

Title:	Agreement between outcomes from different double marking models
Authors & affiliations:	Carmen L. Vidal Rodeiro Research Division, Cambridge Assessment
Abstract: (Your abstract must use Normal style and <u>must</u> fit in this space)	<p><i>Introduction</i></p> <p>In the context of marking examinations, double marking is a means to enhance reliability. However, deciding if it is worthwhile incorporates a dilemma. Intuitively, it increases the reliability of the assessment and shows fairness in marking, but it needs to be proven a benefit in order to justify the additional time and effort that it takes.</p> <p>One factor which affects the re-marking is whether or not the second marker is aware of the marks awarded by the first marker. Higher agreement is observed between two examiners when the second knows how, and perhaps why, the first marked an exam. This may suggest that the second examiner took advantage of the annotations available when trying to judge the best mark for a candidate. An alternative perspective may suggest that the second examiner was influenced by the first examiner's marks.</p> <p>The purpose of this research is to evaluate the extent to which examiners agree when using different double marking models, in particular, blind and annotated double marking. The impact of examiner experience is also investigated.</p> <p><i>Data and methods</i></p> <p>Two GCSE units, English and Classical Greek, were selected (200 scripts per unit, from the 2004 examination). For each subject, a principal examiner, two senior examiners and two assistant examiners participated in the study.</p> <p>Each assistant examiner was allocated 100 scripts with all annotations removed. Each senior examiner was allocated 200 scripts (100 annotated scripts and 100 "cleaned" scripts). Examiners were asked to mark the scripts following the marking instructions used in the "live" marking of the examination. When there were significant differences between marks and re-marks, reconciliation was needed. This was done by the principal examiner who, based on the two marks, produced an adjudication mark.</p> <p>There is little consensus about which statistical methods are best to analyse markers' agreement. In this research, correlation coefficients, proportion of agreement and Kappa statistics are reported.</p> <p><i>Results</i></p> <p>Examiners marked, on average, 5 or 6 scripts per hour. This did not seem to vary whether the scripts were annotated or not. Some examiners originally thought that marking annotated scripts would be swifter but this proved not to be the case. There was no difference between the time employed by assistant and senior examiners in the marking.</p> <p>There was more disagreement between English than between Classical Greek marks. This was probably related to the nature of the questions in English, which allowed for greater freedom in the response and required more subjective judgement by examiners.</p> <p>Removing marks and comments from scripts made a difference. Re-marking agreed more with original marks when these were present on scripts.</p> <p>In this study, reconciliation was carried out when the difference between two marks exceeded 10% of the marking range. For Classical Greek only 16 scripts needed reconciliation (11 scripts were blind re-marked by assistant examiners and 5 by senior examiners). Non-blind re-marked scripts did not need reconciliation. For English, the number of scripts needing reconciliation was 30% higher than in Classical Greek. Among them, 23 scripts were non-blind re-marked.</p> <p>One serious impediment to double marking is the increase in administrative time and costs which it entails. A trade-off has to be made between reliability and cost-effectiveness. Feasibility is a further issue due to the shortage of specialist markers in the UK.</p>