

This essay is copyright under the creative commons Attribution - Noncommercial - Share Alike 3.0 Unported licence. You are encouraged to share and copy this essay free of charge. See for details: <http://creativecommons.org/licenses/by-nc-sa/3.0/>

## **Testing the easy way**

### **Setting and marking examinations, with special focus on essays**

#### **Seminar notes**

Last updated 30 April 2009

This paper should be read in conjunction with the paper on double correction available on this site.

### **Outline**

- A. Introduction
- B. My objectives for examinations
- C. False positives and false negatives
- D. Writing questions
- E. Reducing chance
- F. Analytic versus impression marking
- G. Steps to clear marking - concluding remarks

### **Available material**

1. Lowe I (2001). Statistics and research design: essential concepts for working teachers. *English Teaching Forum* 39(3)36-42
2. Weir CJ (1990). *Communicative language testing*. UK: Prentice Hall, especially chapter 4: Test Methods, and pages 58-73, Testing Writing. Page numbers below refer to this text.

## **B. MY OBJECTIVES FOR EXAMINATIONS**

### **1. Easy to mark**

- a. fast
- b. fair
- c. foolproof - students do as expected, and all markers agree together
- d. fatigueproof- marks are consistent even when hundreds of papers.
- e. free-of-errors. This means that even experienced NS should double check!

### **2. Clear**

- a. Precise questions (example later)
  - 1) The question itself, what is expected
  - 2) Mechanics. Should there be two booklets or one? Should students write on the paper in the spaces provided? Eg, fill in the gap exercises. If only one word is possible for each case, then number the gap, and get students to write their answers in a neat list.
- b. Nothing ambiguous about the instructions
- c. Error free
- d. Hard for students to cheat

### **3. Measure what has been taught**

- a. Content, both knowledge and concepts
- b. Skills, especially inferencing, combining class work, reading, with general knowledge and linking with other subjects (this really tests to see if the material has been internalised or not). Be realistic. The examination should be the same level as the latest material covered in class, neither too easy nor too hard.

### **4. Clearly differentiates good from bad students**

ie separates them. This means a lot of high marks and a lot of low marks.

### **5. Maintains standards**

### **6. Is as fair as possible**

It is this we shall explore in more detail.

How is this checked? Creativity is needed. In laboratory examinations (the part where a tape is played, which is used to test listening comprehension, and often has a section testing their hearing of vowels), the teacher conducting the examination is expected to do the listening comprehension, or the vowel discrimination tests. If for any reason it is not clear, then the question is withdrawn! Not included!

### **C. FALSE POSITIVES AND FALSE NEGATIVES**

1. A 'false positive' in the context of examinations is a student who succeeds when they do not deserve to. A 'false negative' is a student who does not succeed when they deserve to succeed.
2. False positives and false negatives can be reduced by skilful examinations. See section E below: reducing the role of chance.
3. The relative importance of false positives and false negatives varies with the examination. For the driving licence it is very important NOT to pass someone unless they are capable, therefore, even though it is distressing, it is vital for the examiner to be ultra cautious and to fail candidates even if they deserve to pass (false negatives). Similar cautionary tendencies apply to appointing senior staff, awarding medical degrees and so on.

In some examination systems, those who fail are allowed to repeat a few weeks or a few months later. Because repeating is easily organised, the pressure on the first examination will be to avoid the false positives and to risk a higher failure rate than is theoretically fair. In the second session, since the consequences for failure may be that students can never resit, or they have to resit a year later, then the emphasis should turn to avoiding false negatives, or false failures. Yet there are other factors at play. Commonly, teachers are tougher on the resits than on the main session. The argument goes that these students have had time to work harder on the subject therefore can reasonably be expected to achieve a higher mark, therefore the pass mark must be raised.

Factors such as this need to be discussed by the examiners and consistent policy agreed upon and implemented.

## **D. WRITING QUESTIONS**

### **1. All examinations can be improved by:**

- a. Specifying what is expected of the candidate. eg Set a broad question, but give instructions, or an outline, or supplementary small questions which explain what you want.
  - b. The average/good student should never have any doubt about what you expect. There should be no ambiguity or requirement for a student to guess what is wanted.
2. A clearly specified essay topic leads to a clearer marking scheme, and more reliable marking.
  3. Planning is the key. A well set question is much easier to mark and repays the extra effort.
  4. Questions should not just test knowledge. They should test the ability of a student to think for themselves, to apply, to integrate the different subjects.
  - 5.. Number of writing tasks p71 "Reliability of a test score tends to increase as the number of items in the test is increased". At least two compositions from a student reduces marking error - especially when only one teacher does the marking.

## **E. REDUCING THE ELEMENT OF CHANCE**

Let us assume there are three exams, and each teacher gives either 9, 10, or 11. Let us further assume that you have 27 students who are all equal in ability and performance. The exams were all impression marked (even with double correction). But we know that there is a high element of chance. Let us assume there is an equal probability on each exam that a student will get a 9, 10, or 11. These are reasonable assumptions, to show the effect chance can have. Then, for 27 students, these are the scores:

27 marks 1 student  
28 marks 3 students  
29 marks 6 students  
30 marks 7 students  
31 marks 6 students  
32 marks 3 students  
33 marks 1 student

That means, 17/27 pass, solely on the basis of chance, ie 63%.

When there is impression marking and marking close to the average, then the decision as to pass or fail is largely a matter of chance. Double correction using impression marking only reinforces the trend to give a middle mark that does not clearly separate a good student from a bad student.

## F. ANALYTICAL AND GENERAL IMPRESSION MARKING

1. Essays are the ONLY means of testing the ability to develop an extended argument in a logical manner.
2. There are many disadvantages. eg
  - a. The candidate may have no interest in the topic
  - b. There may be different ways the same topic is approached, and these several approaches are difficult to consistently mark
3. Analytical marking means to score different features. Impression marking means to give a global mark. Impression marking means reading 20-30 scripts carefully to get an overall impression, then reading and marking fast all the scripts. It seems that impression marking works best when it is done quickly.
4. All problems (actual mark, and rank order, which is a more serious problem!) can be reduced by using an analytical mark scheme.
5. **Impression marking problems:**
  - a. Too easy or too hard
  - b. Rank order differences. This means that some students are marked too leniently and some are marked too severely. This problem is reduced when double marking takes place  

Rank order error is more serious than being too easy or too hard, and is certainly more unfair.
  - c. Tendency to judge candidates on superficial reasons, to pick on a few points, such as errors in spelling, grammar, facts etc.
  - d. Tendency to be influenced by one or two aspects of the work. For example, I once had this problem when working with another teacher scoring oral performance. Intonation was so important to me that I found it hard to keep to the agreed features to mark.
  - e. One person is NOT consistent. Opinion does change during the marking period, and changes unpredictably. That is why it is so important to set aside quality time when high concentration can be maintained, and to finish the marking quickly.
6. "The greatest increase in reliability occurred in the change from one to two markers". (p65). ie when there are more than two markers.
7. Using a proper mark scheme also increased reliability. This is the easiest step to take and should never be neglected.

8. Generally, double marking is better than one marker. But, Weir argues (p66), not necessarily. It depends on the markers being equally consistent in their judgements.

Weir thinks that one good marker may be better than two poor markers.

For practical reasons, single marking is widely used in UK. BUT, it is always with a clear mark scheme, training, and double checking. And, in many subjects, the composition is only 20% of the marks. Other exam techniques with lower problems with consistency of marking are also used eg MCQ, short answer, and structured papers.

9. There is merit in assessing language while students are concentrating on content. In this way you see the real, natural language ability. For instance, a subject can be examined orally - for content, and the student can also be scored for language. This is easy enough to do in a jury of two teachers. One teacher scores for content and double-checks the language mark. The other teacher scores for language and double-checks the content. The student is concentrating in the oral on the content, therefore the language produced is largely the default non-conscious language.

#### **10. Sub-conclusions**

- a. An analytic mark scheme is a must. In some cases, double marking may not be necessary.
- b. As argued in the paper on double correction, it is better for one teacher to mark all, or one part, and one teacher mark the whole of another part, than to have division of scripts by <fair work allocation>.

## **G. STEPS TO CLEAR MARKING - CONCLUDING REMARKS**

1. The examination should be a clear fair test of what has been taught in class. It should neither be too easy, nor too hard. The types of question should measure the content or skills that were practised in class.
2. The wording of the question should be very clear. In particular, students should not have to guess what the examiner wants them to do.
3. Study various examination formats, and choose a format that matches what you want to test.
4. Try to set a paper that is easy to mark consistently.
5. Only ask as much as you need to obtain a fair impression of the student.
6. Agree in advance on the expected answers, and the mark distribution. Consider putting on the paper the distribution of marks, since this gives a useful guide for students as to how much time they should spend on each question.
7. The aim of an examination is to separate out the good students from the bad students. Therefore a good range of marks should be expected.

If a teacher marks from 9-11, then it is largely a matter of chance who succeeds and who fails. The examination is little more than a respectable gambling game.

8. Try to set a mixture of questions.