to directly combine the input relations in order to obtain the fuzzified output results [9]. The input values for the results of the open question are across the top of the matrix and the input values for the total score of the test are down left in the matrix. We used the conventional Bulgarian grading scale.

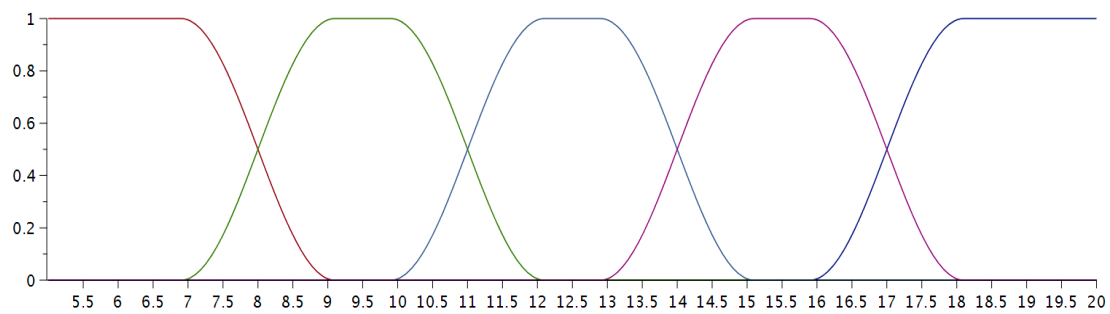


Figure 4. Plots of v_{Fail} , $v_{Satisfactory}$, v_{Good} , $v_{VeryGood}$, and $v_{Excellent}$.

Table 1. Fuzzy associative matrix.

	v_{Fail}	$v_{Satisfactory}$	v_{Good}	$v_{VeryGood}$	$v_{Excellent}$
μ_{Fail}	2	2	3	3	4
$\mu_{Satisfactory}$	2	3	3	4	4
μ_{Good}	2	3	4	5	5
$\mu_{VeryGood}$	3	4	5	5	6
$\mu_{Excellent}$	3	4	5	6	6

The rules for performing set operations of union (AND), intersection (OR), and complement (NOT), which were of most interest to us, are stated below [9].

For union, we looked at the degree of membership for each set and picked the lower one of the two, that is: $\mu_{A \cup B} = \min(\mu_A, \mu_B)$ (Figure 5).

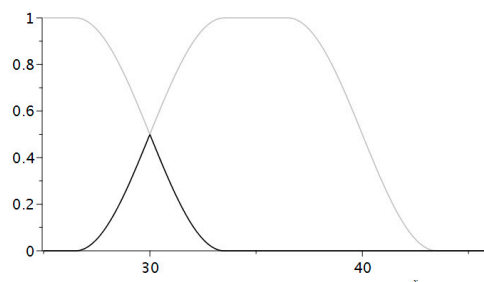


Figure 5. Plots of $\mu_{Satisfactory}$, μ_{Good} , and $\mu_{A \cup B}$.

For intersection, we looked at the degree of membership for each set and picked the higher one of the two, that is $\mu_{A \cap B} = \max(\mu_A, \mu_B)$ (Figure 6).

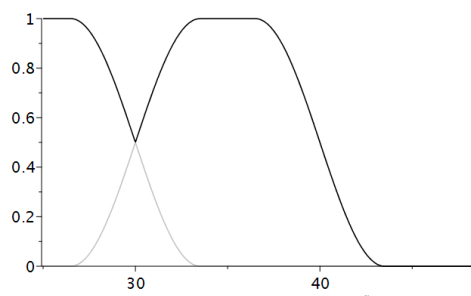


Figure 6. Plots of $\mu_{Satisfactory}$, μ_{Good} , and $\mu_{A \cap B}$.

For complement, we subtracted the degree of membership from 1, that is $\mu_{NOT A} = 1 - \mu_B$.

Let us consider a student with an overall score of 49 points and a score of the open question of 19 points. He/she belonged to the set of Very Good grades with a degree $\mu_{VeryGood}(49) = 0.79$ and to the

set of Good grades with a degree $\mu_{Excellent}(49) = 0.21$. Normally, he/she would get a grade Very good (5). However, the intersection of the two grades—the overall points and the points of the open question, states: He/she belongs to the set $\mu_{VeryGood} \cap \nu_{Excellent}$ with a degree 0.79; to the set $\mu_{Good} \cap \nu_{Excellent}$ with a degree 0.21; to the set $\mu_{VeryGood} \cap \nu_{VeryGood}$ with a degree 0, and to the set $\mu_{Good} \cap \nu_{VeryGood}$ with a degree 0 of the associative matrix. Therefore, we could assign him/her a grade Excellent (6).

In accordance with Reference [9], we needed to calculate the grade for each student, whose test score did not belong definitely to a given set. To do so, we referred to the Table 2 where the function F registered the minimums of μ and ν , i.e., $F = \min(\mu, \nu)$. The maximum membership grade, which was acquired from the table, represented the corrected grade from the associative matrix.

Table 2. The various combinations of the minimums of μ and ν , calculated by use of F .

$F(\mu_F(p), \nu_F(q))$	$F(\mu_F(p), \nu_S(q))$	$F(\mu_F(p), \nu_G(q))$	$F(\mu_F(p), \nu_V(q))$	$F(\mu_F(p), \nu_E(q))$
$F(\mu_S(p), \nu_F(q))$	$F(\mu_S(p), \nu_S(q))$	$F(\mu_S(p), \nu_G(q))$	$F(\mu_S(p), \nu_V(q))$	$F(\mu_S(p), \nu_E(q))$
$F(\mu_G(p), \nu_F(q))$	$F(\mu_G(p), \nu_S(q))$	$F(\mu_G(p), \nu_G(q))$	$F(\mu_G(p), \nu_V(q))$	$F(\mu_G(p), \nu_E(q))$
$F(\mu_V(p), \nu_F(q))$	$F(\mu_V(p), \nu_S(q))$	$F(\mu_V(p), \nu_G(q))$	$F(\mu_V(p), \nu_V(q))$	$F(\mu_V(p), \nu_E(q))$
$F(\mu_E(p), \nu_F(q))$	$F(\mu_E(p), \nu_S(q))$	$F(\mu_E(p), \nu_G(q))$	$F(\mu_E(p), \nu_V(q))$	$F(\mu_E(p), \nu_E(q))$

Let us now take an example with a student whose overall score of the test is 57 points and his/her score of the open question is 18 points ($p = 57$ and $q = 18$). As illustrated in Table 3, the student's grade after the process of fuzzification was Excellent 6, which coincided with the classical grading (56–64 points correspond to an Excellent grade).

Table 3. Fuzzified grade of a student with overall score (p) = 57 and open question score (q) = 18.

0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0.005	0.21
0	0	0	0.005	0.79

If we review another student who obtained 42 points overall (which corresponds to Good 4 using the conventional grading system) and 17 points of the open question, it is evident from Table 4 that after the fuzzification that student should have been assigned a higher grade—Very Good 5.

Table 4. Fuzzified grade of a student with $p = 42$ and $q = 17$.

0	0	0	0	0
0	0	0	0.02	0.2
0	0	0	0.5	0.49
0	0	0	0	0
0	0	0	0	0

To reevaluate the test scores that did not belong definitely to a given set, we only needed to input the test results: first as the overall number of points obtained, and then as the number of points of the open question for all the 78 students that have taken the test. Next, the software (MapleSoft 2016.0) automatically selected the ones that needed to be fuzzified and calculated the fuzzified grades.

Following the calculation, we perceived that there were 42 students out of the 78 test takers, whose overall grade did not belong with a degree 1 to any of the sets: Fail, Satisfactory, Good, Very good, or Excellent. Thus, for any student's score P_k , $k = 1, 2, \dots, 42$, we obtained the set of grades:

$$Y = [6, 6, 4, 3, 5, 2, 5, 4, 4, 5, 4, 2, 4, 2, 5, 5, 5, 3, 4, 5, 3, 4, 5, 4, 4, 3, 4, 6, 6, 6, 4, 2, 4, 3, 4, 5, 5, 4, 4, 2].$$

By using fuzzy logic and the associative matrix, we transformed the set of grades into a new set of grades:

$X = [6, 5, 5, 2, 5, 4, 4, 4, 4, 3, 4, 4, 4, 4, 4, 5, 3, 2, 3, 4, 5, 3, 4, 3, 4, 2, 3, 5, 5, 6, 3, 5, 4, 3, 5, 4, 4, 6, 4, 5, 4]$.

We checked if the new set of grades X were statistically different from Y . Let us denote by $X = (X_1, X_2, \dots, X_{42})$ the random variable, which describes a student's grade after the fuzzification, and by $Y = (Y_1, Y_2, \dots, Y_{42})$ the random variable, which reports the grade of the student before the fuzzification, both in numbers. We used the built-in functions for hypothesis testing in MapleSoft 2016.0.

With the help of the F-test we checked whether the standard deviation S_X^2 after using the fuzzy technique and the standard deviation S_Y^2 before using it were equal. We got the following result:

Two sample F-test ($X, Y, 1$, confidence = 0.95, equal variances = true); F-ratio test on two samples.

Null Hypothesis: *A sample drawn from populations with a ratio of variances is equal to 1.*

Alt. Hypothesis: *A sample drawn from a population with a ratio of variances is not equal to 1.*

Result: [Accepted].

This statistical test did not provide enough evidence to conclude that the null hypothesis was false. Consequently, we could accept that the standard deviations were equal.

We also checked the distributions of the grades before and after using the fuzzy technique for equal mean values.

Two sample t -test ($X, Y, 0$, confidence = 0.95); standard t -test on two samples (unequal variances).

Null Hypothesis: *A sample drawn from populations with a difference of means is equal to 0.*

Alt. Hypothesis: *A sample drawn from a population with a difference of means is not equal to 0.*

Result: [Accepted].

This statistical test did not provide enough evidence to conclude that the null hypothesis was false. Consequently, we could accept that the mean values were equal.

With the help of the paired t -test we checked if the distributions of the changed grades by means of the fuzzy technique were one and the same distribution.

Two sample t -test ($X, Y, 0$, confidence = 0.95); standard t -test with paired samples.

Null Hypothesis: *A sample drawn from populations with difference of means is equal to 0.*

Alt. Hypothesis: *A sample drawn from a population with difference of means is not equal to 0.*

Result: [Accepted].

This statistical test did not provide enough evidence to conclude that the null hypothesis was false. Consequently, we could accept that the two distributions Y and X were equal.

4. Summary

To summarize, we could say that the means of the grades of the group as a whole before and after the fuzzification did not change statistically: the mean of the grades before the process of fuzzification was 4.09, and after the fuzzification 4.04.

The number of the fuzzified grades from the test was 42 out of 78. Nearly half of the grades were fuzzified—namely 53.84%, and the number of corrected grades was 31. The percentage of the corrected grades after the fuzzification from the overall grades was 39.74%, and the percentage of the corrected grades after the fuzzification from fuzzified grades was 73.81%, which indicated that the grading system needed some adjustments and that fuzzy logic was applicable to a large number of students in pursuit of more reasonable grading.

To implement the process of fuzzification, by virtue of the sequence of procedures that we wrote in MapleSoft 2016.0, we only needed to enter:

- the overall points of each student,
- the points of the open question of each student, and
- the boundary values of the membership functions.

The software automatically calculated the fuzzified grades and checked the three hypotheses mentioned above. We could change the boundary values of the membership functions to ensure that the two distributions were equal. If we reduced the intervals, in which the grades were not fuzzified, and increased the intervals, in which we used fuzzification, we obtained results whose distribution was substantially different from that of the results before the process of fuzzification. That is why we would like to emphasize that the teacher's experience in using fuzzification to correct the final grades is of great importance. We illustrate below that by changing the set of points which will be fuzzified. We defined five functions that represent the fuzzy membership bell-like functions to the sets of grades (Figure 7).

$$\begin{aligned}\mu_{Fail} &= \begin{cases} 1 & x < 29.2 \\ \frac{\cos(0.56x-16.38)+1}{2} & 29.2 \leq x < 34.8, \\ 0 & 34.8 \leq x \leq 64 \end{cases} \\ \mu_{Satis.} &= \begin{cases} 0 & 0 \leq x < 29.2 \\ \frac{\cos(0.56x-13.24)+1}{2} & 29.2 \leq x < 34.8 \\ 1 & 34.8 \leq x < 37.2, \\ \frac{\cos(0.56x-20.87)+1}{2} & 37.2 \leq x < 42.8 \\ 0 & 42.8 \leq x \leq 64 \end{cases} \\ \mu_{Good} &= \begin{cases} 0 & 0 \leq x < 37.2 \\ \frac{\cos(0.56x-17.73)+1}{2} & 37.2 \leq x < 42.8 \\ 1 & 42.8 \leq x < 45.2, \\ \frac{\cos(0.56x-25.36)+1}{2} & 45.2 \leq x < 50.8 \\ 0 & 50.8 \leq x \leq 64 \end{cases} \\ \mu_{VeryGood} &= \begin{cases} 0 & 0 \leq x < 45.2 \\ \frac{\cos(0.56x-22.22)+1}{2} & 45.2 \leq x < 50.8 \\ 1 & 50.8 \leq x < 53.2, \\ \frac{\cos(0.56x-29.85)+1}{2} & 53.2 \leq x < 58.8 \\ 0 & 58.8 \leq x \leq 64 \end{cases} \\ \mu_{Excellent} &= \begin{cases} 0 & 0 \leq x < 53.2 \\ \frac{\cos(0.52x-26.7)+1}{2} & 53.2 \leq x < 58.8. \\ 1 & 58.8 \leq x \leq 64 \end{cases}\end{aligned}$$

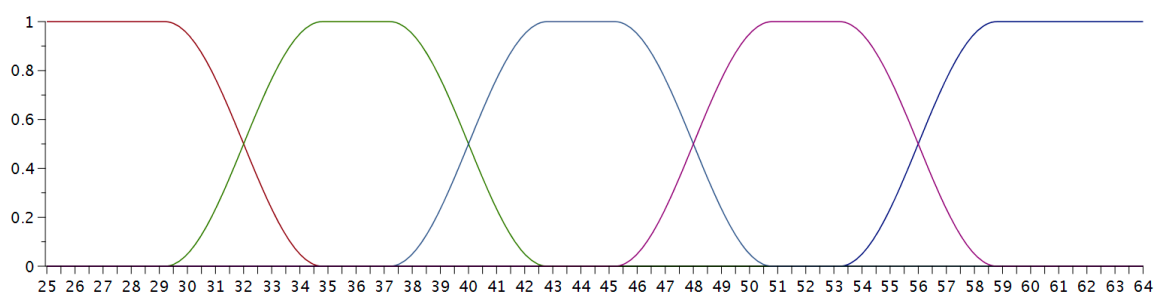


Figure 7. Plots of μ_{Fail} , $\mu_{Satis.}$, μ_{Good} , $\mu_{VeryGood}$ and $\mu_{Excellent}$.

We used the same membership functions v , as in the first case.

We obtained that the set of scores Y that would be fuzzified consisted of 46 items. Performing the same procedure in MapleSoft 2016.0, we got:

Two sample F-test ($X, Y, 1$, confidence = 0.95, equal variances = true); F-ratio test on two samples.

Null Hypothesis: *A sample drawn from populations with a ratio of variances is equal to 1.*

Alt. Hypothesis: *A sample drawn from a population with a ratio of variances is not equal to 1.*

Result: [Accepted].

This statistical test did not provide enough evidence to conclude that the null hypothesis was false. Consequently, we accepted that the standard deviations were equal.

Two sample t -test ($X, Y, 0$, confidence = 0.95); standard t -test on two samples (unequal variances).

Null Hypothesis: *A sample drawn from populations with a difference of means is equal to 0.*

Alt. Hypothesis: *A sample drawn from a population with a difference of means is not equal to 0.*

Result: [Accepted].

Two sample t -test ($X, Y, 0$, confidence = 0.95); standard t -Test with paired samples.

Null Hypothesis: *A sample drawn from populations with difference of means is equal to 0.*

Alt. Hypothesis: *A sample drawn from a population with difference of means is not equal to 0.*

Result: [Rejected].

In short, the number of the fuzzified grades from the test was 46 out of 78. Over half of the grades were fuzzified—namely 63.88%, and the number of corrected grades was 35. The percentage of the corrected grades after the fuzzification from the overall grades was 44.87%, and the percentage of the corrected grades after the fuzzification from fuzzified grades was 76.08%. The means and the standard deviation of the group before and after the fuzzification did not change statistically: the mean of the grades before the process of fuzzification was 4.39, and after the fuzzification 4.04. However, the paired test showed that there was a significant difference in the distributions of the fuzzified marks. Therefore, in this particular case, it cannot be interpreted as obtaining a fairer grading of students' tests.

5. Conclusions

Although research has been made previously to incorporate fuzzy logic and education [6] or to use fuzzy logic as a tool for assessing students' knowledge and skills collectively [7], the focus of our work was the individual student. Our main concern was to verify that the grades each learner obtains are as fair as possible because test scores are important for their final grading and hence for their motivation to study [12].

One of our main contributions is that we have taken into account the reliability of the test when using fuzzy logic and have confirmed that the distributions of the changed grades by means of the fuzzy technique are the same as the distributions before that process.

In our work we wish to add more than two points in the borderline interval (for instance, to change the interval for assigning grade Good 4 from 51–60 to 50–61 points) in order to make assessments fairer and compensate for possible technical errors made by students. Thus, our purpose is to make the interval of borderline scores as large as possible while at the same time guarantee that the test is reliable.

For this reason, calculations were performed automatically to check at which values the test was reliable. In our test this interval included five points: for example, according to the usual grading, the

grade Good 4 was assigned to students who obtained from 51 to 60 points on the test, while by using fuzzy logic this interval added five points and changed it to 47.5–62.5.

Just as importantly, the test results of all the students as a whole did not differ statistically from those obtained by using the traditional scale.

We can conclude that, when assessing students' achievements in tests, the use of the fuzzy technique needs to be handled carefully. We can only implement fuzzification if the mean value, the variance, and the distribution function do not differ statistically when the grades are considered as a set. In this case, we neither increase nor decrease the overall score of the examined group. Taking everything into account, we can deduce that we have found a way of getting a fairer evaluation of students, without changing the overall score of the group and their distribution function.

We are planning to incorporate fuzzy logic in the Distributed e-Learning Center (DeLC) at Plovdiv University during the next academic year. DeLC has been used at the Faculty of Mathematics and Informatics for teaching and testing Bachelor and Master-degree students for years. The tests, created and delivered online, evaluate the students' achievements in a variety of subjects among which English, Software Engineering, Databases, Intelligent Systems, Mobile Applications, elective courses such as Application of Block Chains, Contemporary JavaScript frames for Building Backend Applications, etc. Thus, our goal is to use fuzzy logic to review borderline test scores of a wide range of ability tests on a regular basis.

Author Contributions: The authors worked on this piece in close collaboration and have contributed equally to its development.

Funding: This research received no external funding.

Acknowledgments: The first author wishes to acknowledge the partial support of the National Program "Young Scientists and Postdoctoral Students" of the Ministry of Education and Science in Bulgaria, 2018–2019.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ivanova, V.; Terzieva, T. Criteria for the construction of tests for language assessment and evaluation. In Proceedings of the Doctoral Conference in Mathematics and Informatics MIDOC, Sofia, Bulgaria, 15–18 October 2016; St. Kliment Ohridski University Press: Sofia, Bulgaria, 2016; pp. 58–65.
2. Dunlosky, J. Strengthening the Student Toolbox: Study Strategies to Boost Learning. *Am. Educ.* **2013**, *37*, 12–21.
3. McDaniel, M.; Wildman, K.M.; Anderson, J.L. Using Quizzes to Enhance Summative-Assessment Performance in a Web-based Class: An Experimental Study. *J. Appl. Res. Mem. Cogn.* **2012**, *1*, 18–26. [\[CrossRef\]](#)
4. Roediger, H.L.; McDermott, K.B.; McDaniel, M. Using testing to improve learning and memory. In *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society*; Gernsbacher, M.A., Pew, R.W., Hough, L., Pomerantz, J.R., Eds.; World Publishing Co.: New York, NY, USA, 2011; pp. 65–74.
5. Hsu, W.L.; Chen, Y.S.; Shiau, Y.C.; Liu, H.L.; Chern, T.Y. Curriculum Design in Construction Engineering Departments for Colleges in Taiwan. *Educ. Sci.* **2019**, *9*, 65. [\[CrossRef\]](#)
6. Sobrino, A. Fuzzy Logic and Education: Teaching the Basics of Fuzzy Logic through an Example (by Way of Cycling). *Educ. Sci.* **2013**, *3*, 75–97. [\[CrossRef\]](#)
7. Voskoglou, M.G. Fuzzy Logic as a Tool for Assessing Students' Knowledge and Skills. *Educ. Sci.* **2013**, *3*, 208–221. [\[CrossRef\]](#)
8. Zadeh, L.A. Fuzzy Sets. *Inf. Control* **1965**, *8*, 338–353. [\[CrossRef\]](#)
9. Fahad, S.; Shah, A. Intelligent testing using fuzzy logic. In *Innovations in E-learning, Instruction Technology, Assessment, and Engineering Education*; Iskander, M., Ed.; Springer: Dordrecht, The Netherlands, 2007; pp. 95–98.
10. Ghomshei, M. Fuzzy logic in engineering education and evaluation of graduate attributes. In Proceedings of the Canadian Engineering Education Association (CEEAA-ACEG18)—Conf. Paper 152, Vancouver, BC, Canada, 3–6 June 2018.

11. Rusmiari, N.; Putra, D.; Sasmita, A. Fuzzy Logic Method for Evaluation of Difficulty Level of Exam and Student Graduation. *Int. J. Comput. Sci. Issues* **2013**, *10*, 223–229.
12. Ivanova, V.; Zlatanov, B. Application of Fuzzy Logic in online test evaluation in English as a foreign language at university LEVEL. In Proceedings of the 45th Conference Applications of Mathematics in Engineering and Economics (AMEE'19), Sozopol, Bulgaria, 7–13 June 2019. in print.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).