

Lee Dunn  
Chris Morgan  
Meg O'Reilly  
Sharon Parry

# the Student Assessment

*handbook*

*New Directions in  
Traditional & Online  
Assessment*

**Also available as a printed book  
see title verso for ISBN details**

First published 2004  
by RoutledgeFalmer  
11 New Fetter Lane, London EC4P 4EE

Simultaneously published in the USA and Canada  
by RoutledgeFalmer  
29 West 35th Street, New York, NY 10001

*RoutledgeFalmer is an imprint of the Taylor & Francis Group*

This edition published in the Taylor & Francis e-Library, 2005.

"To purchase your own copy of this or any of Taylor & Francis or Routledge's collection of thousands of eBooks please go to [www.eBookstore.tandf.co.uk](http://www.eBookstore.tandf.co.uk)"

© 2004 Chris Morgan, Lee Dunn, Sharon Parry and Meg O'Reilly.

The right of Chris Morgan, Lee Dunn, Sharon Parry and Meg O'Reilly to be identified as the authors of this work has been asserted by them in accordance with the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

*British Library Cataloguing in Publication Data*

A catalogue record for this book is available  
from the British Library

*Library of Congress Cataloguing in Publication Data*

The student assessment handbook : new directions in traditional and online assessment / Chris Morgan ... [et al.].

p. cm.

Includes bibliographical references and index.

ISBN 0-415-33530-2

1. College students—Rating of—Handbooks, manuals, etc. 2.

Educational tests and measurements—Handbooks, manuals, etc. I. Morgan, Chris, 1957-

LB2368.S88 2003

378.1'67--dc21

2003013321

ISBN 0-203-41651-1 Master e-book ISBN

ISBN 0-203-43948-1 (Adobe eReader Format)

ISBN 0-415-33530-2 (pbk)

I

# The link between assessment and learning

In providing a background to the role of assessment in student learning, it is worth while examining the knowledge base on which it is built:

Theories of teaching and learning focusing on student activity are based on two main theories: phenomenography and constructivism. 'Phenomenography' was a term coined by Marton (1981) to describe the theory that grew out of his original studies with Säljö, and has developed considerably since then (Marton and Booth 1997). Constructivism has a long history in cognitive psychology, Jean Piaget being a crucial figure... And today it takes on several forms: individual, social, cognitive, postmodern (Steffe and Gale 1995).

(Biggs, 1999a: 12)

Biggs explains that, while there are different traditions, the main concern should be to work out what it is that helps to improve teaching and learning outcomes for students. The fundamental principle in either school of thought is essentially what Shuell (1986) has argued: it is not so much what the teacher does that influences learning outcomes but, rather, what the student does. Biggs (1999a) would argue in this vein that learning involves doing; it is 'performative': 'meaning is not imposed or transmitted by direct instruction, but is created by the students' learning activities, their "approaches to learning"' (Biggs, 1999a: 13).

## Approaches to learning

Many researchers have distinguished between the different cognitive levels of engagement with learning tasks. Perhaps most widely referred to is the

distinction between a *surface approach*, in which a relatively low level of cognitive engagement occurs, and a *deep approach*, where a relatively high level of cognitive engagement with the task takes place. In a surface approach to a learning task, the student perceives that it is necessary to remember the body of knowledge. Mostly this would involve the need to rote-learn and then recall the facts concerned. Of course there are many situations where this kind of approach to a learning task is appropriate – such as perhaps learning the chemical tables. At the other end of the spectrum is a deep approach to a learning task, where the student perceives that it is necessary to make meaning of the content concerned, to be able to appraise it critically and to be able to apply the knowledge to other contexts or knowledge domains.

The second main theme concerns the student's *perception* of the learning task. The important work of the phenomenographic researchers (see, for example, Marton and Säljö, 1984; Ramsden, 1992) has shown that students approach a learning task with a surface orientation or a deep orientation according to how they *perceive* the learning task. As exemplified in *The Experience of Learning* (1984), phenomenographic researchers have continued to document students' perceptions of their learning tasks since the early 1980s to the present. The most influential factor in shaping those perceptions turns out to be the nature of the assessment task. If the student thinks that the assessment task requires memorizing and recall, then that is exactly the approach adopted to the learning task. If, on the other hand, the student thinks the assessment task requires analysis, extrapolation, application to other knowledge domains or critical appraisal from a particular theoretical framework, then that is the approach adopted to the learning task. The power of assessment to influence the approach adopted towards a learning task is very considerable.

Underpinning the range of issues raised here is the question of what kind of learning is desirable. The idea of different levels of learning is long established and it fits with the idea of learning tasks at one end of the range that appropriately may be simple memorizing or naming tasks – surface learning. At the other end of the range, learning appropriately may be a higher-order activity that involves making meaning – deep learning. Generally, we would expect that lower-order learning activities would, during the course of an undergraduate programme, lead to their use in higher-order activities. The match between learning objectives and assessment tasks becomes essential if we are to assess in a way that is valid. Learning taxonomies such as Bloom's (1976) taxonomy of learning have been very influential in prompting university teachers to consider what level of cognitive activity is required in different kinds of assessment tasks. Another such taxonomy, which fits with Bloom but advances the importance of emphasizing the performability of intended learning outcomes so that they are more likely to match what is taught and what is assessed, is the structures of learning outcomes (SOLO) taxonomy developed by Biggs and Collis (1982).

There are important differences between surface-oriented approaches to learning and deep or 'meaning' approaches to learning. Researchers mostly agree that students will adopt an approach to learning according to how they

perceive the learning task. But what influences whether a task is perceived as a surface task requiring memorizing or a deep task requiring meaning to be made and applied to a context? Ramsden (1992) argues that students are influenced by a whole nested system of elements, from their prior learning to the quality of the teaching, but also concedes that the most influential is the impact of assessment.

From the student's point of view, getting through the assessment tasks successfully – whether that means earning a pass or achieving a distinction – is the objective. So problems arise when the assessment task requires performance that has not been taught or that does not match the desired learning outcomes.

Learning aims or objectives for university-level study may require students to develop a base level of knowledge requiring naming and memorizing. An example of this would be to learn the names of different periods of prehistory as a basis for then locating particular characteristics within periods of time and geographical zones. However, much of what is learnt in higher education is exactly that – higher order – and it requires synthesis, critique, application, analysis, theorizing or hypothesizing. So a particular problem for students is where learning objectives and study guides direct the student to one kind of learning approach, say at the lower end of memorizing, naming and describing, but the assessment tasks require the opposite – say, much higher-order skills. Of course it might also be the other way around. In any case, the student receives mixed messages.

## Unintended outcomes

Unintended outcomes are likely to arise when assessment tasks don't match the desired learning outcomes. Many different kinds of mismatches are possible but their effect is singular. They result in students trying to guess the agenda of an assessor. Perhaps the most common form of mismatch is where teachers say that they value higher-order thinking skills and that they are trying to promote them in learning activities, but they set assessment tasks that only call for rote-learned facts in their exam questions. Another frequent mismatch is just the inverse of this: learning activities focus on the teacher telling students lots of facts, reinforcing these with study materials, but then setting assessment tasks that call for higher-order operations than students have learnt to perform. Yet another mismatch that is easy to recognize is where students are up- or downgraded according to criteria that are neither explained nor explicit to students, a very familiar example being the downgrading of students for poor written expression or referencing inconsistencies when these criteria have not been made clear to students prior to the assessment task. In these situations, expectations held by individual examiners are not explicit to students. But from the students' point of view, there is more still to go wrong. Orrell (1996), for example, has shown that examiners' conceptions of criteria evolve over time in

assessing the performance of a cohort of students. Selective assessment may be problematic, too, in that students don't know that they should specialize in selected topics rather than cover the entire range.

Unintended outcomes also are likely to occur when what is assessed has not been taught. Where teachers find themselves teaching much larger cohorts of students, where they find there is a great deal of diversity in the student profile, where subjects are heavily content based, such as in anatomy, physiology or chemistry, or where there are several markers who do not normally communicate about their approaches to marking, these kinds of problems can easily arise for students. In these kinds of situations, the mismatch between learning objectives and assessment tasks leads to unintended learning outcomes. From the students' point of view, those unintended outcomes are unfair because they disadvantage the assessment of students' performance. From the teacher's point of view, the assessments are an invalid indicator of their students' performance.

## Side effects of assessment

The impact of assessment practices upon student approaches to learning and learning outcomes is easily underestimated. Not only are there unintended outcomes of assessment, but there are also side effects. A valuable overview is provided by Rowntree (1987: 36), who describes eight side effects of assessment, which are summarized here:

1. The prejudicial aspects of assessment, in which students may be stereotyped by university teachers as having certain levels of ability.
2. The effects on students of knowing they are being assessed, and so their performance becomes a self-fulfilling prophecy or limits the naturally inquiring or motivated student to what will be assessed.
3. The extrinsic rewards of assessment, whereby students who are cue-conscious learn to play the system and selectively neglect topics and skills by minimizing risk and maximizing results for effort.
4. The competitive aspects of assessment, in that many grading systems are competitive. If, for example, only 10 per cent will be admitted to honours, then there will not be enough rewards to go around if the cohort is very able. Rowntree (1987: 52) argues: 'Students generally have all been led to believe that they cannot all achieve a worthwhile level of learning.'
5. The bureaucratic aspects of assessment, where assessment is an instrument of policy, designed to preserve a predetermined standard elsewhere and so depersonalizing the process.
6. The nature of specific assessment techniques, in that many are simply not suited to the kinds of assessment performance to which they are applied.
7. The giving of grades, because they synthesize all of what is known about a student's performance across a range of skills and abilities to a simplistic level.

8. The reporting of assessment results, in that they are not always an accurate indicator of performance or ability to the wider community.

Considering the range of side effects of assessment, it is worth while confronting the real challenges for university and college teachers. What do we have to do to look at assessment from the perspective of the learner? So many of the issues associated with assessment are those of fairness and appropriateness. What can we do to ensure that assessment is fair and appropriate? In this book, we argue strongly for assessment that is aligned closely to the teaching and learning activities in a course. If the assessment tasks are not aligned with the desired learning outcomes or with what students do in their learning activities, how can we make adjustments? Making adjustments might be easy if we are responsible for setting assessment tasks. But what if someone else sets the tasks? At the heart of these challenges is the preparedness of university and college teachers to create learning environments that reward meaningful inquiry and eschew surface or reproducing approaches to assessment tasks. The imperative is to have a conception of learning – and assessment – that is shared by the teacher and by the learner.

## 2

# Roles and purposes of assessment

The question of why students are assessed might at first seem to be simple common sense; it is obvious that their performance needs to be measured. Looking a little further into particular instances, it may perhaps be that the assessment of students has followed disciplinary traditions or regulatory requirements wherein teachers have not had the need or luxury of time to wonder too much about why a particular approach was adopted in that instance. But assessment tasks and strategies vary enormously, both across and within different disciplinary settings. In discussing the rationale for one assessment strategy over another, colleagues may well bring to bear a range of different purposes while others may struggle to resolve questions of whether one approach is likely to be more successful than another.

To be able to make informed decisions about how to assess students, teachers have to take into account the roles and purposes of student assessment, for these may be highly differentiated and they shape our conceptions of what is important. When we consider that the assessment of learning is of interest to a range of different stakeholders, we can also see that the processes and outcomes of assessment have the potential to address a number of different needs.

### **Maintaining standards**

For example, our society expects universities and colleges to uphold standards of excellence in learning. In its turn, the institution is interested in assessment of students in order to maintain its standards and its reputation in the global



marketplace. Individual institutions must meet the needs of their communities. The apparently successful achievement of its graduates ensures an institution's programmes remain attractive to prospective students and to those whose loyalty has been gained enough to return for study of higher degrees or to gain additional qualifications. Professional groups might be interested in our assessment practices in order to ensure integrity of practice, consistency within the profession and compliance with specific legislative and safety regulations.

Our academic supervisors and colleagues are also interested in our assessment practices as a means of benchmarking both the teaching and the curriculum, and they will take an active interest in cases where students mount an appeal or complaint about issues of equity or justice. Our colleagues might also be working with us to ensure that, throughout the programme as a whole, our collective approaches to assessments are complementary and developmental. Considering all these different expectations of assessment, it is possible to identify a range of stakeholders in the assessment process. But in the end, the student must be the first and most important stakeholder. Students themselves demand a quality learning experience that prepares them for new agendas in life. They want value for money and they expect their award to be recognized in the global marketplace.

## Students – the primary stakeholders?

In considering stakeholders in the assessment process, we must consider the main stakeholders – students themselves. The principal aim of assessment, from the students' perspective, is to improve the quality of student learning. According to Biggs (1999b: 68), 'assessment in practice has two functions: to tell us whether or not the learning has been successful, and in conveying to students what we want them to learn'. Unfortunately, when we begin to investigate the experiences of assessment that students remember, more often than not we find a litany of horror stories, of learning by aversion. The assessment tasks that were a struggle for students to handle or that thwarted their efforts to learn are the enduring ones in the minds of students who vow to stay clear of such subject areas or such assessment methods for evermore.

Overall, a range of roles and purposes for assessment includes:

- diagnosing student difficulties;
- measuring improvement over time;
- motivating students to study;
- judging mastery of essential skills and knowledge;
- ranking the students' capabilities in relation to the whole class;
- evaluating the teaching methods;
- evaluating the effectiveness of the course;
- encouraging the tacit learning of disciplinary skills and conventions.

It is possible for assessment to be designed to meet more than one stakeholder's needs. So, how can our assessment practices focus on the needs of both the student and, for example, the institution? Angelo and Cross (1993) argue that classroom assessment can and should be learner-centred, teacher-directed, mutually beneficial to the student and to the faculty, formative and continuous and context specific. More recently, Angelo (1999) argues that assessment can be used to assess performance while at the same time meeting the accountability needs of faculty and institutions. But, in practice, the emphases in assessment – either towards summative, accountability ends or towards formative improvement – are difficult to marry, so students need to know in advance how their performance will be assessed. This is because students take assessment very seriously, and whether there is an opportunity for improvement or not on the same task will motivate very different behaviours.

## Validity and reliability

It is important for all stakeholders in the assessment process that the measurement of performance is valid and reliable. But what do these terms really mean? Throughout this book, these terms are used frequently, so it is important that they are clearly defined. In statistics and measurement generally, validity concerns the extent to which inferences made from test scores are useful and meaningful. When we speak of assessment tasks being *valid*, we are referring to assessment tasks that actually measure the performance of the intended learning outcomes specified, and so meaningfulness and usefulness are central. It is possible for assessment tasks to have low validity, such as when a subject is taught at one level but the assessment tasks require performance of a higher order. It is also possible to construct assessment tasks that have little relevance to what was intended in either the anticipated learning outcomes or the teaching. Assessment is said to be valid when it measures what it is supposed to measure. Validity is the concept that underpins the notion of 'constructive alignment' of the curriculum, as outlined by Biggs (1999a).

*Reliability*, however, concerns the extent to which other assessors would reach the same conclusions. This can be ensured in two ways: either by triangulating with other kinds of tasks to ensure that it is the same performance of learning that is being measured, or through processes of moderation, to ensure that others would consider the assessment task to be a reasonable measure of performance. When colleagues review our assessment tasks, do they consider them to be fair and reasonable methods of assessing performance? If we change the assessment task in a subject, does it still measure performance on the same intended learning outcomes? If the assessment task is reliable, the answer to these questions will be 'yes'.

## Formative assessment

Whether there is an opportunity for improvement or not on the same task is an important consideration. If there is an opportunity for students to improve their performance *on the same task*, then the assessment is essentially *formative*. If, however, the performance on an assessment task indicates the sum of performance on that task, then it is known as *summative*. A summative assessment task cannot be repeated or improved. The trouble is that sometimes teachers confuse formative and summative assessment.

In situations where students know that the assessment task is formative, they understand that they will receive feedback on their performance that will enable them to improve their performance on the same task or on a similar task. For example, students in a creative writing class may be required to write a draft biography as their first task. The teacher and/or the other students provide constructive feedback on the draft, which students then rework as a second task. A second round of constructive feedback is then provided to students so that they can refine their work even further. Each of these three stages is considered formative. There may be a combination of three purposes of formative assessment: 1) diagnosing student difficulties; 2) measuring improvement over time; and 3) providing information to inform students about how to improve their learning. The key here is that students know that the assessment is directed towards providing information about how to improve their performance before the point where a final measurement of achievement is made. This process is known as *scaffolding*, which can be tightly organized, such as in a problem-based course, or it can be less directed. One way of teaching mathematics, for example, is to set small groups to work on problems together, with a tutor working across the groups in class, supervising progress and providing feedback. When the groups find a solution, they write it up in their workbooks for assessment. Each of the stages of the feedback and improvement cycle is part of formative assessment.

In formative assessment, the main purpose is diagnostic but for formative purposes: to enable students to obtain sufficient information to identify their own strengths and weaknesses in terms of current knowledge and skills. In some cases, formative or diagnostic assessment tasks also help students recognize their own attitudes, biases and preconceptions. It is thus possible for students to learn, by increments and gradual insights, the specific nature of their limitations. When the assessment tasks take place in small increments with associated feedback to students, it may be hard to see the difference between assessment and effective teaching.

## Summative assessment

Summative assessment takes place when students undertake a task that measures the sum of their performance. At some point, whether students have had several

opportunities to rework and improve their performance or whether they have had no opportunities at all, the performance must be graded. The allocation of a final grade on an assessment task – or an entire course – is known as summative, and sometimes terminal, assessment. Ideally, summative assessment comes at the end of a systematic and incremental series of learning activities that have formative assessment tasks set at key points during the course. The formative assessment tasks are interwoven with the teaching and learning activities and then a summative task is set. When the interweaving of formative assessment tasks towards a summative event is formalized in a course, it may be called scaffolding. In the mathematics example given above, the summative assessment could be a final exam based on the formulae learnt in the formative assessments or it may be based on students' individual workbooks.

Where scaffolding occurs and students know what is expected of them as part of the teaching and learning activities, there is a formative learning process that leads to summative assessment. The challenge for teachers is to understand the limits of formative assessment. Once an assessment task is allocated a final grade, it is summative, regardless of whether students may use what they learn from it on another task. When students are given a final grade on an assignment together with formative comments, they are highly likely to attend only to the grade and not take the formative comments seriously (see, for example, Butler, 1988). In this vein, Newble and Cannon (1995) have also argued that formative assessment, where students aim to learn from their misconceptions or deficiencies, ought to be kept separate from summative assessments, where a final grade is at stake. Other challenges arise when the summative assessment bears no relationship to the formative assessments and therefore to what has been learnt.

Two more challenges concern workload issues – either for the students or for the teachers. When formative assessments are compulsory and too numerous, student workloads soar while at the same time dampening intrinsic enthusiasm for the learning process. Of course this happens when there are too many summative assessments too, and it is also a problem when the assessment tasks are too demanding, whether formative or summative. The challenge for teachers in these days of shrinking funding and increasing class sizes is to create formative assessment tasks that require a realistic amount of teacher input. Formative assessment can take up extensive teaching hours. In creating realistic formative assessments, the further challenge is to weave them into the teaching and learning activities, so that formative feedback to students is not an additional impost on the department but a way of implementing the curriculum. Black and Wiliam (1998: 5) put this succinctly: 'for assessment to function formatively, the results have to be used to adjust teaching and learning – so a significant aspect of any programme will be the ways in which teachers do this'.

Issues related to workloads lead us to ask 'How much formative assessment is enough?' and 'How much is too much?' The answer is not simple. But one yardstick comes from the much-utilized Course Experience Questionnaire (Ramsden, 1992). We know from this and subsequent studies that students

appreciate teachers who 'make it clear what I need to do to be successful in this course'. The difficulty for university and college teachers is to lead students to share their conceptions of the learning that is expected of them. For this to happen, formative assessment and teaching generally need to be aligned with the ways in which learning is assessed.

## Balancing formative with summative assessment

Some would argue that, when an assessment task is formative, a summative grade should not be given at all. The danger in this situation is that students prioritize their efforts towards achieving an academic award. If an assessment task counts for nothing towards a final grade, why would a student think it necessary to do the work? One way to make formative assessment count is simply to make it compulsory, whether graded or ungraded.

A useful strategy through which teachers have overcome this issue of balance between formative and summative assessment within a course is by designing the curriculum on the concept of *scaffolding* (Biggs, 1990). Scaffolding is designed into the curriculum when formative assessment tasks provide the basis of the teaching and learning activities in a course, and these lead ultimately to summative assessment. Take, for example, a problem-based clinical diagnosis subject in the health sciences where the students are required to follow successive symptoms over a number of cases and provide tentative diagnoses. These are formatively assessed – no grade awarded but just pass or fail. The summative assessment comes when students have to put all their case findings together: over time, symptoms may change and develop, leading to new hypotheses and new directions of inquiry. The overall assessment task comes as a result of the scaffolding of the parts of the problem. The idea of scaffolding is to interweave formative assessment tasks with the teaching activities, so that students have ample opportunities to know how well they are progressing and how they can improve as they work their way through the course. The summative assessment tasks – which are closely aligned with but not the same as formative tasks – form the basis for a final determination of performance.