Half Empty, Half Full - An Examination of Subtractive Versus Additive Assessment

Brett A. Becker and Dr. Kevin Casey
brett.becker and kevin.casey {@gcd.ie}

Faculty of Computing Science
Griffith College Dublin
Dublin, Ireland

Abstract

In this paper, we present a case study comparing subtractive marking schemes with the more common technique of additive marking. Although often accepted for use in oral language and multiple choice exams, subtractive marking has yet to see serious consideration as an accepted technique for assessing other more common paper exams, particularly Computer Science exams with subjective questions. This paper is presented in a number of sections. After a brief overview of previous investigations in the area, we outline an experiment conducted with real examination papers taken by students at Griffith College Dublin. In this experiment, we mark the same set of papers using both additive and subtractive schemes. We then summarise the differences between the two techniques and identify some of the challenges, advantages and disadvantages of subtractive approaches and also the motivations behind them. We also examine how different types of exam questions affect the difference between additive and subtractive marking and make the argument in favour of subtractive marking as a useful QA technique. Finally, we present the results of a student survey regarding their opinions on subtractive marking in order to gain an insight as to how students feel about the concept and what types of students feel strongly for or against it.

Keywords

Additive Marking, Subtractive Marking, Negative Marking, Marking Schemes, Student Assessment, Exam Assessment, QA

1. Introduction and Motivation

Student assessment, in particular the marking of exam scripts and cumulative assessment work, has traditionally been based on predefined marking schemes. These schemes determine how many marks are awarded for correct or partially-correct answers to questions or problems. The correctness of an answer to a given question is often subjective in nature and depends heavily upon the marking scheme. Examples of stipulations in typical marking schemes include mentioning specific keywords, stating required definitions, and drawing comparisons between different topics/techniques/problems/solutions etc. Such marking schemes are commonly *additive* in nature – that is the assessment begins with the student having earned 0 marks for a given script or question, and then points being awarded for adherence to what is expected from the marking scheme (taking into account breadth, depth, insight, etc.).

It is common in oral language assessment (Underhill, 1987), and multiple choice (Holt, 2006), (Scouller, 2006) exams to use a subtractive marking scheme where the student starts out with the presumption of having earned full marks (throughout this paper understood to be 100), and marks are then subtracted for errors or omissions made in answering questions – in other words marks are subtracted for how incorrect an answer is, not awarded for how correct it is. In other examination formats such as a typical Computer Science exam, students may be required to state definitions, answer questions similar to multiple choice questions, solve problems using computer code, demonstrate insight to theoretical questions, and make comparisons between different approaches and techniques of various problem solving methods in essay form. With such a diverse set of question styles, the comparison between additive and subtractive marking techniques is fraught with difficulty and not well studied in the literature. Subtractive marking techniques can be found in some state examinations (State Examinations Commission, 2009) and university examinations (Downing, 2009).

When it comes to the possible differences between the outcomes of additive and subtractive marking there are several potential reasons to believe these differences do exist (Samuelowicz and Bain, 2006). It has been proposed that people perceive more impact when asked to assess whether an action would increase the likelihood or

degree of a potential outcome (mental addition) than when asked whether it would reduce the probability or extent of a potential consequence (mental subtraction). It is hypothesized that this judgmental asymmetry occurs because people give more weight to factors that produce a "positive" outcome as opposed to inhibit the relevant outcome (Dunning and Parpal, 1989). This is an example of the ubiquitous "glass is half full / glass is half empty" adage being applied to student assessment.

Very little work has been done in researching the differences between multiple methods of assessing Computer Science exams. To the authors' knowledge, only one paper has been published in this area, focusing solely on how code is assessed (Denny et. al., 2008).

2. Experimental Technique

To explore the difference between additive and subtractive marking schemes on Computer Science exams, 68 exam scripts taken by students at Griffith College Dublin were studied. The course in question is 'Data Structures, Algorithms and Complexity'. The student cohorts were mixed: second and third year BSc (Honours and Pass levels), as well as Higher Diploma students. In addition, some students were full-time and some part-time. The part-time students were taught and assessed by one author and the full-time by the other. However the exam and marking scheme were written by one author only. The exam was administered in January 2009 and marked by the authors with a standard additive marking scheme. The same scripts were then blindly re-marked by the authors in June 2009 using a subtractive scheme. The subtractive scheme was based on the additive, essentially by 'inverting' the additive scheme (see example questions and schemes below).

The exams were partially subjective in nature with greatly varied styles of questions. Table 2.1 shows brief examples of the breadth of question styles along with the additive marking scheme for each question.

Exams with questions such as those in Table 2.1 are differentiated from multiple choice/essay exams due to the subjective nature of requirements such as *illustrate,*

example, diagram, discuss, explain, write, and *implement.* Table 2.2 shows the same questions with the subtractive marking scheme.

Example Question 1 (Additive - student starts with 0 marks)

(a) Illustrate an example of the non-circular queue data structure using a diagram(s) **AUT 2 marks** Discuss the operations *isfull, isempty, enqueue and dequeue.* **AUT 1 mark each**

Current Maximum (0 + 6 = 6 marks)

(b) Explain the difference between a simple queue and a circular queue **AUT 2 marks** and any advantage each has over the other. **AUT 1 mark**

Current Maximum (3 + 6 [from (a)] = 9 marks)

(c) Write pseudocode to implement a non-circular queue **AUT 2 marks** and a circular queue. **AUT 4 marks** Be sure to include the operations enumerated in (a).

Current Maximum (6 + 9 [from (b)] = 15 marks)

Maximum Possible Total (15 marks)

Table 2.1 – Example exam question and associated additive marking scheme. AUT = "Add Up To"

Example Question 1 (Subtractive - student starts with 15 marks)

(a) Illustrate an example of the non-circular queue data structure using a diagram(s) **SUT 2 marks** Discuss the operations *isfull, isempty, enqueue and dequeue.* **SUT 1 mark for each incorrect (4 total)**

Current Minimum (15 – 6 = 9 marks)

(b) Explain the difference between a simple queue and a circular queue **SUT 2 marks** and any advantage each has over the other. **SUT 1 mark**

Current Minimum (9 [from (a)] – 3 = 6 marks)

(c) Write pseudocode to implement a non-circular queue **SUT 2 marks** and a circular queue. **SUT 4 marks** Be sure to include the operations enumerated in (a).

Current Minimum (6 [from (b)] – 6 = 0 marks)

Minimum Possible Total (0 marks)

Table 2.2 – Example exam question with associated subtractive marking scheme. SUT = "Subtract Up To"

It is easy to see that awarding points for the quality of illustrating an example of something and taking points for the lack of the quality of illustrating an example of something are distinctly different tasks. In theory, the final judgment of how well some aspects of a task have been completed and how poorly other aspects of the same task have been completed should sum to unity. In other words regardless of what marking scheme is used the final mark should be the same. However, as we will

discuss later, due to the subjective elements of some questions and the different psychologies involved in the two approaches this may not always be the case.

3. Results

Figure 3.1 shows a summary of the results marks obtained by comparing the additive and subtractive marking schemes. Each point represents an exam question with the size of the point representing the number of students who chose to answer that question. The average mark per question is plotted on the x-axis while the differences between the additive approach and subtractive approach are plotted on the y-axis. Essentially the higher a point on the y-axis is, the greater the difference in mark from the additive compared to the subtractive approach. All questions are normalised to 1. Therefore if a point appeared at height 1, it would mean that the average mark for that question was 100% in the additive scheme and 0% in the subtractive.

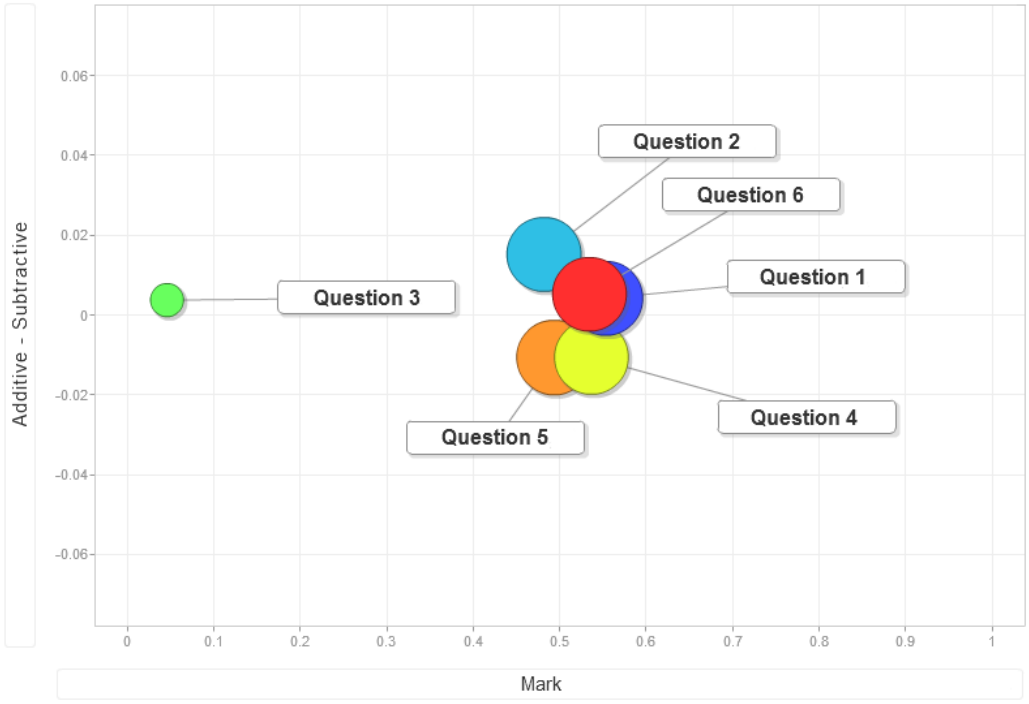

Figure 3.1 – Average marks per question (additive versus subtractive)

A number of features are obvious from the graph. Firstly question 3 looks like a significant outlier but this is true only in the sense that very few students opted to answer that question (i.e. it is an outlier on the x-axis) and the few that did answer it

did so only half-heartedly. Examining variation on the y-axis, we see that there is very little difference between the subtractive and additive approaches. The greatest variation between the two approaches is with question 2, where additive marking gave an average mark that was 1.5% higher than subtractive.

Although the results suggest only minor variations between the two approaches, the authors' decided to examine the small differences that were exhibited. In order to do this, the questions in the original paper were ranked according to their subjectivity. Some questions required mathematical operations or snippets of computer code and were more objective from an assessment perspective, while others asked for student explanations of particular concepts which would be more subjective. These styles of assessment were mixed within questions, but we were able, nonetheless to rank questions according to the level of subjectivity.

The ranking of these questions according to their level of subjectivity is in itself subjective to a degree. To avoid any bias from the authors, a survey of 10 independent Computer Science lecturers was carried out and the average ranking is presented in Table 3.1.

	Question Number	Additive-Subtractive
Most Subjective	2	1.50%
	3	0.04%
	1	0.40%
	6	0.50%
	4	-1.10%
Least Subjective	5	-1.10%

Table 3.1 – Subjectivity of the questions

What we note from the ranking, compared to Table 3.1, is that the most subjective questions seem to have a tendency to have a higher mark from the additive approach than the subtractive. This slight tendency could be explained by a natural bias in favour of the student which exhibits itself more strongly in subjective questions. In

objective questions it is difficult to apply any bias without departing from the marking scheme. The authors found that one of the effects of the subtractive scheme is that they were less likely to apply a bias in favour of the student for more subjective schemes. Of course, this may not be the case for all examiners.

There is a converse effect at work for the more objective questions that could possible yield a higher mark when using a subtractive marking scheme. For questions where there are multiple correct answers, an examiner may have a bias in favour of one correct answer over another. With the subtractive scheme this bias is less likely to exhibit itself since the focus of the examiner is drawn to omissions rather than inclusions in a student's answer.

4. Subtractive Marking as a useful QA procedure:

"Insanity: doing the same thing over and over again and expecting different results."
- Albert Einstein

While there is normally no substitute for getting a second examiner to check some of your assessment, very often it just is not practical. In the authors' college, a percentage of papers are generally selected by the lecturer and re-examined for anomalies. The biggest flaw with this approach is that the same lecturer who originally marked the assessment is now re-marking the assessment using the same techniques as before, and is arguable quite likely to make the same mistakes (if there are any) again.

During the preparation of this paper, the authors' decide to forgo the usual selection of a percentage of papers for re-examination. Instead, the entire set of examination papers were re-marked using a subtractive scheme (as outlined in Section 2). The results were quite interesting in that using an alternative procedure for marking allowed for the successful detection of grading anomalies. After the subtractive marking procedure, the marks for the additive and subtractive procedures were compared and where any substantial differences occurred, the paper was examined for assessment errors. This strategy proved to be highly effective as outliers invariably turned out to be the result of grading anomalies.

To summarise, from a QA point of view: There is no substitute to getting someone else to check at least a selection of one's assessments. However, if this is not possible, then using a different technique that generates similar grades is preferable to applying the same technique twice. Subtractive marking is one such alternative technique.

One noteworthy point here is that the authors found the subtractive approach a little slow and unwieldy initially. This improved considerably with practice. For example, in the case where a student omits a compulsory question, the examiner must consciously remember to include the subtractive mark for the entire mark of that missing question.

5. Student perception of subtractive marking:

While there is much argument (McGuire, 1999) and feedback against (Boyle, 2002) negative marking, there is little, if any, work done on the issue of student perception of subtractive marking.

To get a preliminary feel about whether students might prefer one approach over the other, a survey was carried out by the authors immediately prior to an examination. An explanatory note was given to students detailing the two approaches of subtractive and additive marking. To avoid stressing the respondents unnecessarily, it was made clear to students that only additive marking would be used for the rest of their academic year, regardless of their answers.

Approximately 50% of the 49 students surveyed were asked which approach they preferred, subtractive or additive by circling a number on a scale from 1 to 5 with 1 identifying that the student was highly in favour of subtractive marking and 5 that the student was highly in favour of additive marking. To avoid any left-right bias, a second version of the questionnaire was distributed to the other 50% of the students with the scale switched (i.e. 1 representing a strong preference for additive, and 5 a strong preference for subtractive marking). A qualitative section was also given to all students so that they could supply additional comments.

Initially there was no clear trend in the tabulated results. Roughly equal numbers of students preferred each approach. Even when factoring in student grades, we did not see any substantial trend with respect to the preference between the marking schemes. However, when we looked at the simpler matter of the strength of the students' opinions a trend did emerge. In order to quantify this, we categorised each student on a scale of:

Category	Source
0	No opinion (i.e. the student circled 3 on the questionnaire)
1	Weak opinion (i.e. the student circled 2 or 4 on the questionnaire)
2	Strong opinion (i.e. the student circled 1 or 5 on the questionnaire)

Table 4.1 – Categorising respondents

Then, averaging the grades for each level of opinion, we obtained the following results:

Category	Average Result
0	40.48%
1	30.27%
2	26.61%

Table 4.2 – Average marks in each category of respondent

Stronger students appear to be more confident that their work will stand on it own merits, regardless of the marking scheme used. Weaker students, on the other hand, seem to have a strong preference for one or the other scheme. What this suggests to us is an appreciable correlation between student anxiety and non-performance. Qualitative feedback on the questionnaires seems to support this theory.

6. Conclusions and Future Work

Subtractive marking schemes are not in common usage, except perhaps in Multiple Choice Questions and oral language exams. However, when applied to other “more traditional” examinations (such as those in Computer Science subject areas), several interesting results and observations arise. We have found that in this particular case study, the difference in marks between additive and subtractive marking schemes is negligible, but there are several trends that can be observed.

Firstly, the more subjective a question is, the higher a mark the additive scheme results in, and conversely the less subjective a question is the higher a mark the subtractive scheme results in. Secondly, the stronger a student is the less the student is concerned with what assessment scheme is used. It seems that confident students believe that their work will stand up to examination, regardless of the assessment scheme used. On the other hand, weaker students have expressed a stronger opinion either towards additive or subtractive schemes, possibly out of a lack of confidence and therefore a belief that one scheme or the other would benefit their marks in some way.

Additionally, and precisely because of the negligible difference between the two approaches, the authors found that subtractive marking can be a very useful QA tool. When re-checking assessments for QA reasons it seems advantageous to re-check examination scripts using a technique other than that initially used. Our argument here is that if a different scheme than the original is used, there is a greater probability that assessment errors (if there are any) will be detected.

Future work involves investigating if subtractive methods are more suited for certain problem types over others and consequently if there is a strong case for hybrid schemes.

We conclude that, at least for the type of traditional Computer Science examination used in this case study, there seems not to be a significant difference between the additive and subtractive approaches to assessment. However, as noted in Section 4, this yields the opportunity to use subtractive marking as a QA procedure to recheck examination papers. Psychological differences between the two assessment schemes, both for the student and examiner, have been discussed, but further examination of these differences would be quite interesting. Finally, we suggest exploring if other subject areas exhibit a difference between the two marking schemes. Such exploration can only enhance the quality, reliability and understanding of examination assessment.

7. References

Boyle, A. P., (2002). A reflective study on the use of CAA to test knowledge and understanding of mineralogy. 6th International CAA Conference (ed Danson, M.), pp. 51-62, University of Loughborough.

McGuire B. (1999) Multiple choice examinations. *Anaesthesia*; 54(7): 720.

Denny, P., Luxton-Reilly, A., and Simon, B. (2008). Evaluating a New Exam Question: Parsons Problems. ICER'08, Sydney Australia.

Downing, A. (2009). Personal Communication. Assembly Language Marking Scheme, Dept of Computer Science, The University of Westerrn Ontario.

Dunning, D., and Parpal, M. (1989). Mental Addition Versus Subtraction in Counterfactual Reasoning: On Assessing the Impact of Personal Actions and Life Events. Journal of Pereonality and Social Psychology, 1989, Vol. 57, No. 1,5-15.

Holt, A. (2006). An Analysis of Negative Marking in Multiple-Choice Assessment. 19th Annual Conference of the National Advisory Committee on Computing Qualifications (NACCQ 2006), Wellington, New Zealand.

Samuelowicz, K. and Bain, D. (1998). Identifying Academics' Orientations to Assessment Practice. Higher Education 43: 173–201,

Scouller, K. (2006). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. Higher Education 35: 453–472.

State Examinations Commission (2009). Marking Schemes for Leaving Certificate Examinations. Available at http://www.examinations.ie

Underhill, N. (1987). Testing Spoken Language: A Handbook of Oral Testing Techniques. Cambridge University Press