

Can Students evaluate Teaching Quality?

Matti Wiberg

Department of Political Science, University of Turku, Finland
(eMail: wiberg@utu.fi)

Abstract The paper presents some critical remarks on the usefulness of students' evaluations of the quality of university teaching. It is widely accepted that students periodically evaluate the courses offered and examination books used at university departments. But what, precisely, do they evaluate? How competent are students to evaluate teaching? It is not by itself clear what the proper use of these results should be. What can be learnt from the results? How much power should the university give to the students in evaluating study units? The paper discusses these questions in the light of the Condorcet Jury Theorem of social choice literature and recent results in quality management literature.

Keywords teaching quality, university studies, evaluation, Condorcet Jury Theorem

Numerantur enim sententiae, non ponderantur; nec aliud in publico consilio potest fieri, in quo nihil est tam inaequale quam aequalitas ipsa. Nam cum sit impar prudentia, par omnium ius est.

For votes are counted, their value is not weighed, and no other method is possible in a public assembly. Yet this strict equality results in something very different from equity, so long as men have the same right to judge but not the same ability to judge wisely. (Translated by Betty Radice.¹)

Pliny Letters II. 12, 5

¹In the translation by William Melmoth: Votes go by number, not weight; nor can it be otherwise in assemblies of this kind, where nothing is more unequal than that equality which prevails in them; for though every member has the same right of suffrage, every member has not the same strength of judgement to direct it.

1. Introduction

Do students recognise quality when they evaluate courses or examination books at university departments? How good are they in their judgements? Should we rely on their ability to make the right choices?

The purpose of this paper is to discuss the usefulness of students as evaluators of the quality of teaching in universities. Casual observation, introspection and anecdotal evidence are used along some general results in social choice. We will be sceptical and defend the thesis that students' evaluations should be taken with a grain of salt. One does not have to be a student-hating fascist to raise doubts about students' intellectual competence to evaluate the curriculum or to raise doubts concerning the beneficial effects of strong student participation in the planning of the curriculum. Students should not be treated as consumers of commodities who know better than anyone else what is good for them. A university is (well: should be) seen above all as an information production and transmission organization. A university is no ordinary school. Students study in order to get new, relevant and useful information, and improve their skills of thinking and researchers do research in order to add new, relevant and useful information to the cumulative stock of wisdom of human kind. But these noble goals are by no means the only motivations. Students and teachers are not only students and teachers they are also human beings, citizens, political actors with many other interests than the purely intellectual; they are, among other things, also welfare maximizers. Before taking students' opinions seriously, we should at least ask ourselves about the quality of the intellectual competences on which the opinions are based.

2. Preliminary Definitions

A course, say a series of lectures, or a textbook is hereby called a *study unit* for brevity. Studying consists thus of studying study units. The process of studying can be modelled as a sequence of study units. It should be noted here with all seriousness that good university professors have more to offer than the execution of the contents of a study plan. The real input of good educators are innovation, inspiration, in a nutshell: latest results of research. Too few of us reject the idea of study plans, written years before, that reduce university to higher level school. University education should be based on the very idea of unity of research and teaching. This aspect will not be dealt with in this paper, but it deserves constantly our critical attention.

Informational asymmetry: In order to be meaningful studying should contain the following informational asymmetry before the study unit is stud-

ied: The teacher knows the content of the study unit, the student does not. Sometimes this does not hold true, but as a general rule this should apply. Sometimes the teacher learns more than the student (*docendo discimus*; we learn by teaching), but that is another matter. It is a definitional characteristic that one must know and understand the content of study in order to be able to teach properly. By definition the student does not know the content (intellectual material) of the study unit in advance. If she² would know it, the whole point of studying would be redundant: you study to know things you did not know beforehand. You can only learn things you do not know. You cannot learn things you already know.

The point of education is to remove the informational asymmetry between the instructor and the student.

What is the purpose of studying for instance political science? There are several competing goals. For the sake of simplicity, let's focus on two goals only, the first internal and the second external: (1) Acquisition of information of politics and science and (2) acquisition of a degree in political science.

An individual student might be well able to evaluate the fulfilment of the second goal, but she is not necessarily able to do the first evaluation. A student simply cannot, during her studies, evaluate whether she has learned enough. That is something to be decided empirically only afterwards, i.e. after graduation. She simply does not have any proper measure.

How do students know what they need to know before they have learned it? A priori there is no way students could be knowledgeable enough to signal demand of knowledge. The usefulness is something that one can evaluate only *post festum*.

What do students evaluate? – Take an arbitrary introductory course in political science. It does make sense to ask the students after the course, whether they *thought* that the material was presented in an understandable way or not. (But do they *know* that they understand? To answer this question, is one reason why we do exams.) Does it make sense to ask the students whether the *content* of the course is important, proper and relevant? How could they know? If they were able to tell, why do they participate in the course? It goes without saying that a student takes a course in order to get *more* relevant info on a topic that is of general interest, but it does not follow from this that the student necessarily is able to make prudent evaluations concerning the substantive content of the course.

²In this paper, students are designated to be females, reflecting the growing trend in the Western world.

3. How Clever Are Students?

Every cohort of students should be more intelligent than the previous one. The best students should always be wiser than their teachers. Otherwise there would not be much point in having universities in the first place: the idea of cumulativeness of knowledge, research and teaching is very strong in our culture.

In order to be able to discuss this more properly, it is perhaps helpful to introduce some technical help from the field of social choice theory. Consider an evaluation situation in which students are asked to evaluate whether a particular study unit is good or not good (intellectually satisfactory or not). Intuitively, all true democrats (aren't we all?) are willing to conclude, that a particular study unit is not good, if the majority of students think so. This intuition is very strong, but it could still be misleading.

The Condorcet's Jury Theorem states that a majority of a group is more likely than a single individual to choose the better of two alternatives (see Baker (1976), Black (1958), and McLean and Hewitt (1994). Condorcet (1785) stated that a group of individuals who have to choose one of two alternatives by expressing their individual opinions and the final verdict is determined according to simple majority rule based on these opinions, would likely make the correct choice. Moreover, this likelihood would tend to become a complete certainty if the number of members of this group tends to infinity. There hardly is a better argument for the *vox populi, vox dei* –doctrine.

Now the point here is that there is safety in numbers: if many students claim that a particular study unit is not good, we should trust their evaluation – and consequently remove the unit from the curriculum. The policy recommendation is quite clear: get rid of the bad elements. A majority of students cannot be wrong! –This participatory credo is among the few taboos of the present. We seldom, if ever, hear anyone oppose to the very idea of students being right in some absolute sense.

4. The Condorcet Jury Theorem

In a somewhat more technical presentation: Let $1, 2, \dots, n$ denote the n students ($n > 0$). We assume that there are two states of the world (the study unit), represented by a binary variable X taking the value 1 for 'good' and 0 for 'not good'. The votes of the students are represented by the binary random variables V_1, V_2, \dots, V_n , where each V_i takes the value 1 for a 'good' vote and 0 for a 'not good' vote. The vote of student i is correct *if and only if* the value of V_i coincides with the value of X . The classical Condorcet jury

model assumes:

Competence (C). All actors have at least the probability $p > \frac{1}{2}$ of making the right judgment.

Independence (I). All actors are independent of each other's judgments.

What is crucial here is that *each* member of the group has an ability p to decide correctly such that $p > \frac{1}{2}$ and individuals vote independently, in a sense of a complete statistical independence. In the classical Condorcet jury model, different jurors' votes are independent random variables, where each juror has the *same* probability $p > \frac{1}{2}$ of voting for the correct alternative. The probability that the correct alternative will win under majority voting converges to 1 as the number of voters increases. Hence the probability of an incorrect majority vote can be made arbitrarily small just by increasing the number of voters.

In the original version (not to be explicitly found in Condorcet (1785), but only implicitly included) it is assumed that the individuals have *homogeneous competence*.

The Condorcet Jury Theorem can now be stated (see Ben-Yashar and Paroush 2000): If all members of the jury have the same probability p to make the right judgment, the probability P that the group majority makes the right judgment is as follows

1. If $p > \frac{1}{2}$, then P is strictly larger than p . P grows with the number of n and approaches 1 as n approaches infinity.
2. If $p < \frac{1}{2}$, then P is strictly smaller than p . P decreases with the number of n and approaches zero, as n approaches infinity.
3. If $p = \frac{1}{2}$, $P = \frac{1}{2}$ for all n .

The Jury Theorem holds, though, even if the assumption of homogeneous competence is relaxed, but the average competence of the team to be a constant larger than $\frac{1}{2}$ is imposed. (See Paroush (1998).)

Competence. Clearly, the assumption of homogeneous competence is not true in our context: some students are more able than others to do the right choice. Moreover, it is not true that the average competence of students is *always* larger than $\frac{1}{2}$: there are too many rotten eggs even among university students. There are no general guarantees that the students are competent enough to make a positive contribution to the group decision.

The point of the theorem is that the information of individuals helps to reach a decision superior to that of any individual. But the crucial assumptions are that all individuals have at least a minimal positive competence

above $\frac{1}{2}$, and that each student makes her own choice independently of other students' choices.

The other assumption (independence) will be only shortly discussed in this context. Let us start with one observation: Students do not form and signal their preferences independently; (even) students tend to follow opinion makers. Some student leaders are signaling their individual preferences, and quite a large proportion of students follow their opinion – regardless of the content of these preferences. Rather than forming their own opinions independently, the students' behavior can be modeled as a form of *herding*: there is a strong tendency among students to follow the lead of other people. They follow aggregate market behavior by forming opinions based on collective actions in the opinion forming process (Bikhchandani and Sharma 2000).

Claim: The assumptions of competence and independence do not hold.

Proof: Casual observation of student cohorts during more than three decades.

The university students might very well be *more* competent than their colleagues outside the university, but it does not follow that they are competent in any *absolute* sense of the word. If the cornerstones of the Condorcet Jury Theorem do not hold, then the group of students are not necessarily competent *enough* to correctly evaluate a study unit with a probability greater than $\frac{1}{2}$.

It could very well be the case that paying too much attention to ill-founded signals from students, the university administration is just contributing to the perverse 'dumbing down syndrome', which by the way seems anyway to be going on with an accelerating speed even without the efforts of university teachers. In the long run this means that the rate of information transmission of study units is asymptotically approaching zero. There are no costs to an individual student for $p < \frac{1}{2}$ and for herding.

5. Evaluations

What do students really evaluate when they are asked to evaluate a study unit? This is not a trivial question. Take a series of lectures, for instance. It is by no means guaranteed that students take into consideration exhaustively intellectually important dimensions – and only them. Some students might neglect some crucial dimensions and take into consideration issues that are of no importance given the goal of learning. Sometimes students evaluate more the *persons* delivering the lectures, their style, speed of speech, cloth-

ing, jokes, student friendliness,³ etc. than the intellectual *content* of the lectures. It cannot be guaranteed that students have the same variables in mind as the instructors or those who designed the curriculum. The information transmitted can be of smaller importance than various human interest issues.

The evaluation is a game of three sets of players: (i) the decision makers, i.e. those in charge of *designing* the curriculum, (ii) the teachers, i.e. those in charge of *teaching* the curriculum, and (iii) the students, i.e. those in charge of *learning/studying* the curriculum. There is no way automatically to guarantee the congruence of goals among these players. Their pursuits do not need to be homogeneous. It is almost self-evident that students and teachers do not have identical preferences.

There is no generally agreed upon definition of quality. One quite common idea is that quality is "*fit for purpose*" (Juran 1998). According to his criteria a study unit is good, if and only if it is fit for the purpose the relevant players have in mind. But: Who are the relevant players? The curriculum designers have perhaps one set of goals in mind, a bunch of students might think of something else. And teachers in charge of teaching the curriculum might have something entirely different in mind.

It is noteworthy that the teachers and students by no means need to have identical goals. How could they have? They have different professional and other interests. Teachers might be more interested in, say, their salaries and workload than in transmitting important information, i.e. information that is necessary for the fulfilment of the given pedagogical goals. Students, on the other hand, might be more interested in, say, getting connected to interesting peers in their cohort or in the actualisation of a well-functioning marriage market within the university system than in learning study unit contents.

The traditional definition of quality says simply that quality is the conformance to specifications. According to Taguchi's (1988) definition of quality, quality is the loss imparted to society from the time the product is shipped. The point is simple: the lower the costs are to society after the production of the product, the better quality it demonstrates. If we take this characterisation seriously, we should not allocate much *consequential* free speech to university students before they have graduated.

6. Who Should Evaluate the Quality of Teaching?

We opened the paper with the question whether students recognise quality when they evaluate courses or examination books at university departments.

³ Don't ask for an operationalization!

As it cannot be guaranteed with certainty that the students make competent judgements, their opinions should not be the *sole* basis for decisions concerning future study units. In a society with free speech, we cannot make the students to be silent. But we can ignore their signals. This means that students as they are the informationally handicapped players should not have a strong say in planning the curriculum. However, it is always useful to listen to students, because even when they err, their signals can be helpful in revealing neglected spots that must be addressed by teachers.

This paper supports a more meritocratic, perhaps even a paternalistic view: some things should be decided based on merit and not popular opinion. Some times we as university teachers make better decisions without consulting students. Sometimes we should not even be interested in having the students' opinion on the intellectual content of our study units. Sometimes both the students and the society are better off if we decide which kind of study units students have to study – and the students just study them.

This does not mean that we should trust university teachers without reservation: there are more than plenty of rotten apples among us, too.

References

- Baker, K. M. (ed.) (1976), *Condorcet: selected writings*, The Bobbs-Merrill: Indianapolis.
- Ben-Yasaar, R. and J. Paroush (2000), A Nonasymptotic Condorcet Jury Theorem, *Social Choice and Welfare* 17, 189–199.
- Bikhchandani, S. and S. Sharma (2000), Herd Behavior in Financial Markets, *IMF Working Papers* 48.
- Black, D. (1958), *The Theory of Committees and Elections*, Cambridge University Press: Cambridge.
- Condorcet, N. C. de (1785), *Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix*, Paris
- Juran, J. M. and G. A. Blanton (eds.) (1998), *Juran's Quality Handbook*, McGraw-Hill Professional, 5th edition.
- McLean, I. and F. Hewitt (1994), *Condorcet: Foundations of Social Choice and Political Theory*, Edward Elgar: Aldershot.
- Paroush, J. (1998), Stay away from fair coins: A Condorcet jury theorem, *Social Choice and Welfare* 15, 15–20.
- Taguchi, G. (1988), *Introduction to Quality Engineering: Designing Quality into Products and Processes*, Unipub: White Plains.