

## Thoughts on exams

Professor Mary Beard

I want to say, to start with, that it is very hard to talk objectively about modes of assessment, especially about *exams* (in the traditional sense). For better or worse, exams are, of course, as much a cultural myth as a mode of assessment. They are what many of us still dream about in those classic repetitive dreams (wrong paper, wrong language, pen that wont write), decades after we last sat one. Beyond that, they are a mode of assessment that is, by and large, queried, discussed, worried about and debated by those like us who years ago did well in them. According to the iron logic of academic succession, it is one of the dubious privileges of our early achievement that we get to devise the forms of assessment to be imposed on the next generation. And they are embedded in the communality of a university like this. I'm sure many of you will understand my point when I say that – within an institution that has become ever more digital, ever less face to face, and ever more centralised – the processes of formal assessment have become almost the only times I actually *do something actively* with my colleagues (rather than just *talk* or *tick boxes* with them). Exams have also become something very convenient to blame for any of our problems of pedagogy. I'm thinking here for example of gender. For some people at least (and I know there are more sophisticated analyses around), it's been much easier to put down gender disparity in Tripos results to “female unfriendly” modes of assessment, rather than to wider issues of culture and teaching in what can feel still a very blokeish university. It's rather like trying to blame Russell group selection procedures for a lack of educational achievement among underprivileged kids, which starts before they even go to school.

Even for the most austere professional analysts of assessment theory and practice, this cultural “embeddedness” of the examination system tends to muddy the waters. For me, much more an amateur, it certainly does. And those of you here with a lot more technical knowledge and expertise in this area than I have, will have to forgive me, I'm afraid, for doing what I was asked to do – that is to sound off a bit on the basis of my own experience, mythical and otherwise (it'll involve a bit of stating the bleeding obvious too, but sometime the bleeding obvious could do with a bit *more* stating). It's true that I have dug around a little in the history of exams (and I will have something to say about that in a few minutes), and I've recently put my toe in the water of debates around school exams. But the thoughts I'm going to share mostly come from a lifetime of participant observation of assessment in the Classical Tripos here (with the occasional ventures into History and MML, into external examining elsewhere in the UK and tiny bit of experience in the US and in my first job at Kings College London). For those who don't know Classics well, the Faculty has what I would reckon was a fairly typical humanities repertoire of assessment: a preponderance of so-called traditional written, though incorporating various forms of alternative assessment (such as a portfolio of essays) which have *increased* since 1984 when I took up my lectureship, though weren't *invented* then. When I was an undergraduate in the 1970s there was already a dissertation option for some students. And just

to get the scale in view, we're operating with a bit under 100 candidates in each year between Part IA and Part II.

Now one part of me feels quite up-beat about the assessment practice I've seen and been part of. I think that we do a pretty good job *according to the template of assessment with which we are working* (and that is of course is a very big "according to"). I also think that our "traditional" examination system is testing skills that I value -- and that society, employers, you name it, also value. That goes from the ability to marshal and deploy arguments cogently, or the ability simply to remember stuff, to stamina and resilience. And, notwithstanding the inevitable bits of human fallibility, I think my colleagues operate with a degree of equity and diligence, sometimes I might say *over diligence*, that we can be proud of. I even think that some of the aspects of our system that might seem at first sight least "fair" (a term I shall return to) are no less fair than the things that are proposed to replace them -- but which in fact merely redistribute and reconfigure the inequity (I am thinking here about something like allowing resits. Sure it looks superficially fairer to allow those who, say, fail to have another shot...but it raises all kinds of tricky questions of equity about who exactly should be allowed the privilege of another shot). In general, I like to apply the "eye-ball-to-eye-ball test". Could you look Jenny in the eye and explain to her why she got a 2.1 when Susie who had only one more mark in her overall total got a First, and could you actually *believe* what you were saying to her? There have been very few occasions in my career on which that has not been possible. There have been some, but very few.

Which is to say that I don't think, in my neck of the woods at least, that our regimes of assessment are in terminal crisis. But, of course, another part of me, and I guess another part of us all, seethes with dissatisfaction at what we do and what we claim to do. That goes from tiny discontents about the detail (I always get cross when I see the rubrics on the front of our papers . . . "Candidate are advised of the importance of writing legibly" or something to that effect -- which has always seemed at best a meaningless cliché), to bigger questions of whether we are assessing what we teach, and which of the intellectual qualities that we claim to value are actually tested in the assessment process. At this point I have to underline the fact that exam systems are always discontented with themselves and always have been. There was never a golden age in Cambridge when everyone thought the Tripos absolutely fit for purpose. In the later nineteenth century in the Classics Faculty (when, let me say, there was already a proto- and hugely unsuccessful and short-lived dissertation option) there was intense discussion not just about student performance (low performance, that is), but also about the principles and practice of the assessment process itself. If the students were doing badly, maybe it was because there was something to matter with how they were tested.

Some of this concern overlaps with our own: they tut-tut-ed like we do about mindless "cramming" and "tip-fancying" (that was the 19<sup>th</sup> century word for question-spotting). Some of it seems rather less familiar. One major question 150 years ago was whether the system should be trying test "complete

knowledge” of the ancient world (which ended them up with a series of exam question to be answered in 15 minutes which encouraged a hell of a lot of tip-fancying and looked as mad to some of them as they do now to us (“Name the chief authors who flourished in Rome during the reigns of the four immediate successors of the emperor Augustus and examine briefly the influence of the age upon their writings.” *Briefly* was the word, if you had to do it in 15 minutes.) Or should they be attempting to judge different forms of capacity to handle material and show the kind of historical sophistication that was best suited to an exam essay format in 45 minutes or an hour. Their problem, in this case, was how could you accurately *mark* an essay? And they worried intensely about by what criteria it was possible to distinguish the clever rhetorician -- “the showy and superficial thinker” -- from the good historian; or how you could judge a piece of considered argument produced under time constraints that made considered argument impossible! (Which is not a bad point.)

Insofar as this is still one of our problems (and it might be a good idea if we thought about it more), we have solved it with through spurious transparency and in published marking criteria like my own Faculty’s handbook (“a II.1 essay is characterised by wide reading, interpreted intelligently with some independent thought” . . . it took us ages to hammer that one out). That certainly wouldn’t have convinced the Victorian gents and it’s spurious because we don’t actually mark like that. Maybe I am unusually wicked in this respect, but I mark the papers in my head and then I write the appropriate phrases from the criteria in my mark book (I *don’t* really think I’m unusual – and it’s about as far from true transparency as you could get).

But the further point I want to underline is it’s not just that there have *always* been those doubts and debates, but they are *inevitable*. They are inevitable for the simple and very obvious reason that we *cannot* assess all the skills and qualities we value (we always bound to be debating what *selection* of those skills we will choose to assess); nor do any of us really imagine that can we sum up the peculiar and idiosyncratic characteristics of any human brain or person by a single number or class (“are we making the papers or the man” our 19<sup>th</sup> century predecessors pointedly asked). And what that means is that – although as I have proudly said I think that many *elements* of the process are judged fairly – the very selectivity means that the system as a whole cannot ever be “fair” in any normal sense of that word. It is always playing to someone’s advantage, and the more we are straight about that the better.

So where does that leave us? Well I’m glad to hear that there is a root and branch review of assessment, because I think that we need to nerve ourselves to face much more fundamental questions – about what we think any of our current forms of assessment are *for* and what exactly we think we are testing, and at what cost. I know that I have half confessed to rather enjoying the whole examining process. But I am also well aware that anyone on the outside who saw me and my colleagues devoting literally thousands of person hours to giving 90% of our students an undivided Second in Part IA... traditional or non traditional modes of assessment . . . . would think us barking mad. The effort reward ratio makes no sense. What are we doing it *for*? This

is not a question of *how* we test, but whether we should test *at all* at least in some parts of a student's career. Just imagine a world without Part IA; it looks a nice world to me.

I think too that we have to be careful with superficially attractive buzzwords, like "fairness" and "consistency". I've already said that most attempts to introduce greater fairness are really a way of shuffling the deck-chairs of inequity. And to be honest I think that consistency across the board in the examining process is grossly overrated: does it actually matter if different parts of the university do things differently (if some Faculties penalise disobeying the word limit differently from others, for example)? But the practical issue and danger here is that we spend so much time, and commit so many person hours to plugging the series of holes in fairness and trying to right a series of wrongs, that we simply overload the system... (it's my experience that no reform in this university ever involves less work than before)... and it just implodes.

But that of course may be the only really effective prompt for change. It may force our hands. And here again we might reflect on what happened back in the 1800s, in the early part of that century, when the last really big innovation in assessment practice was introduced here: that is the change from oral to written examination. Why did that happen? It was partly because it was proving increasingly hard to test algebra and geometry without a pencil and paper. It was partly because, with increasing student numbers, the poor examiners were spending months on end simply conducting orals. It was really system overload that drove the change. Maybe it will work for us too,