Instructional Science (2005) 33: 283–309
DOI 10.1007/s11251-005-3002-8

Awareness of the backwash effect of assessment: A phenomenographic study of the views of Hong Kong and Swedish lecturers

DAVID WATKINS[1,*], BO DAHLIN[2] & MATS EKHOLM[2]
[1]*Faculty of Education, University of Hong Kong, Pokfulam Road, Hong Kong;*
[2]*University of Karlstad, Sweden (*Author for correspondence, e-mail: hrfewda@ hku.hk)*

Received: 14 April 2003; in final form: 18 August 2004; accepted: 3 February 2005

Abstract. An effective way to change student learning is to change the form of assessment. This has become known as the backwash effect of assessment. However, academic teachers' ways of understanding the role of assessment in student learning are also important. This paper reports a phenomenographic study of the views of the role of assessment amongst Swedish and Hong Kong university teachers. The results are described in eight categories of conceptions, placed within a two-dimensional outcome space. The two dimensions are (1) the relation between teaching and assessment, and (2) the focus of the backwash effect. The results indicate that two features of the described conceptions are critical for changing teachers' views of the role of assessment. One is the way one understands the significance of ''basic knowledge'' in one's discipline while the other is whether one looks upon the relation between teaching and assessment as being of an internal or external nature. As much research literature points out, to bring about changes in approaches to teaching and learning you must first bring about changes in conceptions of teaching and learning. To utilize assessment to improve student learning, teachers need to be made aware of the need of such improvement and of the role assessment can play in this process. On the basis of research such as that reported in this paper, staff developers could develop workshops or other strategies, which can accomplish this task.

Keywords: assessment, backwash effect, phenomenography, higher education, Hong Kong, Sweden

Introduction

Universities around the world typically state their missions in terms such as developing their students' problem solving and creative skills and encouraging them to become independent, lifelong learners (Biggs, 1996; Nightingale et al., 1996; Watkins, 1998). We also know much about how to bring about such outcomes as there has been a plethora of research, which has done much to improve our knowledge

of the personal and contextual factors that enhance the quality of student learning (Biggs 1987, 1995; Fraser, Walberg, Welch, & Hattie, 1987; Ramsden, 1992).

What this research makes clear is that although such factors are multivariate and interactive in nature, it is "assessment which drives institutional learning" (Biggs, 1996, p. 5) and "the quickest way to change student learning is to change the assessment system" (Elton & Laurillard, 1979, p. 100; see also Crooks, 1988; Frederiksen & Collins, 1989). This phenomenon has become known as the backwash effect of assessment.[1] Armed with this knowledge one might expect that improving tertiary learning outcomes would be easily obtainable. However, as (Boud, 1995, p. 104) laments:

> Despite the good intentions of staff, assessment tasks are set which encourage a narrow, instrumental approach to learning that emphasizes the reproduction of what is presented at the expense of critical thinking, deep understanding and independent activity.

This situation has come about despite clear guidelines about how assessment influences student learning and its outcomes (Crooks, 1988) and numerous books illustrating good practices designed to bring about high quality learning (c.f. Biggs, 1999; Brown & Glasner, 1999; Miller, Imrie, & Cox, 1998; Nightingale et al., 1996). However, as Nightingale (2000, p. 118) comments in her review of the first of these publications: "It has been said before, someone has done it before, and there has been no paradigm shift."

What is the problem? Biggs (1996) argued that the fierce resistance to changing the tertiary assessment system in the UK at that time was due to three main factors: there was the pragmatic argument that as formal assessment is used to make decisions about a person's life it should be as simple to interpret as possible, preferably a single number easily comparable to those of other students; also it could be argued that standardized forms of assessment, such as multiple choice items, are quick, cheap, and reliable estimators of performance which affirm a university's ability to carry out the important social role of accrediting individuals; and finally Biggs considered that views and practices of assessment reflect a person's beliefs about what it means to have learned something.

It is the latter factor that we feel has been the neglected element here and it is research in this area that is the focus of this paper. Recent studies have strongly indicated that how a teacher thinks about teaching influences both how they teach and how their students learn and the learning outcomes they achieve (Ho, 2001; Prosser &

Trigwell, 1999; Tang, 2001). Failure to consider teachers' views has been identified as the missing link in many otherwise well conceived attempts to improve teaching and learning (Watkins & Biggs, 2001).

Conceptions of achievement

Cole (1990) identified two major conceptions of achievement: a quantitative view which is characterized by the accumulation of discrete basic skills and facts and a qualitative view which is characterized by the mastering of higher order skills and advanced knowledge. Both views have their place and the different sorts of learning to be achieved are best assessed by differing assessment methods (Biggs, 1996; Cole, 1990) but it is the latter view that more closely reflects the missions of today's universities described above. Unfortunately, all too commonly our university teachers, often encouraged by their administrators, utilize assessment methods more appropriate for the quantitative view (Biggs, 1996; Crooks, 1998). Research on the backwash effect tells us that use of quantitatively oriented assessment methods influences students to adopt superficial surface approaches to learning rather than the deep meaning-oriented approaches necessary to achieve high quality learning outcomes (Tang, 1996; Thomas & Bain, 1984; Watkins, 1983).

Given the espoused aims of universities and of most individual courses to achieve such higher order outcomes, the abundant examples of good assessment practices and the efforts of staff development units, why do lecturers still use such inappropriate assessment methods? It is our belief that much of the problem lies in the beliefs of lecturers about learning and assessment and in particular their interactive relationship. Studies have indicated that university teachers will only change their approaches to teaching if they are convinced of the need for change and the efficacy of the new approach (Ho, 2001; Tang, 2001; Prosser & Trigwell, 1999). It is thus surprising, given the acknowledged importance of the role of assessment in improving learning, that there has been little research into lecturers' views of assessment in general (Barrie, Brew, & McCulloch, 1999) let alone the backwash effect in particular, or whether such lecturers can or even try to harness this effect to improve student learning. A recent exception concerning general views of assessment is the study by Samuelowicz and Bain (2002), which reports academic teachers' orientations to assessment practice. The results of Samuelowicz and Bain's research did not come to our knowledge until we had virtually

completed our own work. However, not surprisingly, there is some overlapping of their results with ours (see below).

Aims of research

The purpose of this study was to provide a conceptual analysis of the various views university teachers have of the actual – as opposed to the ideal role – of assessment in student learning. That is, how do they understand the role of assessment *as it is actually carried out* in their present teaching practice. We were particularly interested in exploring the variation of conceptions of the backwash effect of assessment. We wanted to see whether academics are aware of this effect and, if so, how they conceptualize it.

In order to maximize the likelihood of obtaining distinctive variations we interviewed university teachers in two rather different contextual and cultural locations: Hong Kong and Sweden. (The questions asked in the interviews are given in Appendix A.) The principal research approach we used was phenomenography (Marton, 1981, 1986), which has been a widely used qualitative method to investigate variations in how a group of people experience various educational phenomena (see Marton & Booth, 1997, for a recent overview of this research).

Assessment practices in Hong Kong and Swedish universities

Before describing the results in detail we first give a brief overview of assessment practices in the Hong Kong and Swedish contexts. Higher education in Hong Kong has expanded rapidly. While 10 years ago there were two universities there are now nine and the percentage of the age cohort studying degree courses has over that period expanded from 5% to 18% (not including another 5% who study overseas). Today at universities in Hong Kong undergraduate courses are mainly taught in one semester modules with course grades awarded at the end of the module on the basis of a formal examination or more usually a combination of exam, assignment(s) and tutorial/practical performance (see Watkins, 1998). Hong Kong universities typically report assessment of undergraduate students on a five-point letter-grade scale and guidelines of the percentage of students expected in each category are often provided (Watkins, 1998). On graduation, students are awarded degrees of honours based on the British system. The assessment system at each university is monitored by external examiners many of whom come from outside Hong Kong. For in-depth accounts of research probing the effects of contextual and

cultural influences on teaching and learning in Hong Kong the reader is referred to Watkins and Biggs (1996, 2001).

In Sweden also the number of undergraduate students has grown continuously during the latter decades. Three former university colleges received full university status in 1999. Undergraduate modules are usually divided into smaller course units, which are assessed separately, at the end of the 5 or 10 weeks. The teacher(s) responsible for the course are generally free to decide which forms of assessment to use, even though there may be informal pressure from the institutional practice that has developed over the years. Grading is comparatively simple. In most fields there are three grades: Fail, Pass and Pass with Honours. In Technical and Engineering subjects, a 1–5 scale is used, in order to be compatible with international grading systems.

Method

The sample's two subgroups

Our sample consisted of two subgroups, one from Hong Kong and one from Sweden. The Hong Kong subgroup consisted of 26 participants who volunteered to be interviewed after having answered a survey questionnaire sent out to 200 randomly chosen academic teachers in Hong Kong (Watkins, 1998). The Swedish subgroup consisted of 20 academics from a middle-sized Swedish university who volunteered to be interviewed. Both sets of interviewees held the position of Assistant Lecturer or higher, and represented a variety of disciplines from the Human, Social and Natural Sciences. Their teaching experience also ranged from 1 to 28 years in both samples.

Naturally, neither subgroup can be considered representative of the general population of academic teachers in Hong Kong or Sweden. Choosing to volunteer for an interview about the significance of forms of assessment presumably means that one has given the subject some thought, and finds it interesting to talk about. Since the two subgroups were not representative, no definite conclusions about the similarities and differences between academic teachers in Hong Kong and Sweden can be drawn.

The interviews

The interviews were conducted by two Research Assistants (RAs), one in Hong Kong and one in Sweden. They were well informed about the style and purpose of phenomenographic interviewing,

especially about the necessity of probing, asking for clarifications, etc. As Prosser (1994) remarks, this is the most important aspect of using RAs in phenomenographic studies.

The interviews were semi-structured. The pre-established questions dealt with various aspects of assessment and student learning. The participants were asked to choose and reflect on one particular course. They were also asked whether this course was typical of their teaching and to explain any differences and possible influencing factors. Apart from the *direct* question of whether and how forms of assessment influenced students' learning, other, indirect questions were also put. These questions included the objectives of the chosen course; the forms of instruction and assessment used; and how examinations were graded. A question about whether "alternative" forms of assessment had been tried was also put.

The interviews were audiotaped and transcribed verbatim by the respective RA. The Hong Kong interviews were conducted in English, the Swedish ones in Swedish. The totality of the transcribed interviews consisted of approximately 600 pages. In the results section, quotes from the Swedish interviews are indexed with an "S," the Hong Kong interviews with "HK".

Analysis

The aim of phenomenographic interview analysis is to construct a range of conceptions held by the group of participants at the time of the interviews. It is not assumed that the interviewees espouse the same conceptions at different times or in different contexts. Nevertheless, *the variation* of conceptions obtained from the analysis is seen as generalisable across contexts.

Naturally, the answer to the *direct* question of whether and how forms of assessment influenced students' way of learning was of central importance for the analysis. However, the whole interview was considered for analysis purposes. The answer to the direct question on the one hand, and answers to other questions on the other, were seen in relation to each other and as mutually illuminating.

This research adapted phenomenographic methods for cross-cultural research. In the first stage of the analysis the first and second authors independently analysed the Hong Kong interviews. They then worked together to produce a mutually agreed outcome space. The second and third authors then collaborated to analyse the Swedish data and an outcome space appropriate for both groups was developed by all authors. In analyzing the Swedish interviews it was found

necessary to modify the initial outcome space to capture the variations found in the Swedish data. Hence, the steps along the two dimensions of the outcome space were somewhat modified. (The major change was the adding of a middle position along the dimension "teaching/assessment relation"). The Hong Kong interviews were then reconsidered in light of these changes. The third author agreed that the structure as a whole was an adequate reflection of the Swedish interview material. However, it was sometimes difficult to agree on *one definite* position of a particular interview in the outcome structure. There were two reasons behind this: (1) the participants often expressed more than one conception; and (2) there was sometimes a lack of relevant and decisive interview statements. In spite of these difficulties, we feel that the outcome space as a whole is a reasonable and reliable description of both the inter- and intra-individual variation of conceptions that exists in this field.

Many phenomenographic studies are the result of the kind of teamwork described above, enhancing their validity. The person primarily responsible for the analysis proposes a system of categories, which is tested and probed by the other members of the group. As Bowden (1994) explains, this procedure "provides the discipline that ensures the full evidence of the transcripts is extracted and used to determine the final categories of description" (p. 51).

Results

The outcome space

Two dimensions of the conceptions expressed emerged as clearly present in the interview data as a whole. These were the focus of the awareness of the backwash effect on the one hand, and the relation between teaching and assessment on the other (see Figure 1 below). It also became clear that the interviews could be placed at definite "steps" along these dimensions. In this way, a two-dimensional outcome space was established, which constituted a structure in which all interviews were roughly positioned.

Figure 1 depicts graphically the two-dimensional outcome space for conceptions of the actual role of assessment in learning, as well as where in this space the different categories are located, derived from analysis of our data. We have not been able to find a suitable name or label for each category. Instead, the characteristics of the conceptions in each category are summarily expressed by that category's position in the outcome space, and described in more detail below.

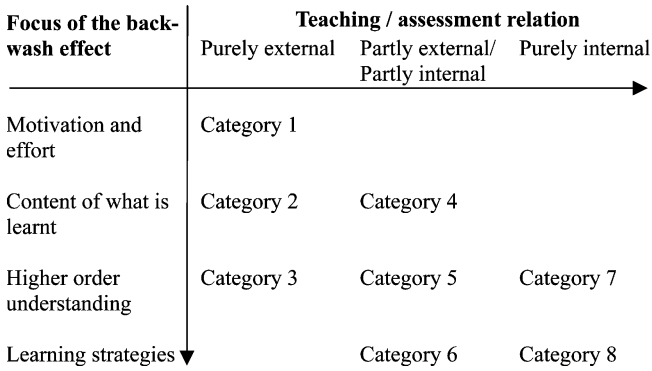

Figure 1. The two-dimensional outcome space for conceptions of the actual role of assessment in student learning (arrows indicate progressions towards more inclusive conceptions).

The dimension "focus of the backwash effect" displays a progression from "Motivation and effort" to "Learning strategies", and the dimension "teaching/assessment relation" a progression from "External" to "Internal". The polarity of external versus a purely internal teaching/assessment-relation is (we believe) clear-cut from the data and is also consistent with theory (Bradley, 1925).[2] In an external relation between teaching and assessment, the process of assessment is clearly separated from that of teaching and learning. Assessment is seen as an "after the fact" measuring of what *has been* learnt. In an internal relationship, processes of teaching and assessment are only analytically distinguished, in practice they are understood as overlapping and interacting. Students learn not only from teaching but also from and in assessments, and teachers may assess their students even in the processes of teaching.

In reality, however, what participants said in the interviews was not always easily categorized into these two types of relationship. For instance, in some Swedish interviews one part of the assessment could be seen as directed towards measuring the acquisition of what was considered to be "basic knowledge" learnt more or less by heart, whereas the purpose of another part could be to induce a more "reflective" learning experience. The latter often took the form of more or less extensive project work, and was more common at higher course levels. Therefore, a middle position was constructed on the teaching/assessment dimension, in which the relation was seen as partly external, partly internal. The necessity of introducing this category may have to do with the rather common notion of "basic knowledge," and the role that memorizing and reproducing was

perceived to play in the acquisition of such knowledge. In contrast to "basic knowledge", which was considered as relatively superficial, real understanding tended to be associated with deeper and more reflective learning processes. We return to this issue in the discussion.

Category 1

In this category, teaching and assessment were seen as externally related. Awareness of the backwash effect was focused on students' motivation and effort, and assessments were often regarded as technical devices for measuring academic achievement. A lecturer in Organic Chemistry recounted how the introduction of brief but more frequent quizzes increased student performance:

> One point is that if we have more quizzes, I can see that the students spend more time during the semester year to do the revisions. [...] At the very last, the final exam, we found the absolute score is higher. (Organic Chemistry, 4HK)

This teacher also regarded it as "the spirit of examination to try to distinguish students according to their ability." The procedure of discriminating between high and low achievers sometimes seemed to be a very technical one, particularly among lecturers in Hong Kong, where the infamous "Bell curve" was often used more or less rigidly (see Watkins, 1998, for details). Assessment, from this perspective, becomes a technique for the presumably fair and objective measurement of student achievement, rather than an intrinsic part of the course curriculum.

Category 2

Here also the teaching/assessment relation was seen as external, but awareness of backwash effect was focused on the content of what is learnt. By "content" we refer to both theoretical knowledge as well as cognitive or practical skills. In subjects like Mathematics or Computer Programming, problem solving and the corresponding skills are essential parts of the curriculum and the course objectives. Such skills have mainly a technical, algorithmic nature, at least at the basic course levels. They can be acquired without any deeper cognitive changes, affecting personal thinking and attitudes. We therefore make a distinction between such skills and the "abilities" which are focused on in the next category.

As an illustration of this category, we quote a HK lecturer who talked about the backwash effect in terms of cue-conscious students:

> The questions that we give them in examination are limited to certain areas in the lectures and in some cases the topics that we choose for exam. When we give a lecture we know that certain parts are very easy to set questions on, others difficult to set questions. So if the students are careful enough they understand that, so that part is more popular. (Computer Studies, 17HK)

For this lecturer, class attendance was not obligatory. This in itself indicated that the relation between teaching and assessment was seen as external. The only relation between assessment and teaching seemed to be that examination results of one year influenced the teaching of next year's students.

Examples of this category could also be found in the Swedish interviews. One lecturer in Mechanical Engineering (7S) talked about how assessment was used to check that students had acquired certain knowledge and skills, and his awareness of the backwash effect was primarily focused on students asking for previous assessment forms in order to familiarize themselves with the kind of questions that would come up in the exam.

Category 3

In this category the awareness of backwash effects was focused on the high order understandings that the students have developed during the course(s). Now we are referring to not purely technical skills (as in category 2), but rather to more internalized, personal understandings of particular phenomena. However, assessment was seen mainly as a measuring device applied after the course.

One HK lecturer in English Literature was concerned that students learned to see the literature of a particular language group "as the culture and the thinking of that group of people in the world, incumbent in words." About assessment this teacher said:

> I am never content simply to set a discursive essay type question. I want that kind of discussion but I also want close reading and specify text and want to know if a student can handle that, handle the vocabulary, handle what it actually means. Because it is quite easy for students to read books of criticism, get general notions about literature from them and spout these in essay without

> actually knowing much literature, without knowing really how to read literary texts very closely .... And of course I'm interested in also finding out who are the "A" students and who are the "B" students, who can do better and who are the "C's" and who are so poor at English that they have to, they really should have remedial English. (English Literature, 6HK)

In this quote, assessment seems to be mainly a means for the teacher to check whether the students have developed a certain ability to handle literary texts, and to diagnose which students need extra help.

A Swedish assistant lecturer in Nursing Science said there were never any disagreements among her colleagues about the form the assessments should take, because:

> ... often one has some thought about what it is that one conceives that the student should reach during this course, so to speak, in skills and in understanding... and as a rule we always agree [...] (Nursing Science, 14S)

Here also, assessment seems to be reduced to testing that the objectives of the course, in terms of acquired abilities, have been achieved.

Category 4

Here backwash effects were again seen as directed towards the content of what is learnt (concepts, facts or skills). However, the relation between teaching and assessment appeared to be conceived simultaneously as external and internal. The internal view was expressed as an awareness of the possibility of using assessment as a learning experience in its own right. As an illustration, the HK lecturer quoted in the previous category at one (and only one) point in the interview said:

> From my experience I can remember some of the questions I actually had on exams, certain exams when I was an undergraduate. That means that you are concentrating so fiercely in the exam that you tend to remember things about an exam, where you might forget other things from more casual attention in classes. So, I try to make the exam part of the learning experience. (English Literature, 6HK)

This lecturer is trying to make the exam a learning experience, but "learning" in this case seems to equal "remembering" some particular

content or aspect of the course. It is not about developing ability in the sense described in category 3.

A Swedish assistant lecturer in Chemistry expressed a similar view:

> A: It's probably a lot of memorized knowledge [that is assessed] since it's on a basic level there is a lot of memorized knowledge, especially in Chemistry 1. But understanding is there too.
> Q: In what way?
> A: That they should put it into context .... one can for instance ask them to solve something ... now that is a very open question, where they themselves have to account for what kind of material is needed, how you do the calculations, that you must do calculations... and here you can... also include for instance how accurately they can or want to work... (Chemistry, 11S)

Even though this teacher talked about "understanding," the focus seemed primarily to be on the *content* of this understanding. That is, assessment was used to check students' understanding of the course content in terms of Chemistry knowledge and laboratory skills. Still, in the area of skills, the described assessment form seems to have the potential for becoming also a learning experience.

Category 5

This is the major category in terms of the number of interviews that contained expressions of it. As in category 4, the relation between teaching and assessment was seen as partly external, partly internal. However, awareness of the backwash effect included the abilities that students were expected to develop.

As an illustration we take some quotes from an HK lecturer in Mechanical Engineering:

> The questions are primarily essay-based questions... . the intention is to make sure that the student understands the technology of the systems, in terms of how they are applied in industrial applications... . I do give students an indication of how I am going to structure the examination, because the examination is not just a case of regurgitating whatever notes I give them... . It is a matter of applying them in maybe some new scenario or recombining them in some alternative way to prove that they can apply that knowledge from different directions. [...] And I have told all the students that examination is a means of testing their understanding

> of the subject rather than testing their ability to reproduce notes. (Mechanical Engineering, HK)

Here, assessment seems intentionally designed to contribute to students' development of the capacity to understand and apply their knowledge in "new scenarios." In other words, assessment's backwash effect was used to influence the abilities that students were expected to develop. The form of assessment also had the potential of being a learning experience in itself. On the other hand, this lecturer also emphasized that an examination is "a means of testing", and in the interview he repeatedly said that he "told" the students that mere "regurgitating of notes" was not sufficient. This merely *telling* the students what is needed seems to implicate that there is still a lack in the integration between teaching and forms of instruction on the one hand, and assessment forms on the other. This, together with the emphasis on "testing," is our basis for saying that in this interview, the teaching/assessment relation was still seen as partly external in character.

Further examples of this category are the following quotes from a Swedish lecturer who said the most important thing for an art historian "is precisely to learn to see and also to describe pictures and to be able to interpret them" (13S). Therefore, at higher course levels, assessment consisted in writing essays with such descriptions and interpretations. However, at the basic course level, examinations had a different purpose:

> ... in order for them to get any base at all to stand on I think it is necessary to memorize lots of factual knowledge because you have to test that in some way. (Art History, 13S)

Category 6

The difference between this and the previous category is that students' learning strategies were included in the backwash effect. Also the grounds of the "partly external/partly internal" categorization were of different kinds. One was the view that higher scores were reserved for those students who showed understanding and abilities to apply. This was made explicit to the students, in order to stimulate deep strategies:

> My point to them [the students] was if you simply have a good memory, if you seem to reproduce the course notes, things that are taken strictly from books – that is not your own original thinking. You haven't got any value added to what is already there. Then you can't expect the top grades. (Public & Social Administration, 20HK)

The second basis for the ''partly external/partly internal'' categorization was that there were cases where some forms of assessment, for instance short answer factual questions, were seen as devices for merely measuring reproduction of the discipline's ''basic knowledge.'' Naturally, such examinations could not be regarded as particularly meaningful learning experiences. However, other assessment forms were seen as designed for stimulating students' development of deeper learning strategies, and as having the potential of being learning experiences. One HK lecturer in Business illustrates this view when he talks about memorizing as the basis for more creative applications:

> I don't want the students to memorize everything, then it will affect their creativity. So that is why I am talking about I want a mix. But I think maybe, you know, in the primary, secondary education the students should memories more. [...] So when they come into this college they already have a very solid foundation of basic knowledge. Then they can be exposed to creativity. And they can sort of forget about all the memorizing things and try to talk about application, trying to relate it to real life. So I think I try somehow to make sure that the students memorize some things. But in terms of assessment I must confess that I don't encourage that sort of thing, because, take for example the case study, it is an open book. The student can prepare everything and bring it into the examination hall because we try to imitate real life. [...] But of course in the objective test [this was a MCQ test], ... the student should remember something. (Business, 1HK)

It seems clear from these quotes that this teacher envisages two kinds of assessment: one which simply measures already memorized ''basic knowledge,'' and another which is designed as real life application in order to stimulate creative thinking. This interview subject was also very keen on trying to change the students' learning strategies. He said he saw assessment as ''a way of helping the student to learn,'' but that HK students most often ''assume they come here to be taught.'' In response to the question how he tried to change the students' approach, he said:

> I try to make use of this sort of project [as a form of assessment]. I try to confront and I try to provoke. ... with a hundred students it is difficult to get students to respond. So we try to develop this sort of small class seminar project, what so ever. Try to confront the student, try to challenge. But most of the time you know you

> would be very lucky if you got 0.5% of the students to challenge you or come up. (Business, 1HK)

Even though he did not seem to be very successful, this teacher seems to do much more than merely *telling* the students what is expected of them: he is trying to make them *do* it in the very process of instruction, not only at the moment of examination.

A similar case was found among the Swedish interviews. This was a lecturer in Political Science. At the basic level courses he gave an examination with what he called ''regurgitate-questions,'' because ''if you are dealing with politics and administration you have to have the structure, you have to know about that [...] and then we have done that [examination] with a lot of a-b-c-questions so that it is easy to correct.'' However, he also used a kind of real life application task, which was discussed in groups:

> ... they get a material that shows the economy of a commune... They get a task that they shall on the basis of this commune's economy [...] they shall deliver a suggestion about how to increase the resources for child-care in this commune. (Political Science, 12S)

This teacher was clear about the backwash effect on learning strategies. The ''regurgitate-questions'' were intentionally designed to get the students to cram a lot of factual knowledge, whereas the group assignments were used in order to stimulate understanding, reflection and application both during course work and in the examination.

Yet another basis for the ''partly internal/partly external'' categorization consists of interview cases where part of the assessment was designed to get away from mere ''memorizing and reproducing,'' but where the more important function of assessment still seemed to be the ''fair and objective testing'' of students' performance. These interviews were characterized by much talk about the grading procedures, and a lot of effort seemed to go into them.

Category 7

This is the first category of conceptions in which the teaching/assessment relation was seen as completely internal. In this view, the purely technical measurement of acquired knowledge, skill or ability was seen as irrelevant for genuine learning. Teaching and assessment were both geared towards developing understanding, even at the basic course levels. In the present category, ''understanding'' refers

to the ability to relate and apply knowledge as an *outcome* of the learning process (not as a learning strategy). There is virtually no talk about the strategies students develop or use during course work.

This category is illustrated by some quotes from a Swedish lecturer in Literature. Similarly to the lecturer in Art History quoted in category 5, she emphasized the importance of being able to reflect and interpret works of Literature, but this was the objective already of the basic course level. In course work as well as in examinations, students were engaged in such activity. Reproducing basic factual knowledge played a very subordinate and insignificant role in her assessments:

> When it comes to Literature there is never a final answer but there are as I said different persons' suggestions for interpretation that you can put up against each other and that is why it is so important that they discuss these texts... and not just memorize from some handbook in Literature. (Literature, 4S)

The following quote was taken as an indication of her view of the relation between teaching and assessment as internal:

> Well, I don't see... an examination as... a goal in itself, but every lesson is really a discussion, since we build instruction very much on group discussions and then make summaries of them. (Literature, 4S)

About the backwash effect she said:

> Q: Do you think the form of assessment influences the students' way of learning? [5–6 seconds silence]
> A: Well, of course, I suppose it does... I mean they usually get hold of old examinations, we know that. And then at least they get an idea about the type of question we give and this thing that we say that it must be written in the form of an essay. That is, that they must follow up their thoughts, that there should be no mere listing of points but there must be a continuous discourse... and I think that makes them put an effort into thinking, even when they read an ordinary book. (Literature, 4S)

Admittedly, this teacher said the assessment form had an influence on students' thinking efforts, but this seems too vague and commonsensical to be taken as an expression of a clear awareness of the backwash effect on learning strategies (cf. also the long pause before answering).

Category 8

As in the previous category, here also the relation between teaching and assessment was seen as completely internal. However, awareness of the backwash effect included specific references to students' learning strategies, not only to the results of learning.

One HK lecturer in Computer Sciences said he let the students bring their textbooks, lecture notes and calculators into the examinations, because

> ... it forces me to set proper questions. I can't set questions like "give me the definition of... " because they can look it up in the book. (Computer Science, 8HK)

He also said that memorizing had no important role in his courses "because of the way in which I do the assessment." In his teaching he used a lot of continuous assessment in order to influence students' learning processes. At one point he had tried problem-based learning, but found that this approach was too "overwhelming" for the students. Still, he tried to use some of the ideas of problem-based learning in a more "structured and teacher directed" way.

Another HK lecturer spoke very frankly about the "children" who wanted answers to memories and who "would never believe me when I told them what the exam was going to be."

> This is a third year final elective International Business, there are no answers! There's judgment, there's perspective, there's risk. It's a question-oriented experience, not answers. [...] I am consistent that my assessment method is similar to my style and to the material that I teach. I teach what is my style. I could not teach factual answer oriented material. [...] I don't look for their answers [in assessment]. I look for how they develop whatever they are saying. Because there is no answer. (Business & Management, 2HK)

It seems clear from these quotes that this teacher used both teaching and assessment in order to influence students' learning strategies. Merely measuring acquired knowledge was not really important to him.

The view that "there are no right answers" recurred in other interviews. In one, which was also placed in this category, the participant noted that the absence of right answers made students very anxious and uncertain. But he reassured them that this approach was "the way to learn. The alternative is rote learning." (Economics, 24HK) In order to change students' ideas about Economics he made them apply

economic theories and concepts to all kinds of social phenomena, such as religion or the institution of marriage. These tasks were part of continuous assessment.

Finally, let us quote a Swedish lecturer in Sociology, who was acutely aware of the backwash effect on student learning strategies:

> A: But I believe that this form of home assignment [writing a paper] quickens this learning. They notice that it doesn't work to just go home and copy the book, ... The teacher does not accept that you merely recount what is in the book, you have to take your position, you yourself have to say what you think is important and why and...
> Q: So your view is that the way you assess can influence how the students, what study strategy they will... ?
> A: Yes, sure, oh yes, absolutely. ... it is a very important determining factor. ... I don't think there is anything wrong with the students. They adapt rationally to the system they have... that is why it is our task to design the examination so that they get the learning style that we want them to have. (Sociology, 15S)

This teacher said he always tried to arrange his teaching from his awareness of "what goes on in the students' heads." From the quote above it seems evident that the mere measuring of already acquired knowledge had no special significance for him, and that he saw an examination in itself as a learning experience for the students. He also questioned the significance of reproducing definitions of "basic concepts":

> Traditionally, Sociology has basic concepts like "social norms," "social roles," "institutions," of course you can memorize those definitions verbatim but by doing that you will never understand what.... If you take a basic concept like "social role," you need an understanding of what question has been put so that one has arrived at this concept. [...] Basic concepts are in some way the results of human knowledge seeking activity. (Sociology, 15S)

The questioning of the notion of "basic concepts" which have to be acquired before one can move on to understanding and deeper learning strategies seems to be an essential characteristic of this category, as compared to category 6. There, the notion of basic knowledge seemed to justify an "after the fact" measuring of how much of this knowledge the students had acquired, even though one was aware that such tests encouraged surface strategies of memorization and reproduction.

Issues of context

Conceptions are often contextual in character. Consider the following quote, which raises the question of the contextual character of the conceptions described:

> We explain very clearly the structure of the course, what will happen and that first you will go into this and so on... and they become very conscious about their learning process, whereas in other courses I have which are not so well structured they are perhaps not as much aware about this [...] it isn't self-evident that it always happens. (Assistant Lecturer, Art & Media, 8S)

This participant talked almost exclusively in terms of category 8. However, the quote above indicates that had she chosen to talk about another course, the conception expressed could have been different.

The disciplinary context

Previous studies have indicated that faculty members from different disciplines often differ in their views of for instance course objectives and standards for grading student progress (Eljamal et al., 1999). A dependence of assessment forms on disciplinary context has also been noted (Glasner, 1999). A plausible assumption would be that this context also influences what functions assessment is seen to have in student learning. In some disciplines, for instance those belonging to Medicine or Engineering, serious consequences would arise if students did not acquire the knowledge and skills necessary for the work they are expected to do. In such cases, teachers may feel obliged to test, check and control that students have achieved the course objectives. This may influence how the teaching/assessment relation is understood (at least partly), making it seem purely external, and students' learning strategies may fall out of focus. In one interview, placed in category 5, a HK Engineering lecturer said about the final examination:

> The questions are primarily essay-based ... the intention is *to make sure that* the student understands the technology in the systems, in terms of how they are applied industrially. (Mechanical Engineering, 16HK; our italics)

In contrast, in the Human and Social Sciences teachers often seem to be aware of the importance not only of the application of general theories to concrete cases, but also of theoretical and conceptual analysis,

self-reflection and interpretation. In such a context it is presumably easier to realize the connection between the development of these abilities on the one hand, and course work and assessment forms on the other. Thereby, students' learning strategies also tend to come more into focus.

These two assumptions about the significance of the disciplinary area for the views expressed were partly confirmed by the interviews. There were more examples of categories 7 and 8 among teachers in Human or Social Science, whereas categories 1–3 appeared more frequently among teachers in "hard core" scientific or technological subjects.

On the other hand, cases pointing in the opposite directions were not completely absent. One HK lecturer in History (5HK) said he used to "mark off heavily" any mistakes in bibliographic entries, or even spelling, indicating that such basic skills were not to be compromised with, because according to his view History is a "public thing," and as an historian one is "working publicly." Another case was the HK lecturer in Computer Science quoted in category 8 above. Although Computer Science is a hard-core subject with many algorithmic skills, which simply must be tested, this teacher had developed a complex conception of the role of assessment in student learning.

Discussion

Dimensions of the backwash effect

Of the two dimensions identified in our analysis, 'focus' is the one that has traditionally been addressed in the literature. It was also the one of which teachers in our research were more aware. However, we would argue that the 'teaching/assessment relation' is crucial for improving learning outcomes. This has recently been recognised in the literature in the emphasis placed on *aligning* the assessment system with the course and the instructional approach (Biggs, 1999). If the relation is understood as external, the importance of this alignment more easily falls out of awareness. That is why we believe this dimension of variation to be significant.

Three basic types of conceptions

From a global point of view, the two-dimensional outcome space described above could be summarized into three general types of conceptions of the backwash effect. The first type would consist of categories 1–3. The common feature of these conceptions is that there is virtually no talk about students' approaches to learning, or learning

strategies. The focus is on the content of what is learnt, or the abilities students are expected to develop as a *result* of the course. Furthermore, assessment is viewed as something "very separate from the teaching and learning process, something to think about once the curriculum has been devised and plans for delivery finalized," as Brown (1999, p. 3) describes what we call the external relation between teaching and assessment.

The second type of conception consists of categories 4–6. Here, teachers seem to have become aware of the teaching and learning *process* of the course, and sometimes even the strategies students use in order to learn. However, they also believe that there is a sort of "basic knowledge" which students simply have to acquire or internalize, *before* more sophisticated learning strategies can be usefully applied.

The third type of conceptions consists of the remaining categories 7 and 8. Here, the notion of basic knowledge to be accumulated has no, or only a very subordinate, role to play. On the contrary, understanding, reflecting, interpreting, analyzing, and relating – all such deeper strategies have an overriding importance already from the start of the basic courses. Another important feature of these conceptions is that assessment is seen as an integral part of the teaching process.

Such a general characterization of our results partly overlaps with the findings of Samuelowicz and Bain (2002). Samuelowicz and Bain analyzed their interviews into six belief dimensions (an indication of the complexity of the views expressed in this field), one of which was the role of assessment in teaching and learning. This dimension varied from the view that assessment was a "means of making the students study" to the view that it was a "means of guiding the students' learning". The latter view correlated with seeing assessment as "an integral part of teaching and learning" (ibid., pp. 183–185).

The critical distinction: "basic knowledge" versus reflective understanding

The general characterization of the findings of our study made above suggests that there are two elements, the awareness of which would be critical for a *development* of teachers' conceptions of assessment: students' learning strategies and the notion of "basic knowledge."

If teachers develop a clear awareness of the kind of learning strategies that various forms of instruction and assessment tend to evoke in their students, they would probably also develop a deeper and wider understanding of the role of assessment in student learning.

Furthermore, as long as teachers have the view that there is a ''basic knowledge'' to be acquired *before* understanding, relating and applying can take place, they will have difficulties in seeing an internal relation between teaching and assessment. This is not to claim that low-level learning is *logically* incompatible with an internal relationship (and vice versa). It means only that such a view is rarely developed among teachers because the perception of ''basic knowledge'' as memorized (more or less fragmentary) facts makes it difficult to do so. In our view, ''basic knowledge'' should be seen as *basic concepts*, which should be *understood* by the students as ''results of human knowledge seeking activity'' (cf. the Swedish lecturer quoted above). From this point of view, it will be easier to develop a conception of assessment, as well as an assessment practice, based on an internal relation between teaching and assessment.

Other studies also point to the potential significance of the common notion of ''basic knowledge.'' Eljamal et al. (1999), in a study of US higher education teachers' goal statements, found that ''intellectual development may be a low priority for instructors teaching introductory classes'' and that ''in introductory classes faculty focus on knowledge acquisition rather than on the relationships and connections with other fields'' (p. 21). These authors considered seeing ''relationships and connections'' as the core of intellectual development. The results of their study indicated that many US teachers in higher education do not see this kind of deeper learning as part of *basic* courses in their disciplines. It is saved for more advanced classes. The same point of view seems to be expressed by many of our interview subjects (see categories 4–6).

The distinction between ''basic knowledge'' and higher order understanding is also related to what Cole (1990) called two conceptions of educational achievement: ''basic skills and facts'' versus ''higher order skills and advanced knowledge.'' The former goes together with a focus on the *outcomes* of learning, and the technical procedures for measuring these. The latter is relatively more focused on the *processes* of teaching and learning. Cole argued that these two conceptions are not sufficiently integrated, neither in educational research nor in teaching practice. In her own words, ''the fact that we have two such different conceptions playing dominant and conflicting roles illustrates a major limitation that our field must overcome'' (p. 5). Overcoming this limitation may also play a significant role in overcoming the somewhat ''double-tongued'' views of categories 4–6 in our results.

Another parallel between conceptions uncovered in other studies and the ones described here, concerns teachers' conceptions of teach-

ing. A number of studies have found that teaching conceptions ranged from those which strongly advocated transmission of information from the teacher to the students, to those which emphasized facilitation of understanding in students (Biggs & Watkins, 2000; Kember, 1997; Prosser & Trigwell, 1999). This basic polarity seems to reflect the polarity between external and internal views of the teaching/assessment relation. That is, the notion of teaching as "transmission of information" is easily assimilated to an external view of this relation; whereas the view of teaching as "facilitating understanding" seems more homologous with seeing assessment as an integral part of teaching and learning.

However, it is also possible, even probable, that teachers adopt one conception or the other, depending on circumstances. We do not claim that teachers' conceptions arise entirely independently of context. The constraints of large classes may for instance "force" the teacher to act according to the information-transmission model, and apply an "after the fact", purely measuring form of assessment. This is related to the distinction that Samuelowicz and Bain (1992) make between "ideal" and "working" conceptions of teaching:

> It seems [...] that the aims of teaching expressed by academic teachers coincide with the "ideal" conception of teaching whereas their teaching practices, *including assessment*, reflect their working conception of teaching. (ibid., p. 110; our italics)

Our interview data suggest the same for conceptions of the role of assessment. In an interview situation, many participants would probably – for various reasons – rather talk about their "ideal" conception than their "working" one. Research may focus more on the factors that prevent teachers from acting according to their ideal conceptions, as suggested by Samuelowicz and Bain (1992). In our interviews, the participants mentioned a number of constraints, which were perceived as preventing them from realizing their ideal forms of assessment. The constraints ranged from lack of resources (time, equipment, physical localities) over student characteristics (cultural influences, established habits and expectations) to institutional policies. Most of these constraints were also related to forms of teaching and instruction.

Conclusion

Our analysis of the views of Hong Kong and Swedish lecturers makes clear that there is a range of possible views of the backwash effect,

from superficial ones which focus on motivating students to work harder to sophisticated ones which perceive an intimate, internal relationship between assessment and student learning.

Prominent educators have warned that a major reason advances in assessment have too often failed to improve learning is that students all too commonly do not perceive what types of outcomes the assessment is intended to require and reward (Biggs, 1996; Ramsden, 1992). However, before the student can perceive assessment in this way the teacher must have the intention and the capability of setting assessment tasks which encourage higher order learning strategies.

Our research strongly suggests that this has been a major failure in attempts of staff developers to use the assessment to improve learning. We would argue that they have been much concerned with providing exemplars of good practice and ways of enhancing student perceptions of course requirements, but they have almost totally ignored how university teachers themselves think about the role of assessment. As the research literature is pointing out, to bring about changes in approaches in teaching and learning you must first bring about changes in conceptions of teaching and learning. We would argue that to utilize assessment to improve student learning teachers need to be made aware of the need of such improvement and of the role assessment can play in this process. We suggest that on the basis of research such as that described in this paper, staff developers need to develop workshops or other strategies, which can accomplish this task.

Notes

1. In this paper assessment refers to summative assessment.
2. The distinction between internal and external relations draws upon Bradley (1925) and is employed by Svensson (1984) in his discussion of the view of the human being in phenomenographic research. If A and B are externally related, they are what they are independently of each other. Their relationship is secondary and purely contingent. If A and B are internally related, their "nature" or essential features mutually constitute each other through their relation. The relation is essential for their being. Seeing things as externally related tends to go with a mechanistic view, whereas seeing them as internally related corresponds to a holistic or dialectic view.

Acknowledgments

The Hong Kong and Swedish sections of this research were supported by grants from the Hong Kong Research Grants Council and the

Committee for Research and Conference Grants at the University of Hong Kong. The authors would like to thank Elizabeth Chan and Bjorn Eliasson for research assistance.

Appendix A

A.1. Interview schedule

Introduction

1. How long have you been teaching here at ... ?
2. What sort of undergraduate courses do you teach?

Description of one course

3. Would you please tell me about one of the undergraduate courses that you teach, what are its objectives and how do you assess it?

Conceptions of learning and assessment

4. What is your conception of learning?
5. How do you think your students learn?
6. Does the way you assess your students affect the way they learn?
7. What role does memorization play in your students' learning?
8. When your students say they understand what do you think they mean?

Methods and aims of assessment

9. How important is it to determine which students pass and which ones fail?
10. Is it important to discriminate between the "A's" and the "B's", etc. here?

References

Barrie, S., Brew, A. & McCulloch, M. (1999). *Qualitively Different Conceptions of Criteria used to Assess Student Learning*. Paper presented at Conference of the Australian Association for Research in Education, Melbourne, November 29–December 2.

Biggs, J. (1987). *Student Approaches to Learning and Studying*. Hawthorn, Vic: Australian Council for Educational Research.

Biggs, J. (1995). Assessing for learning: Some dimensions underlying new approaches to educational assessment. *The Alberta Journal of Educational Research* 41(1): 1–17.

Biggs, J. (1996). Assessing learning quality: reconciling institutional, staff and education demands. *Assessment & Evaluation in Higher Education* 21(1): 5–15.

Biggs, J. (1999). *Teaching for Quality Learning at University*. Buckingham, UK: Society for Research into Higher Education.

Boud, D. (1995). *Enhancing Learning Through Self-assessment*. London: Kogan Page.

Bowden, J.A. (1994). Experience of phenomenographic research: a personal account. In J.A. Bowden and E. Walsh, eds, *Phenomenographic Research: Variations in Method*, pp. 44–55. Royal Melbourne Institute of Technology: Melbourne.

Bradley, F.H. (1925). *Appearance and Reality: A Metaphysical Essay*. Oxford: Clarendon Press.

Brown, S. (1999). Institutional strategies for assessment. In S. Brown and A. Glasner, eds, *Assessment Matters in Higher Education*, pp. 3–13. Open University Press: Buckingham.

Brown, S. & Glasner, A. (1999). *Assessment Matters in Higher Education*. Buckingham: Open University Press.

Cole, N.S. (1990). Conceptions of educational achievement. *Educational Researcher* 19(3): 2–7.

Crooks, T.J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research* 58: 438–481.

Eljamal, M.B., Stark, J., Arnold, G.L. & Sharp, S. (1999). Intellectual development: a complex teaching goal. *Studies in Higher Education* 24(1): 7–26.

Elton, L.R.B. & Laurillard, D.M. (1979). Trends in research on student learning. *Studies in Higher Education* 4: 87–102.

Fraser, B.J., Walberg, H.J., Welch, W.W & Hattie, J. (1987). Synthesis of educational productivity research. *Intrnational Journal of Educational Research* 11: 145–152.

Frederiksen, J.R. & Collins, A. (1989). A systems approach of educational testing. *Educational Research* 18(9): 27–32.

Glasner, A. (1999). Innovations in student assessment: a system-wide perspective. In S. Brown and A. Glasner, eds, *Assessment Matters in Higher Education*, pp. 14–27. Open University Press: Buckingham.

Ho, A.S.P. (2001). A conceptual change approach to university staff development. In D.A. Watkins and J.B. Biggs, eds, *Teaching the Chinese Learner: Psychological and Instructional Perspectives*, pp. 237–252. Melbourne: Comparative Education Research Centre/Australian Council for Educational Research: Hong Kong.

Kember, D. (1997). A reconceptualisation of the research into university academics conceptions of teaching. *Learning and Instruction* 7(3): 255–275.

Marton, F. (1981). Phenomenography – describing conceptions of the world around us. *Instructional Science* 10: 177–200.

Marton, F. (1986). Phenomenography – A research approach to investigating different understandings of reality. *Journal of Thought* 21(3): 28–49.

Marton, F. & Booth, S. (1997). *Learning and Awareness*. Mahwah, NJ: Lawrence Erlbaum.

Miller, A., Imrie, B. & Cox, K. (1998). *Student Assessment in Higher Education: A Handbook for Assessing Performance*. London: Kogan Page.

Nightingale, P. (2000). Assessment matters in Higher Education. *Higher Education Research and Development* 19: 177–119.

Nightingale, P., Te Wiata, I., Toohey, S., Ryan, G., Hughes, C. & Magin, D., (1996). *Assessing Learning in Universities*. Sydney: University of New South Wales Press.

Prosser, M. (1994). Some experiences when using phenomenographic research methodology in the context of research in teaching and learning. In J.A. Bowden and E. Walsh, eds, *Phenomenographic Research: Variations in Method*, pp. 31–43. Royal Melbourne Institute of Technology: Melbourne.

Prosser, M. & Trigwell, K. (1999). *Understanding Learning and Teaching: The Experience in Higher Education*. Buckingham, UK: Society for Research into Higher Education.

Ramsden, P. (1992). *Learning to Teach in Higher Education*. London: Routledge.

Samuelowicz, K. & Bain, J.D. (1992). Conceptions of teaching held by academic teachers. *Higher Education* 24: 93–111.

Samuelowicz, K. & Bain, J.D. (2002). Identifying academics' orientations to assessment practice. *Higher Education* 43: 173–201.

Svensson, L. (1984). *Människobilden i INOM-gruppens forskning*. Report 1984:03. Gothenburg: University of Gothenburg, Department of Education [The image of man in the research of the INOM-group].

Tang, C. (1996). How Hong Kong students cope with assessment. In D. Watkins and J. Biggs, eds, *The Chinese Learner: Cultural, Psychological, and Contextual Perspectives*, pp. 159–182. Melbourne: Comparative Eduction Research Centre/Australian Council for Educational Research: Hong Kong.

Tang, T.K.W. (2001). The influence of teacher education on conceptions of reaching and learning. In D.A. Watkins and J.B. Biggs, eds, *Teaching the Chinese Learner: Psychological and Instructional Perspectives*, pp. 219–236. Comparative Eduction Research Centre/Australian Council for Educational Research: Hong Kong/ Melbourne.

Thomas, P. & Bain, J. (1984). Contexual dependence of learning approaches: the effects of assessment. *Human Learning* 3: 227–240.

Watkins, D. (1983). Assessing tertiary students' study processes. *Human Learning* 2: 29–37.

Watkins, D. (1998). Report assessing university students in Hong Kong: how and why?. In D. Watkins, C. Tang, J. Biggs and R. Kuisma, eds, *Assessment of University Students in Hong Kong: How and Why, Assessment Portfolio, Students' Grading. Project Series, Volume 2*, pp. 5–28. City University of Hong Kong: Hong Kong.

Watkins, D. & Biggs, J. (1996). *The Chinese Learning: Cultural, Psychological, and Contextual Perspectives*. Hong Kong: Melbourne: Comparative Eduction Research Centre/Australian Council for Educational Research.

Watkins, D. & Biggs, J. (2001). *Teaching the Chinese Learner: Psychological and Pedagogical Perspectives*. Hong Kong/Melbourne: Comparative Eduction Research Centre/Australian Council for Educational Research.