2.1 Evolving methods of assessment

Michael Manogue[1*], Mary Kelly[2†], Sonia Bartakova Masaryk[3‡], George Brown[4‡],
Frank Catalanotto[5‡], Teo Choo-Soo[6‡], Elis Delap[2‡], Pavel Godoroja[7‡], Ikuko Morio[8‡],
Jerome Rotgans[9‡] and Mare Saag[10‡]

[1]*University of Leeds, UK;* [2]*University of Dublin, Ireland;* [3]*University of Brno, Czech Republic;* [4]*Nottingham University, UK;* [5]*University of Gainsville, Florida, USA;* [6]*University of Singapore, Singapore;* [7]*University of Chisinau, Moldova;* [8]*Tokyo Medical and Dental University, Japan;* [9]*University of Aachen, Germany;* [10]*Tartu University, Estonia*

This Working Section is concerned with evolving methods of assessment in dental education. It focuses on newer methods of assessment that might have relevance for broader application. Although it cannot provide answers to all the questions it raises, it is hoped that the contribution it makes is of value in the process of the development of a global network in dental education.

Key words: assessment, dental education, methods.

© *Blackwell Munksgaard, 2002*

Introduction

What is assessment?

ESSENTIALLY, assessment is a process of making a judgement or measurement of worth. It consists of taking a sample of behaviours, drawing inferences and making estimates of worth (1). To be effective, the sample needs to be representative of the behaviour skills or attributes being assessed: hence the importance of using several instances whenever possible, as well as using different methods and different observers. The methods of drawing inferences should be consistent across observers or instruments and based on clearly defined, transparent measures. The estimates of worth should be based on explicit values derived from methods of consensual validation. An effective assessment procedure should provide a valid, reliable and practicable assessment of knowledge and understanding, problem-solving performance or professional attributes. Defined skills should be assessed. Tasks should be related directly to course aims and objectives (or learning outcomes) and should be evaluated by explicit criterion referenced methods (2). Criteria should be agreed by staff and understood by students. In the development of more sophisticated criteria, it has been demonstrated that student learning can be enhanced if students are involved in the definition of the criteria to be applied and that a change from 'How did I do?' to 'How can I get better?' may ensue (3). One can see the complexities of the challenge: while technical achievement might be quantified, the assessment of professional attributes, for example, is not simple. This subsection discusses some of the issues that face dental schools in evolving methods of assessment.

Even if an assessment has been designed for judgement, it should be possible to provide feedback to students on their performance, in terms of relative strengths and weaknesses so that they may learn from the experience. Staff should be informed of the outcome of the process, to provide an overview of the strengths and weaknesses of the course in relation to quality of teaching or appropriateness of assessment tasks, and to review how modifications might be made.

Assessment and learning

Assessment is central to the success of any educational programme and is linked very closely with student learning (4). It is well recognized that student learning is influenced and directed by methods of assessment (3, 5, 6). For example, it has been noted that in the assessment of medical students, a change from the award of pass/fail reports at the end of a clinical attachment to a structured, clinical, practical examination led to an increase in the amount of time students spent on the wards (7) as students realized they needed to be able to perform. This Working Section discussed issues in relation to how students learn. There is evidence to suggest that

students' orientation to 'deep' learning (that is, the search for meaning and understanding) declines in undergraduate courses (8). Dentistry has been implicated negatively in this respect (9). Therefore, evolving methods of assessment should be directed towards the encouragement of deeper learning strategies. It is fortunate that there are examples of curricula (for example, problem-based learning, PBL) and assessment practices, that can promote deep, elaborated learning and improved long-term recall (10–12). Critics of PBL, e.g. (13) have claimed that 'traditional' curricula and assessment might allow students to score more highly on knowledge of clinical content (the example given was in medicine), so the aim of educators should be to strike a balance between enhancing relevant knowledge, developing clinical skills and understanding and fostering self-directed learning (14, 15).

Course designers are becoming increasingly aware of the need to develop reflective, lifelong learners who can assess their own performance and behaviour accurately (3, 16, 17). The approach has been recommended as a means of preparation of dental students for continued professional development (18) and is listed in national undergraduate dental curriculum requirements (19).

There is an absence of literature in dentistry as to why particular methods of assessment are used. In essence, lecture-centred academic curricula emphasize the importance of reproductive (superficial) knowledge and assessment that 'fits' this form of curriculum must itself demand reproductive knowledge. Simply changing the method of assessment can cause tensions when traditional teaching methods are still used (20). This working group also discussed the relevance of the curriculum and learning environment on assessment practices.

Parameters within which the section decided to work

In view of the large volume of work already completed and recorded in the DentEd Final Report (21), it was decided that that document should be accepted as a comprehensive reference for extant assessment practices in dental schools. We refer to that report and the information contained therein. It may be necessary to revise this information when the final phase of school visits is completed.

Review of that document enables identification of a number of areas of assessment as follows.

Established methods
a) Assessment of knowledge (e.g. written essays, short answer questions, some forms of multiple choice questions).

b) Assessment of skills (e.g. competence testing, continuous clinical assessment and objective structured clinical examinations or other structured clinical tests). Clinical skills constitute such a vital component of dental practice it is essential that their assessment is as valid and reliable as possible.

While there is evidence to support the validity of competence testing, approaches used in continuous clinical assessment are less clearly defined and standardization of multiple observers is difficult (see 'Best practices and innovations' below).

Evolving methods of assessment based on changing philosophies – the focus for the working group
a) Methods that are not yet fully developed or evaluated, e.g. the use of virtual reality, computer-based tests and other new information technologies (see 'Impact of information and communication technology' below).
b) Assessment of professional aspects of competence (attitudes, patient management, communication and the identification of treatment outcomes). Current curricula in most dental schools are deficient in their attention to the behavioural sciences, as highlighted by Chapter 14 of the DentEd report (21). It is important, therefore, that assessment should only be developed in parallel with comprehensive curricular change (see 'Considerations not otherwise covered' below).
c) Areas where assessment is indicated but has not yet been addressed (e.g. cognition, learning approaches and styles). As understanding of learning theory and its implications for curricular content and design grows and as the need for lifelong learning becomes more obvious, so the development of assessment methods which investigate student-learning processes become increasingly important (see 'Considerations not otherwise covered' below).

Best practices and innovations

In all assessment design, whether summative or formative, traditional or innovative, there are certain basic tenets which should be observed, including the following:
1. The purpose of the assessment should be clear to both assessor and assessed. This will usually embrace one or a combination of the following:
   a) to motivate and direct learning;
   b) to ensure standards for proceeding to the next stage of the curriculum and ultimately graduation;

c) to provide feedback for students on their performance; and

d) to provide feedback on the curriculum and its mode of delivery.

2. The concept of alignment of curricular components is seminal. Assessment tasks must match the course objectives, which in turn must be reflected in the course content.

3. The criteria should be kept simple and should be well understood by students. The criteria should form the basis for feedback and marking.

4. There should be wide sampling using different methods. The closer a method is to recall of knowledge and/or well defined solutions the more reliable is the method, but this does not necessarily make it more valid. The conflicting requirements of validity, reliability, effectiveness and efficiency must always be recognized and reconciled as satisfactorily as possible in the individual circumstances (see later).

5. Feedback to students should be as immediate and comprehensive as possible.

6. Student evaluation of assessment criteria and practices should be elicited, receive serious consideration and, where appropriate, be acted upon.

7. The quality of examiners, whether as sources or instruments of assessment or both, should not be taken for granted. Staff training workshops and peer review of assessment practices and, where appropriate, checklists and marking schemes are an essential part of a good assessment system.

8. Where external examiners are used they often provide a useful, objective view of both the curriculum and the assessment procedures. They also enable inter-school comparisons and exchange of expertise. It is essential that external examiners understand the alignment of curricular design and outcomes, empathize with it and have an awareness of best international practice.

9. Assessment is less a psychometric and more an educational design problem (i.e. how to use assessment strategically for its educational effects). There is a need for more research on educational effects of assessment and less exclusive focus on psychometric properties. This research must necessarily be contextually rich.

10. Assessment schemes, overall, should cover the whole competence pyramid. There is a need to rely more on descriptive/qualitative data and to accept that some subjectivity is both inevitable and acceptable, provided multiple assessors are used. A considered combination of 'controlled' and 'independence' strategies is recommended.

Impact of information and communication technology

An increasing number of higher education institutions use computers to help solve some of the problems associated with the burden of increased student numbers. The application of computer-assisted learning (CAL) is well established and computer-assisted assessment (CAA) is becoming increasingly relevant. CAA includes a range of activities, e.g. the collation, analysis and transmission of examination grades across networks or the automatic marking of answers that have been completed by students at workstations. CAA is, in comparison with computer-aided learning, a relatively new development and is often pioneered by enthusiastic individual academics (22). The successful implementation of CAA is often hindered or abandoned due to time and funding restrictions, or reliance on that individual enthusiast.

Computer-assisted assessment of dental students
In comparison to other educational disciplines(e.g. technical disciplines and/or biological sciences (23), application of CAA in dental education is less advanced and far from widespread. Positive feedback has been reported from medical students (24, 25) showing that they consider this type of learning/self-assessment to be effective.

Forms of computer-assisted assessment
Computer-based tests (CBT) are the most common form and are, generally speaking, electronic equivalents of the more traditional paper-written test. The sort of questions posed in CBTs are almost identical to traditional tests. They consist mainly of:
- Self-assessment questions (SAQs).
- Parts of a computer-aided learning (CAL) package.
- Tests for formative assessment or summative (examinations).
- Diagnostic tests (i.e. 'What stage is the student at now?').
- Multiple choice tests (MCQs) for use with an optical mark reader (OMR).

The basic model has two essential components:
1. The 'no-book' test, which is given in a normal class and reflects topics that have been covered recently – it serves as a feedback mechanism for students and teachers.
2. The 'open-book' test, where information in relation to a test is accessible for several days or weeks in advance (which might be considered as a tool for the evaluation of improvement in knowledge).

Numerous software packages have become available both on commercial and free bases to produce CBT (e.g. QUESTIONMARK, CASTLE). Such packages are easily accessible and user-friendly.

Problem-orientated searching for information on the web

This approach evaluates the ability to find web-based information (text, graphics, video files) on specified topics. The topics may range from simple things, e.g. the location of a dental academic, their postal address, e-mail address, personal page, etc. through intermediate tasks, e.g. how to find relevant information on specific dental topics such as dental implants or guided bone regeneration, to more complex scenarios such as how to find a case study that is similar to one specified, and to gain alternative solutions to a dental problem.

Problem-orientated searching for information requires a certain 'outlook' from academic staff. They must be PC/web-orientated and they must be willing to prepare, electronically, for such a session. The approach is quite new.

Student assessment could follow along the following lines:

- Can the student find the required information within a time limit?
- Can they interpret the relevance of test results?
- Are they able to handle information (download, save, print out, book-marking, etc.)?

Rank ordering of performance seems not to be a good idea because of the possible skewing effects of unstable web resources, e.g. changes in amount, availability and accessibility of information. Further information on this approach is available at the following sites:

- The American Academy of Periodontology.
- ADA Online/the American Dental Association.
- DERweb/Dental Education Resources at the University of Sheffield.
- Dental X Change, GlobalDent, Dental Study Club/Swiss universities base.

Exploiting databases

This consists of searching web databases that provide information of bibliographic quality. It evaluates the ability to find dental monographs and papers on specified topics by using:

- International searchable databases, e.g. ISI, MEDLINE and NIH.
- Local and national-related library resources.
- Personal bibliographic databases.

A more advanced alternative is to ask a student to find dental publications that are not indexed in international databases, e.g. final reports, 'electronic-only'

dental medical/dental journals, etc. This requires an ability to browse through national institution websites, such as those belonging to governmental bodies, commissions, etc. For students from non-English speaking countries, a useful alternative of the approach is to find English equivalents for dental terminology. This has a double effect in that it improves the ability to work efficiently in 'English-speaking' web resources as well as improving general knowledge of English.

The future of CAA in dental education

Implementation of CAA appears to have been hindered by a lack of institutional commitment, strategic direction and easy-to-use and established methodologies. The potential of CAA is, however, high. When employed correctly, it can ensure that curricular modifications take place at a time when students can benefit (26). In terms of quality assurance, CAA could drive dental educational institutions to reconsider their existing student assessment methods. For academic staff, CAA can provide the stepping-stone to greater use of computers for teaching.

CAA will challenge the organizational structures of dental schools. Future promotion of CAA should result in a greater collaboration between technically orientated support and academic staff. This process may be challenging in the first instance, as disparate groups are forced to find ways of working together, but such collaboration is essential to the progress of CAA, as its future depends on both pedagogical and technological advances.

How to converge towards higher global standards

This is the focus for the subsection entitled 'Towards global convergence of education, training, quality and assessment' and, consequently, only a précis is given here. Our group was aware that there are two areas of concern in relation to the concept of convergence: the first is a fear that convergence may result in a reduction of standards towards a minimum universally achievable level; the second relates to academics who may be concerned that they will be required to achieve levels of performance which are impossible in their particular circumstances. Both these reservations are real and care must be taken to reassure both groups. A major factor in enabling continuous quality improvement in education is the availability of information on best practices and access to the evidence for their acceptability. A network such as DentEdEvolves is ideally placed to act as a source for this information. It will also be important to

facilitate its adoption by engaging in discussion with schools that wish to develop new practices but which are unsure how to progress.

Important regional and continental differences

The European Dimension (based on DentEd, Chapter 17)

From the 25 DentEd visits the following best practices and innovations were recorded.

Ten schools used formative assessment in addition to summative assessment. In most of those schools, formative assessment is an integral and ongoing activity of all courses. Self-assessment and peer-assessment, in particular, have been shown to increase the motivation to learn and change student attitude from one of 'how have I performed' to that of 'how can I get better?' (3). In a few schools, discussion and feedback takes place during oral assessment or after written examinations so that the examination comprises not only an assessment but also an opportunity to learn. As indicated by the schools themselves, most of the assessment concepts match the course objectives.

Assessment methods

In 23 schools a variety of assessment methods were used in order to assess different aspects of students' competence. These multiple-method combinations are important since no single assessment method can adequately measure clinical competence.

Fifteen schools used case presentations as a trigger for assessment of knowledge but there is no indication that this occurs early in the programme, to reflect authentic situations and the kinds of processes that are central to the profession.

Various skills are assessed by Objective Structured Clinical Examinations (OSCE) in nine schools.

Competence tests, which comprise preclinical and clinical tests of the student's ability to perform designated procedures have been implemented in 19 dental schools; it could not be concluded from the information available that these tests take place in the familiar clinical environment or in a non-threatening manner.

The organization of assessment

The use of external examiners from professional bodies, such as the Public Dental Health Service in Sweden, gives an independent view. For the same reason, in German schools every licensed dentist can take part in the first two State Examinations. Representatives of the 16 State Dental Councils can take part in the third, final State Examination. Members of staff attend conferences on medical and dental education and assessment in order to ensure continuing development of the curriculum and assessment methods, which drive and ensure quality in student learning.

Innovations

Included in new approaches to assessment or non-traditional assessment methods are:
- Trends towards formative assessment.
- Self- and peer-assessment.
- Use of portfolios or reflective log-books.
- Performance-based assessment.
- Practical competence tests.
- OSCEs or other structured clinical tests.
- Scientific meetings.
- Presentations of scientific projects.

The American dimension

There are three instances of external assessment of dental education in the United States.

a) The Joint Commission on National Examinations, sponsored by the American Dental Association (ADA), the American Dental Education Association (ADEA) and the American Association of Dental Examiners (AADE). An organization of state licensing boards (see item (c) below) administers the National Board Examination at the end of the second year (part 1) and the middle of the fourth year (part 2) of the Dental curriculum. The 'part 1' examination assesses basic science knowledge using standard short answer questions. The 'part 2' examination assesses clinical knowledge using a case-based approach.

b) United States dental schools are accredited by the Commission on Dental Accreditation, sponsored by the ADA, ADEA and AADE and authorized by the US Department of Education. This is a voluntary process but, without accreditation, dental students and schools cannot receive federal loans and grants and graduates cannot sit state licensing examinations. The Commission sets a number of standards relating to outcome measures, educational parameters, student support services and the research mission. Every 7 years, the dental schools perform a self-assessment/self-study and are site-visited for 3 days by a team of Commission consultants who verify the findings of the self-study, in a process similar to the recent DentEd visits. Deficiencies result in recommendations that must be corrected within 2 years under threat of loss of accreditation.

c) Individual states conduct practical and written examinations for dentists seeking to practise in that state, in order to assess their competence. The purported rationale is to protect the public but some would argue that the main objective is to restrict mobility.

There is no state which assesses continued competence in their practising dentists.

The Singapore dimension

The Faculty of Dentistry in Singapore believes in the new philosophies of learning and assessment and accordingly has introduced innovative assessment methodologies such as open-book examinations, which assess students at a deeper level of critical thinking and application. Clinical programmes are competence-based, with a strong emphasis on multidisciplinary and holistic management of the patient.

The Japanese dimension

In Japan there is a growing awareness among dental academics that assessment is an integral part of the curriculum and should be closely linked to the educational aims, objectives and methods. The assessment methods most frequently employed, however, are dominated by traditional written tests. The introduction of OSCEs is being considered in some dental schools. The Japanese National Board Dental Examination, consisting solely of MCQs, has a direct effect on student learning. Serious consideration must be given to developing a more effective combination of assessment methods, both formative and summative.

Considerations not otherwise covered

Under this heading, we discuss parts (b) and (c) of 'evolving methods of assessment', as defined above.

Assessing attitudes (part b)

An attitude is a mixture of beliefs, thoughts and feelings that predispose a person to respond, in a positive or negative way, to other people, objects or institutions. Clearly, attitudes are a directional force, so when considering attitudes one has to ask: attitudes towards what? Attitudes summarize past actions, they influence future actions and they may be used to predict future actions. However, just as past educational achievement is not necessarily a good predictor of future educational achievement, so too attitudes are not necessarily good predictors of future behaviour. One cannot measure attitudes directly, so one infers a person's attitudes from his/her actions and uses this to predict future

actions. There are, however, other predispositions to act. These are understanding, knowledge, skills, motivation and habits. Hence, inferring attitudes from actions or behaviours is complex and requires careful attention to the sample of behaviours and the methods of assessment and an examination of the underlying assumptions of the methods of assessment. Finally, attitudes can also be conceived as occupying the middle ground between personality traits and opinions or mood swings. They are more stable than opinions and less stable than personality traits. The closer the attitudes are to personality traits, the less amenable they are to change.

Can attitudes be changed? There is plenty of evidence that attitudes can and do change. The evidence is drawn from personal experience, naturalistic studies, experimental studies and laboratory studies. Attitudes may be changed through:

- personal experience (direct contact);
- reflection on personal experience;
- group interaction; group membership;
- professional identity;
- chance conditioning;
- media influences (persuasion); and
- cult influences (closed communities).

Attitudes may be resistant to change and a change in attitude does not necessarily lead to a change in behaviour. Other attitudes, predispositions, motives, emotions or habits may be more potent.

Changes in attitude can be brought about by changing knowledge, understanding, skills, actions and context. It is assumed that these changes can be brought about through lectures, small group work, practicals, clinics and projects. Probably, an equally potent force is the 'hidden curriculum'. The hidden curriculum might be described as the unintended consequences of the structure of the course, its methods of assessment and teaching and the attitudes of its teachers. Even the unintentional remarks of a tutor may reveal attitudes that have an effect upon a student's attitudes. This form of 'chance conditioning' is perhaps more common in laboratories and clinics than is realized.

Why assess attitudes?

Attitudes in Dental Education may be assessed in order to:

- ensure dental students and dentists are safe professionals;
- help dental students and dentists to develop;
- estimate change;
- improve interpersonal relationships;
- change contexts and organizations;
- satisfy demands of accountability and control.
- assessment of attitudes can itself change attitudes.

Each of these purposes requires a different emphasis in content and approach. The underlying assumption is that attitudes provide a measure of future behaviour, but as indicated above, this assumption is problematic.

How can one assess attitudes in dentistry?

One can assess attitudes through direct observation of actions or observations of video-recorded actions. The usual rules apply: trained observers, the use of usable explicit criteria and an adequate sample of behaviours. Alternative approaches are self-reports, portfolios, reflective practice assignments and audits of practice- or qualitative-based projects. These approaches are particularly useful for formative assessment. However to be effective, the methods do require trust between the student and tutor. Methods that are normally used for assessing knowledge and understanding may also be used to assess attitudes. Here the danger is that students provide 'socially acceptable' answers. Similar remarks apply to questionnaires and attitude inventories such as a dentist–patient attitude scales. It is suggested that such a scale should be developed and field-tested. Broadly speaking, the more remote the method of assessment is from actual behaviours, the less valid the method is likely to be. However the reliability, validity and practicality is determined by the specifics of the method used rather than the method per se. A good attitude inventory is better than a bad set of observations.

Is assessing attitudes worthwhile?

The answer is a cautious 'yes'. Attitudes are an important ingredient of professional expertise and behaviour. Attitudes in dentistry can be assessed – although their assessment is in its infancy. Just as all methods of assessment of student learning have strengths and weaknesses, so too do different methods of assessing attitudes. The important points are:

- Be clear why you are assessing attitudes.
- Use methods of drawing inferences that are fair, reliable and valid.
- Be aware of the limitations of the methods that you are using.
- Be aware of the assumption that you are making of attitudes as predictors of future behaviours.
- Monitor your methods of assessing attitudes.

Learning (emphasizing process rather than content) (part c)

This section includes discussion of the development of the assessment of critical thinking and decision-making, its relationship to curricular strategy and its relevance to professional practice.

A greater understanding of cognitive processing, particularly in relation to learning, coupled with the realization that lifelong learning skills are essential for continuing safe effective practice has highlighted the need for a new dimension in the assessment of health-care professionals. It is now agreed that it is important to measure not only what is learned, but also how it is learned. It is possible that this new concept will have implications for the selection of students for courses and for modification not only of curricular strategies but also of learning practices at both undergraduate and postgraduate levels.

A review of the understanding of learning theory shows that it can be considered as encompassing two parallel approaches. These are described by Biggs (27) as:

a) The information processing theory deriving from cognitive psychology.
b) Approaches to learning theory based on student personality, learning styles, individual learning context and motivation as reported by the student. Much of the work in both areas relates to medical education but is clearly relevant to all health-care professionals.

Background literature

Information processing theory

Researchers such as Schmidt, Norman, Holyoak, Patel, Schwartz and Barrows have performed extensive investigations into cognitive and metacognitive processing during problem analysis and decision making since the early 1980s. Norman (28) referred to the statement of the American Board of Internal Medicine (1979) that 'the ability to define and manage clinical problems is viewed as central to clinical competence in medicine'. Norman reflected further that none of the variety of evaluation methods then available demonstrated a high correlation between performance on one problem and on the next. He concluded that the low correlation suggested that something other than the general skills of data gathering, problem solving and clinical judgement, as they were then perceived, was having a significant effect on performance.

Patel et al. (29), again addressing the dilemma of the precise components of successful problem analysis and decision making, demonstrated that the process involved in the memory for clinical cases is complex and involves the ability to make inferences from a highly developed knowledge base. Further work by these same researchers and others in the 1990s has helped to clarify a number of these issues. Schmidt and Norman (30), reviewing the publications of the

previous two decades, proposed that medical expertise is not so much a matter of either superior reasoning skills or in-depth knowledge of pathophysiological states, but is based rather on cognitive structures that describe the features of prototypical or even actual patients. They suggested that these cognitive structures contain relatively little knowledge about pathophysiological causes of symptoms and complaints but contain a wealth of clinically relevant information about disease, its consequences and the context under which illness develops. In relation to assessment they suggested that clinical competence will be properly assessed only if tools are used which focus on both the knowledge base and 'on the qualitative transition from a conceptually rich and rational knowledge base to one comprised of largely experiential and nonanalytical instances'. They admitted that measurement is difficult but suggested a two-stage strategy where the first stage provides limited information in a limited time period and probably includes visual representation and the second stage supplies more data and allows for detailed information processing.

The studies of Whittlesea et al. (31, 32) also indicate that it is possible to identify many of the interactive factors involved in the accessing and application of knowledge to a prescribed problem/task/decision. They asserted that it is possible to assess these factors individually and demonstrated, in their 1994 paper, how this may be carried out. The methods which they used are complex and not immediately applicable to general assessment in undergraduate curricula but should become so if further refined and modified.

New developments in assessment of clinical reasoning suggest that, as knowledge and clinical reasoning are closely linked (33), we should not be afraid to examine for knowledge, provided the assessment, whatever the format, focuses on clinical decisions and uses multiple case scenarios.

Examples of stimulated recall techniques described by De Grave et al. (34) and, more recently, Barrows (14) seem to offer readily accessible assessment tools for clinical reasoning including both knowledge processing and conceptual change during problem analysis.

Approaches to learning theory

This alternative approach to the understanding of learning theory is based on the analysis of learning in the learning environment. It relies largely on the use of inventories and questionnaires completed by students involved in curricula. In the 1960s and 1970s researchers such as Biggs and Entwistle, investigating student-learning approaches, began to develop and continue currently to modify, questionnaires such as the Study Process Questionnaire and the Approaches to Studying Inventory. Using these tools three basic approaches to learning, i.e. surface, deep and strategic, have been identified. Evidence shows that the deep approach to learning results in a much better understanding of the material and a better performance in immediate examinations and on later recall (5, 8).

Researchers generally agree that the student approach to learning is dependent to some extent on student learning styles but is also much affected by interaction with teaching, curricular strategy, context and learning environment. The Biggs model of relationships in education (3P model), i.e. Process, Presage and Product, provides a comprehensive exposition of all aspects of current learning theory. Because assessment is such a strong motivator of learning it is essential that it is designed to encourage a deep approach. There have been some disquieting results of studies that suggest that students entering third level education as deep learners may change to strategic or even superficial learners during the undergraduate years. McManus et al. (35), reported on a study of the relationship between medical students clinical experience and their final examination results and found that there was a lack of correlation between examination performance and clinical experience. This study did, however, find a correlation between study habits and learning approaches. Once again the work of this group of researchers highlights the need to expedite the development of new assessment practices to take account of methods of information processing and learning process profiles, both of which must be identified so that they can be modified by curricular strategies.

Implications and potential for emerging countries

Course objectives and assessment methods and rates of curricular development must reflect cultural, demographic, financial and environmental circumstances of each individual school and country. Personal interaction and outreach programmes designed for areas with limited accessibility must support networking of technological advances. See also the report of Theme 4, 'Web-based interactive learning programmes'.

Core values applicable to all

Miller, in his invited review (36), stated the following:
It is important to start with the forthright acknowledgement that no single assessment method can provide all the

data required for judgement of anything so complex as the delivery of professional services by a successful physician.

He suggested a framework on which to base assessment which is known as the Miller pyramid. The pyramid is designed with 'knowledge' as the basal section and moves apically through 'knows how' (competence) and 'shows how' (performance) to 'does' (action).

The main general characteristics of instruments in assessment are reliability, validity, educational impact, acceptability and cost (35).

In a selective assessment situation educational impact and validity are reduced in weight to the advantage of reliability, whereas in formative assessment the reverse may be the case. Reliability is related to sampling across content and testing time. Adequate reliability requires substantial sampling and multiple examiners, patients and resources.

Validity can be addressed by reference to Miller's pyramid as above. Miller (36) recommended that factual tests be applied to the 'knows' stage, clinical context-based tests to the 'knows how' stage, performance assessment *in vitro* to the 'shows how' stage and performance assessment *in vivo* to the 'does' stage.

Educational impact is related to the assumption that assessment drives learning. It undoubtedly does this, but there is a need for more research into ways in which it can be used to encourage desirable learning habits.

Range of assessment methods available

Cases and open problems
This approach has potential for measuring application of knowledge, analysis, problem-solving and evaluative skills. Short cases are relatively easy to design and mark. Design of more complex cases and their marking schemes is more challenging. Marking for grading and feedback is about as fast as essay marking.

Computer-based assessment
This is much discussed. (See section on 'Impact of information and communication technology'.) Reliability is high but validity (the match with outcomes) needs careful attention.

Direct observation
This is useful for immediate feedback, for developmental purposes and for estimating performance provided a simple, structured system is used. The presence of an observer can change performance, so the method should be handled sensitively. Impressionistic observation can be useful, if supported by constructive feedback. This method can be used by a group of peers to provide feedback as well as assessment. Intensive, lengthy training is required for high reliability if detailed checklists are used. Reliability, validity and manageability are fairly high when structured observation is used.

Essays
These constitute a standard method. There are several types of essays that test different styles of writing and different types of thinking. Essays usually measure understanding, synthesis and evaluation provided that the right questions are asked. They are relatively easy to set. Marking for grading, based on impressionistic marking, is fast. Marking for feedback can be time-consuming. Overall, the advice is to keep the criteria simple. Variations between assessors can be high – variations are also demonstrable in repeat marking by individual assessors.

Learning logs/diaries
There is a wide variety of formats ranging from an unstructured account of each day to a structured form based on tasks. Some training in reflection is recommended. Inevitably, it is time-consuming for students and requires a high degree of trust between assessors and students. Measuring reliability is difficult. It may have high validity if the structure matches learning outcomes.

Mini-practicals
This is interpreted as a series of mini-practicals undertaken under timed conditions. There is potential for sampling a wide range of practical, analytical and interpretative skills. The initial design is time-consuming. Some, if not all of the marking can be done on the spot so it is fast. Feedback to students is fast. A reliable method, but training of assessors is necessary.

Modified essay questions (MEQs)/structured questions
This relates to a sequence of questions based on a case study. After answering one question, further information and new questions are given. The procedure continues, usually for about 1 hour. It is relatively easy to set up and may be used in teaching or assessment for developmental or judgemental purposes. It can be computer- or paper-based. It may encourage reflection and analysis and has the potential for high reliability, validity and manageability.

Multiple choice questions (MCQs)
This approach may be used to sample a wide range of knowledge quickly. It has the potential for measuring

understanding, analysis, problem solving skills and evaluative skills. There are a wide variety of formats from true/false to reason–assertion. More complex formats are not recommended as they confuse students unnecessarily and they are time-consuming to design. More demanding MCQs require more time to set. Better ones are based on case studies or research papers. They are easy to mark and results are analysed easily. They are also useful for self-assessment and screening and have potentially high reliability, validity and manageability. Feedback to students is fast. There is a danger of testing only trivial knowledge. To save time, look for banks of items on the Net or in US textbooks. A team of assessors, working to the same learning outcomes, can brainstorm and produce several questions in an afternoon.

Orals
These help to test communication, understanding, capacity to think quickly under pressure and knowledge of procedures. Feedback potential is good. Marking for grading can be fast but some standardization of interview procedure is needed to ensure reliability and validity.

Objective structured clinical examinations (OSCEs)
Initially these were used in medicine and only latterly gained interest in dentistry (20). They are particularly useful for assessing practical and communication skills quickly but are fairly hard to design and organize, although easy to score and provide feedback. They may be used in the induction phase to estimate key practical skills. Group OSCEs are useful for teaching, feedback and developmental purposes. OSCEs can be used towards the end of a course to provide feedback or to test performance against outcomes. Reliability, validity and manageability are potentially fairly high. They are probably less labour intensive than other forms of marking but several assessors are required at any one time. Initially, they are time-consuming to design – but worth the effort. It is essential that a large number of stations be used.

Portfolios
There is a wide variety of types ranging from a collection of assignments to reflection upon critical incidents. The latter are probably the most useful for developmental purposes. They may provide a basis for orals and have rich potential for developing reflective learning if students are first trained in these techniques. Effective use of the portfolio requires a high level of trust between assessors and students. Measuring reliability is difficult, although they may be high on

validity if the structure matches the objectives of training.

Poster sessions
These test the capacity to present findings and interpretations succinctly and attractively. There is a danger of focusing unduly on presentation methods but this can be avoided by the use of simple criteria. They offer good opportunities for feedback from tutor, self and peers. Marking for grading is fast and the use of criteria reduces variability.

Presentations
Presentations test various aspects of preparation, understanding, knowledge and capacity to structure information and oral communication skills. They also have feedback potential from tutor, self and peers. Marking for grading is based on simple criteria and is fast and potentially reliable. They measure their ability to respond to questions and also manage discussion.

Problems
The use of problems has a potential for measuring application, analysis and problem-solving strategies. Complex problems and their marking schemes can be difficult to design. Marking for grading of easy problems is fast. Marking of complex problems can be slow. Marking for feedback can also be slow. Variation between markers is fairly low when based on model answers or marking schemes. Allow for creative, valid solutions by bright students.

Projects
This includes group projects and dissertations. They have good all-round testing ability and the potential for sampling a wide range of practical, analytical and interpretative skills. They also provide a measure of project and time management. Group projects can provide a measure of teamwork skills and leadership, motivation and teamwork. Marking for grading can be time-consuming. Marking for feedback can be achieved through peer and self-assessment and presentations. Learning gains can be high particularly if reflective learning is part of the criteria. Projects test methods and processes as well as end results. There can be variations between markers. The use of criteria reduces variability but variations of challenge of project or dissertation can affect reliability.

Questionnaires and report forms
Structured questionnaires may elicit the information wanted but semi, or open-ended, questionnaires may

provide the information needed. A mixture of structured and open-ended questions is therefore recommended. Criterion reference grading is recommended for judgemental purposes. Broad criteria are more reliable and valid than highly detailed criteria. Detailed criteria tempt users to react negatively or disdainfully.

Reflective practice assignments
This measures student capacity to analyse and evaluate experience in the light of theories and research evidence. It is relatively easy to set. It provides feedback potential from peers, self and tutors. Marking for feedback can be slow. Marking time for grading is about the same as for essays. Use of criteria reduces variability.

Reports on practicals
This is a standard method. It has potential for measuring knowledge of experimental procedures and analysis and interpretation of results. It measures 'know-how' of practical skills but not skills themselves. Marking for grading using impressions or simple structured forms is relatively fast. Marking for feedback with simple structured forms is faster than without them. Variations between markers, without structured forms, can be high. The method is sometimes over-used. To reduce student workload and the assessment load, different foci of assessment for different experiments are recommended.

Self-assessed questions based on open learning (distance learning materials and computer-based approaches)
Strictly speaking, this is a method of learning, not of assessment, but it could be used more widely for assessment purposes. Self-assessed questions could form an integral part of open learning. These could be based on checklists, MCQs, short-answer questions, MEQs and other methods. Their primary purpose is to provide feedback and guidance to the users. They can be used to integrate open learning and practice-based learning if students visit general dental practices or community health clinics. Reliability and validity is probably moderately high and manageability is high in the long term, but initially it is low.

Short-answer questions
This standard method has potential for measuring analysis, application of knowledge, problem-solving and evaluative skills. Questions are easier to design than complex MCQs but still relatively slow. Marking to model answers is relatively fast compared with marking problems but not compared with MCQs. Marking for feedback can also be relatively fast.

Simulated interviews
These are useful for assessing oral communication skills and for developing ways of giving and receiving feedback on performance. Video-recorded sessions take more time but are more useful for feedback and assessment. Peer and self-assessment can be used. Sensitive oral feedback on performance is advisable. Assessment by simple rating schedule or checklist is potentially reliable if assessors, including students, are trained.

Single essay examination
This is usually structured on a 3-hour response on a prepared topic. It is relatively easy to set but attention to criteria is needed. This method tests a wider range of ability including capacity to draw on a wide range of knowledge, to synthesize and identify recurrent themes. Marking for feedback is relatively slow. Marking for grading is relatively fast providing the criteria are simple.

Work based assessment
This includes a variety of methods including learning logs, portfolios, projects and structured reports from supervisors or mentors. It is important to provide training for supervisors and mentors in the use of criteria. Work experiences can be variable so reliability can be low. Validity, as usual, is dependent upon clear learning outcomes.

Open book assessments
This format focuses on skills of information sourcing and ability to perform critical analysis, prioritization and summarization of information. It is increasingly relevant in an age where easy access to information sources reduce the need for memorization of large volumes of information.

Conclusions

- There is general agreement that the quality of the assessment is a major determinant of undergraduate learning.
- The range of assessment methods available for monitoring knowledge is wide and well documented. Educators who follow the basic guidelines summarized in this report and widely available in the literature should be relatively reassured about their assessment of knowledge.
- There is widespread and growing concern about the integrity of the assessment of many components of clinical performance in dental education. Even the

assessment of a student's physical skills, long recognized as the core of dental practice, is a source of concern. Consistency, objectivity, training and performance of assessors are among the topics that need to be addressed. Questions still need to be answered about what 'assessment of competence' really means, the (lack of) reliability of day-to-day continuous clinical assessment and the value of traditional, unstructured forms of assessment, such as the viva voce examination. Some essential components of clinical competence (e.g. professional behaviour, attitudes and communication skills), which are central to effective clinical practice, have always been regarded as difficult to measure. Many educators are now recognizing the need for a structured approach to the teaching and assessment of professional behaviour, communication skills and attitudes in the curriculum.

- New emphasis on the importance of the learning process and the need for active lifelong learning has inevitably led to the search for tools to measure this important aspect, although much work remains to be conducted.
- Careful organization and synchronization of the many disparate initiatives currently in progress within dental education and the health care professions in general should enable more rapid solutions to the many challenges which remain within the area of assessment.

Building and growing a thematic network

This was not discussed in any detail. Please refer to the reports of 'Towards global convergence of education, training, quality, outcome and assessment' and 'Web-based interactive learning programmes'.

Recommendations, realistic goals and a time frame

1. There is irrefutable evidence that assessment is central to the success of education programmes. Therefore, reference to 'Best practices and innovations' and 'Core values applicable to all' in this report (above), dealing with best practices and core values in assessment, is strongly recommended.
2. It is recommended that dental educators ensure that assessment practices, which are already established within their curriculum, are reviewed and modified as appropriate in the light of the best current evidence of efficacy.

3. It is further recommended that educators involved in the design and implementation of new curricular strategies explore the evidence for the need to introduce new forms of assessment that are in alignment with curricular initiatives.
4. We propose the development of a multidisciplinary (intradental and non-dental) resource group including those from the present group who have the time and interest to continue, but augmented by recruitment of the international education community.
5. It is recommended that an open access resource be put in place to facilitate the development of maximum achievable standards in assessment practices for all dental schools wishing to be involved at any level. In establishing such a resource, care should be taken to identify other national and international groups with similar objectives and liaise with them in order to reduce duplication of effort. Development of an effective resource will require that individual members of the group choose and concentrate on specialist areas (initially these may be based on the parameters chosen in this present report). Accumulated information on best assessment practices and literature relating to each specialist area will be accessed either on the web or through an electronic distribution list.
5. The group is conscious that workshops and personal communication play a vital role in development and dissemination of good practice. It is therefore recommended that the group plans assessment events to be held in association with, or as satellite workshops to, established educational meetings such as ADEE, ADEA and IADR in 2002 and with the second phase of the DentEd Evolves project in 2003. The proposed format would be a combination of keynote lectures by experts on assessment and workshops run by members of the working group.
7. Consideration should be given to a longer-term aim, to maintain and develop a more permanent group dedicated to the promotion of good practice of assessment in dental education. This should include the establishment of an annual or biennial assessment meeting either in isolation or by arrangement with an existing assessment meeting such as the Ottawa Conference.

Additional reading

Ashworth P, Saxton J. Managing work experience. London: Routledge, 1992.

Biggs J. Enhancing teaching through constructive alignment. Higher Educ 1997: 32: 347–364.

Biggs J. Teaching for quality learning at university. Buckingham: Open University Press, 1999.

Bloom BS. A taxonomy of educational objectives. Handbook I: Cognitive domain, 2nd edn. New York: McKay, 1965.

Brennan J, Little B. A review of work based learning. London: Department for Education and Employment, 1996.

Heywood J. Assessment in higher education: student learning, teaching, programmes and institutions. London: Jessica Kingsley Publishers, 2000.

Murphy RJH. Removing the marks from examination scripts before remarking them. Br J Educ Psychol 1979: 49: 73–78.

Newton PE. The reliability of marking of general certificate of secondary education scripts: mathematics and English. Br Educ Res J 1996: 22: 405–420.

Partington J, Brown G, Gordon G. Handbook for external examiners in higher education. Sheffield: Committee of Vice Chancellors and Principals, UCoSDA, 1993.

Partington J et al. Improving examiners meetings: a video-training package. Sheffield: Committee of Vice Chancellors and Principals, UCoSDA, 1995.

Quality Assurance Agency (QAA). Code of practice for external examiners. Gloucester: Quality Assurance Agency for Higher Education, 1999 (available at http://www.qaa.ac.uk).

Quality Assurance Agency (QAA). Handbook for academic review. Gloucester: Quality Assurance Agency for Higher Education, 2000 (available at http://www.qaa.ac.uk).

Ramsden R, Entwistle NJ. Effects of academic departments on students' approaches to studying. Br J Educ Psychol 1981: 51: 368–383.

Wisker G, ed. Making the most of the external examiner. Birmingham: SEDA, 2001.

References

1. Brown G, Bull J, Pendlebury M, eds. Assessing student learning in higher education. London: Routledge, 1997.
2. Godfrey J, Heylings D. Guide to assessment of students' progress and achievements. London: Queen Mary and Westfield College, 1997.
3. Boud D. Enhancing learning through self-assessment. London: Routledge, 1995.
4. Gibbs G. Improving the quality of student learning. Bristol: Technical and Educational Services, 1992.
5. Entwhistle NJ. Styles of learning and teaching. London: Routledge, 1981.
6. Rowntree D. Assessing students – how shall we know them? London: Kogan Page Ltd, 1987.
7. Newble DI, Jaeger K. The effects of assessment and examinations on the learning of medical students. Med Educ 1983: 17: 25–31.
8. Biggs J. Student approaches to learning and studying. Melbourne: Australian Council for Educational Research, 1987.
9. Davis JEC, Sales GD. Dental and life science students: a comparison of approaches to study and course perceptions. Med Educ 1996: 30: 453–458.
10. Martenson D, Eriksson H, Ingelman-Sundberg M. Medical chemistry. evaluation of active and problem-oriented teaching methods. Med Educ 1985: 19: 34–42.
11. Walton HJ, Matthews MB. Essentials of problem-based learning. Med Educ 1989: 23: 542–558.
12. Schmidt HG. Foundations of problem-based learning: some explanatory notes. Med Educ 1993: 27: 422–432.
13. Albanese MA, Mitchell S. Problem based learning. a review of literature on its outcome and implementation issues. Acad Med 1993: 68: 52–81.
14. Barrows HS. A taxonomy of problem-based learning methods. Med Educ 1986: 20: 481–486.
15. Chiodo GT, Bullock WW, Creamer HR, Rosenstein DI. An application of the patient-oriented problem-solving system. J Dent Ed 1991: 55: 327–331.
16. Schon DA. The reflective practitioner. London: Temple Smith, 1983.
17. Gordon MJ. Self-assessment programs and their implications for health professions training. Acad Med 1992: 67: 672–679.
18. Carrotte PV. The assessment of dental students – a University education? Br Dent J 1993: 175: 220–223.
19. General Dental Council. The first five years: the undergraduate dental curriculum. London: General Dental Council, 1997.
20. Manogue M, Brown GA, Nattress BR, Fox K. Improving student learning in root canal treatment using self-assessment. Int Endod J 1999: 32: 397–405.
21. Rohlin M, Roger-Leroi V, Miotti F et al. Assessment Methods. In: Shanley DB et al., eds. Dental education in europe, towards convergence. Budapest: Dental Press kft, 2001: 100–112.
22. Stephens D, Bull J, Wade W. Computer-assisted assessment: suggested guidelines for an institutional strategy. Assess Eval Higher Educ 1998: 23: 283–294.
23. Kingsley J, Leonard L, Sutcliffe R, McAteer E. Development and evaluation of computer based assessment for 1st year biology students. ALT-C97. Telford: University of Wolverhampton, 1997.
24. Gunn C, Maxwell L. CAL in human anatomy. J Comp Assist Learning 1996: 12: 205–215.
25. Sewell RDE, Stevens RG, Lewis DGA. Multimedia computer technology as a tool for teaching and assessment of biological science. J Biol Educ 1995: 29: 27–32.
26. Bull J. Computer-assisted assessment: impact on higher education institutions. Educ Technol Soc 1999: 3(2). Online journal available at: http://ifets.ieee.org/periodical/Vol, 399/joannabull.html
27. Biggs J. What do inventories of students' learning processes really measure. Br J Educ Psychol 1993: 63: 3–19.
28. Norman GR, Tugwell P, Feightner JW, Muzzin LJ, Jacoby LL. Knowledge and clinical problem-solving. Med Educ 1985: 19: 344–356.
29. Patel VL, Groen GJ. Knowledge based solution strategies in medical reasoning. Cogn Sci 1986: 10: 91–116.
30. Schmidt HG, Norman GR, Boshuisen HPA. A cognitive perspective on medical expertise – theory and implications. Acad Med 1990: 65: 611–621.
31. Whittlesea BWA, Brooks LR, Westcott C. After the learning is over – factors controlling the selective application of general and particular knowledge. J Exp Psychol Learn Mem Cogn 1994: 20: 259–274.
32. Whittlesea BWA, Dorken M. Incidentally, things in general are particularly determined. an episodic processing account of implicit learning. J Exp Psychol Gen 1993: 122: 227–248.

33. van der Vleuten C, Newble D. Guidelines for assessing clinical competence. Teach Learn Med 1994: 6: 213–220.
34. DeGrave WS, Boshuizen HPA, Schmidt HG. Problem based learning: cognitive and metacognitive processes during problem analysis. Instruct Sci 1996: 24: 321–341.
35. McManus IC, Richards P, Winder BC, Sproston KA. Clinical experience, performance in final examinations, learning style in medical students: a prospective study. Br Med J 1998: 316: 345–350.
36. Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990: 65(Suppl. 9): S63–S67.

Address:
Michael Manogue
School of Dentistry
University of Leeds
Clarendon Way
Leeds
LS2 9LU
UK