Fair exams in a mass university: can technology get us out of trouble?
2021 Ethical Forum of the University Foundation
Brussels, 9 November 2021

Concluding remarks by
Philippe Van Parijs
(UCLouvain & KU Leuven)

Can technology help us address the challenge of designing fair exams in our mass universities? This is the subject we announced in the title of this Ethical Forum. But the subject we treated was wider. Not only the designing of fair exams occupied us, but more broadly the designing of effective evaluation, whether or not linked to exams.

Feedback versus certification

The evaluation of the competence of our students has two functions. When it takes the form of formal examinations, it serves the function of *certification*, i.e. it helps us decide which of our students have achieved the level of competence that justifies awarding them a degree. But evaluation, especially but not only when it takes the form of continuous assessment, also serves the function of providing students with *feedback*, i.e. with information about what they got right and what they got wrong, and about how they are doing overall, relative to each other and relative to expectations. Both the certification function and the feedback function contribute to the learning process, in one case by providing incentives, in the other by providing information.

Our question, then, was whether and how technology could help us achieve a more effective evaluation in this broad sense, given the irreversible massification of our universities. Unavoidably reference was repeatedly made to tools we discovered and experimented with during the pandemic, especially for the sake of remote evaluation. However, remote evaluation can use rudimentary technology, such as a phone, while on-site evaluation can be technologically sophisticated. Remoteness and sophistication can come together, for example in the case of the controversial proctoring methods touched upon repeatedly in our exchanges, but they don't always do so.

So, can technology help our evaluation methods cope with a number of students that makes the use of traditional methods prohibitively costly? As regards continuous assessment, two main routes were discussed. One is learning analytics, i.e. the use of traces left by the students in order to improve their learning: algorithms applied to big data hold the promise of providing effective feedback on a large scale, at least on condition that a teacher supervises the whole process. The other route is peer assessment, which is of course possible without sophisticated technology but whose use can be made more effective and less labour-intensive thanks to appropriate digital tools.

Multiple Choice and its alternatives

What about exams, the part of evaluation that is geared towards certification? The risk that the use of technology could increase the unfairness of exams was emphasized, especially from the students' side: most obviously, substandard equipment, precarious network connections, uncomfortable home environments may make it harder for students to take remote exams. On the other hand, technology also enables us to reduce some inequalities suffered by students with disabilities: most obviously, computers can be adapted for people with very bad sight.

Unsurprisingly, however, much of the discussion focused on multiple-choice questionnaires (MCQs) and alternatives to them, as they provide the most common and most straightforward way of automatizing the marking. I confess that I remain prone to some nostalgia thinking back to the time when I examined orally a couple of hundreds of first-year students, eye to eye, each in turn as an individual person. This was a sort of Socratic luxury that included the occasional satisfaction of managing to get them to understand at long last — and a bit late — what I had tried but failed to explain to them in the lecture room. I realize, however, that one should not idealize this traditional form of exam, which avoided the cold anonymity of MCQs but was more vulnerable to subjective biases and above all was, in comparison, extremely time-consuming.

Thinking about how to correct the defects of MCQs is therefore more fruitful than dreaming about an unretrievable past. For example, Tinne De Laet suggested replacing the standard yes-or-no MCQs by "elimination marking" that rewards partial knowledge, discourages guessing and limits the bias of risk-aversion; Jean-François Rees experimented with replacing the ticking of options by the filling in of maps; and Yves De Ville advocated using the Graderscope, a software that consists in scanning answers to open questions and in processing them so as to facilitate the final marking by the human examiner. In light of my experience of endless struggling with the handwriting of my students, I am not sure how well this would perform without forcing students to type their answers. Nor do I see how it could deal with answers meant to show how well students can think on their own. However, each of these proposals is well worth considering seriously, and they just a few illustrations of the sort of tools that must be imagined and tried so as to enable us to assess our students in a less labour-intensive yet no less reliable way.

Fairness between students versus fairness to society

By way of closing remark, I want to stress one single point. The fairness of exams is not only a matter of minimizing biases and securing reliability in order to achieve fairness between our students, between those who have the privilege of getting access to higher education. It is also a matter of fairness to society, of making good use of the scarce material and human resources that are made available to higher education out of public funds. This can justify using technologies that enable us reduce our direct involvement in marking exams and thereby to keep more of our time and energy for the rest of our job: our teaching, our research, our other forms of service to society.

However, it is no less important that what is being tested through the exams should be the competence that is important for students to acquire and for society to see certified with the awarding of a degree. Technologically enhanced exams may be both fair between our

students and cheap in examiners' time, and yet be the opposite of what we owe to society. The capacities that we most need to develop in our students and to certify with our degrees may not be among those that are easiest to test in a way that is both fair and cheap.

Pressed by the challenge of massification and stimulated by technological innovation, we shall need to keep using our imagination and our critical faculty in order to design and implement as well as possible testing methods that are fair in this comprehensive sense. The illustrations discussed in the course of this Forum are a good start.