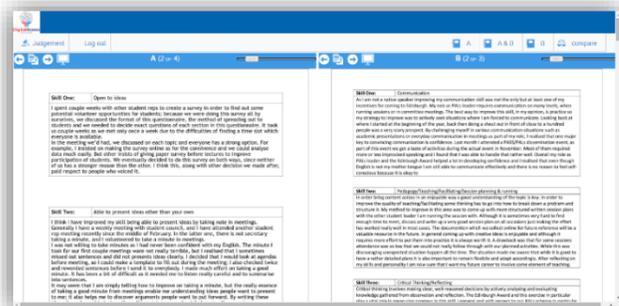


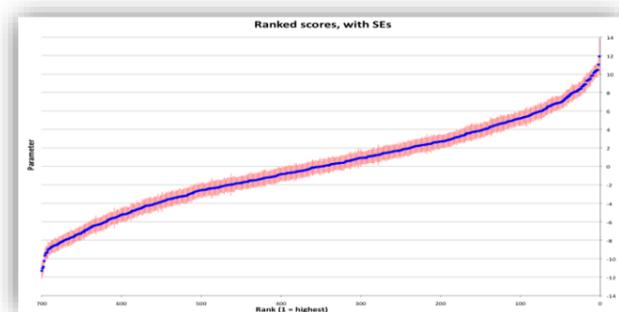
Trialling Adaptive Comparative Judgement in English A Literature

Background

A number of assessment components can prove challenging to mark reliably. Adaptive Comparative Judgement (ACJ) offers a radical alternative by presenting examiners (renamed judges) with pairs of candidate work and simply asking them to choose which is 'better'. The results of these pairwise comparisons are then used to create a scaled rank order of scripts from best to worst which can then be converted to more conventional 'marks' and grades. In spite of the clear promise of ACJ, concerns remain surrounding the feasibility of such a process and the time it would take to create a stable rank order. This study seeks to investigate this and considers other issues that the IB would need to overcome in order to bring ACJ into a live examination session.



Importance Statement (excerpt)
 "... candidates will engage in depth with a wide range of texts to discern literal meaning and broader implications based on clear evidence from the given texts. Successful candidates will develop ideas logically, using fluent language. Authentic, thoughtful engagement with the texts and/or question at hand is the hallmark of a successful response."



Pollitt (2015)

Trial design

Software provided by Digital Assess

May 2014 English A Literature P1 (timezone 1)

700 candidate responses

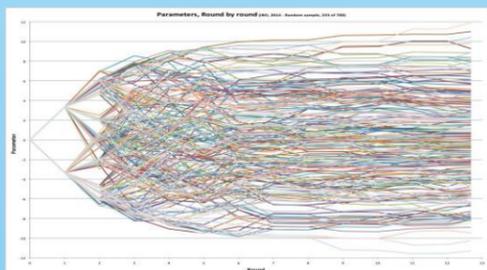
29 examiners (judges)

4442 pairwise judgements (12.69 'rounds')

Initial analysis provided by Alastair Pollitt

Results

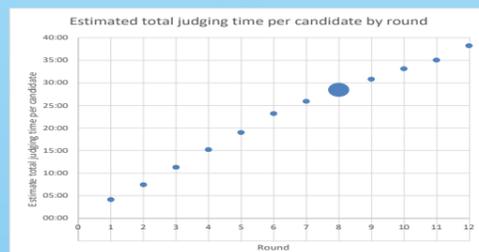
How many judgements are needed?



The parameter values had mostly 'settled down' by round 8, suggesting that an average of 8 judgements per script was necessary to create a stable rank order.

Pollitt (2015)

How long does it take to produce stable rank order?



Estimated total judging time for 8 rounds is 28.5min per candidate.
 Estimated average marking time is 15-17 minutes per candidate.

Further findings
 The quicker examiners were no less reliable than slower examiners (Pollitt, 2015).
 Some examiners tended to over-select the second script they looked at.
 Correlation between the ACJ results and the marks awarded during the exam session was low ($r=0.53$).

Examiner feedback

Feedback Question	Yes	No	Unsure	N/R
Did you feel that ACJ allowed you to suitably reward what is important in English A Literature?	16	1	4	3
Would you recommend ACJ as the future method for grading candidates' English A Literature scripts?	14	4	6	0

Examiners' comments

"I think ACJ is an excellent method of grading because it allows for the grader to rely on his/her professional knowledge and experience, instead of following specific rubric guidelines. This freedom will allow students to express themselves in effective ways, rather than in very limited ways defined by a rubric."

"I would recommend it. I think it is very fair as it would be difficult I think to mistake the best essay and of course markers have professional experience which shows them what a successful response to text should look like. I found it easier and more satisfying than other marking I have done, which has been over many years, across different exam boards and levels."

"I prefer the method of marking currently used rather than ACJ. To me it seems more fitting to judge each script on its individual merits rather than in comparison to another script. While marking is always subjective, I feel that marking individual scripts according to a mark scheme does allow for more objective grading."

Discussion

Is ACJ feasible for the IB?

Whilst ACJ took longer than marking for this trial, ACJ could still be considered feasible for the following reasons:

- Qualification could take place before the exam and be based on the previous year's exam
- Examiner retention should be much better
- Examiners could be trained to make quicker judgements without compromising reliability

Issues to consider

- Why were the results so different from those produced through live marking?
- Just how reliable are the ACJ results produced from the study?
- How would Enquiry upon Results work with ACJ?
- What would the feedback mechanism to schools be, if any?
- The results of ACJ are much more consensual than marking and based on the entire community of examiners involved in the process. Should the IB support such a philosophical shift?

Reference

Pollitt (2015), Analyses of the Adaptive comparative Judgement scoring of IBO English A Literature scripts, Report for the IB by Camexam.