Portfolios as a method of student assessment

AMEE Medical Education Guide No 24

This AMEE Education Guide was first published in Medical Teacher:

Friedman Ben David M, Davis M H, Harden R M, Howie P W, Ker J and Pippard M J (2001) AMEE Medical Education Guide No 24: Portfolios as a method of student assessment. Medical Teacher 23(6).

The Authors

Miriam Friedman Ben David is Professor of Medical Education, formerly Co-Director of the Clinical Skills Certification Programme at the Educational Commission for Foreign Medical Graduates (ECFMG) and Associate Dean at the Medical College of Pennsylvania, Philadelphia. Professor Friedman Ben David is currently at the Centre for Medical Education, Dundee, Scotland. Her main areas of interest are physician and student assessment, programme evaluation and development of leadership in the medical profession.

Margery Davis is Senior Lecturer in Medical Education at the University of Dundee's Centre for Medical Education. She has provided educational support for the Dundee Medical School since 1992.

Ronald M Harden is Director of the Centre for Medical Education, University of Dundee and Director of the Scottish Council for Postgraduate Medical and Dental Education's Education Development Unit.

Peter Howie is Professor of Obstetrics and Gynaecology at the University of Dundee and Chairman of the Scottish Council for Postgraduate Medical and Dental Education. He has had a long-standing interest in medical education and the development of portfolio assessment in particular.

Jean Ker is a Clinical Lecturer in Medical Education in the School of Medicine, University of Dundee. She is a practising General Practitioner and provides support to the undergraduate curriculum in clinical skills and assessment.

Martin J Pippard has been Professor of Haematology in Ninewells Hospital and Medical School, University of Dundee, since 1989. He trained in Oxford UK and Seattle USA and was formerly Consultant Haematologist at the MRC Clinical Research Centre and Northwick Park Hospital, Harrow UK. As convenor for Phase 2 (years 4 and 5) of the Dundee undergraduate medical curriculum, he has coordinated the implementation of this phase of a new curriculum over the last 3 years.

Guide Series Editor: Pat Lilley

Production and Desktop Publishing: Lynn Bell

ISBN: 1-903934-04-4

Copies of this guide are available from:

AMEE, Centre for Medical Education, Tay Park House, 484 Perth Road, Dundee DD2 1LR, UK
Tel: +44 (0)1382 631953; Fax: +44 (0)1382 645748; Email: amee@dundee.ac.uk

Contents

Purpose of this guide

This guide informs medical teachers about the use of portfolios for the purpose of student assessment. It looks at:

- What are portfolios?
- Why use portfolios and the contributions that portfolios can make to student assessment and the psychometric issues relating to use of portfolios in assessment.
- Guidelines for the implementation of portfolio assessment.

The educational/theoretical considerations presented in this guide along with principles of practical applications should enable the reader to design and implement a portfolio assessment.

Summary

This guide is intended to inform medical teachers about the use of portfolios for student assessment. It provides a background to the topic, reviews the range of assessment purposes for which portfolios have been used, identifies possible portfolio contents and outlines the advantages of portfolio assessment with particular focus on assessing professionalism. The experience of one medical school, the University of Dundee, is presented as a case study. The current state of understanding of the technical, psychometric issues relating to portfolio assessment is clarified. The final part of the paper provides a practical guide for those wishing to design and implement portfolio assessment in their own institutions. Five steps in the portfolio assessment process are identified: documentation, reflection, evaluation, defence and decision. It is concluded that portfolio assessment is an important addition to the assessor's toolkit. Reasons for using portfolios for assessment purposes include the impact that they have in driving student learning and their ability to measure outcomes such as professionalism that are difficult to assess using traditional methods.

A: Portfolios as a method of student assessment

1 *Background to the use of portfolios for assessment purposes*

Recent changes in medical education and training emphasise the development of teaching and learning programmes which meet the needs of the medical profession and society as a whole. Professional organisations and the public at large demand demonstration of professional attributes which ensure doctors' fitness for practice while adhering to high standards of care (GMC 1997).

The medical schools and postgraduate training programmes have responded by introducing new teaching and learning strategies which will enhance doctors' accountability. One important innovation in curriculum reform is the development of an outcome-based educational framework (Harden et al 1999). Learning outcomes of undergraduate and postgraduate training are defined broadly to allow students to recognise their progress along the scope and depth of their professional competences. The use of such frameworks facilitates the development of traditional learning outcomes, such as clinical skills, as well as learning outcomes defining the doctor as a professional which, in the past, have been ignored. The growing interest and concern with doctors' professional attributes is shared worldwide as indicated by the vast number of medical education papers published on the topic of professionalism during the year 2000 alone (Ginsburg 2000).

Furthermore, the concept of professionalism has filtered into teaching and learning strategies where students are encouraged to take responsibility for their own learning, and personalise their learning experiences. Student-centred programmes increasingly incorporate the learners' needs, as well as listen to the learners' voice. Consequently, medical schools' curricula are seeking to broaden student experiences inside and outside the hospitals to allow better appreciation of the multi-context system within which doctors are now practising. The new direction of medical schools' curricula results in somewhat less structure and more authentic experiences for the trainee which, in turn, increases self-directed learning.

Concern with the lack of continuity between undergraduate and postgraduate education is another important factor. Transition from the undergraduate to the postgraduate phase of medical education should be consistent with progression from technical discrete abilities to full integration of professional competencies, which is the ultimate outcome of medical training.

Concurrent to these educational reforms, new assessment strategies are being developed to meet the needs of recent innovations in the health professions. The search for new assessment tools is a reaction against existing methods of assessment, which often have had adverse effects on the learner, the teacher and the curriculum as a whole. It is recognised that assessment tools should enhance and support learning as well as measure performance. Much current interest is in authentic, performance-based assessment (Koretz et al, 1998) which also encourages learners to take responsibility for their own learning and guides the learners to accumulate evidence of learning, while incorporating a criterion-referenced interpretation of their performance.

The portfolio is an attempt to counteract the limitations of a reductionist approach to assessment. It facilitates assessment of integrated and complex abilities and takes account of the level and context of learning. It provides an assessment solution for a curriculum that designs learning along broad educational and professional outcomes. It personalises the assessment process while incorporating important educational values. It supports the important principle of "Learning Through Assessment" (Friedman 2000).

This guide describes the use of portfolios as a tool for assessment and explores the contributions they can make. It examines how they can be implemented in practice and looks at their psychometric properties. A previous AMEE guide, Portfolio-based Learning and Assessment (Snadden et al 1999), introduced the concept of portfolios in undergraduate, postgraduate and continuing medical education and focused on their content and construction.

2 *What are portfolios?*

Definitions

The idea of a portfolio is drawn from the study of art, "where the term signifies a purposeful collection of work". *Stecher 1998*

"A portfolio is a collection of papers and other forms of evidence that learning has taken place".

Davis et al 2001

"It is a collection of student work that exhibits the student's efforts, progress and achievements in one or more areas"… "This collection represents a personal investment on the part of the student – an investment that is evident through the student's participation in the selection of the contents, the criteria for selection, the criteria for judging the merit of the collection and the student's self reflection".

Gisselle O, Martin-Kniep 2000

These definitions imply that the purpose of student assessment will guide the content of the portfolio, i.e. best work, typical work, most diverse work, work on a theme, or work on one project as it evolves over time (Stecher 1998). Definition of purpose, from an assessment perspective, will determine the portfolio content, the process of creating it and the interpretation of the evidence. Therefore, portfolios will vary from context to context depending on their specialised purpose. Most portfolios will share a common characteristic of being cumulative, i.e. contain work completed over a period of time relevant to students' ongoing work or studies.

3 *Types of portfolios*

Portfolios are used in all stages of education: elementary, higher education and in professional and vocational programmes.

Portfolios are used in elementary school classrooms as part of ongoing assignments (Herman & Winter 1994). Students prepare draft written assignments over time and, at a pre-determined date, choose the best work for a "show case". Such portfolios could be scored at a school level or at state level (at a state scoring site) - for high stakes examinations or for quality control procedures.

In the USA, the State of Vermont introduced in 1992 a large-scale portfolio assessment for grades 4 to 8 in two subjects, mathematics and writing. Students select from their on-going assignments in mathematics, five to seven best pieces of work which are scored on seven dimensions of performance, three pertaining to communication and four to problem solving. The students submit a portfolio with one best piece of work and other specified pieces. The best piece was scored separately and the remaining pieces were scored as a set. The writing pieces were scored on five dimensions: purpose, organisation, details, voice/tone, usage/mechanism/ grammar. Portfolio materials were scored by teachers to provide individual scores and by the State to provide school profiles in those two subjects (Koretz 1998).

The State of Kentucky developed a similar large-scale portfolio assessment in writing for 4th grade students. Over a period of one year, the students compiled a portfolio comprising six pieces: a personal narrative, a poem, a play, a piece of fiction, one information or persuasive piece, one piece from any subject area other than English and language arts, a best piece and a letter to the reviewer about the best piece and student's growth as a writer (Koretz 1998).

In the professions, portfolios are used, for example, in recording nurses' career and professional development through formal and experiential learning for periodic registration with the UK Central Council (Jasper 1995). In medicine, portfolios are used to study critical thinking and self-directed learning in the daily practice of general practitioners (Al Shehri 1995).

The Royal College of Physicians and Surgeons of Canada introduced a learning portfolio, The Maintenance of Competence Program (MOCOMP), using computer software for physicians to create a personal learning portfolio. The collection of

information shows how physicians learn, which factors affect changes in practice and what sources of learning physicians use. It also creates databases for establishing standards of physicians' performance and for continuing medical education programmes (Bashook et al 1998).

The National Council for Vocational Qualifications established assessment procedures for qualification in England, Wales and Northern Ireland in which portfolios play a central role. It is aimed for young people in upper secondary schooling and young people and adults in work or in government training. Wolf (1998) reports that the portfolio approach to assessment has affected the desired learning styles, but major problems have arisen relating to the manageability of the approach and reliability of assessors' judgements.

The General Medical Council in UK has put forward a revalidation proposal for practising physicians (GMC 2000). The 3-stage proposal suggests a portfolio (folder) which will include information and evidence regarding doctors' cumulative performance to be assessed by groups of doctors.

4 *What portfolios might contain*

Portfolios contain students' ongoing work over time which may provide evidence for learning and progress towards educational and professional outcomes or learning objectives. These may include maintenance of competences, keeping up to date with professional practice, fitness for practice, adherence to professional standards and more. Portfolios may also contain student's or candidate's reflection on the submitted evidence.

By definition any material which provides evidence for the above mentioned educational and work related criteria can be included in portfolios. This may include:

- best essays
- written reports or research projects
- samples of evaluation of performance, e.g. tutor reports from clinical attachments
- video tapes of interactions with patients or with peers
- records of practical procedures undertaken (log books)
- annotated patient records
- letters of recommendation
- CVs
- written reflection on the evidence and on professional growth.

Evidence included in portfolios is limited only by the degree of the designer's creativity.

5 *Why use portfolios?*

The use of portfolios for students' assessment enables students and teachers to engage in a process of learning through assessment. This implies that the assessment procedure not only measures and reinforces the desired learning outcomes but rather enhances the development of strategies, attitudes, skills and cognitive processes essential for life long learning. Consequently, the use of portfolios not only broadens the scope of assessment but also introduces a number of educational benefits.

Portfolios' contribution to assessment

- *The assessment of learning outcomes*:
  This includes outcomes not easily assessed by other methods, e.g. personal growth, self-directed learning, reflective ability, self-assessment of personal growth, professionalism and more. An extended discussion of the use of portfolios for assessment of professionalism appears at the end of this section.

- *The provision of evidence of performance*:
  This evidence may be collected from a range of sources, e.g. student's on-going work over time, over settings and over subject matters.

- *The representation of evidence collected over a period of time*:
  Portfolios provide evidence of student development over time and not just a snapshot at one time or a series of snapshots, as in traditional assessment methods.

- *Student progression towards the learning outcomes*:
  Portfolios allow assessment of progress towards the learning outcomes by using chronological work samples collected at different points of time.

- *Summative and formative assessment*:
  The evaluation of the portfolio content generates summative statements regarding student performance for promotion or pass/fail decisions. It also provides an in-depth profile of student abilities. Summative evaluation statements may also contain information regarding students' strengths and weaknesses, thus integrating summative decisions with formative profiles.

  The current challenge facing portfolio assessment is to judge the qualitative and quantitative evidence and yet maintain reliability and validity

of the judgement. Section C in this guide expands the discussion of this issue.

In addition to the contribution that the portfolio makes to student assessment, it also reinforces other important educational aspects central to effective teaching and learning.

Focus on personal attributes

- Provides personal and professional educational evidence for student learning (Glen and Hight, 1992).
- Makes provision for students to receive feedback about their own personal values, feelings, styles of interaction and ways of handling significant experiences (Beattie 1991)
- Provides a personalised sensitive portrait of what students know and are able to do (Herman & Winter, 1994).
- Contextualises learning and links experience with personal interpretation.

Enhances interactions between students and teachers

- Allows dialogue between students and educators and ensures individual protected time.
- Reminds students that learning is a two-way process between learner and educator.
- Mirrors students' and teachers' work and stimulates teachers to re-assess teaching strategies (Aschlacher 1993).
- Raises teachers' expectations in relation to thinking ability and problem solving (Finlay et al 1998).

Stimulates the use of reflective strategies

- Facilitates the use of past experiences to define learning and to recognise progress.
- Stimulates the use of reflective skills in order to analyse and synthesise experiences. Students can also describe and analyse learning strategies thus engaging in metacognitive processes.
- Allows educators to separate the quality of evidence from the student ability to reflect on the evidence.

Expands understanding of professional competence

- The variations in student perception and interpretation of their experiences increases understanding of students' professional growth. Those newly acquired insights inform medical education and practice (Herman & Winter 1994).
- Encourages a holistic and integrative approach to medical practice.

6 *The assessment of professionalism*

Why assess professionalism?

The acquisition of the professional behaviour required to practise medicine is slowly emerging as a central focus in the new undergraduate medical curriculum (Cohen 2001) and with it increasing consensus on professional behaviour (Swick 2000).

The introduction of new methods of assessment to expand professional horizons (Friedman 2000) presents the profession with the opportunity to realign curricula to incorporate the development of professionalism as a central outcome and find ways to assess professionalism and professional growth.

What is professionalism?

One of the unique problems for medicine in defining professional identity is that it develops from practice which in recent times has seen such enormous changes both in biomedical advances and in shifting role boundaries in the delivery of care (Shapiro 1999). There has been little appreciation that professionalism is a dynamic concept which needs to be defined in the context of changing expectations of society and changing times. Recently, however, clearer definitions of professionalism have started to emerge. For example, reflective ability is increasingly being identified as a key component of medical professionalism (Grundy 1987, Lundberg 1991, Parboosingh 1996, Wear 1997, Hafferty 2001). This builds on Schon's work in relation to the development of the reflective practitioner for professional practice (1983).

Reflection requires both cognitive and humanitarian attributes. It is a process that enables practitioners to determine their own actions, critically review these actions and act on the outcome in the best interest of the client or patient.

Project Professionalism, which was introduced by the American Board of Internal Medicine (ABIM) (1994), defined medical professionalism as an ideology consisting of the following elements: a commitment to:

- The highest standards of excellence in the practice of medicine and in the generation and dissemination of knowledge (cognitive attributes)
- Sustain the interests and welfare of patients
- Be responsive to the health needs of society

These elements were further defined as the humanitarian qualities of altruism, accountability, excellence, duty, honour, integrity and respect for others. These are required of all candidates seeking certification and recertification from the ABIM.

The Association of American Medical Colleges (AAMC) (1998), in their Medical Schools Objectives project, have also focused on professional development in relation to both humanitarian and cognitive qualities. They have identified four core attributes that all medical students should be able to demonstrate on graduation. The four attributes include altruism, knowledge (cognitive) skill and dutifulness. They further identify learning objectives for each of these attributes. In relation to altruism AAMC include characteristics such as compassion, honesty, integrity, team relations, commitment to patients interests, understanding threats to medical professionalism and an awareness of one's own limitations.

In the UK, the Scottish Deans Medical Curriculum Group (2000), which consists of five medical schools has identified a Scottish medical graduate as a competent and reflective practitioner. In their new outcome based model, they have identified an aptitude for personal and professional development as one of the twelve curriculum outcomes of a Scottish doctor at graduation. The key features of this outcome include self awareness, (developing reflective practice to be aware of one's own competence), self learning (identifying one's own strengths and weaknesses) self care (being aware of internal and external factors on one's own professional development, career choice and commitment).

Why are portfolios a useful means of assessing professionalism?

The development of multiple sampling in the assessment of clinical practice is receiving increasing attention particularly in relation to the assessment of professional behaviour. Turnbull (2000) has used a system of clinical work sampling to assess students during a rotation and found it to be a reliable, valid and feasible form of assessment.

Portfolios, which sample evidence over time, also demonstrate a potential and reliable tool for the assessment of acquisition of knowledge, skills, attitudes, understanding and achievements and "...reflect the current stages of development and activity of the individual" (Brown 1995). The use of portfolio assessment has recently been introduced at the University of Dundee School of Medicine (Davis 2001) as a key component in the final undergraduate degree examination, in order to assess aspects of medical practice such as professional development.

Portfolios present an ideal method of assessing professional development as they enable the assessment of the doctor's progress matched against professional criteria over time.

Portfolios also enable professional development to be assessed at different stages of a practitioner's career.

The sampling of work-related activity reflects the reality or authenticity of an individual's professional clinical practice. Having a variety of authentic events recorded in the portfolio enables the examiner to identify any cognitive or humanitarian deficiencies that candidates may have in relation to their professional development.

The portfolio assessment process allocates time for the examiner to focus on the students' development, during the oral component of the assessment, by asking the examinees to reflect on their professional role and related experiences.

The portfolio provides an ideal context to assess reflective ability as a key component of professional development. Time is provided during the portfolio assessment for examiners to assess the ability of students to reflect in relation to their portfolio evidence. In addition, reflective writing for a portfolio assessment demonstrates the examinee's experience separated from his/her personal interpretation of the experience. Organising students' experiences through reflective writing modifies their perception of the experience and enables it to be integrated into prior learning or used to initiate new learning (Eisner 1991).

The portfolio assessment process also provides a safe opportunity to explore errors in professional medical practice and strategies to avoid future errors. These sensitive matters can be discussed within the outcome of an aptitude for personal and professional development.

B: The Dundee case study

Dundee medical school moved to outcome-based education in the 1996-1997 academic year with 12 learning outcomes identified. A three-circle classification model was developed for the presentation of the learning outcomes (Harden et al 1999). The inner circle describes "**what the doctor is able to do**" and this includes seven competencies: 1) clinical skills, 2) practical procedures, 3) investigating a patient, 4) patient management, 5) health promotion and disease prevention, 6) communication skills and 7) information handling and retrieval.

The middle circle describes "**how the doctor approaches the task**" and it includes 3 competencies: 8) understanding of basic, clinical and social sciences, 9) appropriate attitudes, ethical understanding and legal responsibilities, 10) appropriate decision making, clinical reasoning and judgement.

The outer circle describes the "**doctor as a professional**" and it includes: 11) the role of the doctor within the health service and 12) aptitude for personal development.

In Dundee University School of Medicine, the learning outcomes provide the framework for student learning, teaching and assessment. Individual students' progress towards the 12 outcomes, is the focus of the portfolio assessment. In their portfolios the students provide the evidence that they have achieved the outcomes at the appropriate level expected of a graduating student. The outcome definitions are the criteria for the portfolio assessment.

The final examinations were redesigned to meet the needs of the Dundee new curriculum (Davis et al 2001). The curriculum incorporated an integrated, task-based approach that is interwoven with the outcome model. The final examinations are administered in two parts: Part 1 occurs in the end of year 4 when students have completed their core clinical attachments. It includes a knowledge and problem solving component measured by extended matching items (EMI), and constructed response questions (CRQ); and an OSCE to assess clinical skills. Part 2 of the final examinations takes place towards the end of year 5 by employing a portfolio assessment. Students and faculty were given the opportunity to evaluate, in a holistic approach, student progress towards all the outcomes including those dealing with the doctor as a professional and personal development which are not easily assessed by traditional assessment methods. The content of the portfolio includes students' summaries of progress towards each of the outcomes and a variety of materials prepared during years 4 and 5 (Table 1). The notes to examiners and students contain detailed descriptions of student work which are listed in Table 1.

The portfolio examination in Dundee is part of a sequential assessment process in that students need to pass the year 4 final examinations before they are permitted to sit for the portfolio. End of year 5 students, who pass the portfolio examination, require no further assessment and thus fulfil requirements for graduation. Students who present with minor or major deficiencies are referred either for further remedial work or for a year 5 final examination comprising an OSCE and review of the portfolio after appropriate correction or additions.

Figure 1 outlines the process of portfolio examination in Dundee from the examiners' perspective. Two examiners independently read each student portfolio a few days prior to the date of the portfolio oral review with the student. The two examiners independently grade the portfolio content employing criteria and guidelines set out by the medical school. They use a marking sheet, the portfolio assessment scoring sheet (PAS) to insert their outcome grades based on their reading of students' portfolio work. The examiners' gradings of the 12 outcomes are independent but take account of the already pre-marked student work. They focus on the overall patterns of achievement and progress over the two years rather than the grading of specific pieces of evidence. They make judgements on how the different sets of portfolio evidence relate to the student progress towards the 12 outcomes. In a pre-review session on the day of the oral review, the two examiners meet and discuss the students' portfolios. They agree on the students' strengths and weaknesses and specific issues that should be explored during the oral review of the portfolio with the student. During the oral review of the portfolio with the student, which lasts 40 minutes, the examiners pursue the issues of concern and, at the end, they independently assign grades for each outcome based on students' performance during the oral review. The examiners then reach consensus on a third set of outcome grades which constitutes the final grading based on which the pass/referred decisions are made. They also make recommendations for distinction. At the end of each examination day, all examiners for this day meet as an Examiners Committee to consider the referred and distinction students, to challenge the decisions

Student work	Year	Number	Pre-marked*
Student's personal summary of progress towards each outcome.	5	12	No
Patient presentations: short summaries of patients seen by the student, selected from the 100 core clinical problems on which teaching and learning is based on this phase of the curriculum.	4	10	Sometimes marked
Case discussions: reports of approximately 1500 words, each analysing a patient's history and findings in terms of one of the curriculum themes.	5	7	Yes
Year 4 assignment: a project report with a grade awarded by the project supervisor together with feedback for the students.	4	1	Yes
Record of achievement: a record of procedures that the student was expected to have completed or observed during the phase.	4&5	1	Signed by Faculty
GP special study module assessment form: a report on the student by their general practice supervisor with a grade awarded.	5	1	Yes
Clinical special study module assessment form: theme special study module assessment form: a report on student performance during the module with a grade awarded for relevant outcomes.	5	2	Yes
PRHO apprenticeship in medicine assessment form: a learning contract between the student and his/her educational supervisor with grades awarded for each learning outcome.	5	1	Yes
PRHO apprenticeship in surgery assessment form: a learning contract between the student and his/her educational supervisor with grades awarded for each learning outcome.	5	1	Yes
Elective report: a report completed by the student after the elective period. This was read by one of the two members of staff responsible for elective studies and written feedback provided.	5	1	Yes

* The pre-marked student work is scored using the University grading system (A-G) and contains comments provided by Faculty.

Table 1: Description of portfolio contents

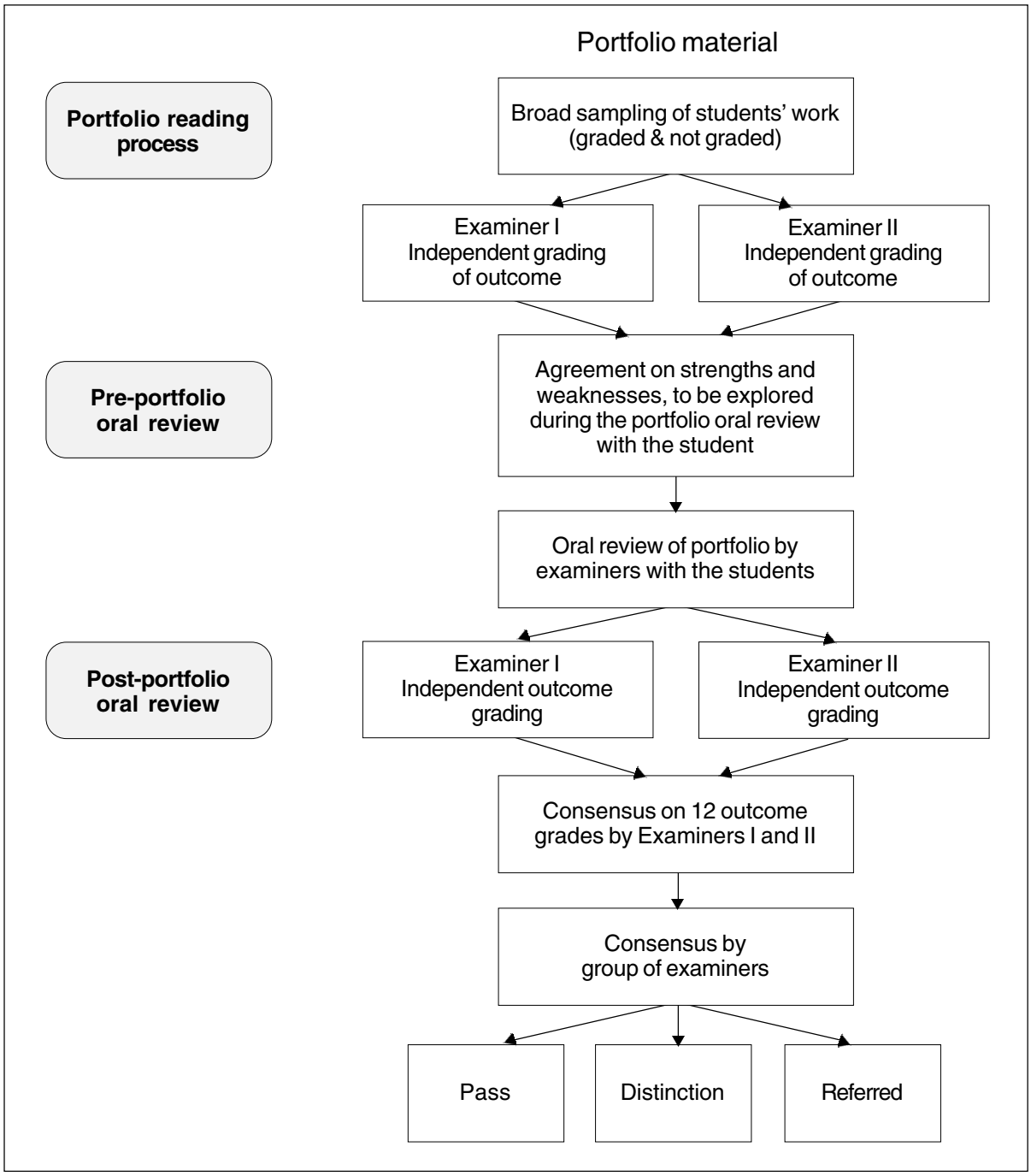


Figure 1: The portfolio examination process: the examiners' perspective

and reach consensus for the final list of the passed, referred and distinction students.

Grades awarded for each outcome correspond to the following criteria:

Grade A – excellent
Grade B – very good
Grade C – satisfactory
Grade D – borderline pass
Grade E – marginal fail
Grade F – definite fail
Grade G – bad fail

Following the Examiners' Committee meeting, students are allocated to one of the following categories:

- Exempt from further pass/fail outcome-based assessment, (all grades are D or above).
- Exempt from further pass/fail outcome-based assessment, but to be considered for distinction. These students will sit for a distinction examination which will determine who is awarded a distinction.

- Potential exemption from further examination (only one outcome is graded E with all others graded D or above). Further work will be assigned to address the limited deficiency identified in the portfolio examination.
- Proceeds to further examination (about 6 weeks later) if two or more outcomes at Grade E or below or one or more outcomes at Grade F or G. Those students will sit an OSCE and a further portfolio assessment.

In summary, the portfolio examination in Dundee is one sequential component of the final examination process which starts at the end of year 4. Students need to pass the year 4 final examination components before they can sit the portfolio assessment. Students who pass the portfolio assessment, meet the final examination requirements. Students who fail the portfolio assessment proceed to either remediation or further testing via an OSCE and further review of the portfolio.

Figure 2 demonstrates the portfolio process from the student perspective. For each phase of the process there is an associated educational value which indicates how the portfolio process supports student learning.

In the beginning of year 4 students learn in advance what is expected of them. As they select the evidence during years 4 and 5, they decide what information they need in order to meet the 12 outcome criteria. When they compile the evidence in the portfolio, they demonstrate their organisational skills, time management and prioritisation ability. The 12 outcome summary sheet allow students to reflect on their progress and prepares them to "think on their feet" during the oral review. Students are asked to complete an evaluation form at the end of the portfolio assessment which allows them to further reflect on the whole process.

Students who pass the portfolio complete the cycle of demonstrating performance achievements. Students who were selected for distinction have further opportunities to demonstrate problem solving ability, information handling and retrieval and knowledge. The ones who were referred to further work or testing have to act on their own deficiencies.

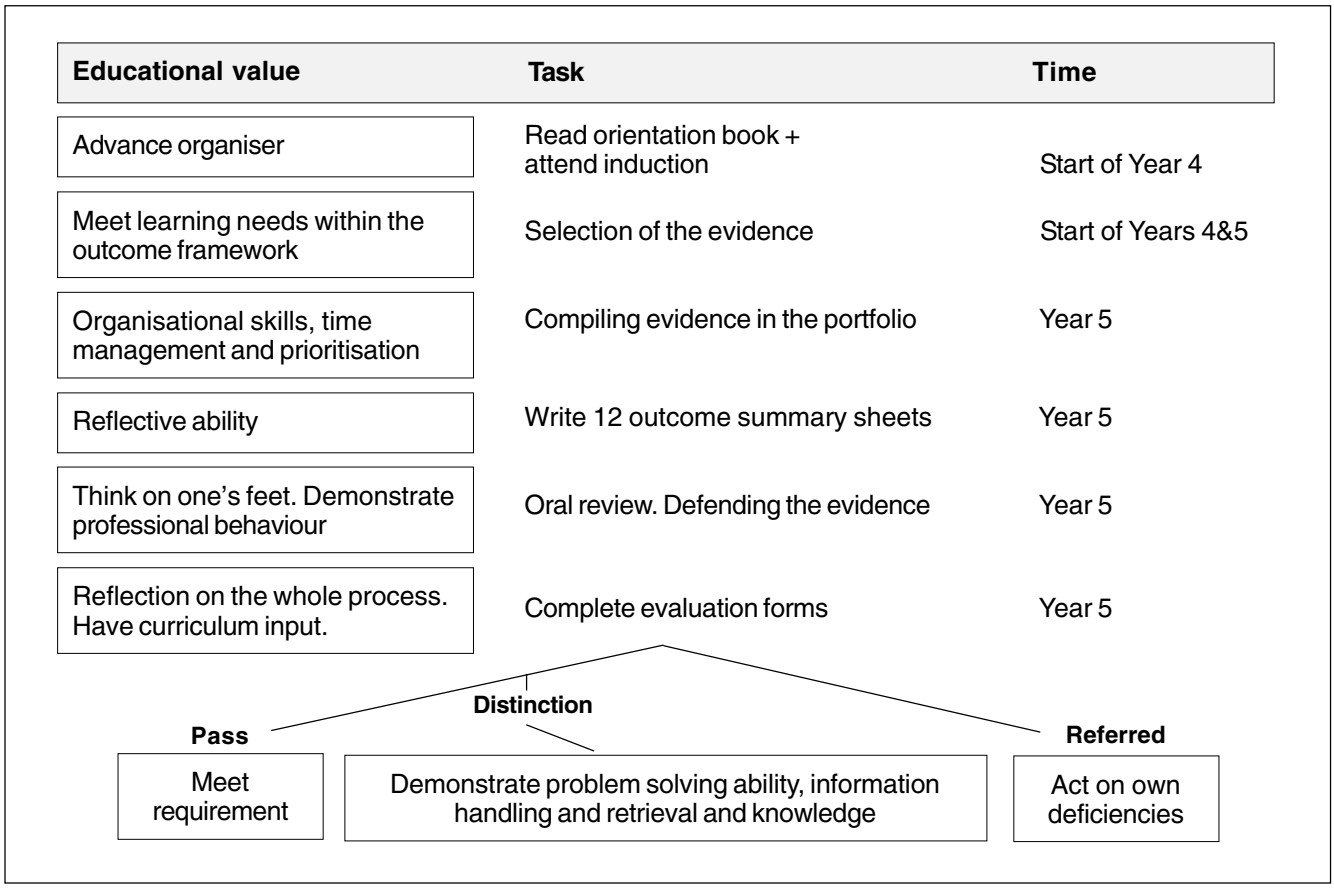


Figure 2: The portfolio process: student perspective

C: Psychometric issues relating to the use of portfolios for assessment

1 *Summary of the portfolio's main assessment features*

Table 2 provides a summary of the portfolio's main assessment features.

1	Formative and summative
2	Qualitative and quantitative
3	Personalised
4	Standardised
5	Authentic

Table 2: The portfolio's main assessment features

Formative and summative

The main assessment characteristics of portfolios focus on the FORMATIVE and SUMMATIVE aspects in which students collect evidence of their own performance for pass/fail decisions, feedback, diagnostic purposes and intervention. The formative value of the portfolio, when linked to summative decisions, present a very powerful assessment tool. In this section, a forward strategy is discussed for the use of portfolios for summative decisions such as promotion, graduation, certification and other high stakes applications.

Qualitative and quantitative

An important feature of portfolios is the combined QUALITATIVE and QUANTITATIVE approach, unlike many other performance tests that rely mainly on the quantitative aspect of performance. The portfolio contains descriptive recorded material as well as graded evidence. This approach to assessment allows the integration of qualitative judgement with quantified information to generate a more comprehensive interpretation of student achievement.

Personalised

The STUDENT CENTERED aspect of portfolios ensures that the student assumes an active role in directing the evidence while teacher involvement is essential in monitoring the process. By personalised, we imply the student individualised selection of evidence, the individualised experience contained in the evidence, the review process by which the examiner learns the unique student's characteristics and the opportunity given to students to reflect and defend their work. This process creates a highly PERSONALISED assessment method. When the portfolio approach is contrasted with the common forms of an OSCE clinical station, the OSCE will focus mainly on STANDARDISATION of the encounter while providing an individual assessment profile.

The portfolio seeks to balance the personalised approach to assessment and the standardisation of the process.

Standardised

It is difficult to achieve maximum control of standardisation for the portfolio content and process. It is possible to determine the type of work to be submitted (case studies, projects, essay, etc). However, human performance with regard to structure, depth, volume, material exhibition is highly variable (Pitts et al 2001).

In their arguments Pitts et at (2001) fear that a too structured approach to developing guidance for using a portfolio as an assessment tool might diminish its strength. The dilemma is how to maintain the personalised aspect of portfolio but at the same time standardize the approach to enhance the portfolio's reliability. Some suggestions for standardisation are provided in Table 3.

1	Same portfolio's units of evidence are assigned to all students.
2	Tasks and criteria for assessment are defined and made clear.
3	Instructions to students provide clear guidelines.
4	The portfolio reading process and rating of material follow standardised guidelines – mainly by written instructions and training workshops for examiners.
5	The probing in an oral review of the portfolio with the student follows standardised guidelines.
6	Pass/fail decisions follow a pre-determined policy.

Table 3: Suggestions for standardisation of portfolio assessment

In cases where portfolio assessment incorporates an oral portfolio review component (note this is not a traditional oral examination) one has to consider strategies to avoid the many pitfalls of oral examination (Missin 1985). Factors which account for problems in oral examinations include: taxonomic level of questions (Des Marchais 1995), the subjectivity of examiners, the lack of standardised criteria employed by examiners either for probing or for grading, lack of shared interpretation of examinees' behaviour and varied levels of expectations from examinees. All may contribute significantly to lack of standardisation and low reliability.

However, some of these problems could be rectified by training and by extensive reading of individual student's portfolios which helps prepare the examiners for a contextualised and personalised oral encounter while maintaining a holistic framework of student performance.

The standardisation process is central to the reliability and the validity of the portfolio. The validity of grading relates to the interpretation by the examiners of students' work, the expectations, shared understanding of criteria for assigning the grade and the use of the results (Messick 1994).

Authentic

Authenticity is defined as "the extent to which the outcomes measured represent appropriate, meaningful, significant and worthwhile forms of human accomplishments". (Archibald and Newmann 1988). Characteristics of authentic achievements as judged in portfolio assessment are provided in Table 4.

1. Production of knowledge rather than reproduction or citing others' work
2. Disciplined enquiry, dependent on:
   - A prior knowledge base (to be used to help produce knowledge).
   - In-depth understanding
   - Integration – the production of knowledge requires the ability to organise, synthesise and integrate information in new ways.
3. Value beyond assessment – aesthetic, utilitarian or personal value.
4. Higher order thinking.

Table 4: Characteristics of authentic achievements as judged in portfolio assessment (Cumming et al 1999)

The portfolio is authentic in the sense that it incorporates past academic and work-related experiences which form the foundations for the above-mentioned criteria. Messick (1994) states that a portfolio is an authentic simulation in that it does not simulate a particular task, but rather requires that the examinees by conceptualizing about the case will generalize to other occasions. The authentic aspect of the portfolio is central to the validity of the portfolio material and the scores' interpretations.

It is also an essential aspect of developing professional attitudes while reflecting on one's own experiences (Schon 1987).

For this reason Case (1994) states her concern for the loss of portfolio authenticity in large scale, more standardised, less personalised examinations. The personalised approach of portfolio assessment is the unique component which contributes to its authenticity, i.e. its validity. The lack of it may increase its reliability but may also decrease its validity.

2 *Raters' consistency*

Discussions of reliability are often centered around terms of "generalisability"; that is, the extent to which performance is consistent across various sources of error labelled "facets". (Koretz 1998). In the portfolio case, raters are considered as one facet of error that threatens generalizability of performance.

The raters' effect includes issues of raters' stability over time (Test - retest), stability over raters (Inter-rater) and reproducibility or decision consistency of pass/fail marks (Livingston 1995).

Given the complexity of portfolio assessment, the amount of time involved in processing the test and the actual knowledge learned about examinees from the first administration, present difficulties for studying raters' stability over a relatively short time. As for inter rater agreement, Le Mahier et al (1993) reported .60 to .70 inter-rater correlations and generalizability estimates between two raters for each unit of portfolio evidence in the range of .80. Herman et al (1995) reported average correlations between pairs of raters of .82. Percentages of absolute agreements of all pairs averaged resulted in 98%.

Koretz (1998) states that much of the total error in scores was attributed to factors other than raters in the portfolio. High agreement depends on clear criteria, adequate examiners' training, communicating criteria to students, good student orientation materials, examiners' familiarity with the context and shared understanding of expected

student performance and the assessment purpose. In the Dundee case study, as in other portfolio assessments, the stability over raters is mainly related to the consistency of their interpretation of *patterns* of performance over time and over tasks, rather than agreement on a single unit of performance or on a single grade. The shared understanding of student progress towards the curriculum learning outcomes in Dundee is the focus of examiners' training. Authentic criteria, which incorporate progressional benchmarks towards the global outcomes, is the Dundee current challenge for achieving high inter-rater reliability. We will further discuss development of criteria in the section regarding implementation. However, one needs to question the meaning of inter-rater consistency in portfolio assessment. For example, in a portfolio which contains one sample of "best work", for the purpose of assessing an ability such as writing, high absolute inter-rater agreement will be essential if decisions are based on the writing score attained. However, for portfolios which contain broad, complex, multiple tasks, one needs to question the likelihood of two independent examiners arriving at identical grading of individual sets of evidence. In this case, it might be important that examiners agree on the consequences of the grading rather than on the actual grade obtained.

3 *Decision consistency*

Misclassification of students into pass/fail or referred categories due to raters' source of error, is a main concern regarding disagreement among raters. The lack of reproducibility as a function of low inter-rater agreement, is a major threat to the reliability of the portfolio examination in high stakes examination or any portfolio examination with summative consequences.

Inter-rater agreements for units of portfolio may result in high correlations. However, what is important is the reproducibility of pass/fail decisions among raters. For example, do the pass/fail standards/ criteria or other policy decisions assigned to the marks produce the same classification consequences across raters? If correlations are high among raters, while one rater consistently rates lower, students who obtained the low ratings may fail based on the examination standards, while students who achieved higher ratings may pass even though the two raters' grading correlated very highly. Consequently, agreements expressed in correlation statistics are not sufficient if pass/fail standards are employed. Correlation statistics and percent agreements should accompany the following questions: 1) What is the extent of agreement for individual units among raters in reaching the same pass/fail results? (Livingston 1995). 2) Irrespective of the grades assigned, do raters reach the same pass/fail decision?

Data available from the first year of the Dundee portfolio administration, shows 98% pass/referred agreement between two pairs of examiners, who independently examined the students. Generalizability coefficients will define the amount of error attributed to raters. It will also assist in estimating the number of raters needed to achieve high reproducibility (Brennan 1983).

4 *Sampling*

Sampling of tasks is another important "facet" of measurement error in portfolio assessment. In fact, given the complexity of behaviours measured by the portfolio assessment, and in order to achieve consistency of performance across a number of tasks, sampling needs to be extensive.

Linn (1994) raises the issue of limited sampling in performance assessment and urges the use of sampling of student work over time and over tasks to reduce sampling error due to task, by employing portfolios and/or multiple sources of performance assessment. Appropriate sampling is a common problem in performance assessment due to limited examination time. The Dundee approach allows such sampling through the preparation of evidence over time, over occasions, employing multiple tasks while addressing the 12 outcomes. This approach is different from a portfolio in which one unit of "best work" is presented as evidence. In the "best work" case, when a limited number of competences are assessed, the support provided by more evidence of the same ability as in the Vermont example, will increase the portfolio validity due to broader sampling.

5 *The consensus approach*

In the Dundee portfolio examination, judgements on student performance are generated by incorporating a consensus approach. Can one employ a consensus approach for pass/fail decisions? How scientific and accurate is the consensus process between two raters or among committee members? There is evidence that scientific theories are determined by a consensus approach rather than by a critical thinking process, or by consistency among independent research findings (Hand 1999).

It is common practice in assessment centres to incorporate a consensus approach for judging trainees performance (Gagler et al 1987). In practice, the selection of trainees and other higher level

personnel is conducted by selection committees which employ a consensus process rather than inter-rater agreements. Juries also operate on a consensus principle. In both cases evidence is presented, the panel argues the case and decision is granted following a consensus process.

Relevant to personalizing assessment in a performance domain, is the issue of how one guards against subjectivity of raters. The portfolio approach, by its complex nature, calls for judgement of multiple performance aspects, cognitive, attitudinal, professional and technical skills that are best expressed at the personalised level. The threats to reliability among raters, in such a complex and personalised undertaking, raises the question: should consensus procedures be recognised as a legitimate and reliable assessment strategy and can consensus procedures replace independent ratings? Linn (1994) mentions the work of Moss (1994) on validity versus reliability and brings examples of search committees examining the credentials of candidates and arriving at an integrated decision. Delandshere and Petrosky (1994) talk about confirmation of prior judgement rather than replication, analogous to a second opinion from a physician. Linn (1994) states that the confirmation approach may be a "useful way of looking at student portfolios and exhibitions?" However, in such cases examinee performance is judged by internal examiners who are familiar with the educational context and the outcome framework as it is practised, for example, in Dundee.

Due to the highly qualitative nature of portfolios, it seems reasonable to maintain independent rating and strive for high inter-rater agreement. At the same time, we may ask: how can we equip examiners with the right tools and the appropriate definitions to judge examinees independently and further submit these judgements to a rigorous consensus process?

The use of examples and evidence to support raters' judgements against pre-determined criteria is one way to achieve a meaningful ratings and consensus process. Justification of grading within defined frameworks and shared criteria may create a meaningful and reliable consensus process.

In the Dundee example, two stages of independent ratings provide the basis for the final consensus process, which leads to an overall judgement and agreement between the two raters. It seems that the three forms of reliability: agreement among independent ratings, reproducibility of pass/fail decisions and rigorous consensus procedures may all provide an integrated solution for achieving high reliability and for the use of portfolios for summative purposes.

When raters provide their independent grading they own this decision. In the consensus process they must negotiate the grades by supplying evidence of student performance and explain their interpretation of the performances. Inter-rater agreements could be studied as well as reproducibility of pass/fail decisions to establish the overall reliability of the portfolio examination. A stronger design may incorporate two pairs of examiners, each providing individual ratings as well as consensus judgement.

This approach was implemented in Dundee in the first year of the portfolio examination. Each pair of examiners conducted a 20-minute oral review of the portfolio with the student. However, it was felt that some students needed more time with each pair and, on many occasions, students faced the same questions by the two pairs albeit concerning different components of the portfolio. There was also an issue of manpower. Currently, Dundee employs only one pair of examiners for the 40 minute oral review of the portfolio with the student, but this policy is under review. One proposal is to have two pairs of examiners with a student for 30 minutes each.

Work needs to be done to further study the validity and reliability of the consensus approach as complementary to other forms of reliability.

6 *Forms of validity*

Discussions of validity include the appropriateness of score interpretation and the specific inferences made from the test (Messick 1994). The validity of the inferences depends also on the reliability of the test. If test scores suffer from low accuracy due to low inter-rater agreement or poor sampling, inferences cannot be made.

The authenticity aspect of the portfolio strengthens the predictive validity of the portfolio assessment. If portfolio examinations are adequate to measure high level, complex abilities applicable in real life situations, then its predictive validity should demonstrate strong relationship with later professional or career performance.

Such long-term studies were not found. Finlay et al (1998) conducted a randomized controlled study and has shown better results for the portfolio (experimental) group on Oncology factual knowledge. Those submitting portfolios had overall higher marks than those who have selected not to submit. This study may also indicate differences in motivation.

As for construct validity, Koretz et al (1993) found moderate correlations (.47-.58) between writing

portfolio scores and other direct writing assignment. Where pass/fail decisions are considered, the dilemma of low correlations between different measures becomes more problematic. One has to ensure that students will achieve mastery on both the portfolio and other forms of writing ability to establish comparability of measures. Also it was found that different scoring approaches to the portfolio may produce different classification results. (Koretz 1993).

Currently, a central validity aspect of the portfolio assessment is its face validity. The process, the evidence and the personalised approach are attractive to both students and educators. However, work is needed to further establish its predictive validity and its relationship to other measures of student performance.

7 *External vs internal examiners*

It is argued that portfolio assessment is more suitable for internal examiners (Linn 1994) where familiarity with the educational programme and with student progression ensures consistency of marking. The personalised approach of portfolio assessment requires internal knowledge of the educational milieu. External examiners at the state or district level may be recognised more as examiners for quality control of the portfolio process. This may have implications for judging portfolio work at a central site without seeing the examinee or lack of familiarity with the educational context. The purpose of such central marking of portfolio assessment should be clarified and the criteria employed for judgement should incorporate standardisation procedures as well as generic criteria suitable to all examinees independent of educational programme. External examiners who sit in a personalised portfolio may differ in their interpretation of examinee performance due to a different educational philosophy in their own schools. External examiners, however, are crucial in the assessment systems of UK medical schools. When external examiners participate in portfolio examinations, steps should be taken to ensure that they do not bias the results. Specific training programmes should be designed for external examiners to familiarise them with the curriculum, student assessment and educational philosophy of the school.

The contextual effect on portfolio performance as well as the personalised approach to students are forces which attempt to localize the portfolio rather than use it for external testing.

The large number of studies on portfolio implementation indicate that internal as well as external uses are reported. The difference in the application lies in the purpose of the portfolio and the criteria employed for judgement.

External uses may employ general criteria for students' work which is independent of educational or work-related context. The external marking of student work may generate group data on school profiles, students' level of achievement and the quality control of educational programmes.

D: Implementation of portfolio assessment

Implementation of portfolio assessments commonly incorporates a sequence of steps. These steps are identified in Table 5. For each step we will bring examples to illustrate some practical considerations.

1	Defining the purpose
2	Determining competences to be assessed
3	Selection of portfolio material
4	Developing a marking system
5	Selection and training of examiners
6	Planning the examination process
7	Student orientation
8	Developing guidelines for decisions
9	Establishing reliability and validity evidence
10	Designing evaluation procedures

Table 5: Steps in developing portfolios

1 *Defining the purpose*

Portfolios are used for summative and formative decisions, for different trainees' levels and in a variety of settings. The variation in its content, process and application requires that the purpose will be clearly defined. The non-traditional assessment approaches employed by portfolios may introduce confusion and uncertainty regarding its purpose. The clearer the purpose, the higher the quality of the portfolio assessment.

An example of a purpose definition

"The purpose of the portfolio assessment is to determine, in line with the GMC guidelines, that the graduating student is ready to proceed to the Pre-registration House Office (PRHO) year and that the

student has achieved the required leaning outcomes. The portfolio is only one component of the final examination. Successful performance on the portfolio is a requirement for graduation."

The purpose statement directs the test designers to consider the competencies essential for entering the PRHO year, to determine which competencies are already measured by the other final examination components, and how the portfolio should supplement the other assessment methods. In this statement, the portfolio is one hurdle which necessitates a summative decision in order to fulfil graduation requirements. The purpose defines the assessment system and the relationship between the portfolio and the other assessment components.

2 *Determining competences to be assessed*

Faculty, medical educators, assessment committee and other medical school staff will meet to discuss the purpose of portfolio assessment. They will review the existing assessment procedure and will identify gaps in assessment which should be supplemented by the portfolio.

In Dundee, where the portfolio is defined as a requirement towards graduation, the first 11 outcomes are assessed at the "knows" and "knows how" level (Miller 1991) by EMIs and CRQs, and at the "shows how" level by the OSCE. At the level of "does", however, outcome 12 is only assessed by the portfolio.

Table 6 shows the grid of final examination and the associated methods for each outcome.

The portfolio, through students' cumulative work, identifies strengths and weaknesses in all outcomes with emphasis on outcomes 6-12. The portfolio designers may also emphasise the interrelations among the outcomes and students' overall progress towards all outcomes.

The process of identification of competencies to be assessed signifies a systemic and systematic approach to assessment (Friedman 1999) which ensures a comprehensive strategy to determine qualification for graduation. The principle that only one assessment method is inadequate to determine fitness for practice guides this design.

3 *Selection of portfolio material*

The identification of competencies to be measured is a starting point for selection of portfolio material. It is important to identify key performance behaviours relevant to the competency measured. The identified behaviours will guide the portfolio content selection.

For example, in assessing trainees' patient management skills, one of the selected performance behaviours could be patient education. The portfolio designers will search for a work sample of patient education. Evidence might include written outline of a patient education programme in the community, or a video of individual patient education session with patient discussing smoking cessation. It is important to remember that evidence is selected from ongoing trainee's work and is not created in an artificial way just for the portfolio examination. It is also essential that the compilation of evidence will serve as learning and educational experience for the trainee. For example, in selecting a video for patient education, trainees will employ criteria for selecting the best patient education encounter. In defending the selection, the trainee demonstrates understanding of criteria for optimal performance.

For outcome 9, appropriate attitudes, ethical and legal understanding and responsibility, the portfolio material in Dundee contains the elective report – which provides evidence if the student shows ethical understanding of issues inherent in the elective. In a case discussion on ethics, the students will provide evidence of ethical judgement and moral reasoning – and could be questioned about the case.

In the Dundee case study, where the purpose of assessment is progress towards the 12 outcomes, the evidence should indicate progress over time. A selection of samples of student work will facilitate discussion of patterns of performance and overall strengths and weaknesses.

4 *Developing a marking system*

In the Vermont application of portfolios, generic criteria were developed specifically for grade 4 in mathematics. Students' work is judged by criteria which will specify the level of their academic achievements and will determine their progress towards state standards. This approach is possible when one unit of portfolio content is evaluated. However, in the Dundee case study, the 12 broadly defined outcomes do not allow the application of specific criteria, but rather general statements regarding student performance relative to the outcome specification. For example, under the outcome Information Handling the following dimensions are specified: "the doctor is competent in recording, retrieving and analysing information using a range of methods including computers". Criteria for marking may include evaluation of

Outcomes		EMI[1]	CRQ[2]	OSCE	Portfolio
1	Clinical skills	xxx	xxx	xxx	xx
2	Practical procedures	xx	xx	xxx	xx
3	Investigating a patient	xx	xx	xxx	xx
4	Patient management	xxx	xxx	xxx	xx
5	Health promotion and disease prevention	xx	xx	xxx	xx
6	Communication skills	x	x	xxx	xxx
7	Information handling	xx	xx	x	xxx
8	Understanding of basic, clinical and social sciences	xxx	xxx	x	xxx
9	Appropriate attitudes ethical and legal responsibility	xx	xx	xx	xxx
10	Decision making, clinical reasoning	xxx	xxx	xxx	xxx
11	Role of the doctor	x	x	xx	xxx
12	Personal development	x	x	xx	xxx

[1] EMI (Extended Matching Items), [2] CRQ (Constructed Response Questions)
x – considered for use, xx – appropriate and is currently used, xxx – most appropriate assessment method

Table 6: Final examination grid in Dundee outcome-based curriculum

patient records (including specific marks for quality of records). Criteria for search strategies (including marks for formulating research questions and generating key words for multiple databases) and criteria for analysing information (including marks for critique of a research paper).

The portfolio material should direct the examiner to consider student progress according to the outcome specification and should enable the examiner to identify strengths and weaknesses. As mentioned earlier, highly specific criteria could be employed if one competence is assessed. For multiple competences, general standards should be developed.

It is also possible as in the Dundee case study to assign grades to student global work. However, prior to grades' assignment, examiners must share understanding of the specific or general criteria and consistently negotiate what constitutes an A level grade (excellent) versus an F level (fail) and further define and refine the points on the continuum.

Schools may be bound by university grading systems. The best approach will be to follow the above procedures. Definition of criteria for performance and identification of level of performance which will correspond to the university grading system.

5 *Selection and training of examiners*

If one considers only an internal examiners' system for portfolio assessment, selection of examiners will follow the internal structure and philosophy of the educational program. Each portfolio according to its purpose will dictate the appropriate examiners. These will include a wide range of staff; teachers in the basic sciences and laboratory-based disciplines, clinicians, faculty who indicate special interest in education and in student development, faculty who are motivated to spend the extra time needed and faculty who are good oral examiners (mainly student centred). Another selection issue relates to the seniority of examiners. Pairing of new examiners with more senior colleagues with experience of portfolio assessment has been found to be helpful in Dundee. Whatever the selection process employed, the training of faculty examiners and maintaining them in the examiners' pool is a key point for the success of the programme. Naturally the amount of time and effort invested in the examiners' pool through training and orientation and their accumulated experience with portfolio examination, should be preserved and reinforced. Consequently, they should be treated accordingly (dinners, lunches) and given help to form a group with recognised academic respect and appreciation. The fact that a faculty member says " I can't join you for tomorrow's meeting – I am a portfolio examiner", should bring pride to the speaker and respect from the listener.

Faculty in general appreciate their participation in portfolio examinations (Davis 2001). It provides an opportunity to know the students and to share opinions with their colleagues. They obtain better perspectives on student accumulated work and progress. Although the amount of reading might be

excessive (a problem if one decided to maintain the broad sampling principle) realistic approaches to reading plans can keep this burden at manageable levels.

6 *Planning the examination process*

Fortunately, the use of OSCEs has introduced the concept of individualised student assessment. This concept, together with the time needed, faculty manpower, logistics, administration issues and availability of support staff are all issues shared with portfolio examinations.

In Dundee, with a class of 150, each pair of examiners sees five students for a total examination time of 200 minutes (40 minutes for each oral review of the portfolio with the student). If we include lunch, coffee breaks, discussion time between candidates, filling up forms and the Examiners Committee meeting at the end of the day – five students per day is the maximum number. If 10 pairs of examiners are recruited for each day – all 150 students could sit the oral review examination in three days with different sets of examiners. The number of examiners needed for such an assessment operation is 60 with each examiner being committed for just one day. There is, of course, the additional time examiners spend reading the portfolios before the meeting with the students, approximately one hour per portfolio in Dundee.

Faculty needs to buy into the concept of portfolio to be willing to participate in such an important undertaking. As with all aspects of reform, strong leadership is needed within the medical school to facilitate the "launch" of the first administration and keep the momentum in subsequent years in spite of any reservations among faculty, students and administrators.

7 *Student orientation*

Students must be informed at the beginning of the undergraduate course about the portfolio examination and kept up to date with any changes in subsequent years.

A special orientation booklet for students is issued to inform them of the purpose and content of portfolio building, the portfolio assessment process, the examination day, the marking system and the use of results.

The more complex the task of portfolio building, the more anxious the student becomes. Problems may occur if the guidelines and criteria for judging performance are not entirely clear as may happen when the guidelines are too broad. However, in the Dundee case study, students use their on-going work for selection of material – and usually, if they demonstrate good progress in their achievements, they have confidence in their ability to pass the portfolio.

The more information given to students the more positive they become towards the portfolio. Following the portfolio examination, students are pleased that senior examiners have given detailed attention to their work over their last two years of medical school. The students, in general, like this personalised, student centered, contextualised assessment approach.

8 *Developing guidelines for decisions*

Once a grading system is developed, the portfolio designers must develop policy and guidelines for standards of performance. If the portfolio is used for summative pass/fail decisions, standards should specify what constitutes a fail or a pass. If portfolios are used for formative decisions, standards for strengths and weaknesses should be defined along with due process for intervention and remediation. The process should also incorporate the procedures for evaluating the intervention and the decision mechanisms to further promote or to hold back/expel the student. The decision making process may be outlined by the following flow diagram.

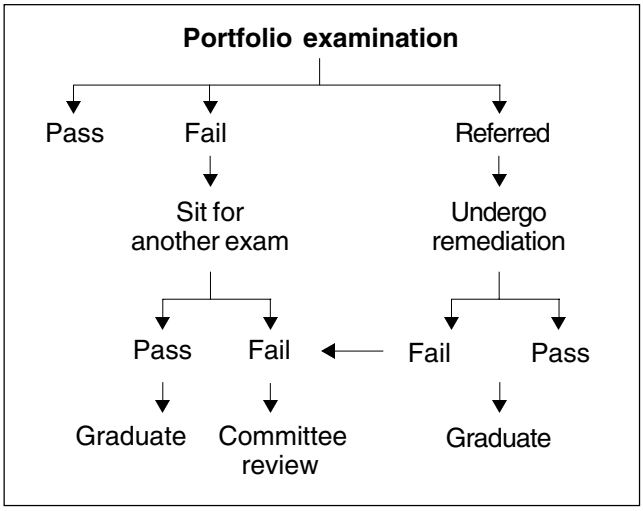


Figure 3: Flow diagram for the decision-making process

9 *Establishing reliability and validity evidence*

It is important to determine, prior to the implementation of the portfolio, what will constitute good reliable evidence and plan the examination

accordingly. For example: 2 pairs of independent examiners, 1 pair of examiners, independent rating, consensus or both, minimum desired reliability or generalisability co-efficient. Define desired correlations or absolute inter-rater agreement and set the minimum standards for tolerance of misclassification error. The degree of reliability in pilot studies, prior to actual administration, should direct faculty to make decisions regarding whether the portfolio should be used as a formative or a summative assessment.

Triangulation of portfolio results with other forms of assessment will increase the validity of the decision and will guide faculty as to the use of the portfolio results. If portfolio assessment does not correlate with any of the other forms of examination – its results could be aggregated as a weighted component towards graduation. Recent studies support the use of a battery of tests to generate composite reliability for test components combined (Wass 2000).

10 *Designing evaluation procedures*

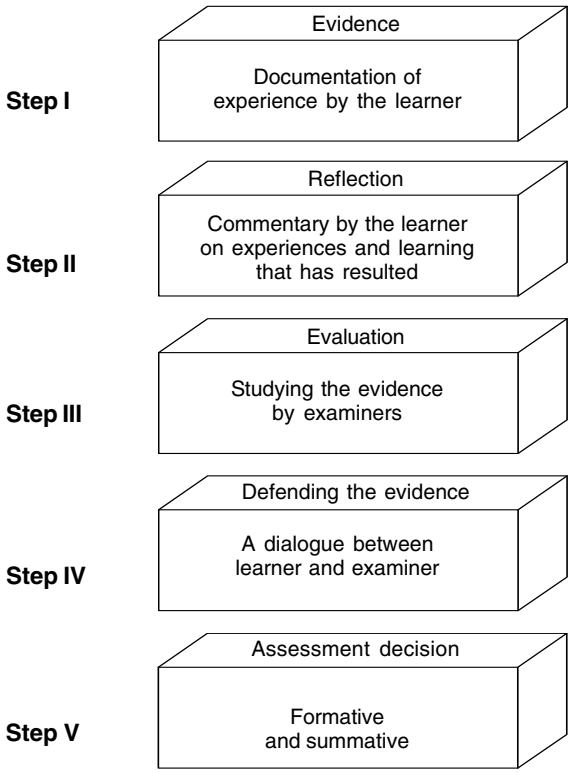
Students' and examiners' opinions on the portfolio's strengths and weaknesses should be sought in order that feedback can introduce changes and improvements. Questionnaires, focus groups, individual interviews and requests for written comments can all be employed. The appropriate committees may initiate changes based on evaluation results and will inform the subsequent years about the areas of satisfaction with the portfolio in order to market its usefulness.

Students' performance on the portfolio with relation to performance on other assessment tools are important evaluation strategies in order to detect if students who are generally high academic achievers do not fail the portfolio or vice versa. Too many surprises may not be good news for the medical schools. Many concerns may arise: either the undergraduate course does not detect early signs of deficiencies, the other assessment tools are not adequate, or the portfolio examination is not valid for its purpose. With any form of new examination, it is important to conduct pilot studies and use the procedures as formative until the necessary evaluation data is obtained.

E: Conclusion

The added value of portfolios for students' assessment is increasingly recognised by the medical profession. In the UK, medical schools have reported a variety of forms of portfolio application (McKimm 2001). The portfolio method fits the triangulation approach in assessment in which not one method but rather a combination of assessment methods can capture the complexity of professional competencies. With the addition of innovative methods, like portfolios, we are increasing the "menu" from which we may choose appropriate assessment methods. The portfolio assessment process can be summarized by a number of steps.

Traditionally, assessment procedures may incorporate only few steps, for example, Steps I, 111 and V only. Thus, the learner takes a passive role. The portfolio process allows learners to take an active role in the assessment process through selection of material, reflection on progress and learning and, defending the evidence, similar to a dissertation in which the candidate assumes responsibility for the entire process.



Use of portfolio in all stages of medical education will be most beneficial for the education of professionals. Each phase of the medical continuum: undergraduate, postgraduate and continuing medical education, should consider how to use portfolios for determining fitness for practice. The portfolios may provide the methodological aspect which focuses on professionalism. The Quality Assurance Agency stipulation that UK universities introduce personal development planning is another potential opportunity for introducing portfolios both for education and assessment.

It is important to consider the implication of portfolios as a new assessment method in the undergraduate curriculum. Students are normally prone to demonstrate anxiety and stress with the introduction of new assessment methods. The development of explicit criteria for portfolio content selection, with examples of portfolio evidence, and explicit criteria for judging the material, will certainly alleviate some of the anxiety.

Students must feel that the content selection is meaningful and that they are not involved in a paper chase exercise. Strong staff support is another important aspect of implementing a portfolio assessment.

As was shown in the Dundee case study (Davis, personal communication) students change their perception of portfolios with time. Acting on student feedback is a very important strategy to orient students not only to the assessment process but also to the design.

However, portfolios for examination have yet to win public and professional acceptability. They are used successfully for formative purposes but summative applications of portfolios are yet to be further developed. The portfolios' psychometric aspects of validity and reliability are currently being explored for high stakes examination as well as for summative purposes. However, in seeking technical rigor we need to be sure not to lose the appeal of the portfolio concept. It is hoped that this Guide will stimulate further studies and innovative approaches that will strengthen the psychometric aspects of portfolios as well as its educational implications.

References

Al Shehri, A. (1995). Learning by reflection in general practice. A study report, *Education for General Practice* 7: 237-245.

American Board of Internal Medicine (1995). Project Professionalism. Philadelphia PA.

Archibald, D.A. & Newmann, F.M. (1988). Assessing authentic academic achievement in the secondary school, Reston, V.A., National Association of secondary school principals.

Aschlacher, P.R. (1993). Issues in innovative assessment for classroom practice, barriers and facilitators (Tech. Rep No 359) Los Angeles University of California, CRESST: Center for the Study of Education.

Association of American Medical Colleges (1998). Learning Objectives for Medical Student Education: Guidelines for Medical Schools: Medical Schools Objectives.

Bashook, P.G. & Parboosingh, J. (1998). Recertification and the maintenance of competence, *British Medical Journal* 316.

Beattie, A. (1991). Curriculum Co-ordination. National Board for Nursing, Midwifery and Health Visiting for Scotland 1(5): 47-59.

Brennan, R.L. (1983). Elements of generalisability theory. The American college testing program.

Brown, R. (1995). Portfolio Development and Profiling for Nurses, *Journal of Clinical Nursing* 4: 249-255.

Case, S.H. (1994). Will mandating portfolios undermine their value? E*ducational Leadership*.

Cohen, R. (2001). Assessing professional behaviour and medical error, *Medical Teacher* 23(2): 145-151.

Cumming, J.J. & Maxwell G.S. (1999). Contextualising Authentic Assessment. *Assessment in Education* 6(2): 177-194.

Davis, M.H., Friedman Ben-David, M, Harden, R.M., Howie, P., Ker, J., McGhee, C. Pippard, M.J., Snadden, D. (2001). Portfolio assessment in medical students' final examinations, *Medical Teacher* 23: 357-366.

Delandshere, G. & Petrosky, A. (1994). Capturing teachers knowledge, performance assessment a) and post-structuralist epistemology, b) from post-structuralist perspective, c) and post-structuralism, d) none of the above, *Educational Researcher* 23(5): 11-18.

Des Marchais, J.E. & Jean, P. (1995). Effects of Examiner Training on Open-ended, higher taxonomic level of questioning in oral certification, *Examinations Teaching and Learning in Medicine* 5(1): 24-28.

Eisner, E. (1991). Forms of Understanding and the Future of Education, *Educational Researcher* 22: 5-11.

Finlay, I.G., Haughan, T.S., Webster, D.J.T. (1998). A randomised controlled study of portfolio learning in undergraduate cancer education, *Medical Education* 32: 172-176.

Friedman Ben-David, M. (2000). The role of assessment in expanding professional horizons, *Medical Teacher* 22(5).

Friedman Ben-David, M. (1999). Assessment in outcome-based education, *Medical Teacher* 21(1): 23-25.

Gahler, B.B., Rosenthal, D.B., Thoronton, 111, G.C., Benston, C. (1987). Meta analysis of assessment center validity, *Journal of Applied Psychology* 72(3).

Ginsburg, S., Regehr, G., Hatala, R., McNaughton, N., Frohna, A., Hodges, B., Lorelei, L., Stern, D. (2000). Context, conflict and resolution: A new conceptual framework evaluating professionalism, *Academic Medicine* 75(10).

Gisselle, O. & Martin-Kneip. (2000). Becoming a Better Teacher. ASCD, Alexandria, Virginia, USA.

Glen, S. & Hight, N. (1992). Portfolios: an Affective Assessment Strategy? *Nurse Education Study* 12: 415-423.

GMC (2000). Revalidating Doctors: ensuring standards, securing the future.

GMC (1997). The New Doctor.

Hand, R.J. (1999). Science Education: Consensus vs. Critique, *Teaching in Higher Education* 4(4): 501-510.

Grundy, S. (1987). Curriculum: Product or process. Lewis Falmer.

Harden, R.M., Crosby, J.R., Davis, M.H., Friedman Ben-David, M. (1999). Outcome-based education from competency to meta competency: a model for specification of learning outcomes, *Medical Teacher* 21(6).

Herman, J.L., Gearhart, M, Baker, E.L. (1995). Assessing writing portfolios: issues in the validity and meaning of scores, *Educational Assessment* 1(3): 201-224.

Herman, J.L. & Winter, L. (1994). Portfolio research a slim collection. *Education Leadership*.

Jasper, A.U. (1995). The potential of the professional portfolio for nursing, *Journal of Clinical Nursing* 4: 249-255.

Koretz, D. (1998). Large scale portfolio assessments in the US: Evidence pertaining to the quality of measurement, *Assessment in Education* 5(3): 309-334.

Koretz, D., Broadfoot, P., Wolf, A. (1998). Editorial, *Assessment in Education* 5(3).

Koretz, D., Stecher, B., Deibent, E. (1993). The reliability of scores from the 1992 Vermont portfolio assessment program. (Tech. Rep. No. 355). Los Angeles University of California CRESST: Center for the Study of Evaluation.

Le Mahieu, P.D.H., Gitumer and J.T. Evesh (1993). Portfolios in Large Scale Assessment: Difficult but not impossible. Unpublished manuscript, university of Delaware (principal author).

Linn, R.L. (1994). Performance Assessment: Policy Promises and Technical Measurement Standards, *Educational Researcher December* pp. 4-14.

Livingston, S.A. & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores, *Journal of Educational Measurement* 32(2).

Lundberg, G. (1991). Promoting Professionalism through Self Appraisal in this Critical Decade, *Journal of the American Medical Association* 265: 2859.

McKimm, J. (2001). Using portfolios in medical education. Imperial College School of Medicine preparedness to practice project. Internal report.

Messick (1994). The interplay of evidence and consequences in the validation of performance assessment, *Educational Research* 2(2): 13-23.

Miller, G.E. (1991). The Assessment of Clinical Skills/Competence/Performance, *Academic Medicine* 65(9): S63-67.

Missin, L.J. & Hart, L. (1985). Oral examinations in assessing clinical competence, Neufeld, V.R., Norman, G.R. (eds):.71-93. New York. Springer Publishing Company.

Moss, P.A. (1994). Can there be validity without reliability, *Educational Researcher* 23(2): 5-12.

Parboosingh, J. ((1996). Learning Portfolios: Potential to Assist Health Professionals with Self-Directed Learning, *Journal of Continuing Education in the Health Professionals* 16: 75-81.

Pitts, J., Coles ,C., Thomas P. (2001). Enhancing reliability in portfolio assessment - shaping the portfolio. In press, *Medical Teacher*.

Schon, D. (1983). The Reflective Practitioner - How Professionals Think in Action. Basic Books. Harper Collins.

Schon, D. (1987). Educating the reflective practitioner, San Francisco, C.A., Jossey-Bass.

Scottish Dean's Medical Curriculum Group (2000). Learning Outcomes for the Medical Undergraduates in Scotland: A Foundation for Competent and Reflective Practitioners.

Shapiro, J. & Talbot, Y. (1999). Applying the Concept of the Reflective Practitioner to understanding and teaching Family Medicine, *Family Medicine 32*(6): 451-457.

Snadden, D., Thomas, M., Challis, M. (1999). Portfolio-based learning and assessment, *AMEE Guide No. 11*.

Stecher, B. (1998). The local benefits and burdens of large-scale portfolio assessment, *Assessment in Education* 5(3).

Swick, H.M. (2001). Towards a Normative Definition of medical Professionalism, *Academic Medicine 75(6):* 612-616.

Turnbull, J. (2000). Clinical Work Sampling, *Journal of General Internal Medicine 15*: 5556-5561.

Wass, V., McGibbon, D., Van Der Vleuten, C. (2001). Composite undergraduate clinical examinations: how should the components be combined to maximise reliability? Medical Education 35, 4, 326-330

Wear, D. (1997). Professional Development of Medical Students: Problems and Promises, *Academic Medicine 72*:1062.

Wolf, A. (1998). Portfolio Assessment as National Policy. The National Council for Vocational Qualifications and its quest for a pedagogical revolution, *HSS in Education* 5(3).