

Ah! ... So *That's* 'Quality'

Royce Sadler, Senior Assessment Scholar in the School of Education, University of Queensland, and Professor Emeritus of Higher Education, Griffith University

This Briefing tells the story of a university student who was puzzled by not receiving the high grade the student had expected. The work was independently reviewed, but the original grade was confirmed. The rest of the story is about the power of concrete examples of work of the highest quality as a means of conveying what quality meant, in this learning context.

It originally appeared in 2002 as chapter 16, pp. 130-136, of P. Schwartz and G. Webb, eds, *Assessment: Case studies, experience and practice from higher education*, London: Kogan Page. It is reproduced here by kind permission of the author and the publishers.

In Queensland, Australia, a person may qualify as a high school teacher by undertaking a post-degree programme in education. Part of one such programme was the course on assessing student learning that is the focus of this case. It was taught by two people, one of whom was me, using a combination of large mass lectures (250 students) and small tutorial groups of about 20 students each. Assessment for grading purposes involved two components, a mid-semester written project–assignment of about 2000 words and a formal short-answer examination at the end of semester.

Do students produce better quality work when they understand the standards that a university teacher works to?’ I had often wondered about that, no more seriously than when I was teaching (and assessing) the course on ‘assessing student learning’ for a group of post-degree students who were training as high school teachers – where they would in turn teach (and assess) their own students. I knew that many students find having their work assessed stressful. I always felt obliged to minimize stress and to demonstrate within the course itself how assessment principles could be translated into good practice. So I put a lot of effort into being as precise as I could about the assignment topic that was part of the assessment for the course – the nature of the task and what I expected. I did not want students to have to puzzle out or guess what the task required. That much at least should be clear. As I saw it, the students’ challenge was to develop a quality product.

The specifications that I produced for the assignment ran to about one page. My own challenge was to find the optimum balance between spelling out exactly what I was looking for and leaving enough scope for originality. The last thing I wanted were cloned assignments.

My usual practice was to set out a number of criteria for the assignment so that the students would know in advance how their work would be judged. The criteria in this case were:

- Relevance
- Comprehensiveness
- Coherence
- Logical reasoning and
- Presentation.

I put together a brief glossary to explain the meanings of these terms. *Relevance* meant the degree to which the assignment stuck to the topic and addressed the task, without extraneous material. *Comprehensiveness* referred to how well all bases had been covered, with nothing essential left out. *Coherence* meant how well the entire piece hung together. And so on.

The students duly completed their assignments, and I spent a week or so trailing through what they had written. It was relatively easy to judge the high quality submissions, and just as easy the weak ones. The ones in the middle, as usual, gave me the most trouble. There seemed to be two main reasons.

First, there were different patterns of performance. Some pieces of work were sparkling on one or two of the criteria, abysmal on some others, and middling on the rest. To arrive at a mark, I mentally traded off excellence on some dimensions against weaknesses on others. However, the same overall score also arose for assignments that were reasonably good on all the criteria. How fair was it to use the same mark to represent very different sub-patterns of achievement? What purpose was served by collapsing all the information into a single score?

I had another problem with using separate criteria and reaching an 'on-balance' decision. I knew that some university teachers handled this by specifying maximum scores for each of the criteria, then allocating sub-scores for each and adding them up. I found these numerical schemes to work only moderately well. Typically, I felt uncomfortable with the conceptual overlap that emerged between two or more criteria when I had to decide on a sub-score for each of the separate dimensions. Some criteria that seemed to be discrete and distinct in the abstract turned out to blend into one another when I came to actually use them. The boundaries became fuzzy. This seemed to be more of a problem with some assignments than others. I also had a problem with the occasional discrepancy between the story told by the sum of the sub-scores and my global judgement. Which should I trust?

The second reason that I found it more difficult and time consuming to appraise the middle-level assignments was that I felt morally obliged to give students written comments as feedback on their work. In reality, there was always more to be said about mediocre work, because there were so many ways in which it could have been improved. For any assignment that I thought was potentially salvageable, I put a lot of effort into making detailed suggestions. I hoped that students would take these on board for future assignments. I mostly had little that was diagnostic to say to the best performers, because they had got their projects well and truly

together. From time to time, it was great to come across one that was better than I could probably have done myself. At the other end of the scale, the weak projects had so many deficiencies that I hardly knew where to start with the feedback. Privately I wondered how these students could have already completed degree studies.

Once the assignments were marked, I distributed them to students during one of the tutorial sessions. Students tended to sit in the same seats each time. On one occasion, a student in the front row said almost immediately that the score I had given her, 13 out of 20, was far too low. I was slightly taken aback by the insistence in her voice, especially when she was unaware at that point of the scores or comments on anyone else's assignment. Most of the other students by this time were looking through their own work and took no notice of this student's comment.

'My assignment is worth a lot more than 13', she said. 'I never receive scores as low as this. Never.'

I asked her what she thought her score should have been.

'I always get 19 or 20 on a 20-point scale. This one deserves the same. My assignment certainly met all of the criteria you specified.'

She was very definite about it – and I was left wondering how to respond.

My first reaction was that this seemed to be shaping up as a standard attempt at bargaining for a higher grade. I had experienced this from students occasionally in the past. Some students are accomplished players at haggling. They know how to use body language, have all their ammunition ready, and set out to win. I wondered, 'What is the deal here?' In any case, how could the student know realistically what her assignment was worth? I had had years of experience behind me, setting and marking similar assessment tasks.

On the other hand, maybe I had for some reason simply been too harsh. Perhaps hers was the last one I had marked one night before I dozed off to sleep. My usual self-alert for tiredness was when I became conscious of having read the same paragraph three or four times, still without a clue as to what was in it. Maybe I had passed that point without knowing it.

I offered to grade her work again. I also said that I would have my teaching colleague make an independent judgement without any communication from me as to the reason. My colleague and I had co-taught for many years, so I knew that getting a second opinion would be easy to organize. I said that to make this arrangement fair, I would need two unmarked copies of the assignment so that neither of us would be influenced by my earlier written comments.

She was happy to provide me with these and I agreed to have a fresh look at her work within one week, in time for the next tutorial session. I found that the work was clearly on task. It was also well written, smooth and very readable. It was, however, lacking in penetrating thought and was descriptive rather than analytic. She had engaged with the topic, but not in a way that got to the fundamental issues and principles.

My colleague's independent conclusion was almost exactly the same as my first one. So was my second. I expected that the student might interpret this as defensiveness or stubbornness on my part and collusion with my colleague.

That year, as it turned out, I had given two students a score of 21 because of the superb quality of their work. It was just brilliant. I approached both of them to see if they would let me show their assignments to another student to demonstrate what superb work was really like. I explained that I would do it anonymously and that I would need fresh copies of their assignments, without my comments in the margins. They were a bit surprised but were happy to go along with the arrangement.

When I returned the first student's work, I explained that my and my colleague's separate judgements were basically consistent with my first. She was somewhat stunned and immediately expressed concern. Then I offered to let her see the other two assignments if she wished. I explained that I had obtained the appropriate permissions. She jumped at the opportunity.

'Read the other two assignments and compare them with yours. If you cannot see why theirs are an order of magnitude better than yours, I will be happy to talk with you and explain what makes the difference in terms of quality.'

The next week she returned the two assignments and said simply: 'No contest! I can see what you mean'. This time she made no reference at all to the criteria.

'Do you want to discuss the two assignments with me?'

'No. I'm quite satisfied.'

That was the end of the episode as far as she was concerned.

Sitting beside her was another student who had heard the conversations over the previous two weeks. He asked me, 'Please could I read those assignments as well?' Because I already had permission to show them to another student for illustrative purposes, I could see no ethical or practical reason to deny his request, so I agreed. One week later, he brought them back with the comment, 'I had no idea that this is what you were expecting. I am positive I could do as well as this

myself, now that I can see what you were looking for. Can I have another go?'

Instinctively I wanted to agree, but I was conscious of then having to decide what, in the interests of fairness, I should do about the other 248 students in the course. These students had not seen the two exceptional assignments, but I would have no grounds in principle for denying them access. The option of reworking their assignments had not occurred to them, or to me for that matter. I shuddered at the thought of possibly having to scale another mountain of marking once word got around.

In the end, I agreed to give the neighbour of the first student another opportunity to demonstrate what he could do. He accepted that he would not be given a higher grade on the second attempt even if it were warranted. His main aim was to demonstrate what he believed he was capable of. Our arrangement was that he would work independently, without further access to the two sample assignments and complete the work within two weeks. He had claimed that he could do as well as the sample assignments. As it turned out, he was not far wrong.

Discussion

Simple though this incident was, and despite the fact that it involved only four people at the time, it had a profound effect on my view of teaching and assessment, and what I built into future strategies. It drove home to me just how many assumptions university teachers make about how our expectations can be communicated.

One of the key assumptions is that specifying the criteria to be used in appraising students' work does two things. It sets students on the right course for shaping the content and structure of their work, and it is sufficient to explain how we intend to score or grade it. This case study shows that, whereas students may well believe that they have 'met' all of the criteria, the fundamental issue has to do with quality.

In a restricted sense, content and structure can be audited. Student and assessor alike may be able to detect whether (as in the example above) all of the content is relevant to the task (with little or nothing that is irrelevant included), whether that content covers all or most of the important issues, whether the treatment hangs together and is not disjointed, whether the reasoning is sound, and whether the presentation is consistent with academic norms.

But in describing the role of criteria in this way, I have expressed the issue in too simplistic a fashion. I used forms of words that broadly imply a two-state situation: *relevant* or *not relevant*, *coherent* or *not*

coherent, and so on. The first student's comment that she had 'met' all of the criteria followed the same line. In practice, most judgments of complex outcomes do not seem to me to be adequately made using discrete criteria. Listing criteria separately invites students to think about *qualities* rather than *quality*. In any event, there is always the possibility that additional criteria should be invoked for judgments about particular cases. Which of us has not felt constrained by a fixed set of criteria that we are confident cannot do justice to a particular assignment? The real issue is that of quality. Quality is determined by how the specified – and the unspecified – criteria are invoked in practice, and how all the various qualities contribute together in concert.

How could I as a university teacher improve my ability to convey to my students a concept of quality? Initially, I had tried to define content and structure in reasonably factual ways, as propositional knowledge. But that was not all there was to 'quality'. The old adage: 'I cannot really describe what quality is, but I know it when I see it' has more than a grain of truth to it.

If quality has to be *recognized* rather than defined, my practices had to be modified accordingly. The only way to recognize something is to 'experience' it in some way. That is what recognition is about, more or less by definition. The way to recognize 'quality' – that is, high quality – is to see it standing out against a background of the ordinary. That is what accounted for the potency of the exceptional assignments in the case study above as vehicles for conveying the concept of quality as it applied to the set task. It was *show and tell*.

The criteria that I specified were relevant to the assignment, of course, but neither of the two students at the centre of this case study expressed the need to debate them. The fundamental problem was one of *standards*, not *criteria*. Any statement as to whether something has or has not 'met' a criterion presupposes some kind of threshold. Telling the students about the criteria identified the key dimensions of interest to me as assessor, but not the thresholds.

Although this case is not about peer assessment as such, it convinced me of the necessity of focusing explicitly on the issue of quality. I now design peer assessment activities to provide students with opportunities to develop not only production skills but also clarity and an improved personal knowledge of what constitutes quality. I expose students routinely to a range of works that display the quality continuum. Those works are authentic and come from other students.

Students tell me that making judgments about the quality of work of the same kind that they are working on themselves is difficult. It is. But by developing these skills, they are better able to monitor and control the quality of what they are producing during the production process. That is, of course, precisely when it matters.

There is another side to this case as well. The documentation for the project assignment in this course was, I still judge, quite thorough. But by itself it did not go far enough. Giving students specifications for tasks, no matter how detailed those specifications, can never go far enough. If I were to formulate a theorem in teaching and learning similar to those in mathematics, this would be it:

Exemplars convey messages that nothing else can.

As a university teacher, I have long been aware of the clarifying power of concrete examples, illustrations, stories, case studies and metaphors. They help students to understand abstract concepts and to appreciate the relevance of theory. 'Telling' students about assessment requirements often turns out to be fairly abstract to the students. In the past, when normal telling failed to carry the message adequately, I resorted to more elaborate telling. I now try to show them as well. I realize that the same pedagogical devices that I use with respect to the subject matter in the courses I teach make equally good sense with respect to communicating my expectations about the quality of students' work.