# Should the best mark count when resitting at A-level?


**Summary**

- According to modern test theory the 'true' ability of a candidate is the theoretical average of all of their scores over an infinite number of attempts at equivalent tests. The accuracy of a test can then be measured by how closely it orders a number of candidates compared to their true scores.
- Here we use the analogy of the high jump event in the Olympic Games – where athletes are allowed more than one attempt at a jump – to illustrate the impact of 'retakes' on the final outcome.
- The results from the high jump example are mirrored in A-levels:
  - The very best candidates still come out with the best grades whether resits are allowed or not.
  - Average performance is lower if no resits are allowed.
  - Allowing resits means candidates with lower 'true' ability could end up outperforming better candidates.
- If candidates are allowed to keep their best mark after resitting, the accuracy with which all candidates are classified into grades falls, because some resitters may perform particularly well on a specific exam, thus outperforming their 'true' grade.
- If the candidates are given their final mark after resitting rather than being allowed to keep the better of the two, then retakes have no impact on the accuracy of grading. However, the impact of such a rule change on students' and teachers' decision marking needs careful consideration.


**Second chances and the legitimacy of final results: Was Stefan Holm robbed of the silver medal in the men's high jump event at the 2008 Beijing Olympics?**

In the men's high jump event of the 2008 Beijing Olympics, Andrey Silnov sailed over the bar at every height first time until it was set at a world record beating height of 2.42 metres, at which height he failed three times. By that time he was already assured of the gold medal as no-one else had cleared the previous height of 2.36 metres. No-one would doubt that he was the worthy winner of the Olympic gold medal. There may be more disagreement, however, over whether the silver medal was truly deserved.

Germaine Mason had failed his first jump at 2.29 metres, a jump which three athletes who were placed after him, had flown over at their first attempt. These athletes included the Olympic gold medal winner from 2004, Stefan Holm, whose average best jump in major competitions that year had been six centimetres higher than Mason. According to Olympic rules, the rank order of the athletes is determined by their highest jump, and they are allowed up to three consecutive failures in the course of the competition. Mason, who went on to clear 2.34 metres, was therefore placed second despite failing in his first attempt at 2.29 metres (Table 1).

**Table 1: Men's High Jump results from the 2008 Beijing Olympics**

| | | 2.15 | 2.2 | 2.25 | 2.29 | 2.32 | 2.34 | 2.36 | 2.42 |
|---|---|---|---|---|---|---|---|---|---|
| Andrey Silnov | Russia | – | o | o | o | o | o | o | xxx |
| Germaine Mason | Great Britain | – | o | o | x– | o | o | xxx | |
| Yaroslav Rybakov | Russia | – | o | o | o | xxo | o | xxx | |
| Stefan Holm | Sweden | – | o | o | o | o | x– | xx | |
| Jaroslav Bába | Czech Republic | o | o | o | o | x– | x– | x | |
| Tomáš Janků | Czech Republic | o | o | o | xo | x– | xx | | |
| Tom Parsons | Great Britain | o | o | o | xxx | | | | |
| Martyn Bernard | Great Britain | o | o | xo | x– | xx | | | |
| Jessé de Lima | Brazil | o | o | xxx | | | | | |
| Filippo Campioli | Italy | – | o | xxx | | | | | |

Is it fair to allow high jumpers a second chance when they fail at a certain height? Understanding whether it is or not can lend some insight into whether it is fair or not to allow candidates to resit A-levels.

According to modern test theory the true score (ability) of a candidate is the theoretical average of all of their scores over an infinite number of attempts at equivalent tests. The accuracy of a test can then be measured by the extent to which it rank orders candidates according to this true score. We do not have multiple equivalent tests for exam candidates, but for high jumpers who compete throughout the year we have a number of scores (jumps in other competitions) that can be averaged to approximate the athletes' true ability. As we can estimate the athletes' true ability from previous competitions we can therefore test whether allowing high jumpers three fails more accurately reproduces their rank order according to their true ability than if they were permitted no fails.

Table 2 shows the rank order of the high jumpers according to their true ability, the rank order according to the current Olympic rules and the rank order if they had not been allowed any fails. In this case, their true ability has been calculated as the average of their competitive jumps that year.

Neither set of rules reproduces the true rank of the athletes, suggesting that neither set of rules is perfect: Jessé de Lima, for example, clearly had an off day. To determine which set of rules is better we can use correlations to test which set of rules best reproduces the true rank order. A perfectly reliable test would produce a correlation of 1, a test that pronounced the worst jumper the best would have a correlation closer to -1, and a test that ordered them randomly would have a correlation close to zero. If no re-jumps are allowed the correlation is 0.44 while allowing three re-jumps produces a correlation of 0.34. If this result were replicated across events, and if the sole purpose of the Olympics was to crown the best high-jumper in the world, then there would be a case for removing re-jumps.

**Table 2: Ranks of high jumpers in the 2008 Beijing Olympics according to different rules**

| | True Score | True Rank | Olympic Best Jump | Rank | No Fails | No Fails Rank |
|---|---|---|---|---|---|---|
| Andrey Silnov | 2.33 | 1 | 2.36 | 1 | 2.36 | 1 |
| Stefan Holm | 2.32 | 2 | 2.32 | 4 | 2.32 | 2 |
| Jessé de Lima | 2.31 | 3 | 2.20 | 9 | 2.20 | 8 |
| Yaroslav Rybakov | 2.30 | 4 | 2.34 | 3 | 2.29 | 3 |
| Martyn Bernard | 2.28 | 5 | 2.25 | 8 | 2.20 | 8 |
| Germaine Mason | 2.26 | 6 | 2.34 | 2 | 2.25 | 5 |
| Tom Parsons | 2.24 | 7 | 2.25 | 7 | 2.25 | 5 |
| Filippo Campioli | 2.24 | 7 | 2.20 | 10 | 2.20 | 8 |
| Tomáš Janků | 2.23 | 9 | 2.29 | 6 | 2.25 | 5 |
| Jaroslav Bába | 2.23 | 9 | 2.29 | 5 | 2.29 | 3 |

**What insight can high jumping give us into whether or not candidates should be allowed to resit A-levels?**

Firstly, Silnov would have won whatever the rules. This is also true for A-levels (see *What is the impact of resitting at A-level* in the CERP A-level reform series). Secondly, the average performance would have been lower had no re-jumps been allowed. This is also true of A-levels: resitting does increase the overall level of performance. Thirdly, by allowing retakes, better jumpers can be outperformed by worse jumpers: according to true score theory Stefan Holm should have been awarded the silver medal rather than Germaine Mason. The argument would seem to suggest, therefore, that retake opportunities should be removed.

Before we create a class of wandering samurai, the notorious *ronin* of Japan, shamed by their failure to pass their examinations first time (Ono, 2007), we may wish to consider whether a change in the retake rules would be preferable. A theoretical simulation of A-level results (Wheadon, 2010) shows that if candidates are allowed to keep the best mark when they retake, the accuracy with which all candidates are classified into grades falls. This is because the retake candidates may outperform their true grade. If the retake candidates are given their final mark rather than being allowed to keep the better of the two, then retakes have no impact on the accuracy of grading. However, the impact of such a change to resitting policy on students' and teachers' decision marking needs careful consideration.

**Bibliography**

Ono, H. (2007). Does examination hell pay off ? A cost–benefit analysis of 'ronin' and college education in Japan. *Economics of Education Review*, *26*(3), 271–284. doi:10.1016/j.econedurev.2006.01.002

Wheadon, C. (2010). *The impact of re-sitting on the fairness of grading*. Manchester: AQA Centre for Education Research and Policy.