

The obsession with exam fairness

Mathieu Bouville*
University of Cambridge

It is widely agreed that exams must be fair; yet what this exactly means is not made clear. In particular, the meaning of ‘fair’ is seldom made explicit. Even though the concept of fair grading is rather straightforward, what it means for the exam itself to be fair is far from obvious. One may mean that the same rule must be applied to all students, but this merely propagates the fairness or unfairness of pre-existing rules. On the other hand, any attempt to account for the various talents of the students faces the problem of the diversity and of the incommensurability of these talents. In fact, it is not obvious that fairness should be our main concern at all. © Mathieu Bouville, 2008

Keywords: assessment; bias; education; examinations; grading; justice; tests

Article available at <http://www.mathieu.bouville.name/education-ethics/Bouville-exam-fairness.pdf>

The unbearable vagueness of ‘fairness’

There is a wide-spread consensus —amongst both teachers and students— that exams must be fair: “what students hate more than anything else are examinations that they perceive as unfair,”¹ “unfair and poorly graded exams cause student resentment,”² “individual perceptions of instructor fairness had strong effects on student satisfaction,”³ etc. If it is self-evident that exams must be fair, what this exactly means is not. An American may be pissed because he does not have a rubber (imagining the possible circumstances is left to the reader as an exercise) but no Briton will ever be pissed because he does not have a rubber.⁴ Plainly, it is not because one uses the same words that one means the same thing — if people do not agree on what ‘fair’ means then the consensus that ‘exams must be fair’ is illusory.⁵

In chapter II of *The Adventures of Huckleberry Finn*, Tom Sawyer sets up a band of robbers which would rob and kill people, “except some that you bring to the cave here, and keep them till they’re ransomed.” When asked what this means, Tom replies “I don’t know. But that’s what they do. I’ve seen it in books; and so of course that’s what we’ve got to do.” And someone with more logic points out “But how can we do it if we don’t know what it is?”. Follows a discussion of what ransoming may mean, giving rise to absurdities such as “We’ll keep them till they’re ransomed to death” and “Why can’t a body take a club and ransom them as soon as they get here?”. If we cannot say what it means for an exam to be fair, we are as likely to design fair exams as Tom Sawyer’s Gang is to ransom people (even if we have clubs).

It is rather common for words such as ‘unfair’ to have their meaning replaced with a vague negative connotation — ‘this exam was unfair’ standing for the less precise ‘there was something wrong with this exam.’ The meaning of the word can drift further: since getting course information in a user-friendly form is construed by students as a matter of fairness,⁶ the word ‘fair’ may have the vague meaning of ‘stuff students like.’ Scholars try to be more specific in their use of the word ‘fairness.’ Here is a typical example:

Fairness encompasses a broad range of interconnected issues, including absence of bias in the assessment tasks, equitable treatment of all examinees in the assessment process, opportunity to learn the material being assessed, and comparable validity (if the test scores underestimate or overestimate the competencies of members of particular groups, the assessment is considered unfair).⁷

The meaning of ‘fairness’ is still imprecise: (i) it is a non-exhaustive list; (ii) it relies on concepts, such as ‘bias’ and ‘equitable’ that are not defined (for instance, it is unclear whether “absence of bias” and “equitable treatment” are different); (iii) it otherwise assumes knowledge that may not exist: in order to know whether one “overestimate[s] the competencies” one needs to know independently what these competencies are, which may require another test (which would have to meet the same criterion of fairness); and (iv) it does not specify what these “particular groups” are or if there is a rule to determine what they are. This kind of definition does not really help us to write fair exams.⁸

Applying the same rule to all students

A common construal of fairness is the application of the same rule to all students; for instance, the same answer should get the same points. In particular, one may say that an exam is fair if it is based on explicit class objectives — say what you do and do what you say. The ability to solve simple problems quickly is then an acceptable criterion if it was an explicit objective of the class. But what if some students complain

that such an objective is unfair because it favors students who are naturally fast or because students whose strength is their ability to cope with complex issues are at a disadvantage? One will reply that this ability is not part of the objectives, and the student will ask why. Saying that we tested on the objectives, that objectives were clearly defined, and that we taught students what would be on the test will not convince them. The tests are based on the objectives; but are objectives fair?

Let us imagine that we are trying to organize a match-up between American swimmer Michael Phelps (best swimmer in the world) and Jamaica's Usain Bolt, best sprinter in the world. If they compete in water and the same rules apply to both athletes Phelps will win. But is this fair? After all, choosing swimming in itself gives Phelps an overwhelming edge. This match is unfair because it takes place in a swimming pool. The rules may be the same for the two contestants, they are still the rules of the sport of one of them.

Giving credit to students because their name starts with a B is not any less fair than giving them credit because they have the correct answer. In both cases the same rule is applied to all students identically, but it happens to be more convenient for some students than for others — all students are treated fairly but some are treated more fairly than others. Applying the same rule to all students is no fairness at all when it means applying rules that are themselves unfair: it merely propagates the fairness or unfairness of pre-existing rules. It may be relevant to grading but it is not applicable to exam design as a whole. (Those who equate fairness with fairness of grading are missing something and therefore cannot be right.) Saying “Assess As Ye Would Be Assessed”⁹ is of no help — even though I may personally like the idea that those whose name starts with a B should be advantaged, this is not an acceptable policy. One needs a richer concept, one that could generate the necessary criteria (which chanting ‘treat all students the same’ will never achieve).

Diversity and incommensurability of talents

Some teachers “value students with a diverse set of academic talents” and do not want to give advantages to those with one talent rather than another.¹⁰ What if a student claims that his specific talent was not taken into account, so that the exam is unfair? This is just what happens when Phelps and Bolt compete in a swimming pool: the gift for running of the Jamaican is not taken into account at all. Let us imagine that we came up with some weighted average of their times on 200 m dash and 200 m butterfly that, we hope, values their different gifts. Let us further say that Michael Phelps still wins. Can we conclude that he is superior to Usain Bolt? The latter would contrarily conclude that it shows that our scheme was still unfair to him. And if he wins then it is Phelps who will consider that he was unfairly disadvantaged. Unless the match-up results in a tie, the loser will claim that the match was unfair, and there would be no way to prove him wrong.

Of course, this Solomonian outcome is unsatisfactory. But it gets worse when one considers a match between either champion and someone who is in no way a great athlete, say me. This should also be a tie, otherwise I would complain that it was unfair to me. Does it mean that I am as great an athlete as Michael Phelps or Usain Bolt? Not at all. It simply means that one cannot assess both the fairness of the match and which contender is greater (one cannot solve for two variables based on a single equation). If there is a single thing which I can do better than him then I can argue that his victory is unfair as it undervalues my one talent. The difficulty springs from the number of possible talents and their incommensurability — how to rank people who are better in some respects and weaker in others?

Let us consider the example of partial credit in math or science assignments involving calculations. If it is granted generously, students who are conscientious and careful will complain that this is unfair to them: obtaining the correct answer or not makes little difference, so that a talent of theirs is undervalued. If little partial credit is given, on the other hand, students who knew how to solve a complex problem but made a small mistake will complain against such unfairness: since they get about the same grade as students who could not solve the problem at all, a talent of theirs is undervalued. Of course, it is obvious that —whatever the grading policy— all students are treated in an identical manner, so that ‘applying the same rule to all students’ is of no use to make a decision regarding partial credit. Whatever we do, we are bound to give an advantage to someone. (In fact there is another solution, as already mentioned: we can give the same grade to all students — but this obviously undermines the point of assigning grades.) The weakness of the extant talk on fairness is such that few teachers would be able to make a meaningful response if accused of writing unfair exams (or even to understand why there is such an accusation in the first place: after all, they did give the same points for the same answer).

Fairness cannot be the central issue

It does not hurt to say that “fair test design should provide examinees comparable opportunity, insofar as possible, to demonstrate knowledge and skills they have acquired that are relevant to the purpose of the

test”¹¹ but, since cautiousness and the ability to solve complex problems may both be relevant, it does not help either. If students have different kinds of abilities one cannot tell which student deserves a better grade unless one can tell which ability is more valuable. And nothing is said about the criteria to be used for this comparison, i.e. how to decide what criteria are relevant, fair to use. Actual exams may account for only a limited set of talents, and oftentimes focus on unimportant ones — students with little aptitude who have enough memory and are willing to waste their time filling it are often favored.

The main issue is not fairness (beyond the obvious ‘same answer same grade’) but rather that what is exactly being assessed is seldom thought through. I may hold that the ability to solve complex problems is more important (which is true for scientists and engineers for instance) and someone else may value accuracy more (which would be the case for example in accounting). What is an important ability depends on the context: being tall is good for basketball players while being short is good in the case of jockeys. Primary and secondary students may end up doing about anything, so that about anything may *a priori* turn out to be a relevant skill — if opposites can *a priori* be both good, to which of the two would it be fair to give an edge? It does not seem that pro-fairness platitudes contribute to coping with such issues.

The *British Medical Journal* recently asked “Are national qualifying examinations a fair way to rank medical students?”¹² The question had to be framed in terms of fairness to comply with tradition but, since there is little to be said on the fairness of the matter, the responses had nothing to do with fairness: they dealt with economies of scale and quality of education. Why contrive everything in terms of fairness? why make fairness the only concern? Fairness is a moral requirement and as such is supposed to trump mere technical issues (students may complain that exams are unfair but are unlikely to complain that an exam does not efficiently test the knowledge and skills it is supposed to assess). No evidence that its moral nature can really give fairness a special, superior status is provided.¹³ And, as I pointed out in this article, discussion of the fairness of examinations is bound to drift towards non-moral questions, such as what skills are important. Since an advantage must necessarily be given to someone, the question is to whom, i.e. which competence is more valuable. And this, obviously, cannot be determined by repeating ‘exams must be fair.’
© Mathieu Bouville, 2008

* Electronic address: mathieu.bouville@gmail.com

¹ R. M. Felder, “Designing tests to maximize understanding,” *Journal of Professional Issues in Engineering Education and Practice* **128**, 1–3 (2002). Also see H. Vos, “How to assess for improvement of learning,” *European Journal of Engineering Education* **25**, 227–233 (2000): “For the student, the most important question to be answered in order to be content with the outcomes of the assessment is ‘what is a fair assessment?’”

² P. Wankat and F. Oreovicz, *Teaching Engineering* (New York: McGraw-Hill, 1993), p. 214. Note that it is not a sufficient condition: a good grade, however unfair, will never make a student complain.

³ C. A. Wendorf and S. Alexander, “The influence of individual- and class-level fairness-related perceptions on student satisfaction,” *Contemporary Educational Psychology* **30**, 190–206 (2005). Also see P. A. Koushki, A. Christoforou, A. M. Larkin, and A. Al-Roomi, “Engineering students’ attitudes towards teaching and teachers,” *International Journal of Engineering Education* **21**, 424–433 (2005).

⁴ An English–English bilingual dictionary can be found for instance at http://en.wikipedia.org/wiki/List_of_words_having_different_meanings_in_British_and_American_English.

⁵ Also see M. Bouville, “Is diversity good? Six possible conceptions of diversity and six possible answers,” *Science and Engineering Ethics* **14**, 51–63 (2008).

⁶ A. Lizzio, K. Wilson, and V. Hadaway, “University students’ perceptions of a fair learning environment: A social justice perspective,” *Assessment and Evaluation in Higher Education* **32**, 195–213 (2007).

⁷ National Research Council — Committee on the Foundations of Assessment, *Knowing what Students Know* (Washington, DC: National Academy, 2001), p. 39.

⁸ In fact, no exam can be perfectly fair. Imagine that student D spent days doing problems from various textbooks to prepare for the exam while student L did only one such problem. Imagine further that the problem L did is on the exam while none of those done by D is. This is unfair — student D would have *deserved* the good grade that L received out of *luck*. Yet, nobody can be blamed: the instructor did not favor L, luck did. In such a case, D complaining that the exam was unfair is both justified and pointless. There are always situations that are unfair without anyone doing anything unfair. Part of the unfairness of exams is inexpugnable. Also see M. Bouville, “Why is cheating wrong?” (available at <http://arxiv.org/abs/0803.1530>).

⁹ D. A. Payne, *Applied Educational Assessment* (Belmont, CA: Wadsworth Publishing, 2nd edition, 2003).

¹⁰ L. J. Everett, R. M. Alexander, and M. Wiene, “A grading method that promotes competency and values broadly talented students,” *Journal of Engineering Education* **88**, 477–484 (1999).

¹¹ W. W. Willingham and N. S. Cole, *Gender and Fair Assessment* (Mahwah, NJ: Lawrence Erlbaum Associates, 1997), p. 10.

¹² C. Ricketts, J. Archer, and I. S. G. Noble, “Are national qualifying examinations a fair way to rank medical students?” *British Medical Journal* **337**, 546–547 (2008).

¹³ M. Bouville, “Crime and punishment in scientific research” (available at <http://arxiv.org/abs/0803.4058>).