

# On Using Fuzzy Logic for Grading Highly Subjective Assessment Material – a Case Study

**Séamus C. McLoone**

*The Callan Institute, Department of Electronic Engineering,  
National University of Ireland Maynooth,  
Maynooth, Co. Kildare, Rep. of Ireland.  
email: seamus.mcloone@eeng.nuim.ie*

---

**Abstract—** It is difficult to remove subjectivity from assessment, especially when the assessed material is highly subjective by nature. While one assessor deems work to be of high quality, a second may view it as poor. Moreover, although two assessors may agree that the work is of satisfactory quality, they may nevertheless proceed to grade it significantly different, based on their interpretation of what is meant by satisfactory. While this is part of human nature, it is important that we strive for an assessment mechanism that is fair to all students, particularly when multiple personnel are involved in the grading process. This paper investigates the suitability of using fuzzy logic, and its ability to model linguistic terms, for the purposes of achieving fair assessment. A case study is carried out in which both the fuzzy logic and the standard numerical grading approaches are applied to a postgraduate poster competition. The results from this case study are presented and discussed within.

**Keywords –** Fuzzy Logic, Linguistic Variables, Subjectivity, Assessment Grading

---

## I INTRODUCTION

Assessment consists of many forms, ranging from simple multiple-choice questions and fact-based answers to large, written reports and essays. Those assessments that require simple, factual, mathematical, logical or even code-based answers are typically straightforward to grade. The answers are either correct or incorrect. In some cases, marks can be awarded for the methodology which, again, is either right or wrong. Some marks can be awarded for attempt also. In general, most assessors will grade such material in a consistent manner, based on a common marking scheme, as the assessment material is effectively devoid of subjectivity.

In contrast, assessments such as written reports, essays, posters, presentations, etc. can be more difficult to grade. Right and wrong answers no longer suffice. The material can be correct, but to what level is it correct? One assessor may find the material to be of a high standard while another may feel otherwise. As such, obtaining consistency in grading such material across different assessors is more challenging. Agreeing some sort of suitable marking scheme can help alleviate this issue to some degree. However this is not a trivial task and does not, in general, remove the issue of subjectivity.

Good assessment practices strive for fairness and consistency in grading, by reducing the impact of subjectivity [1]. It is important to recognise that two forms of subjectivity exist in the context of grading. One relates to assessors having different opinions of the same piece of work while the other

relates to assessors agreeing that the work is of a particular standard, but subsequently award different marks for it. Consider for example two examiners marking an essay. One may be impressed by the essay while the other is not. This reflects different subjective opinions by the examiners, due to any number of reasons such as differences in experience, expectations, etc. For ease of readability, this will be known as subjectivity form I for future reference. In contrast, consider the case where both examiners agree that the essay is of a satisfactory standard, yet can award significantly different marks for the essay. This reflects the examiners' differing viewpoints of what is meant by satisfactory in this case. This will be referred to as subjectivity form II.

Fuzzy logic [2] offers the potential of modelling linguistic variables and, hence, provides a mechanism for capturing the uncertainty introduced by subjectivity. Fuzzy logic has seen widespread use ranging from controlling vehicles [3] to evaluating journal grades [4]. It has also found its way into the area of education with fuzzy sets being employed in student evaluation [5] and assessment of student-centred learning [6].

In previous work by the author [7], a two-tier fuzzy grading system was proposed in an effort to eliminate subjectivity form II. Preliminary results, focusing solely on the viewpoints of the staff of the Electronic Engineering Department at the National University of Ireland Maynooth (NUIM), tentatively confirmed the validity of this grading system. This paper extends this work by carrying out a case study

where the proposed grading system is applied to a postgraduate poster competition, in which both the students and the judging panel consist of personnel from numerous different departments within NUIM. The results and observations from this case study are presented in detail later in the paper. Furthermore, the issue of what is perceived as ‘hard’ and ‘easy’ markers is also discussed.

The rest of this paper is structured as follows. The key concepts of fuzzy sets and linguistic variables are outlined in the next section. For the sake of the reader, section III presents an overview of the two-tier fuzzy grading system, as outlined in detail in [7]. The case study is presented in section IV while the results and observations from this study are given in section V. The issue of ‘hard’ and ‘easy’ assessors is also discussed here. The paper ends with some conclusions in section VI.

## II FUZZY SETS & LINGUISTIC VARIABLES

Boolean logic dictates that everything is classified as either 1 or 0, true or false or, in the context of grading, right or wrong. Values either belong to a particular set or they do not. This is acceptable in certain circumstances such as grading multiple-choice questions. However, the concept falls short when uncertainty or subjectivity is involved. The correct answer is no longer black or white, but usually lies within the gray area between. This is particularly evident when linguistic variables are used to describe an attribute. For example, terms such as wet, hot, satisfactory, etc. are all well known and commonly used expressions, but what specific quantity do we associate with such terms?

Words, or linguistic variables, such as 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

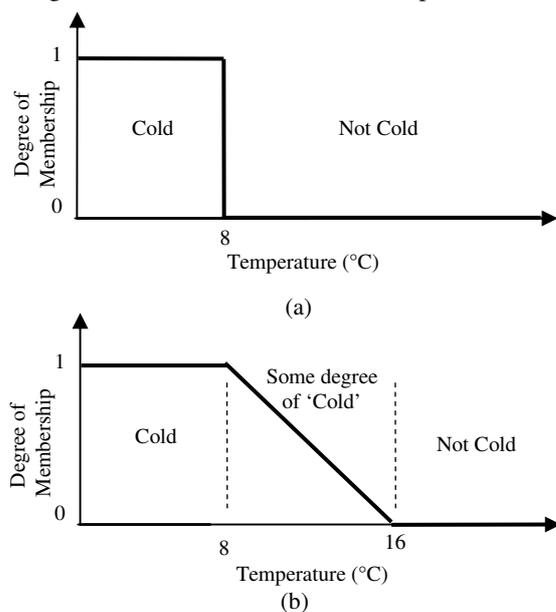


Figure 1: (a) Boolean and (b) Fuzzy logic representation of set of temperatures regarded as cold.

person considers cold, another person considers not cold. These varying subjective opinions exist as part of human nature but they are not easily measured and cannot be expressed objectively using standard Boolean-based methods.

Fuzzy logic [2] softens the boundary between black and white to incorporate the gray 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 cold. For argument sake, let's say that cold is defined as all temperatures below and including 8°C. Thus the Boolean set for cold would be represented as shown in Figure 1(a). In this case, a temperature of 8.01°C is absolutely not cold. Fuzzy logic adopts a more rational approach and regards the latter as belonging to the cold set to some degree. The degree of membership is directly related to how close the temperature is to the predefined value. Hence, 8.01°C has a very high degree of membership to the cold set, while 15°C has a very low one, as illustrated by the fuzzy set in Figure 1(b). In this example, temperatures greater than 16°C are absolutely not cold.

Fuzzy logic and fuzzy sets convey the typical imprecise nature of our language, providing a scientific means to deal with subjectivity. This is of particular relevance in grading assessments that are highly subjective by nature. The next section outlines a two-tier fuzzy-based grading system to alleviate the issue of subjectivity.

## III TWO-TIER FUZZY-BASED GRADING SYSTEM

The reader is referred to the author's previous work [7] for a detailed presentation of the two-tier fuzzy-based grading system. A summary is provided here for the sake of convenience.

The first tier of the grading system consists of the 7 linguistic variables *Excellent*, *Very Good*, *Good*, *Satisfactory*, *Poor*, *Very Poor* and *Extremely Poor*. An appropriate range was assigned to each of these variables, reflecting the views of the staff of the Department of Electronic Engineering at the National University of Ireland Maynooth (NUIM). 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. Figure 2 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. It is worth noting that these values reflect the overall group mentality. In other words, Figure 2 reflects the opinion of the majority of staff and smoothes out any extreme individual contribution.

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 *Satisfactory* is 50%. Now, if two or more assessors choose the same linguistic term, then the same exact numerical mark is awarded, removing the aforementioned issue of subjectivity form II.

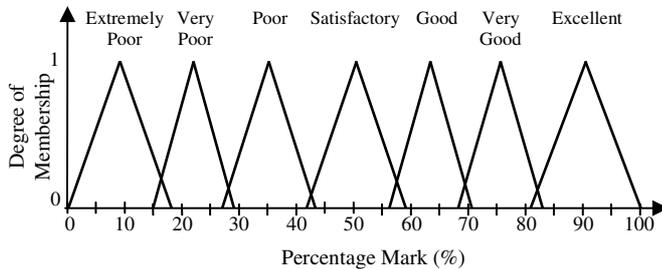


Figure 2: First tier of the fuzzy-based grading system.

This grading system effectively quantises the range of marks to one of seven possible numerical values. Clearly this is a very coarse level of quantisation and unsuitable for assessment purposes. Hence, a second-tier was included in the grading system, as follows. 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 3. This results in a further division of the numerical range, resulting in a total of 21 divisions, which is a more acceptable quantisation level for the purposes of assessment grading.

The next section outlines a case study, which is used to evaluate the performance of the outlined fuzzy-based grading system. The results from this study are discussed in the subsequent section.

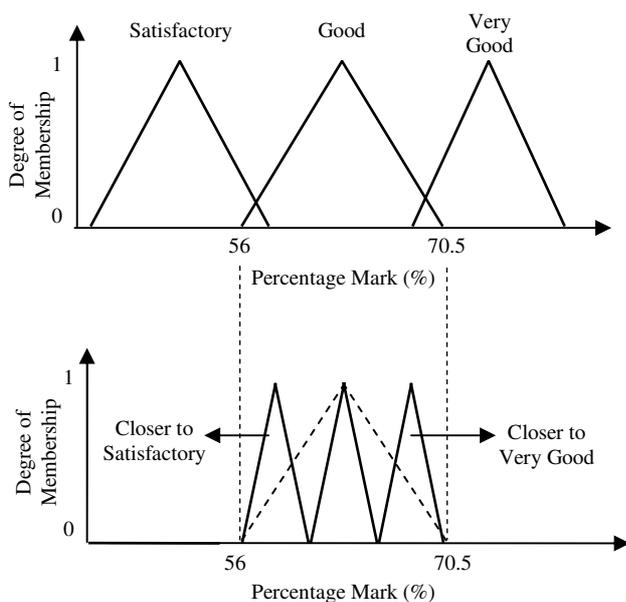


Figure 4: The two-tier fuzzy-based grading approach.

#### IV CASE STUDY – THE POSTER COMPETITION

A faculty-wide poster competition was organised in which postgraduates were able to showcase their research work. This involved a total of 49 postgraduates from the Departments of Biology, Chemistry, Computer Science, Maths and Psychology. The judging panel consisted of 8 different university staff members, 6 from the aforementioned departments and 2 from other departments within the university.

Each judge had up to 13 posters to grade, allowing for 2 judges per poster. Posters were graded according to three aspects, namely presentation, content and the ability of the presenter to answer questions. As part of the grading process each judge was required to use two different marking systems. The first involved providing a single numerical percentage value to the poster, i.e. the conventional approach. The second employed the two-tier fuzzy-based grading scheme as outlined in the previous section. For the purposes of the competition, the numerical grade was solely used to determine the overall winner. The latter grading system was used for research purposes only.

This poster competition provided a suitable case study for evaluating the performance of the fuzzy-based system for two main reasons. Firstly, posters offer a highly subjective form of material, where an assessor's viewpoints can significantly differ. In addition, as the posters related to research from different departments, this meant that judges would be more familiar with the content of posters closer to their area of expertise and less familiar with others. This, in itself, introduces an element of subjectivity, as one tends to be more critical or appreciative (depending on quality) of work that they can relate to, as opposed to material that is effectively unknown to them.

Secondly, the fuzzy-based grading system was designed for the staff of the Electronic Engineering (EE) Department. This competition provided an opportunity to see how the chosen numerical ranges for the linguistic variables would compare with the mindset of staff from other departments.

#### V RESULTS AND DISCUSSION

Due to unforeseen circumstances, fuzzy-based grades were not received for 7 of the 49 candidate posters. As such, these were removed from the collected data and the results presented here relate to the remaining 42. Table 1 shows the average percentage mark awarded to each poster for both the conventional (i.e. numerical) and the fuzzy-based grading schemes, in columns 2 and 3 respectively. Note that, while a numerical value is given in each case, the latter is only evaluated after the process, based on Figures 2 and 3. In other words, the judges only choose the appropriate linguistic variable. They do not know what numerical value is associated with

their choice. Column 4 shows the different between both sets of marks. Negative differences indicate that the fuzzy-based grading scheme resulted in a lower numerical value than that of the conventional one and vice versa.

In order to explore the issue of subjectivity in detail, the actual linguistic variables chosen for each poster needs to be examined. These are presented in Table 2. It should be noted that in some cases, three grades are presented for a given poster. This occurred due to some judges marking additional posters in error. However, this does not affect the analysis in this paper.

Examining Table 1, it can be seen that the difference in the grading schemes varies from 11 for posters 17, 18 and 27 to -9.5 for poster 30. These extreme values reflect the differences in opinion of some members of the judging panel and the staff of the EE Department in assigning numerical values to the linguistic variables. By way of example, consider poster 18. Here, both assessors agreed that the poster was excellent in standard, choosing the mid value of

Poster No.	Conventional Grading (CG)	Fuzzy-based Grading (FG)	Difference (FG - CG)
1	63.5	72	8.5
2	67.5	71	3.5
3	77.5	80	2.5
4	60	69.5	9.5
5	72.5	69.5	-3
6	72.5	72	-0.5
7	68	73.5	5.5
8	63.5	69.5	6
9	79.5	82.5	3
10	68.5	72	3.5
11	71.7	75	3.3
12	64.5	69.5	5
13	67	72	5
14	67.5	69.5	2
15	75	78.5	3.5
16	70	65.5	-4.5
17	76.5	87.5	11
18	80	91	11
19	67.7	69	1.3
20	64	69.5	5.5
21	69.5	74.5	5
22	75	77.5	2.5
23	78.5	82.5	4
24	66	73.5	7.5
25	65	73.5	8.5
26	63	63.5	0.5
27	80	91	11
28	65	59	-6
29	69	67	-2
30	60	50.5	-9.5
31	71.5	69.5	-2
32	71.5	67	-4.5
33	67.5	67	-0.5
34	61.7	60.3	-1.4
35	57.5	56.5	-1
36	57.5	63	5.5
37	62.7	66.3	3.6
38	60	63	3
39	55	56.5	1.5
40	62.5	69.5	7
41	71	73.5	2.5
42	76	80	4

Table 1: Calculated percentage averages for posters for both conventional and fuzzy-based grading schemes.

the 'Excellent' range in both cases (see Table 2). This corresponds to a numerical grade of 91%, as derived from Figure 2. However, the two assessors awarded 82% and 78% respectively. In other words, these assessors felt that an almost perfect poster was worth no more than 82%. Similarly, in the case of poster 30, the two assessors (different from those for poster 18) agreed that it was of 'Satisfactory' standard (one choosing the low end, the other choosing the high end, as in Table 2) and both gave it a 60% grade. This is in contrast with the calculated fuzzy-based grade of 50.5%.

Interestingly, the issue here is not one of subjectivity between assessors, as both are in general agreement on the standard of the poster and provide similar grades. Instead, this is possibly a reflection of the fact that the assessors of poster 18 can be regarded as the 'harder' set of markers within the judging panel, while those of poster 30 can be regarded as the 'easier' set of markers. In the former case, there is also the possibility that some of the assessors do not make use of the full range of marks, i.e. from 0 to 100%.

It is worth noting, that although both schemes gave different sets of grades, both produced the same set of joint winners, i.e. posters 18 and 27. However, at the other end of the scale, the worse poster is different depending on the chosen grading scheme. The conventional scheme (and hence the actual decision) indicated that poster 39 is the worst, whereas the fuzzy-based scheme chooses poster 30.

In the case of poster 39, the marks awarded reasonably matched the linguistic terms chosen, as evident from the small difference between them. On the other hand poster 30 was assigned worse linguistic terms but better grades. As already noted, this likely reflects the fact that the assessors involved were both 'easy' markers. This example provides a good argument for the use of the fuzzy-based scheme, where a student is awarded the mark that reflects the average opinion of a group (in this case the EE staff) as opposed to the mark that reflects the (possibly extreme) viewpoint of just one or two people. Thus, in the author's opinion, poster 30 should have been deemed the worse poster.

#### a) Subjectivity form I

Subjectivity form I refers to the case where assessors differ on their opinions of a poster. This is clearly evident from Table 2. Consider, for example, poster 25. One judge feels that it is of an 'Excellent' standard while the other feels that it warrants no better than 'Good'. The assessors involved are marking the same poster but clearly differ in their opinions of its quality. Similar observations can be extracted for several of the posters in the case study.

This form of subjectivity provides a more difficult challenge than subjectivity form II and neither of the grading schemes currently deal with it effectively. In fact, they simply don't deal with it.

*b) Subjectivity form II*

Subjectivity form II refers to the case where the assessors share the same opinion of a poster but award it different grades. Table 2 shows that in many cases the judges shared very similar, and sometimes identical, opinions of the poster involved. However, analysis of the conventional grades awarded (the full set of grades are omitted due to lack of space) reveals that, in several cases, they marked the poster differently. In the case of poster 2, for example, both judges assigned the poster as ‘Very Good’ but graded the same poster 73% and 62% respectively, a difference of 11%. Poster 9 offers one of the worse case scenarios where, once again, the judges have very similar opinions on the standard of the poster and yet grade it 69% and 90% respectively, a difference of 21%. These results provide additional support for the use of the fuzzy-based grading system, where the poster is marked based on the view of the majority of assessors as opposed to the possible extreme view of the individual. The conventional method of grading does not alleviate this issue.

*c) ‘Hard’ and ‘Easy’ assessors*

In the current grading scheme, one poster could be marked by two ‘hard’ markers while another could be marked by two ‘easy’ markers. Clearly, students would prefer the latter scenario. The ideal solution is to have all assessors grade all posters and so the issue of ‘hard’ and ‘easy’ markers is avoided, as everybody is effectively treated the same. However, this is not a very practical solution as it requires each member of the judging panel having to correct all 49 posters. This is both time consuming and resource intensive.

The issue of ‘hard’ and ‘easy’ markers is relatively common and exists for various reasons. As humans, we naturally have different levels of experience, different expectations and different personalities, all of which can lead us to having different opinions when grading a poster, for example. As such, some will mark it easier while others will invariably mark it harder.

Arguably, a group of assessors could express their views on what makes a good or bad poster in an effort to come to a general consensus. However, this is not a trivial task and is unlikely to remove the inherent subjective opinions of the various assessors. An alternative solution is to accept the fact that ‘hard’ and ‘easy’ markers exists and to determine a mechanism by which grades can be adjusted to allow for such markers. For example, grading statistics, collected over several years can potentially indicate which category an assessor may belong to.

In this case study, the numerical values chosen by a judge for a set of linguistic variable can be compared to the averages given in Figure 2. If, in general, their values are less than the average, then it can be argued that they are ‘hard’ markers. If their

values are higher, then they are ‘easy’ markers. Finally, if they are close to the chosen averages then they can be considered ‘fair’ markers. By noting the numerical values and linguistic terms recorded by each of the 8 judges, it was possible to categorise each judge as either ‘hard’, ‘easy’ or ‘fair’. Of course, it should be noted that this is relative to the views of the EE staff. A similar study for a different department could yield a higher or lower numerical range for the specified linguistic terms, given in Figure 2. In other words, the EE staff could be viewed as ‘easy’ or ‘hard’ markers from the viewpoint of another department.

Poster No.	Excellent			Very Good			Good			Satisfactory		
	H	M	L	H	M	L	H	M	L	H	M	L
1				X				X				
2						X,X						
3			X		X							
4					X			X				
5					X			X				
6					X			X				
7			X					X				
8					X			X				
9			X	X								
10					X			X				
11		X				X		X				
12					X			X				
13				X				X				
14					X			X				
15				X	X							
16				X								X
17		X	X									
18		X,X										
19					X		X	X				
20					X			X				
21					X			X				
22			X			X						
23			X	X								
24					X	X						
25			X					X				
26						X					X	
27		X,X										
28							X					X
29						X		X				
30										X		X
31						X	X					
32						X		X				
33						X		X				
34								X	X,X			
35								X			X	
36								X,X				
37							X,X	X				
38								X,X				
39								X			X	
40					X			X				
41					X	X						
42	X							X				

Table 2: The linguistic variables chosen for each poster.

Consider for example, a sample of 3 judges as presented in Table 3. Here, the numerical values and the linguistic terms chosen by each of the judges are given. Ideally, it would be useful to have all judges assign values to all the linguistic terms, but unfortunately, this was not carried out as part of the case study. However, sufficient information is available to argue a case for each judge. The last row in the table reflects the average values as chosen by

the EE staff. These are used as a point of reference. From this table, it is clear that judge A closely reflects the reference values and, as such, can be regarded as a ‘fair’ marker. Judge B consistently assigns lower numerical values to the linguistic terms and is, therefore, regarded as a ‘hard’ marker. Finally, Judge C consistently assigns higher numerical values (although, in some cases, the difference is quite small) and is, thus, an ‘easy’ marker. Using this principle, the 8 judges were determined to consist of 4 ‘hard’, 2 ‘fair’ and 2 ‘easy’ markers.

Once, a marker has been categorised, it is now possible to adjust grades to allow for the nature of the marker. Using Table 3, a heuristic adjustment of +10 for a hard marker and -5 for an easy marker is applied to the conventional grades in column 2 of Table 1. The adjusted conventional grades were calculated, along with a new set of differences between these and the fuzzy-based grades. For ease of illustration, the absolute values of the old and new set of differences are presented in graphical form in Figure 4. The solid line represents the old differences (i.e. between the conventional and fuzzy-based schemes) while the dashed line represents the new differences (i.e. between the adjusted conventional and fuzzy-based schemes).

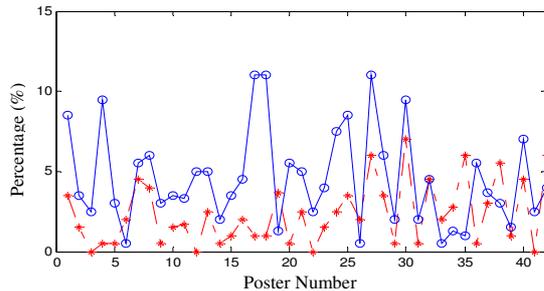


Figure 4: The absolute percentage difference between grading schemes; solid line represents difference between conventional and fuzzy-based schemes; dashed line represents difference between adjusted conventional and fuzzy-based schemes.

Figure 4 shows that, in general, by adjusting for the perceived ‘hard’ and ‘easy’ assessors, the new conventional marks are more aligned with the fuzzy-based grading scheme. In other words, the difference between the two is now smaller. There are a few exceptions to this, however. This is largely due to two key issues. Firstly, a heuristic adjustment measure was used and applied to all judges equally. However, consider the 4 ‘hard’ judges as an example. In this case, some of the judges were harder markers than others and yet all were adjusted in the same fashion. Perhaps a sliding scale could be employed to address this issue. Secondly, a ‘hard’ marker was assumed to award lower than average marks across the range of linguistic variables. This was not always the case. In some instances, a marker awarded lower than average marks for one linguistic

variable, but closer to the average for another. A similar analogy applies to an ‘easy’ marker. This issue requires some further study. Nevertheless, the results show that the fuzzy-based scheme can certainly help to alleviate some of the issues associated with different levels of markers, as the grades from the fuzzy-based grading system are more aligned with the adjusted set of conventional grades.

Judge	Excellent			Very Good			Good		
	H	M	L	H	M	L	H	M	L
A						73		63	
B		81		68	65	62			
C			90	85	80	75	70		
EE staff	97	91	84	81	76	71	68	63	59

Table 3: The percentage values chosen for certain linguistic variables by 3 different judges A, B and C. The EE staff values are shown for reference purposes.

## VI CONCLUDING REMARKS

This paper has carried out a case study, examining the use of a two-tier fuzzy-based grading scheme using linguistic variables. The case study has shown that this grading method offers several advantages over the traditional method of assigning a numerical value. It can be argued that the fuzzy-based scheme alleviates the issues associated with ‘hard’ and ‘easy’ markers. In addition, the inherent subjectivity in grading, where two markers agree on a specific standard but award differing grades is also eliminated. This results in a fairer assessment for those on the receiving end of the grades. In addition the fuzzy-based system also eases the assessment process for the assessor as it is simpler to choose a linguistic variable than it is to assign a specific numerical value. Unfortunately, neither the conventional nor the fuzzy-based grading systems adequately deal with the other aspect of subjectivity, where two assessors have varying opinion on the standard of a piece of work. This remains a key aspect of future studies.

## VII REFERENCES

- [1] P. Race, S. Brown and B. Smith, *500 Tips on Assessment*, RoutledgeFalmer, 2<sup>nd</sup> edition, 2005.
- [2] L.A. Zadeh, “Fuzzy sets”, *Information and Control*, vol. 8, pp. 338-353, 1965.
- [3] S. G. Kong and B. Kosko, “Adaptive fuzzy systems for backing up a truck-and-trailer”, *IEEE Paper on Neural Networks*, vol. 3, no. 2, pp. 211-233, 1992.
- [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*, 43 no. 2, pp. 237-241, May 2000.
- [7] S. C. McLoone and N. O’Neill, “Can fuzzy-logic remove subjectivity from final year project assessment?” *Irish Signals and Systems Conference*, IET, Dublin, Ireland, pp. 67-72, June 2006.