

What is Wrong with the PPE Admissions Test?

I would like to suggest that the PPE Admissions Test, in its current form, is “unfit for purpose” and ought to be discarded as soon as possible in favour of some alternative – for example either the *LNAT* used for admissions in Law, or the Cambridge *Thinking Skills Assessment*. Its faults are many:

1. Lack of clear criteria for marking

Question 1, the “Politics” question, which usually involves comparison of two passages or extraction of an argument from a passage, lacks clear criteria that would enable different markers to agree. This problem has at least two roots:

(a) Variety of markers

Where the test is marked within colleges, usually involving a number of markers in each, this could involve well over 100 distinct individuals, each marking in their own way. More scholarly markers are likely, perhaps, to put greatest emphasis on precision of content; others may be more concerned about clarity of expression (and these two criteria can easily conflict, because clarity is much easier to achieve if one ignores subtle points in the text).

Last year, on re-marking all of Hertford’s “borderline” cases, including seven from other colleges, I found that my own marks differed by as much as 24/100 from those previously allocated! One of the principal factors was that others had given high marks for representations of the Question 1 vegetarian argument that seemed clear and elegant, yet involved steps that simply were not there in the given text. Some markers, I concluded, were basing their judgement on an overall “gestalt” rather than precise analysis.

(b) Lack of clear requirements

The passage(s) set can be tricky to interpret, such that analysing them properly could easily absorb a couple of hours. In asking students to summarise or compare them within an unrealistically short time, it is very unclear exactly what is being asked for, or what criteria should be applied in assessment.

This year, in consultation with Suke Wolton and John Hyman, I worked out a detailed marking scheme for Question 1, listing explicitly the various themes and contrasts that could be found in the Aristotle and Russell texts to be compared. Even using this common scheme, however, a small-scale comparison between different markers of the same scripts found an average difference between them of around 7 marks out of 32! It seems that the question is just too difficult, so markers end up looking for things that “deserve a mark”, and they have different levels of toleration for things that are just hinted at or vaguely expressed.

2. Linguistic/cultural bias

Question 2, the “Philosophy” question, which involves definition or illustration of subtly different word-meanings, is strongly biased towards those whose native language is English, and who have had the good fortune to be brought up in a highly literate environment (either at home or at school). Such a bias is particularly inappropriate in any test that might be used as a basis for desummoning, because in that case no interview may be available to correct the biased impression.

This year Hertford is accepting two Europeans, one German and one Swede, whose marks on the test were significantly inferior to the analytical ability and understanding evinced in their interviews. Had we been desummoning rigorously using the test, it is entirely possible that both would have been desummoned.

3. Inappropriate task

Speaking as a philosopher, I am doubtful about the appropriateness of Question 2 independently of any problems with bias. Thinking quickly of word definitions or suitable illustrations of them can be extremely hard, and being unable to do this in the very short time available does not necessarily imply any incapacity in understanding. Another problem is that sometimes the words given can be virtually synonymous (as in last year’s example of “equitable”, “fair”, and “just”), the differences reflecting slight variations in usage rather than anything that someone could be expected to learn from a dictionary. In other cases the meanings can be debatable, with some dictionary definitions being prescriptive while others are merely descriptive (e.g. this year’s example of “infer”, one of whose

contemporary acceptable uses according to some dictionaries is equivalent to “imply”). Quite apart from this, many words have a range of rather different meanings, so that someone might well have more difficulty in answering this question owing to *greater* knowledge of the language.

In my view, the tasks set in any very short test should be much more clearly circumscribed, testing the ability to identify specific flaws in arguments, or presuppositions, or logical relations between propositions etc. rather than something as open-ended as giving examples to illustrate word similarities and contrasts.

4. Economic knowledge bias

Question 3, the “Economics” question, is the only one of the three that is very explicitly directed towards its corresponding discipline. Here I must defer to Economists, but will nevertheless express a serious doubt about this question’s appropriateness. In particular, the problem that is set seems often to be one that presupposes some understanding of the Economics approach, or even specific knowledge (e.g. of decision trees or the notion of “payoffs”), which however is itself very straightforward rather than theoretically deep. So on the one hand, students who lack relevant experience are at a serious disadvantage, often enough to wreck their prospects of getting anything at all from the question. But on the other hand, those who do have relevant know-how can potentially pick up marks extremely easily, without this indicating very much about their intellectual ability. Given the very different background of the students concerned, many of whom have A level Economics while many do not, I would suggest that it is more appropriate to have a general numerical problem-solving question (or questions) rather than one skewed towards specific understanding of Economics or its methods.

In last year’s test, the Question 3 marking scheme gave 10 marks in several cases to parts that were almost trivial for anyone who understood the context; in this year’s test, 29 of the 376 centrally marked students scored 0 on Question 3, 75 of them scored less than 10%, and 182 scored less than 25%.

5. Unrealistic time constraints

The problems above are compounded by the very limited time allowed for the test, and because three quite distinct tasks are being fitted within it. First, it is very unclear that speed of performance on these tasks is an appropriate measure of ability (as an historian of Philosophy, for example, I would see accuracy as far more important than speed in Question 1, but the two conflict; cf. also my comments in §3 above). Secondly, the very fact that the candidates have to make strategic choices of time allocation between the question confounds even our ability properly to compare them against each other on those tasks for which this might be appropriate.

One student, faced with difficulties in understanding Question 3, say, might (as advised) spend 20 minutes on it, without gaining any marks. Another, more strategic, might ignore Question 3 entirely and thus have 50% more time available for Questions 1 and 2. It then becomes dubious to compare these students’ performance on any of the questions. In respect of those students who are better at Economics a different problem of time viring and consequent incomparability emerges: last year, for example, as many marks were given for a trivial piece of arithmetic in Question 3 as for defining five words in Question 2!

6. Lack of systematic monitoring or learning from experience

As far as I am aware, no systematic monitoring of the PPE test has been carried out, either in terms of the consistency of its questions from year to year, their fairness in the context, their comparability, the adequacy and appropriateness of the marking schemes provided, variations between markers, etc. This is in stark contrast with tests such as the LNAT and the Cambridge *Thinking Skills Assessment*, and might be thought to reflect very badly on the professionalism of our procedures. Even worse, as far as I know there is no arrangement for “learning” from year to year: the onerous burden of designing the test moves from college to college, with entirely new individuals being required to do it purely on the basis of previous tests. This may be fair as far as the academics are concerned, but it positively hinders any learning from experience, and could almost be designed to generate inconsistencies over time. To be frank, I do not believe that our current practice is defensible by the standards now widely expected.