

53 Powerful Ideas All Teachers Should Know About

Graham Gibbs

Idea Number 10, June 2014

Students' marks are often determined as much by the way assessment is configured as by how much students have learnt

When students appear to perform unexpectedly badly the spotlight tends to fall on the students. Did they lack the necessary background knowledge? Were they not working hard enough? Did they not understand the criteria? When students do unexpectedly well, the spotlight may fall on the brilliance of the cohort – or of the teacher. But often it is in the way the assessment regime has been configured that determines average marks.

One of the most important factors is sampling. An unseen examination with unpredictable (though fair) questions, without too wide a choice of question, with questions that between them randomly sample the entire curriculum, is capable of producing a mark that, even though the students only answer a small number of questions, is a reasonable indication of what students have learnt across the whole curriculum. Exam marks are usually lower than marks for coursework for precisely this reason. Coursework usually samples the curriculum in a different way. Students choose perhaps two topics to study in depth and the mark they get may be a fair indication of their learning about those two topics, but it certainly isn't about all the other topics. The fact that a student averaged 70% on the coursework cannot be assumed to mean that, had they unexpectedly been asked to tackle questions on other topics, they would have

done as well – they almost certainly would not. Coursework marks are therefore rarely valid as indicators of learning across a curriculum. Institutions where 90% of student marks come from coursework cannot make the same claims about what degree classifications mean about overall student learning as those where only 25% of marks come from coursework. It is one of the most likely reasons that there has been such rampant grade inflation: assessment no longer samples the curriculum as rigorously as it did and marks no longer mean what they used to mean.

It is clear that on coursework-assessed courses students are highly strategic in their use of time and allocate it primarily to the topics they will tackle assignments on, and not to topics that they do not tackle assignments on. As they know which topics these will be it is easy to do this without risk. In contrast, with unseen exams a student has to study all topics in some depth or they are potentially taking a very great risk. Question-spotting and some risk taking obviously takes place, but nevertheless effort is distributed more widely across topics, so each topic gets less time. Exam marks reflect this lower allocation of effort per topic, even though more effort overall is allocated to more topics than with coursework assessment and students are therefore likely to have learnt more.

53 Powerful Ideas All Teachers Should Know About

Graham Gibbs

If you wanted to sample the curriculum with coursework assignments rather than with exam questions this would be easy to arrange. You simply have to set assignments for every topic and then only mark two of them at random: the mark will now be a fair indication of learning across the course. In addition sensible students will have taken every assignment equally seriously in case it were to be marked, and so you will get more effort and more even effort across topics than you did before and students will have learnt more than they did before. Nevertheless you will probably get lower average marks as students will not be able to allocate most of their effort to only those two topics that they know will be marked. Sampling of coursework seems perfectly fair and much more valid, but students may howl in protest, even though they do not protest when exams sample the curriculum. They would argue that it is not 'fair' that they put work into assignments that were not, in the end, marked. But that is exactly what happens when not all topics turn up on the exam paper. You would simply need to explain the logic of the arrangement and tell them to get used to it.

In some subjects 'sampling' is not a central concern. For example in English Studies, the main concern may be for the sophistication of students' use of the discourse of critical analysis, and how wide a range of literature this has been demonstrated on may not be considered to matter much (though longer practice on more varied literature seems likely to develop more sophisticated critiques).

With exams, marks may depend on the unpredictability of the questions, the proportion of questions that are compulsory or optional and the breadth of choice, the extent to which questions are recognisable variants of questions addressed in seminars or completely new, and so on. Some such issues are invisible to external examiners – all they see are the questions and the answers. If you have lots of choice of predictable questions that are very similar to ones that have already been discussed in seminars, then the examiner is likely to be impressed with students' answers, even if in reality they may not reveal much about the breadth of student learning. If a teacher has been criticised for the poor performance of their students it is very easy to arrange for performance to go up next time without anyone spotting how it was achieved. When I was gaining publicity for conducting research on the negative effects on student performance of large classes I used to receive letters (it was a long time ago...) from Professors at Russell Group universities explaining how they were being expected to increase the proportion of students gaining top grades, each year, while the students stayed the same and class sizes increased, by teaching to predictable questions and by other such shenanigans. Grade inflation has been highest at [Russell Group Universities](#) despite their greater use of examinations.

Coursework marks, for a single piece of coursework, may be influenced by how large and difficult the assignment is, how much original work and independent thought is required, how much integration of related

53 Powerful Ideas All Teachers Should Know About

Graham Gibbs

topics is involved, and so on. Again it is easy to manipulate the demands to rig the system, even if the learning outcomes and criteria are apparently exactly the same. I have little confidence in the ability of all but a very small minority of teachers to draft learning outcomes in ways that assure standards. The QAA pinning all their trust in standards on the specification and assessment of learning outcomes seems to me to be naive to the point of recklessness – it is simply too easy to ‘game’ the system and get away with it.

External examiners are also sometimes not fully aware of the size or level of the curriculum that is being sampled in assignments or exam questions. It can be very hard to gauge from course documentation. In addition the level that subject matter is dealt with in teaching often differs markedly from what documentation describes without anyone but the teacher and the students being aware of it. An assessed sample might be of a very small curriculum or of a very broad curriculum, with the same list of course descriptors being a tolerable summary of either.

It has also been demonstrated that the degree classification students get is determined to an extent by the institutional rules, or departmental rules, for adding up marks from assessed work and from courses. In some contexts it is a strict average with no weightings. In others the lowest marks can be dropped. In some only the best marks count. In others students can keep on taking modules

until they can accumulate enough high marks (and dropping out of modules they were not doing well on) so that they can get a ‘good degree’, where less choice and flexibility would have produced a much lower average or even failure ... and so on. If you take the same marks from students in one institution and process them using the addition rules from another institution, it has been demonstrated that this would often produce a different set of degree classifications. Institutional mark averages vary widely in ways that cannot be easily explained by the quality of their students or any other variable, and it has been argued that the best predictor that an institution has a particular pattern of performance is that they have always had this pattern: it is an institutional characteristic underpinned by procedures and culture that is independent of student learning. Teachers learn to fit in with it, to some extent regardless of student performance.

Sometimes the differences have pedagogic justifications. For example In Fine Art, students often present only their very best work, right at the end, and that carries much of the weight of assessment. Imagine what would happen in Medicine if that assessment regime operated! In Humanities at Oxford I often heard academics justify some assessment regulations because they ‘give students an opportunity to shine’ and they often ignored or discounted poorer performances elsewhere.

To comment or contribute your ideas, see SEDA’s blog: thesedablog.wordpress.com