



Table of Contents

Assessment

Prof Hang Chang Chieh

Purpose

Change

Active Teaching

Creativity

Open-Book Exams

Question-&Answer

Prof Hang Chang Chieh & Participants at the Seminar on Assessment

The Bell Curve

Rigid Exam Rules

Change

Types of Assessment

Norm-Refernce vs Criterion-Reference

Tutorial Assessment



On Saturday 3 April 1999, CDTL organised a seminar on student assessment which was led by Deputy Vice-Chancellor Professor Hang Chang Chieh. In this issue of CDTL Brief, we present to you a summary of Professor Hang's discussion and the Q & A session that followed.

Assessment

Professor Hang Chang Chieh
Deputy Vice-Chancellor

Purpose

The purpose of assessment is very important. Is it to assess our students or is it to assess the effectiveness of our teaching? And is it correct to have the same criteria for all kinds of assessment? We have to think deeply about these, for it is part and parcel of our professional duties and responsibilities.

Change

As recent as two years ago, we still had the British external examiner system in which we had to send our exam papers for vetting way ahead of time, even before teaching began. Hence, we had always been constrained by assessment. But most faculties have now changed the system. The examination questions are now vetted by the external examiner only after the event. As far as the registrar's office is concerned, the deadline for submission of exam papers is one month ahead for printing and processing. However, if the deans and the heads are "kiasu" and bring this deadline forward by one or two months, we would be stuck with the old system again. We should also aim to decentralize our examination process. We have decided to gain some valuable experience by first decentralizing the postgraduate examination process.

It is always painful to have to change. But it will be ridiculous if we do not. We are now experimenting with more empowerment, I would really like to see a total empowerment like the system which I had experienced as a visiting professor at Yale University in 1983. I went to the head of department and asked him when I would have to submit my exam paper. His reply was: "I don't have time to read your exam paper. What I had time to read in detail was your C.V. and when the university

council approved of your appointment as visiting professor, that was it. You decide how to set your exam questions; you decide the number of questions to set and the format of the paper. You decide the grades."

That was full empowerment. I liked that as a teacher and I think we should set it as our target. If you think about it, one day when you become a head of department, would you really have time to read all those exam questions? Are you an expert in that particular area to critique one of your experts and say that the question is too difficult or too easy? And when a student has done only half the question and the professor decides to give him 80% of the mark, can you challenge him?

So what we have been doing is not very effective. Time is better spent in a session like this when we discuss what we are trying to achieve through assessment. We have to think about it. If we do not, we are not fair to ourselves, the department and the system. The university depends a lot on our teachers to provide feedback to us on the effectiveness of their teaching, on how they find it in the field. Are we wasting our teachers' time? Are we wasting our students' time? Students spend three or four years with us as undergraduates—very precious time. Are we fair to them?

Assessment can help the students to learn, particularly continuous assessment, class test, assignment; less so for the major exam. Perhaps the major exam could be used to tell the society whether a particular student deserves a scholarship or whether he is a fairly weak student. But if we really think about it, the

fundamental purpose of assessment is to help a student know whether he has really understood the subject.

Active Teaching

I have circulated a paper on active teaching. Here is an example of active teaching/learning:

A Harvard professor thought that he had been doing a good job teaching first year physics because he had done it for more than ten years. One day, he read a paper about the ineffectiveness of the conventional teaching method and the advantage of active learning. He did not believe and decided to test it. He was teaching Newton's third law. This is High School physics—Action Equals Reaction—common sense. Normally, he would set the tutorial questions and 90% would get them right. But this time he decided to test whether the students really understood. So he devised a simple problem: One big truck and one small truck collided. Is the big truck exerting a bigger force on the small truck? Half the class said: "Yes", and half said: "No". Yet they were just taught the same principle.

To implement active learning, he then formed four small groups for students to discuss and argue among themselves. Ten minutes later, 90% got it right. This raises a question over the conventional way of teaching and assessment. With the conventional system, weaker students would have a superficial understanding of the subject and only the very bright ones would get into the deeper level of understanding. It would help tremendously if the students could discuss and help each other understand better.

We have to ask ourselves whether the way we assess, teach and engage students in discussion is effective. Do we compromise on our teaching because we have to complete the syllabus? Or is it because the exam questions have already been set? Think about it.

There is a real case which someone wrote to me about. To qualify for engineering, our students usually need a minimum of two A's and one B. Most of these students going to the U.K. will end up with a minimum of a second upper, if not a first class honours. And every year, we are dealing with eight or nine hundred of them, yet only thirty or so get a first class honours, a hundred or so will get a second upper, while the majority will be second lower, and some will go into third class and so on. We thought this bell curve was very good.

One of our second lower students went to do the GRE and the TOEFL. He had an almost perfect TOEFL score and 2100 points for the GRE. When we have overseas applicants with such scores, we would offer them a post-graduate scholarship. But this is our second lower student. Are we too strict in our assessment? This student who is labelled second lower may one day be the managing director of a large engineering company and try to fund our research. Are we caught by the bell curve? Not every student will be a professor and this is certainly not our goal. Are we assessing them based on who has the potential to be a professor?

Creativity

I will touch on two more topics very briefly; first, about creativity. Increasingly, we are setting our purpose to educate in such a way that our students will not only have sufficient knowledge to work in the society and continue to learn, but they must also be creative. Our country has reached an advanced stage of development where the conventional solutions may not work anymore. Some of our students are engaged by top companies and they have to produce world-class products. Some will become entrepreneurs competing with the best in the world. They have to come up with creative ideas and solutions. But during their years with us, have we engaged them in this kind of learning and given them the opportunity to think of creative solutions? Teaching creativity is a challenge. Assessing creativity is another.

The first thing to do is to read more about creativity. We should then ask ourselves some questions: Have we read the top three books in this area? Have we read the top ten papers about creative teaching? Better still, have we been practising creative teaching and have we been confident enough to share our experience with colleagues?

Open-Book Exams

The second and associated topic is open-book exams. There is a lot of resistance against open-book tests and exams. They are difficult to carry out. Unlike the conventional way of assessing, there are few examples available for reference.

If we really think about it, open-book exams is real life. Today I am standing here; it is an open-book exam for me. I can bring books and notes but they will not really help me when I do not know the topic that I am talking about. There is nothing wrong if I forget some details and refer to my notes and a few sheets of data. So in

open-book exams the books are not really very helpful although they could help psychologically. By the time one reads them to seek new understanding, two hours would have gone.

So, what is holding us back? If we say that it is unfair because students can refer to chapter five and then copy the first two paragraphs and score an A, then our exam question is set wrongly in the first place because we are testing memory recall. We should assume that if a student is training to be a medical doctor, he must know the fundamentals. An engineer designing a bridge must also know the basics. So we do not test this type of knowledge anymore. But they may not remember 100% of the material, they may remember 90%; so they could refer to a book for the details they may have forgotten. But if they do not understand such knowledge, even if they refer to books, it is useless. We should be testing how they apply the concepts and whether they know which concepts to apply. So by that understanding, every paper can be open-book.

Looking at the statistics, we are hardly reaching 20% in open-book exams. I would argue that postgraduate should be 100%; final year should be 100%; and we can aim at 50-50 for first year. But we are far from these figures. We are nervous because it is new to us. We should set ourselves a five-year target. It will be embarrassing to be caught in a situation where we are said to be a world-class institution, yet there are still major practices that are not world-class. We are preparing our students for the 21st century, but more than 80% of our exams are still closed-book.

I hope I have achieved what I have set out to achieve—to stimulate you to think about the subject of assessment. Please ask questions and continue to challenge me by sending me e-mails. You have to be sporting, knowing that I am still experimenting with new ideas in order to pioneer changes. But I am single-handed unless you give me more input and feedback so I could use these to strengthen my case when I discuss the issue with the deans and the heads. Hopefully together, we will move in the right direction. ■

Question-&-Answer

Q Some of the points you raised actually point to the administrative hurdles which are experienced by teachers. One difficulty is the bell curve. That is okay with essay type questions. You can give marks of 80% or 60% depending on what you expect. But if you have a large number of problem-solving questions, there is no way of predicting what the distribution of marks will be. What you need is a kind of politically flexible relation between marks and grades. In the absence of that, you will find the curve shifted a little to the left or right and then you have to do all kinds of tricks to adjust the marks to reflect the bell curve. Actually, I do not understand why there should be a fixed marks-grade relationship. That can be left to the teacher.

The Bell Curve

A The problem lies with the fact that this is not Cambridge University, U.C. Berkeley or Stanford, where every single teacher can be safely empowered and the student abilities are also more homogeneous. Here, past experience has shown that the moment we relax, there are misuses and problems. Let me give you some indirect evidence to show that people need time to adjust to the flexibility of empowerment. When I was a head of department, I removed the need to record xeroxing. One month later, xeroxing cost went up by 300%. Some staff were actually bringing their children's music books to copy. The problem is that the moment you give that kind of freedom some staff will suffer. While staff who are fair will spend 50% more of their time grading, others will spend more time on consulting. I think our spectrum of quality of teachers is not as uniformed. But I am just painting you the problem. It can be done, but with proper management. The first thing is to train the heads of department on how to deal with difficult staff. All these are systemic problems that we have to address.

Q In many departments, there are very rigid rules about the number of questions and the type of questions and so on. For example, there is a rule that you have to have five questions and only one of them can be obligatory. I do not see the rationale for such rules particularly when each department is different.

Rigid Exam Rules

A I think we have to provide training for heads of department to be able to accept that kind of situation. You see, if the head of department just follows the convention, he is there just to safe-guard the conventional approach. If we are going to train creative people, some of these will become creative lecturers and some will become creative heads. A creative head will realise that there is a group of very good teachers who should be given more freedom; the weaker teachers should be guided by the system. So we should introduce a flexible system where if a teacher is voted best teacher, or there are a few top researchers, you would want to empower them with flexibility. Then these few teachers will be the ones who will spearhead innovation.

They could try out their own types of questions and format. After a year or two, they can then share their experience with others and there will be team-learning. The better, more creative staff who are top researchers and able to integrate teaching with research should be encouraged to go ahead. However, anyone who wants to deviate should be prepared to take a risk. If he fails badly, he should apologise and tell the head of department his new thinking and plans to adjust. Staff should be allowed the chance to fail without penalty so long as they are responsible. Otherwise, who would dare to stick their neck out? Again, the head has to be trained on this.

Q Change here has been very slow. Even if heads were willing to go through training, it would take a great effort to change. And also, when you say we should stand up and say things, change takes place only when implemented from the top. A few years ago, I ended up having a fight with an external examiner and the head of department when I tried to make changes to the exam. It is not easy to change because the people are still there—the same people who have been there a long time.

Change

A I am sending a paper on Learning Organisations to the deans, asking them to share it with their vice-deans and heads and expecting the more dynamic deans and heads to share that paper with the staff. In that paper, I raised the subject of upward appraisal which I started in the EE Department in 1989. I allowed staff to carry out confidential evaluation of their division heads. All these would come to me, and I would then provide feedback to the division heads on what their colleagues thought about them. I went a step further to allow all the forty-five staff to evaluate me as a head of department. It was confidential, and the assessment went to the dean. The dean would summarise the rating and discuss it with me. This is upward appraisal. Till today, not all engineering departments have adopted it. So, it is not easy to change. Without training, incentives and clear guidance, the heads cannot do it. But we want to share with the heads and the deans that those who are daring and ready can go ahead.

Q I was looking at the types of assessment in different universities. There is one assessment system used by the University of Queensland. They call it criterion-referenced assessment system as opposed to what we are using—the norm-referenced system. In the norm-referenced system, we measure students' achievement with respect to their peers'. As a result, if we have very good students with two A's and two B's going through our system, we will eventually drop some of these first class into second class and some of the third class to pass degree. But in the criterion-based system, we first set our criteria and conditions for grades A, B, C, and students will not have to pitch against themselves but will try to measure up to our yardstick. Also, in this criterion-based system, we have to think very deeply about assessment because we have to do a lot of planning before we assess.

But of course, this criterion-based system has drawbacks as well because, it is more difficult to sort out first class, second class and so on. So a lot of thought is needed if we adopt this system for NUS. Secondly, in this system, the setting of conditions for the different grades can be abused. Finally, adopting this system will be difficult for our staff at present because they are so used to the norm-referenced system. But I think this system addresses a lot of our concerns.

Types of Assessment

A Agreed. But this is a major issue from a management perspective. It means the head of department would know where he wants to go. A very strong department would set a target of three or four years, allow half the staff to try it and then they will become role models for the rest. So in five years, the department as a whole would reach the goal. A weak department would allow five or ten percent of its best staff to try it out and then evolve slowly. It may take ten or more years. But it will not be acceptable if nothing much is done after ten years. Again, it boils down to the training of heads and deans.

Q The issue about norm-referencing and criterion-referencing is more urgent because the crush is already being felt. For example, in CA's we set the learning goals and we set the criteria—what would get an A, what would get a B and then according to the criteria, students work very hard. Many of them actually achieve an A. As a result, most of the students get A's and B's. But when it comes to the final exam, we are told to satisfy the bell curve. In that sense, those students deserving A have to be marked down and I find that rather difficult to take. The bell curve may be true in statistics, but it can be done as a description after the event. Yet we say that there should always be a bell curve, therefore we should fit whatever population we have got into that bell curve. In this case, the already deserving A and B students are not properly assessed.

We have to think what sort of information we get from this norm-referenced system. In norm-referencing, we can only tell that out of a batch of students, who are the top thirty, who are the second group from the top. But it does not tell us what the students can do. In criterion-referencing, we can tell what a student has achieved; we know exactly what he can do, or what he cannot do. I think top international companies would be more interested in what the student can do rather than what percentage of NUS he has achieved. Criterion-referencing is the trend in the world today; sticking to the bell curve in norm-referencing seems backward.

Norm-Reference vs Criterion-Reference

A Let me tell you that there are some structural constraints which I find very hard to remove. I am starting to raise this question: In the knowledge economy and the age of life-long learning in the 21st century, the concept of a first class honours may not hold water anymore. So you may find that one day when the British abandons the honours system, NUS, because of tradition, might hang on. The issue about the old titles “lecturer” and “senior lecturer” is the same. The world had shifted to the U.S. style and we were still hanging on to the old system. Our very senior person was only called “Associate Professor” and I was queried by my Japanese counterpart for sending such a ‘junior’ person. It took us a few years of discussion before we switched to the U.S. titles recently. So these are the things we are willing to change but I am afraid that there is still hesitation regarding the honours system. If you really think about it, the Grade Point Average (GPA) scoring is good enough.

Many people who rise up to become bosses or leaders were second upper or second lower students. Our first class honours may not be the boss or leader. So all these are only academic achievements. I would suggest a halfway solution in the first instance. That is, instead of a bell curve, we will have a modified bell. That means because our students in some departments are better, relative to other universities, our middle section should be larger. Then the first class honours should be those who are very outstanding. So now, what is outstanding? Scoring an A is not automatically outstanding; to have produced something publishable or something unique would be. For example, a student who takes the initiative to tell me what he is interested in, and that he has formulated a problem and wants to solve it is outstanding. Actually, the Undergraduate Research Opportunity Programme (UROP) gives you a lot of opportunity to single out the outstanding students who will deserve first class honours. I am actually challenging the heads and deans to see how many on their list of first class honours have had the UROP experience.

In relation to that I am also thinking of the 'O' levels. There are so many straight A students. But how many of them did literature? The ones who take literature are not really scoring straight A's. The straight A students avoid literature. Something is wrong there. They go for the score and not for the challenge and broadening of mind. What is the purpose of education? So all these things have great implications.

Firstly, I would like to say something positive. In my ten years here, I have seen changes for the better. Some have been pushed down from the top and some through discussion and I think both methods are necessary to bring about change. And I do feel happy about the changes that I see. I see more life, vitality, discussion, openness, and I think that is good.

Secondly, I would like to invite your comments on assessment during tutorials. We have talked about wanting to teach students to be creative, to question and be more open but we do most of our assessment based on written work, projects and output. I would like you to comment on the assessment of what goes on in tutorials in terms participation, involvement, creativity and preparation. If this is done in the course of the semester on a periodic basis and fed back to students, then they would have the opportunity to get more involved in the rest of the semester.

Tutorial Assessment

I think there are two considerations. One is that you have to shift your attention from the result to the process. In the past, we have been trying to use a very big feedback loop, that is to look at the end result and then try to feedback to students thereafter. Bright students would know how to adjust and weak students would not. So what you could do is to have intermediate measurement and feedback. This will measure process, the rate they participate in discussion and so on. So you might want to call the student in and say: "Look, over the last six weeks I have observed that you have not contributed to discussions. You should not feel hesitant to raise questions because that is your right, that is how you'll learn."

Secondly, you have to be creative in the way you approach things. Let me give you an example from the Engineering Faculty. Some have said that it is impossible to have interaction in a class of twenty students. But a creative way to solve the problem is to split the class into five groups. Each week, a different group leader is elected. The leaders will come to the front seats and the tutor will only interact intensely with five persons—the rest are largely observers. Their turns will come in subsequent weeks. Then each time the tutor can assess who will speak and contribute more.

So, I support your proposal to measure the interim outcome and deviate from measuring the final result. ■



The Centre for Development of Teaching and Learning (CDTL) provides a wide range of services and facilities to support the teaching, learning and research programmes of the National University of Singapore.

These include teaching and learning support, research on educational development issues, as well as instructional design and development.

contributors

Hang Chang Chieh
and participants of the CDTL
seminar on *Assessment*
on 3 April 1999

advisor

Daphne Pan

editor/asst editor

Christina Low/Verena Tay

layout

Ma Lin Lin

© 1999 *CDTLBrief*

is published by the
Centre for Development of
Teaching and Learning.
Reproduction in whole or in part
of any material in this publication
without the written permission of
CDTL is expressly prohibited.

The views expressed or implied
in *CDTLBrief* do not necessarily
reflect the views of CDTL.

Previous issues of the *CDTLBrief*
are available at our web site:
www.cdtl.nus.edu.sg

Comments, suggestions
and contributions
should be addressed to:

The Editor, *CDTLBrief*
Centre for Development of
Teaching and Learning
Central Library Annexe, Level 6
National University of Singapore
10 Kent Ridge Crescent
Singapore 119260

Tel: (65) 874-4692
Fax: (65) 777-0342
Email: cdtlowc@nus.edu.sg