

THE IMPACT OF RE-SITTING ON THE FAIRNESS OF GRADING

Christopher Wheadon

One consideration in allowing re-sitting is that examinations are not perfectly reliable instruments. On any occasion a candidate may perform slightly better or worse depending on a number of chance factors related to their preparation and the test content. From this perspective it may seem fair to allow candidates to re-sit to compensate for the unreliability inherent in measuring ability through examinations.

The extent to which candidates' performance is likely to differ on notional parallel tests and the impact this will have on their grades can be measured through measures of classification accuracy. If a test were perfectly reliable then all candidates would be classified into their 'true' grade. As the test becomes less reliable the grades they receive are likely to vary. They may perform worse than their true grade (a false negative) or better (a false positive).

It is possible to simulate the performance of candidates on parallel tests using measures of the ability of the candidates and the difficulties of the items (see Wheadon & Stockford, 2010 for details). To simulate the impact of re-sitting, the following steps were taken.

Measures of the ability of the candidates and the difficulties of the items were derived from a typical AS unit. These were then used to simulate performance of 1000 candidates on the unit. Candidates who achieved lower than their true grade were then allowed to re-sit up to three times until they had achieved their true grade or higher. They would only re-sit if they hadn't achieved their true grade or higher.

Table 1 illustrates the results. As candidates re-take they are less likely to receive a false negative result. The false negative rate falls quite dramatically after a single re-take from 15 per cent to 5 per cent, with no candidates achieving two grades lower than their true grade. This would, in itself, increase classification accuracy. As a number of candidates re-sitting will actually outperform their true grade, however, the false positive rate rises. After a single re-sit it rises from 20 per cent to 35 per cent. The net outcome is that grading accuracy falls after a single re-sit from 65 per cent to 59 per cent.

Attempt	Difference from true grade						
	-3	-2	-1	0	1	2	3
1	0%	2%	13%	65%	17%	3%	0%
2	0%	0%	5%	59%	30%	5%	0%
3	0%	0%	2%	52%	39%	7%	0%
4	0%	0%	1%	44%	46%	8%	0%

Table 1: The difference between true grade and achieved grade following one to four attempts at the test

A fall in accuracy may seem to suggest that re-sitting reduces the quality of the decision making; however, the relative importance placed on false negative results and false positive results should also be considered. Is it educationally more harmful that 15 per cent of candidates worthy of a grade did not receive that grade when they first sat the examination? Or is it worse that an additional 15 per cent of candidates overachieve following a re-take? It is also worth considering the impact on different grades.

Figure 1 shows the impact of re-taking on grading accuracy by true grade. Grade A candidates are virtually unaffected by re-sitting. The real impact is on the lower grades as false negative results disappear after subsequent re-takes.

This model considers only one source of unreliability inherent in the tests, the sampling of the curriculum. If other sources of unreliability are added, such as marking reliability or variations induced by factors such as examination stress then the impact would be even more pronounced. Of course this simulation only considers one aspect relevant to re-taking. There are other concerns such as the impact of re-takes on motivation that are beyond the scope of this simulation. The model also assumes that candidates know their true grade and will only re-take if they fail to achieve this grade.

Overall it would seem safe to conclude, however, that allowing a single re-take has a substantial impact in reducing the false negative error rate for candidates outside the grade A range while having little perceptible impact on candidates in the grade A range. Removal of re-takes would seem to perversely deny less able candidates of the right to achieve their true grade. That said, allowing more than one re-take opportunity adds little in terms of reducing the rate of false negatives, increases the cost and burden of assessment and is of less obvious educational value.

Christopher Wheadon
22 September 2010

Figure 1: Difference from true grade by re-take attempt and true grade

