Can Fuzzy Logic Remove Subjectivity From Final Year Project Assessment?

Séamus McLoone and Niamh O'Neill

Department of Electronic Engineering,
National University of Ireland Maynooth,
Maynooth, Co. Kildare, IRELAND
E-mail:seamus.mcloone@eeng.nuim.ie

Subjectivity is an important issue in the assessment of final year projects. Naturally, students will want to be assessed by the more generous marker when given the choice between two assessors. Intuitively, if two assessors examine the same project and both feel that it is of the same standard then both should award it the same mark. However, this is generally not the case and often the marks awarded can be significantly different. This occurs because human beings naturally have different interpretations of the same thing. It is purported that fuzzy logic is adept at overcoming the issue of subjectivity. Thus, this paper explores the use of fuzzy logic in removing subjectivity from final year project assessment. A novel approach to the assessment procedure is outlined and some preliminary analysis is presented.

Keywords – Fuzzy Logic, Student Assessment, Subjectivity

I INTRODUCTION

Final year projects are very different from most other forms of module assessments, as different students are evaluated by different assessors. Ideally, every student should be marked by the same set of evaluators but, due to resource limitations, this is neither practical nor feasible. Instead, it is common practice for every student to have their project marked by two different examiners from a pool of selected examiners, typically the academic staff of the relevant department. Hence, the issue of subjectivity arises. Are all projects equally and fairly assessed if different projects are marked by different evaluators? The answer is clearly no as the assessment is inherently linked to the subjective feelings of the assessor.

Good assessment practices encourage eliminating subjectivity if possible [1]. For the purposes of this paper, it is important to clarify that subjectivity arises at two distinct levels in project evaluation. The first is where the evaluators can have differing opinions on the quality of the same piece of work, while the other is where they can agree on the quality of the work and yet mark it quite differently. For example, consider two assessors grading a poster presentation.

One may consider the poster to be simply poor while the other is suitably impressed by it. Clearly, this highlights differing subjective opinions by the examiners, which may be a result of a varying number of factors, including differences in experience, matter of taste, etc. In contrast, consider the case where both assessors actually agree that the poster presentation is of a 'good' standard, yet one may decide to give it 50% while the other is more generous and awards it 60%. Here, the difference in marks occurs as a result of the examiners' varying subjective views of the linguistic variable 'good'. It is this latter case that is addressed in the remainder of this paper.

Since its introduction in 1965 [2], fuzzy logic has appeared on numerous occasions throughout the research literature and has been applied in a wide variety of applications, ranging from obstacle avoidance in robotic vehicles [3] to evaluation of journal grades [4]. Furthermore, it has found extensive use in the field of education, including student evaluation and assessment of student-centred learning [5, 6].

Fuzzy logic has an inherent ability to model natural behaviour and linguistic terms. As such, this paper explores its use in removing the aforementioned subjectivity associated with final

year project assessment. A novel approach is adopted in developing a suitable grading system based on the concept of fuzzy logic. Here, a set of fuzzy grading functions are developed that reflect the average viewpoint of the academic staff of the Department of Electronic Engineering at NUI Maynooth in Co. Kildare. We propose replacing the existing numerical grading system with the new fuzzy-based one. Preliminary analysis and results are outlined and discussed later in the paper.

The rest of this paper is structured as follows. A brief description of the basic principles of fuzzy logic and fuzzy sets are outlined in the next section. The novel approach adopted for marking the final year project is presented in section III, while some preliminary analysis is discussed in section IV. The paper ends with some concluding remarks and suggestions for future work in section V.

## II FUZZY LOGIC AND FUZZY SETS

For the convenience of the reader, the relevant principles of fuzzy logic and fuzzy sets are now outlined.

Fuzzy logic was first introduced in 1965 by Zadeh [2] as a means of overcoming the inadequacies of Boolean algebra to many real world problems. With Boolean logic, everything is classified as black or white, true or false, right or wrong. Values either belong to a given defined set or they do not. In contrast, fuzzy logic softens this boundary to incorporate the grey area in between. Values no longer need to belong to one set or another but can in fact have partial membership to more than one.

In order to illustrate the concept, consider, for example, the basic set of temperatures regarded as hot. First of all, what defines hot? For argument sake, let's say that hot is defined as all temperatures exceeding and including 25°C. Thus the Boolean set for hot would be represented as shown in figure 1(a). In this case, a temperature of 24.99°C is absolutely not hot. Obviously, this logic is more suited to non-continuous phenomena. Fuzzy logic adopts a more rational, human decision making approach and regards the latter as belonging to the hot set to some degree. The degree of membership is directly related to how close the temperature is to a predefined value, i.e. 25°C in this example. Thus 24.99°C has a very high degree of membership while 21°C has a very low one, as illustrated by the fuzzy set in figure 1(b). Note, in this case, temperatures less than 20°C are certainly not hot.

The last example clearly highlights the usefulness of fuzzy logic and fuzzy sets. Furthermore, it conveys the typical imprecise nature of our language. Words, or linguistic terms, such as hot and cold are inherently vague and imprecise. It is words such as these that allow us to express our subjective feelings about a measurement or concept.

What one person considers hot, another person considers not hot. Conversely, if asked to specify a lower limit for hot, one person may give 25°C while another gives 21°C.
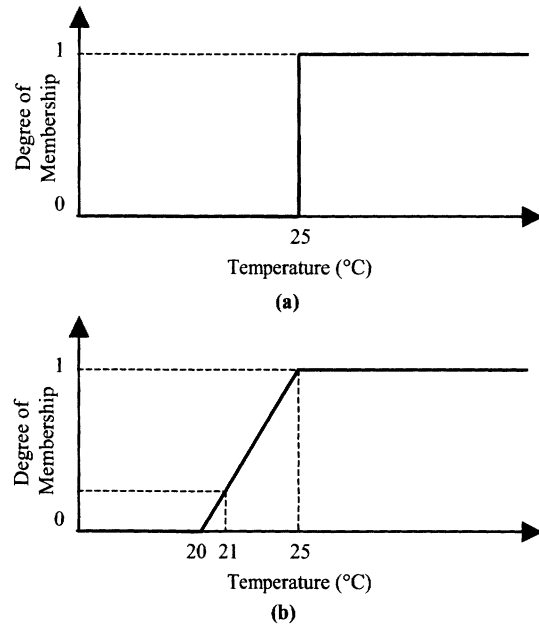
Figure 1: **(a)** Boolean representation of set of temperatures regarded as hot and **(b)** Fuzzy Logic representation of same

These varying subjective opinions exist as part of human nature but they are not easily measured and cannot be used objectively using standard Boolean-based methods. Instead, as we have shown, linguistic variables can be quantified using fuzzy sets. In other words, fuzzy logic and fuzzy sets provide a means to deal scientifically with subjectivity and thus can be used to model real world problems.

This is of particular relevance in grading systems where subjectivity can affect the results. Two different examiners can assess a 'good' piece of work and award it two very different marks. This is clearly unfair to the student that receives the 'harder' marker for their final year project. In the next section, we propose a solution to alleviate this issue of subjectivity.

III FUZZY-BASED GRADING

Traditionally all components of the final year project are assessed on a numerical scale from 0 to 100. How do we ensure that marks from different examiners are independent of subjectivity and, thus, fair to the students? This is almost impossible to achieve, as long as we continue directly using the numerical scale. However, adopting a fuzzy logic scale that consists solely of linguistic variables such as good, poor, etc. has potential to alleviate this problem.

The use of a fuzzy logic scale also has the added advantage of making it easier for the individual

examiner to provide marks in the first place. In other words, it is a lot simpler for the assessor to choose a linguistic term such as 'good' rather than having to assign an exact percentage mark – how does one realistically choose between awarding 62%, 63% or even 64%, for example?

Rather than simply choosing an assessment scale based on the standard letter grades A to F, we wanted to obtain a grading scale that accurately reflected the views of the staff involved in grading final year projects - in this case, the staff of the Department of Electronic Engineering, NUI Maynooth.

The first step involved deciding on a suitable number of linguistic variables. If there are too many variables then it becomes almost impossible for the assessor to distinguish between them in any meaningful fashion. On the other hand, too few variables can result in limited information in terms of student performance. Clearly, a compromise is required between obtaining detailed information and providing meaningful linguistic terms. Hence, a preliminary survey was conducted on a number of final year students. They were given four lists, each with a different number of linguistic variables and asked to assign a suitable numerical range to each variable in each list. They were also required to provide comments on the process.

The feedback obtained clearly confirmed what we expected. The shorter lists proved easy to fill out while the longer lists caused problems for most of the students, as they found it very difficult to distinguish between the terms given. Furthermore, while the shorter lists caused no such problems to those completing the survey, it nevertheless does not provide the evaluator with enough detail regarding the student's performance. Also, it is worth noting that Miller [7] states that, as humans, we can optimally perceive up to 7 different states. Thus, a list of 7 variables was chosen, as a suitable compromise, for the fuzzy logic grading scale. The list consisted of the following linguistic terms: *Excellent, Very Good, Good, Satisfactory, Poor, Very Poor* and *Extremely Poor*.

The second step required obtaining an appropriate numerical range for each of the linguistic variables in the specified list. These ranges had to reflect the views of the academic staff in the department. Hence the staff were presented with the list and asked to assign a suitable numerical range, in their opinion, to each term. Each numerical range consisted of a minimum and a maximum value and the staff were informed that they had to ensure that the complete list of terms covered the full numerical range 0 to 100. The results are presented in figures 2 and 3 below.

Figure 2 shows all the different minimum and maximum values chosen by the staff for the linguistic variables *Good* and *Excellent*. Similar results are obtained for the other variables. Here, each different range of values, chosen by the staff, is

represented by a triangular membership function. The base of the triangle defines the range of marks chosen while the apex is simply taken to be the midpoint of the numerical range. In the case of the linguistic term *Excellent* all staff chose 100 as the maximum value, as expected. In contrast, the minimum value varies from 70 by one staff member to 85 by another. The shaded area highlights the wide range of minimum and maximum marks awarded for each linguistic term. This clearly shows the varying opinions of the staff and, thus, the inherent subjectivity in decision making.

(a)

(b)
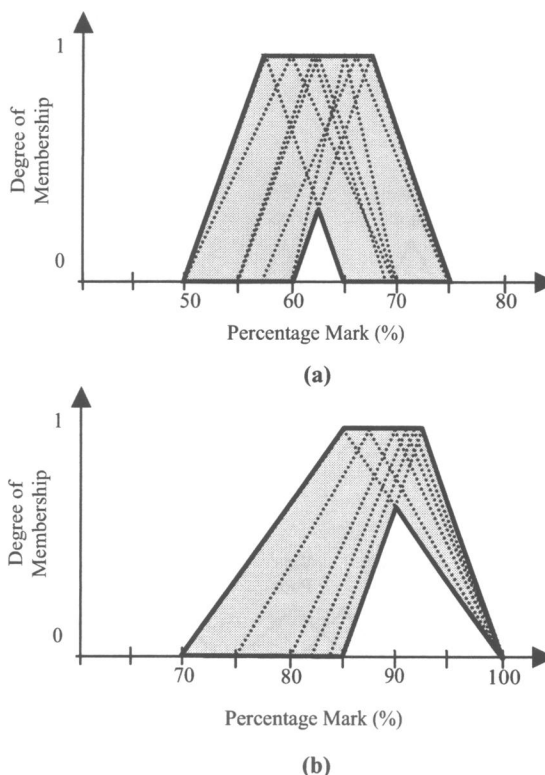
Figure 2: The different range of values chosen by the staff for the linguistic terms **(a)** Good and **(b)** Excellent. In each case, the dotted lines indicate the chosen range of values by different members of staff.

Figure 3 (a) shows the membership functions for all the linguistic variables on one set of axes. Here, the limits of each one are obtained by calculating the average value of those chosen by the staff for that particular variable. An alternative representation of the results is given in figure 3 (b). In this case, the limits of each linguistic term are determined by using the respective minimum and maximum value from all of the values chosen by the staff for that term. Thus, this figure represents the extreme limits.

Once again, the significantly differing opinions of the staff are clearly illustrated. Consider, for example, the linguistic term *Very Poor*. The average range is given by 15 – 29 whereas the extreme case shows a corresponding range of 0 – 40. However, it is worth noting that the extremities are a result of

individual staff opinions and don't reflect the overall group mentality. The results in figure 3(a) are clearly more reflective of the staff's opinions as they are less susceptible to individual contributions. Thus, these values are used in our proposed grading system.

The exact percentage value for each linguistic variable is represented by the midpoint of the associated range. This point is easily identified as the apex of the triangle. Thus, for example, the mark associated with *Good* is 63%. Now, if two or more assessors choose the same linguistic term, then the same exact numerical mark is awarded. Therefore, this removes the aforementioned issue of subjectivity and maintains the average departmental outlook on assessment values.

The same survey was also carried out by 12 final year project students, the results of which are shown in figure 4 below. It is interesting to observe, from an educational point of view, that the average numerical values chosen by the students is almost identical to those of the staff, within reason. This would suggest that the proposed marking range would also be acceptable to the students themselves.

At this stage, it is worth noting that the proposed approach quantises the range of marks for grading. However, one problem exists. We have only 7 different linguistic terms and therefore effectively end up with only 7 possible numerical values to choose from. Obviously this is a very coarse level of quantisation, unsuitable for assessment purposes. In order to obtain a finer level we need to increase the
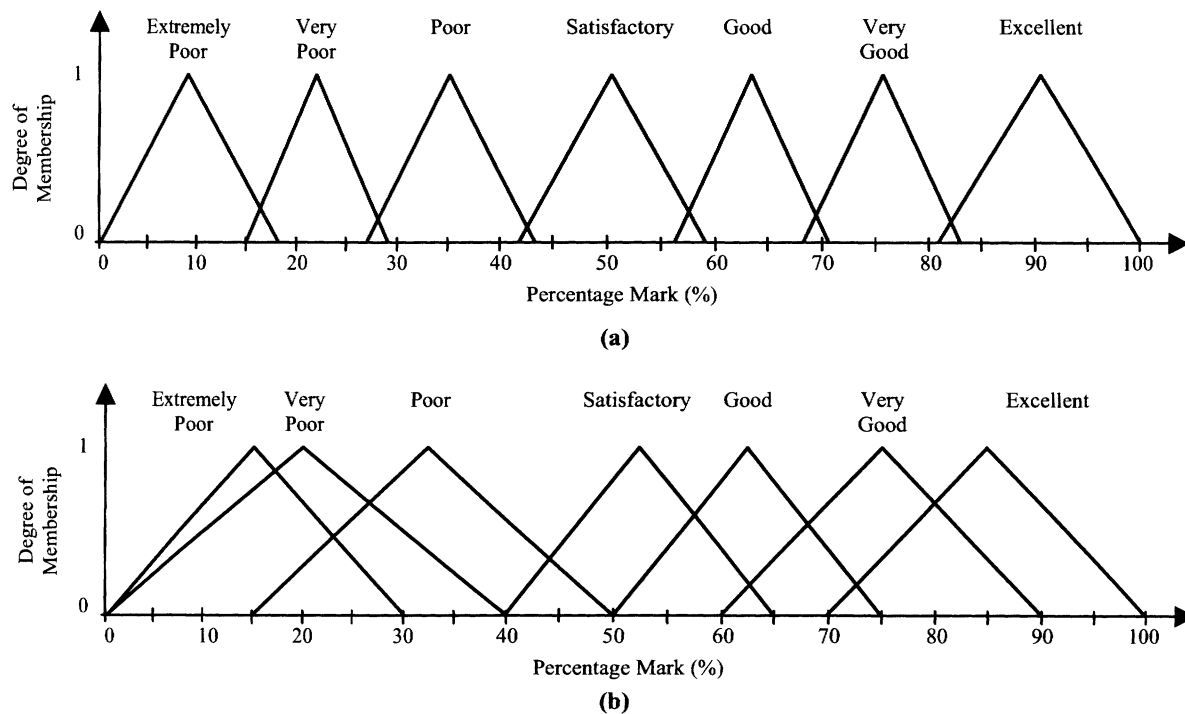
Figure 3: Numerical ranges chosen by staff, evaluated using **(a)** the average of each limit and **(b)** the minimum value of the lower limit and the maximum value of the upper limit in each case

Figure 4: Numerical ranges chosen by students, evaluated using the average of each limit

number of terms, but as we have stated already, this leads to the problem of having to distinguish between very similar linguistic variables.

We propose a simple solution in the form of a hierarchical approach to grading. After the assessor has decided on a suitable linguistic term, from the list of seven, they then have to decide on a further subdivision, by choosing one of three options within this range. For example, consider the case where an assessor chooses the linguistic term *Good*. They then have to choose if the term *Good* is nearer to *Very Good*, nearer to *Satisfactory* or simply in the middle, as illustrated in figure 5. This gives us a further division of the numerical range, resulting in a total of 21 divisions. This is a more acceptable quantisation level for the purposes of assessment grading.
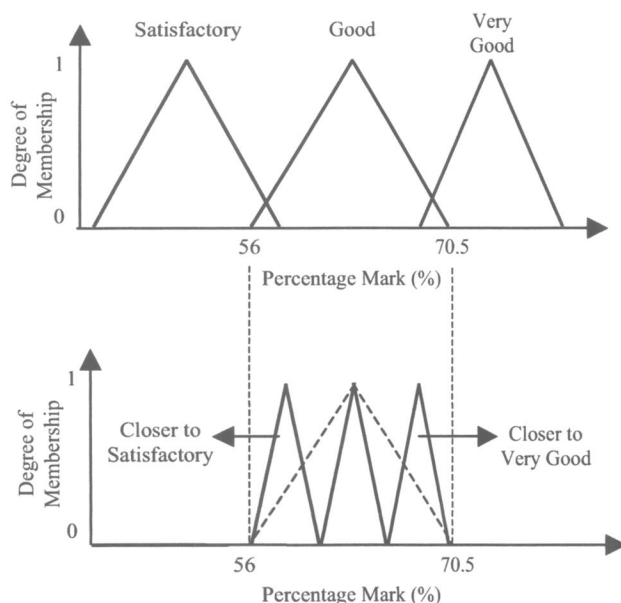

Figure 5: Illustrating the hierarchical approach to grading

Finally, it is worth mentioning that while this grading scheme is specifically designed for the Department of Electronic Engineering at NUI Maynooth, a similar one can easily be developed for other departments. Furthermore, a survey incorporating numerous departments has the potential of giving a single grading scheme that all departments could use.

In the next section, we conduct a very simple experiment to analyse the proposed fuzzy-based grading scheme.

IV   RESULTS & ANALYSIS

In order to investigate the validity of the proposed scheme, a simple experiment was conducted. For the sake of convenience, three sample posters were generated. Note that a poster presentation is one of the assessed elements in the final year project. Both the staff and final year students were asked to grade each of these posters solely in terms of its presentational quality, i.e. how well it looked, how the information was structured, etc. They were required to use the two different grading schemes for each poster, namely the proposed hierarchical fuzzy-based scheme and the currently used conventional approach of providing a percentage mark. The results are given in tables 1 and 2.

	Staff Average			Student Average		
Marking Scheme	P1	P2	P3	P1	P2	P3
Conventional	68	74	77	67	72	73
Fuzzy-based 1	69	77	80	63	71	71
Fuzzy-based 2	69	77	79	64	73	71

Table 1: Calculated Averages for Posters P1, P2 and P3. Note: Fuzzy-based 1 refers to grading scheme where only the 7 linguistic variables are consider while Fuzzy-based 2 refers to the hierarchical approach.

Table 1 shows the average mark awarded for each poster, by both the staff and the students, for the two different grading schemes. In the case of the proposed fuzzy-based scheme, results are presented for two versions, i.e. the version where only the first level of linguistic variables is used (Fuzzy-based 1) and the one where the hierarchical approach is adopted (Fuzzy-based 2). In all cases, the average mark does not vary by more than 3% for the staff and 4% for the students. This shows that the proposed scheme does not distort the marks and thus appears to be a valid approach for assessment.

Table 2 presents a detailed breakdown of the linguistic variables chosen by staff and students for each poster. There is a wide variety of opinions over each poster by both staff and students alike. This clearly highlights the issue of subjectivity where different people have different opinions over the quality of the poster in question.

The other issue of subjectivity also exists, though it is not directly evident from table 2. Take for example the case where 5 staff members chose the high end of *Good* for poster 1. The actual marks awarded by the same staff varied from 63% to 70%, a 7% difference. Note that these values are not presented in this paper due to lack of space. Numerous examples of this type of subjectivity are evident in the results. The use of the proposed fuzzy-based grading scheme can effectively eliminate this problem. Thus, for the case where 5 staff members chose the high end of *Good*, the same mark of 68% would now be awarded by each one. Interestingly, the actual average mark of the 5 assessors in question is 66%. Now, if two different students get two each of these 5 assessors, then they don't have to worry about their fellow classmate getting the 'easier'

		Poor			Satisfactory			Good			Very Good			Excellent		
		L	M	H	L	M	H	L	M	H	L	M	H	L	M	H
Staff (12)	Poster 1							1	2	5	2		1	1		
	Poster 2					1			1	1	1	2	2	1	3	
	Poster 3									2	2	3	1	2	2	
Students (12)	Poster 1	1				1		4	3		1	1	1			
	Poster 2							4	3		1	2			2	
	Poster 3		1		2							3	3	2		

Table 2: A breakdown of the linguistic variables chosen by staff and students for each poster. The letters L, M and H indicate the lower, middle and upper parts of the range chosen (refer to figure 5). Each digit represents the number of staff or students that chose that particular value.

marker. Of course this is all based on the fact that the assessors agree on the quality of work in the first place. Unfortunately, this is not always the case.

V   CONCLUDING REMARKS

This paper has explored the use of fuzzy logic in removing subjectivity from final year project assessment. The subjectivity in question relates to the case where two or more assessors agree that a project is of a certain standard. However, they award it very different percentage marks due to their varying opinions. By adopting a fuzzy-based grading scheme that employs the use of natural linguistic terms, this subjectivity can be alleviated.

Here, we proposed using a novel hierarchical approach. Initially, assessors must convey their opinion by choosing one of 7 pre-defined linguistic variables. These variables have already been pre-assigned a percentage mark that reflects the average opinion of the group of assessors. The assessors then have to choose a subdivision of their chosen variable, thus allowing for a more accurate reflection of their opinion.

This proposed approach has two key advantages over the conventional marking scheme of simply assigning a percentage mark from a scale of 0 – 100. Firstly, it is easier for the assessor to choose one of 7 linguistic terms (and subsequently one of 3) than it is to attempt to assign a specific mark. Secondly, the marks awarded should, in general, reflect the average mark of the group of assessors and thus ensures a fairer grading system for the students involved.

While this new fuzzy-based grading scheme alleviates one aspect of subjectivity in final year project assessment, it still does not prevent one assessor choosing *Good* and a second choosing *Poor* for the same project. Unfortunately, this is a much bigger issue in final year project assessment.

A possible solution may involve using linguistic descriptive phrases in association with the linguistic terms. In this paper, assessors were simply presented with a list of 7 linguistic terms. However, it should prove beneficial to have a list of associated criteria associated with each term. Clearly, these can differ for different assessed components of the final year project. The criteria need to be carefully considered and defined by the department as a whole. This idea has great potential of reducing the subjectivity in quality and remains the subject of future work.

VI   ACKNOWLEDGEMENTS

VII   REFERENCES

[1] P. Race, S. Brown and B. Smith, *500 Tips on Assessment*, RoutledgeFalmer, second edition, 2005.

[2] L.A. Zadeh, "Fuzzy sets", Information and Control, vol. 8, pp. 338-353, 1965.

[3] J. W. Baxter and J. R. Bumby, "Fuzzy control of a mobile robotic vehicle", *Proc. Institution of Mechanical Engineers,* vol. 209, pp. 79-91, 1995.

[4] D. Zhou, J. Ma, E. Turban and N. Bolloju, "A fuzzy set approach to the evaluation of journal grades," *Fuzzy Sets and Systems*, 131, pp. 63-74, 2002.

[5] R. Biswas, "An application of fuzzy sets in students' evaluation," *Fuzzy Sets and Systems*, 74, pp. 187-194, 1995.

[6] J. Ma and D. Zhou, "Fuzzy set approach to the assessment of student-centred learning," *IEEE Trans. on Education*, vol. 43 no. 2, pp. 237-241, May 2000.

[7] G.A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information," *Psychol. Rev.*, vol. 63, pp. 81-97, 1956.