Structured Abstract

Paper Title: Literature and Medicine: A Problem of Assessment

Author's Name: AYELET KUPER

Author's Degrees: MD, DPhil

Author's Affiliations: Donald R. Wilson Centre for Research in Education, University Health Network, and Division of General Internal Medicine, Department of Medicine, Sunnybrook and Women's Health Sciences Centre, both in the Faculty of Medicine, University of Toronto, Ontario, Canada

Problem Statement and Background: 'Literature and medicine' is increasingly common in medical schools but not within medical education research. This absence may relate to it not being problematizable in the quantitative way in which this psychometrically-oriented community tends to conceptualize research questions.

Methodology: Databases were searched using relevant keywords. Articles were evaluated using methodologies appropriate to their fields. The resulting information was structured around a framework of construct-appropriate assessment methods.

Synthesis of Findings: Literature and medicine is intended to develop skills as potential proxy outcomes for important constructs. Proposed tools to assess these skills are difficult to evaluate using the field's traditional quantitative framework. Methodologies derived from the qualitative tradition offer alternative assessment methods.

Conclusions: The medical education research community should take on the challenges presented by literature and medicine. Otherwise, we run the risk that the current evaluation system will prevent important constructs from being effectively taught and assessed.

Literature and Medicine: A Problem of Assessment

AYELET KUPER

Physicians are being exposed to a steady stream of articles about the field of literature and medicine. Such pieces have been appearing in high impact medical journals including the New England Journal of Medicine,[1,2] the British Medical Journal,[3,4] the Journal of the American Medical Association,[5] Annals of Internal Medicine,[6-10] and especially the Lancet.[11-28]. Literature and medicine has a well-established journal of its own, aptly titled Literature and Medicine, and has a substantial presence in journals such as the Journal of Medical Ethics: Medical Humanities[29] and the Journal of Medical Humanities. Within the field of medical education, the journal Medical Education launched a new section in February 2002 called 'Arts and Humanities' to highlight an 'academic discipline concerned with research and education'[30] – a discipline that includes literature and medicine as a major component. Academic Medicine has published two theme issues on the medical humanities.[31] Interested education-oriented readers can also turn to an anthology, *Teaching Literature and Medicine*.[32]

Practically speaking, literature and medicine courses have been flourishing across the English-speaking world. By 1994 about a third of American medical schools were known to be teaching literature within their medical curricula.[7] In 1998 74% of them offered it as an elective, while 39% required it as part of at least one course.[33] In 2003-2004 the medical humanities as a whole were represented by at least one required course at eighty-eight of 125 American medical schools and in an elective course at fifty-five schools.[34] In the United Kingdom the number of humanities courses in general, and of literature and

medicine courses in particular,[35-38] grew in response to the General Medical Council's endorsement of the humanities as appropriate selective courses for medical students in their 1993 report, *Tomorrow's Doctors*,[39] and their reiteration of this concept in 2003.[40] Many descriptions of such courses have been published in medical and medical education journals[13,37,41-50] as well as in on-line databases.[51,52]

Yet, despite this seemingly thriving field, questions regarding its value and legitimacy continue in the medical education literature. For example, of the twenty-four articles in the Arts and Humanities section of Medical Education over the last four years, seven[37,41,43,49,53-55] dealt directly with literature and medicine, but another five[56-60] were largely occupied with legitimizing and justifying the medical humanities. Even Medical Education's current Arts and Humanities editor believes that the humanities are 'not yet part of the mainstream of medical education'.[61] This echoes Friedman's anxiety, expressed in an article published in Academic Medicine in 2002, about the still-precarious place of the humanities, including literature, in medical training.[62] It is curious that Charon and her co-authors can write increasingly confidently in high-impact medical journals about the utility of literature in medicine and in medical education,[1,5,7,9] yet they appear to have to strenuously justify its importance in Academic Medicine twice in articles published five years apart.[63,64] Although Academic Medicine published the aforementioned medical humanities theme issue in 2003, recognizing and describing a plethora of medical humanities courses for students at different levels of training, this issue does not include any academic articles or research papers[42] despite the primacy of such genres to that journal's main audience. Meanwhile, although there have been books about the medical humanities[65,66] and narrative in medicine[67] from general medical

publishers, many of the first generation of books that incorporate the theory and practice of the use of literature in medicine and medical education have instead been published by general academic publishers[68-70] or (in the case of *Teaching Literature and Medicine*) by the Modern Language Association of America.[32]

It seems that literature and medicine is being accepted in medicine (and in literature,[50] where it is a well-recognized subdiscipline[64]), and in medical schools, but not within the academic field of medical education. Why not? One potential reason for the struggle that the field of literature and medicine is facing within medical education research relates to the two disciplines' very different discourses. Fundamentally, the role of literature in medical education may not have been addressed from within the medical education research community because it has not been problematized in the way in which that community tends to conceptualize research questions. In other words, it may be that it is not easily incorporable into the current medical education research agenda because the constructs derived from literature and medicine are not amenable to being addressed by the tools most commonly used within medical education research. This paper, therefore, is an attempt to bridge that divide.

Background

It is by now a truism that medical education research is increasingly driven by outcomes and by evaluation.[71] In the United States, the Outcome Project of the Accreditation Council for Graduate Medical Education (ACGME) has put a strong emphasis on educational outcomes in residency training, with a clear link between objectives and assessment,[72] while the Medical School Objectives Project (MSOP) has focussed on similar issues in undergraduate medical education.[73] Other influential bodies, such as the

Royal College of Physicians and Surgeons of Canada[74,75] and Britain's General Medical Council,[40] have also adopted an outcomes-based approach to medical training. This link between measurable objectives and assessment has been driven by public expectations,[72,73,76] by the accountability agenda,[77] and by pragmatism in face of the fact that medical trainees guide their learning to meet the requirements of the evaluations that they will undergo.[78-80] There has also been considerable spill-over from the evidence-based medicine movement, so central to current medical research and practice, in the promotion of evidence-based medical education and in the tools being endorsed for this process.[81-84]

The issues around student assessment which are thereby foregrounded have generally been addressed, in the context of medical education research, using the discourse of psychometrics.[85-87] Statistical concepts used in educational testing, such as reliability and validity,[88] have framed much of the discussion.[71,86,89] It has been shown both that it is possible to create tests with holistic rating scales that are valid and reliable[90] and that the breakdown of complex tasks into objectified component parts can trivialize the overall construct, thereby decreasing the validity of the assessment.[91] We now know that reliability and subjectivity are not mutually exclusive nor are reliability and objectivity inextricably linked.[80,92] Nonetheless, there has long been a movement towards using more objectively scored[93] and hence more 'granular' examinations in an attempt to improve reliability.[94] There is also an implication 'that the value of the assessment can be researched and described in numbers only.'[87] The overall ethos of medical education research, then, incorporates a tendency to break down into component parts that which it is studying, whether competencies, constructs, or content, and to reduce it to numerical

values that can be manipulated and classified. Similarly, the search for the 'true' score, made explicit in any discussion of reliability, generalizability, and error,[88] highlights the underlying positivism in medical education research that reflects a paradigm which pervades much of modern medical education.[58]

The study of literature, even when placed within the medical context, is resistant to simplified analysis and is not compatible with straightforward positivism. Grappling with meaning in literature, in the context of a constructed reality, inevitably encounters a major problem with assessment in this area: there can be no one right answer when discussing a text, and, indeed, this very ambiguity is one of the lessons which may be learned from literature and carried forward into the context of patient care.[60] That is not to say that one should not discriminate between students' abilities and achievements in a literature curriculum, but rather that the means of this discrimination in higher education has not engendered significant debate. Whatever the reason, it must be noted that at least two published calls for research related to such assessment in the medical humanities[35,95] seem to have largely gone unheeded. A separate call for papers related to pedagogy in the medical humanities, for a special issue of the Journal of Medical Humanities that has never been published, simply assumes that only critical descriptions of individual courses will be forthcoming.[96] Yet, within the medical education research community, we should not scorn literature and medicine because it does not yet have the techniques with which to respond to our evaluation-driven curricular criteria. Rather than ignoring the problems of assessment in this area of medical education, it is time for the medical education community to develop, adapt, or recognize rigorous methods of student evaluation that respect and reinforce the important competencies intended to be attained through the

study of literature and medicine. To that end, this paper will outline these competencies, discuss assessment measures in current use, and survey the performance of select tools with respect to our traditional evaluation criteria. It will then explore other rigorous methodological structures that have been proposed for evaluating such measures of student assessment.

Method

A computerized search was undertaken of the following electronic databases: Medline (1966 to November 2005); Scholars Portal Search – Social Sciences subject area, which includes among the databases it searches the Educational Resources Information Center (ERIC) database (1966 to November 2005), Education – A SAGE Full-Text Collection (1968 to November 2005), Education Abstracts @ Scholars Portal (1983 to November 2005), and a range of other databases in areas such as sociology and psychology; Scholars Portal Search – Arts & Humanities subject area, which includes among the databases it searches the Modern Language Association (MLA) International Bibliography (1963 to November 2005, plus JSTOR's Language and Literature collection back to 1881), BHI: British Humanities Index (1962 to November 2005) Humanities Abstracts @ Scholars Portal (1984 to November 2005), and a range of other databases in areas such as art and philosophy; and Google Scholar. Search terms used included the following, alone and/or in combination: student assessment, student evaluation, humanities, medical humanities, literature, literature and medicine, narrative, story, higher education, professional education, medical education, medical education research, qualitative methods, quantitative methods, reliability, validity. Keywords from relevant retrieved texts were iteratively incorporated into new searches. Abstracts and full-text

articles and books retrieved were assessed and their references searched for further sources; where technically possible, an electronic 'forward' search for articles citing or similar to relevant articles was also performed. In a series of discussions over the past months, scholars in the fields of health professions evaluation, medical humanities, and the sociology of medical education also pointed out a number of other texts which were similarly mined for further sources. The University of Toronto library system was searched by title and keyword for paper and electronic texts using the above search terms. Recent tables of contents of the journals Literature and Medicine (May 1995 – July 2005), Social Science and Medicine (January 2000 – December 2005), Journal of Medical Ethics: Medical Humanities (June 2000 – December 2005), and Journal of Medical Humanities (Spring 1997 – Winter 2005) were searched manually, as were syllabi posted at the Literature, Arts, & Medicine Database[51] and the Medical Humanities Resource Database.[52]

Given the wide-ranging and disparate nature of the items retrieved, no attempt was made at a formal meta-analysis. Individual articles were evaluated using methodologies appropriate to their fields, including tools from quantitative methods, qualitative methods, and hermeneutic[§] textual analysis. The resulting information was structured with the competencies to be assessed as a framework within which to explore the most construct-congruent evaluation methods and to test these for appropriateness and rigor.

[§] As a method of textual interpretation, hermeneutics involves iterative analysis of the parts of a text against the whole until all of those parts contribute to a single consistent meaning. The reader must take his or her own sociohistorical position and intellectual tradition, as well as the context in which the text was originally created, into account in this interpretation.[97]

Results

Curricular Objectives for Literature and Medicine

Charon *et al*'s conceptual framework[7] provides a useful starting point for delineating explicit objectives being claimed for literature and medicine curricula. It outlines five rationales for introducing literature to medical students. One of these, narrative ethics, is presented in contrast to traditional precept-based ethics but is for our purposes an alternative pedagogic approach to the subject of medical ethics, which is already widely taught[98] and which has its own evaluatory frameworks. The current paper, therefore, will not address this particular rationale. Another rationale, the study of literary theory, offers interesting 'new perspectives'[7] on physicians, their patients, and their practice, but student learning in this area would require an *a priori* grounding in the methods and texts of literature. This rationale will thus also not be addressed in this paper. Instead, this paper will focus on the evaluation of objectives drawn from the remaining three rationales: the ability to respond to the patient experience, the ability to reflect on the physician experience, and the ability to develop and make use of narrative skills in practice. Each of these rationales will now be explored in turn.

Patients (and doctors) live their lives as narratives.[4,10,99] A significant Narrative-Based Medicine movement has emerged, specifically problematizing the need to foreground patients' narratives in order to imbue medicine with a holistic understanding of patients' emotional and existential responses to their illnesses.[4] Within the realm of medical education, it has been posited that literary texts selected for realism and relevance 'can help bridge the gap between *knowing* the facts about the disease and *understanding* the patient's illness experience [emphasis in the original]',[100] including illuminating its

important socio-economic and cultural contexts.[26,55] Stories about illness could therefore enhance physicians' abilities to imagine and understand the experiences of their sick patients,[7] potentially contributing to their capacity to provide empathic[50] patient-centred care. 'Narrative competence', defined as 'the competence that human beings use to absorb, interpret, and respond to stories'[5] (whether derived from texts, from patients, or from non-professional encounters), is taught through the close reading of and engagement with literary texts.[5] It is thought to contribute to the development of both professionalism[101,102] and empathy[5,8] in the physician-patient relationship. Thus, the construct of 'narrative competence' might be conceptualized as a surrogate endpoint for these more complex constructs, with, for example, students' abilities to identify and reflect on the emotions and experiences of characters in stories as potential surrogate markers for their later understanding of and responses to the experiences of patients and their loved ones. In short, although this has not yet been tested, the evaluation of narrative competence may be a classroom-based proxy outcome for anticipated empathy in clinical practice.

Similarly, literature provides trainees with 'a vivid means of understanding the physician's often quite lonely job'.[68] Physicians, especially during the intense years of student and residency training, live outside the realm of the commonplace. Their everyday experience of death, suffering, and healing is situated outside the boundaries of everyday language. Regular encounters with emotionally challenging situations, combined with academic stressors, are reflected[103] in the increased rates of stress and depression among medical students as compared to their peers.[103,104] Stories and poems may perform tasks otherwise missing in medical education, wherein they 'can stimulate

important personal introspection about and examination of all that the physician is called on to do.'[7] In other words, literature may provide trainees with the language and the tools to reflect, not about their patient care abilities, but about themselves[50] and their own emotions,[100] and thereby may help to heal the nascent healer.[18] Nothing in the medical curriculum adequately prepares trainees for 'the moment after' – the moment they walk out of a patient's room and realize that they have just told someone that they are going to die, the moment when they must have a framework for recognizing and responding to their thoughts and emotions in order to be able to move on to the next encounter and to carry on with their own lives. By creating narrative competence, this process of emotional self-reflection could be practiced in order to provide trainees with a set of narrative tools to use in their own lives. Narrative competence and emotional self-reflective ability are therefore potential classroom-based proxy outcomes for the resilience to emotionally challenging situations that may protect trainees from existential distress and from the development of callousness and cynicism.

The study of literature could also provide trainees with useful clinical abilities that may be grouped under the general rubric of 'narrative skills'. Most prosaically, reading stories and writing about them can enhance more general communication skills.[41] Other narrative skills are components of the larger construct of narrative competence, which is believed to contribute not only to empathy but also to the physician's ability to organize and meaningfully integrate the complex stories to be gleaned from patients' histories, physical findings, and other ancillary data.[7] These skills make explicit use of the narrative structure that underlies clinical knowledge.[16,68,105] The ability to write about stories taken from literature, then, might become a useful surrogate outcome for the ability to construct

and communicate a coherent and rhetorically sound plan for patient diagnosis and treatment. The study of literature, in which there are myriad possible interpretations, also increases students' exposure to the concept of ambiguity.[28,60,106] This may help prepare them[12,64] for the uncertainty[107] and ambiguity[99] they will have to face as professionals in clinical practice by exposing them to ways of knowing other than the 'Positivist epistemology of practice' of professional training in general[107] and of medical training in particular.[58] Their grasp of this concept of ambiguity, as assessed through their responses to literary texts, could therefore be examined as a proxy for their preparedness for encountering it in the context of patient care.

It is possible, then, to interpret the objectives for courses in literature and medicine as being the development, in a safe, classroom situation, of a set of critical skills relevant to clinical practice. These skills are potential proxy outcomes for the higher-order objectives intended to be developed by literature and medicine. The development of narrative competence is a skill necessary for empathic understanding of patients' experiences of illness and treatment. The development of emotional self-reflective ability is a skill necessary for the ability to maintain a caring and connected professional approach to patient care. Finally, the development of narrative skills is necessary for the construction of coherent and comprehensive clinical pictures. Unfortunately, the ability to 'test these hypotheses' and 'validate these measures' using the usual methodologies of North American medical education research will likely be limited by the ability to measure these proxy outcomes using the field's traditional reductionist approaches to quantifying individual skills. The challenge therefore becomes the identification of appropriate evaluation methods that can be rigorously assessed with respect to their success in

measuring these surrogate outcomes. The following sections describe several techniques that are currently used in the assessment of abilities related to literature and the humanities within both medical and other higher educational contexts. These include both assessment methods that have been used more generally in medicine in the past and newer methods that are struggling to achieve legitimacy in medicine today.

Assessment Techniques in Literature and Medicine

Literature and medicine courses currently use a wide variety of assessment tools, including long or short essays, essay examinations, portfolios, oral presentations, posters, case write-ups, journals, response papers, creative projects such as poems, short stories, narratives in the patient's voice, and even OSCE stations.[13,37,41-43,45-49,51,52] Few course descriptions comment explicitly on the rationale for selecting the method(s) of evaluation to be used. What seems clear, however, is that multiple choice questions and their ilk, which are the most commonly used forms of written assessment elsewhere in medical education because of their reliability and ease of administration and marking,[89] have not been seen as suitable for student evaluation in this domain.

Squier,[100] one of the only authors who discusses the nature of appropriate assessment for courses in literature and medicine, endorses the use of written assignments as both formative and summative assessments. However, she provides no evidence for this endorsement, nor does she propose a marking scheme or guidelines other than to 'avoid an excessive focus on grading' and to encourage with such grading 'self reflection, interpretation, and creativity,' focussing on giving feedback and comments rather than making '[f]ine distinctions between students'.[100] Similarly, Downie writes: 'there is no difficulty about evaluating a medical humanities course. It can be assessed by

examination or essay or other project. This has happened in Arts faculties for centuries and there is a great deal of experience in Arts faculties of this sort of evaluation.'[106] He goes on to discuss how a text (in this case, a poem) might be taught, and an essay about it would then be marked, in a standard arts class – by the presence of coherence and of an argument grounded in the text and its historical background, rather than by the direction of that argument.[106] Again, he does not justify or provide evidence for his assertions. This opinion of the proper way in which to evaluate a humanities discipline is also shared by the British Quality Assurance Agency for Higher Education (QAA). Regarding undergraduate instruction in English literature and language, its subject benchmark statement, *English*, mandates that essays be 'an essential component in the assessment process' and comments on their appropriateness to the demonstration of the skills required in this discipline.[108] Benchmark statements for related subjects, such as *Languages and Related Studies* (which includes the study of literature in languages other than English),[109] *Philosophy*,[110] and *History*[111] also prominently feature the use of essays and related written assignments. The *History* benchmark statement even includes suggested criteria for assessment of timed essay examinations, with the specific attributes in the categories of structure and focus, quality of argument and expression, and range of knowledge appropriate to each class of mark.

Although they are not as prominent in the study of literature and in related higher education domains, portfolios, like essays, have been used in humanities-related curricula in medical education.[52] They are also being used more generally in medical schools to evaluate constructs similar to those being advocated for literature and medicine, such as empathy, comfort with ambiguity, and the ability to reflect on one's own emotional

needs.[112] A current general textbook on portfolio development describes a portfolio as a compilation of works, at least partially student-selected, that shows what the student has accomplished (or tried to accomplish) over time; this includes a particular emphasis on 'the centrality of student self-evaluation and reflection and the opportunity to portray the processes by which the work in the portfolio is achieved.'[113] The meaning of the term 'portfolio' has been the subject of recent debate in the medical education literature, particularly with respect to the requirement for reflection, which some have viewed as difficult, time-consuming, and potentially unnecessary[114] and others believe to be a unique and fundamental aspect of the tool.[115,116] Given the otherwise widespread recognition of self-reflection as part of the definition of a portfolio, both in the education literature in general[113] and within medical education in particular,[117-120] within this paper the use of the term portfolio will refer to a collection of works which includes evidence of reflection.

These literatures neither engage in nor acknowledge the discourse of validity and reliability, nor do they look explicitly to qualitative methodology for sources of evidence. They discuss neither reproducibility, accuracy, nor ease of marking, and they have no criterion reference. They are nonetheless related to rhetorical arguments that have emerged from within medical education. Norman *et al*,[91] reviewing the literature in 1991, concluded that subjective and objectified tests of the same construct were highly correlated. These authors foregrounded the risk of trivializing certain constructs using either multiple-choice and short-answer questions, as opposed to question types, such as essays, 'which require students to handle several aspects of knowledge in relation to each other.'[91] They therefore accepted the use of the latter as a legitimate option for testing,

particularly of higher-order constructs.[91] Moreover, Schuwirth and van der Vleuten have gone so far as to say that if the goal of a test is 'to set up a reasoning process or summarise information, or […] to apply a known principle in different contexts' then the only appropriate type of written question is an essay,[121] particularly if one is also concerned with writing ability.[122] As the following summary will illustrate, there has also been a significant amount of psychometric research, much of it controversial, on the value of essays and portfolios as evaluative techniques.

Essays and Portfolios

The construct validity of essay assessments has been specifically studied in the medical context. For example, in a 1990 study an essay test was validated for the assessment of clinical judgment at the postgraduate level.[123] In terms of essays' inter-rater reliability, several studies have given conflicting results.[94,124-126] There has also been research into their generalizability. For instance, the generalizability of an essay test of clinical judgment improved (and its required testing time decreased) when three non-physicians using a detailed checklist were replaced by three physicians marking holistically.[127] Frijns *et al*[128] showed that open-ended questions could be marked by physician-raters in a reproducible manner, although achieving a generalizability coefficient of 0.80 or above with one or two raters required between four and six hours of testing time. In keeping with our current understanding that psychometric criteria are not inherent qualities of an instrument's format,[80] the issue of reliability for marking essays is not an intrinsic problem with the test type but rather a question of the availability of adequate testing time and of expert markers. Multiple means of improving reliability, and thereby decreasing the need for those resources, have been suggested in the

literature,[88,121,122,125,126,129-131] although Schuwirth and van der Vleuten caution against over-structuring rubrics for marking essays to avoid trivializing the construct being assessed.[121,122]

Portfolios are newer to medical education, as well as more individual and process-oriented. It has thus far been difficult to establish their validity with respect to the constructs for which they have been studied in the medical context. The best established form of validity for portfolios is face validity for constructs including reflection[119,132] and performance over time.[119] Establishing their predictive, criterion, and construct validity for such constructs will be challenging.[117,119,132,133] The authors of a study published in late 2001 of portfolios used as part of the final examination for medical students in Dundee, Scotland claimed evidence of divergent validity for constructs, such as attitude and diligence, which are not assessed by their more traditional examination components.[134] A more recent study of portfolios in psychiatry residency education demonstrated modest convergent validity for psychiatric knowledge and level of training.[135] The reliability of portfolio assessments has also been studied, with estimates of interrater reliability ranging from 0.1 to 0.82.[119,136-138] Generalizability and decision studies have also generated a wide range of numbers of items and/or raters required for a generalizability of at least 0.8.[119,135,137] Multiple suggestions of strategies to improve reliability in portfolio evaluation have been made.[120,132,139] These include objectification through specific criteria and standardization.[120,132] The concern remains, however, that the standardization of content and the development of specific criteria present threats to the validity of the assessment by limiting the range of student reflection and learning,[119,120] perhaps without actually eliminating the problem of reliability.[140]

So we are now faced with a conundrum. Essays and portfolios are promoted as the most appropriate tools for the evaluation of literature and medicine, and they may allow us to tap into competencies that we cannot easily assess but which are becoming increasingly important, like empathy, personal reflection, and professionalism. However, as tools, they are not readily analyzable using the granular techniques of our traditional psychometrics, and by some measures they are not 'good' enough to use for summative decisions. Nonetheless, abandoning these important constructs, and a curriculum that is designed to promote them, is too radical an option. Not summatively evaluating these competencies is out of the question as well, if for no other reason than the message it would send about their true importance in the hidden curriculum. As Cannings *et al* summarizes: 'In our efforts to find a truly reliable assessment, we must not lose sight of the need to occasionally assess a 'subjective' piece of work […]. We then have to accept that there will be some loss of reliability in the marking that follows.'[130] Otherwise, Snadden warns us, if we 'continue to struggle to measure the unmeasurable, […we] may end up measuring the irrelevant because it is easier.'[141] Fortunately, the rigor of more subjective evaluation tools can be assessed in other ways, without relying on measurement, and examples of such assessment are beginning to enter the medical literature.

Qualitative Methods and Hermeneutics as Assessment

Rather than attempting to apply the rules of quantitative rigor to the qualitative, individualized world of patient experience, physician experience, and ambiguity, we can look to the increasing published recognition that some forms of assessment can be better studied, and some question better answered, using qualitative methodology.[80,87,92] Reflected in this trend is a growing understanding of the importance of 'interpretation –

the discernment of meaning'[4] in the narrative worlds of medical practice and medical education. This may be particularly true of portfolios and essays, given the wealth of qualitative information that they provide. Analyses of essay evaluations in other disciplines have taken qualitative approaches.[142] Since portfolios have been described as embodying a qualitative[113] or a mixed quantitative and qualitative[119] approach to student assessment, some have argued that the evaluation of portfolios as an assessment tool in healthcare education could also benefit from an approach founded in qualitative research.[138,140,141,143,144]

This has recently been tried at the medical school at Maastricht University.[145] The researchers rooted their intervention in 3 basic premises: (1) that the value of portfolio evaluation stems from its basis in the richness of authentic personal experience, which would be lost by standardization; (2) that rater training and checklists cannot compensate for this lack of standardization to produce adequate reliability as assessed by traditional psychometric methods; and (3) that qualitative (and subjective) methods, derived from the qualitative research tradition, can offer novel approaches to student assessment. They carried out both formative and summative evaluations of portfolios intended to contain reflections on, and evidence of, personal strengths and weaknesses in relation to four physician roles ('medical expert, scientist, health care worker and person'[145]) as well as learning plans to address these areas. In the summative evaluations, each student's mentor used multiple global criteria such as 'the quality of the analysis of strengths and weaknesses' and 'the clarity and feasibility of the learning objectives'.[145] The grade assigned by the mentor, which could be either distinction, pass, or fail, was then discussed with the student, confirmed by one or two other readers, and in cases of

continuing disagreement reviewed by a committee of 13 assessors (including the student's mentor).

Having introduced two methodological criteria from the constructivist tradition[146] within qualitative research, credibility and dependability, which can be used to parallel validity and reliability, they then used accepted strategies from the realm of qualitative methods for ensuring the credibility (triangulation, prolonged engagement, member checking) and dependability (audit trail, dependability audit) of their summative assessment. In terms of further research, it was suggested that their evaluation methods could be further supported through having the portfolios assessed by other committees of assessors, in the manner of an external dependability audit.[145] Some might argue that such reassessment could then be used to calculate more traditional measures of interrater reliability. However, in the absence of a true hierarchy of methodologies we should be careful to avoid imposing the criteria of one tradition onto the already rigorous methodology of another valid discourse, wherein the 'mechanistic decision based on a standard of performance on a single assessment is replaced by a professional judgement based on accumulated and triangulated information across multiple sources of assessment information.'[87]

A third methodological criterion which is often discussed in the context of qualitative methods in assessment research is that of authenticity, defined as 'the extent to which the outcomes measured represent appropriate, meaningful, significant and worthwhile forms of human accomplishments'.[147] Portfolios in particular have been advocated as authentic forms of assessment[119] in that they allow the evaluation of 'performance in practice over a period of time, in other words they assess the application of theory and the performance

of the student or doctor.'[117] Essays can similarly be argued to be authentic evaluative tools for competencies such as written communication skills and narrative structure. Such authenticity allows 'optimal congruence between assessment on the one hand and educational goals and the demands of future practice on the other.'[92] However, the standardization of assigned tasks and the structuring of their assessment, which, as we have seen, are often advocated in the psychometric discourse to increase reliability, present a significant threat to authenticity.[119]

Another interesting non-psychometric approach to the evaluation of written texts comes from Moss's hermeneutics of assessment. This uses less rigidly constructivist methodology than the Maastricht group's qualitative criteria, drawing instead on the classic tradition of hermeneutic textual analysis (see the Methods section, above). Moss describes hermeneutics in education as a practice based on progressive integration, in which human phenomena (whether literary works or students' tests) are deciphered by trying 'to understand the whole in light of its parts, repeatedly testing interpretations against the available evidence until each of the parts can be accounted for in a coherent interpretation of the whole.'[148] Her methods highlight context and promote discussion and debate around the assessment of a series of texts, such as might be in a portfolio, carried out in a documented, stepwise manner.[113] Interestingly, many of the qualitative research strategies proposed by the Maastricht group (such as triangulation, prolonged engagement, audit trail) are also advocated, albeit using different terminology, by Moss's approach.[148] Her emphasis on consensus building is echoed in the assessment approach being tried, within a more psychometrically-oriented framework, to evaluate medical student portfolios in Dundee.[119]

Discussion

In spite of their quantitatively-determined psychometric flaws, essays and portfolios may remain suitable tools for the assessment of many of the objectives of a literature and medicine curriculum. For example, essays are appropriate for the assessment of constructs such as reasoning and writing skills, while portfolios have been specifically developed for the promotion and evaluation of reflection. Essays and related written assignments therefore present an appropriate mechanism for the evaluation of narrative skills, including narrative competence and written communication skills. Most simply, they can be used to assess the ability to extract a story from a text by close reading in the way that physicians fashion a coherent history from disjointed pieces of clinical information. Their subjectiveness, although problematic within a psychometric framework, also lends itself to the exploration of ambiguous texts, concepts, and feelings. Essays written in the pre-clerkship, for example, about stories presented from patients' and physicians' points of view, can allow a student to show an understanding of a patient's experience, to explore possible emotional repercussions of a patient's illness on her physicians and other caregivers, and to reflect on potentially difficult professional dilemmas. By emphasizing process rather than content, this can help develop skills related to empathy and emotional self-reflection before having to experience such situations in real life. The lack of a 'right answer' for such an essay can, if framed appropriately, introduce the idea of uncertainty both in literature and in medicine.

For practical reasons, formal essays are hard to assign during the clinical years. However, a portfolio of short reflections (or a section of a more general portfolio, depending on the rest of the curriculum), which builds on skills originally taught in the pre-clerkship, can

continue to encourage and assess empathy, the process of self-reflection, and narrative skills. Having learned the appropriate language and tools with carefully selected texts, students can be taught to transfer their knowledge to the clinical setting, writing pieces about the patients they encounter and reflecting on their own responses. Since portfolios focus on personal, individual attributes and experiences,[119] they are ideal for expressing the 'local and particular understandings about one situation by one participant or observer'[5] encouraged by narrative knowledge. Given the importance of student reflection in this context, care must be taken to create the conditions necessary for 'successful reflective use of portfolios': coaching by mentors; initial structure, especially for weaker students, with the freedom for students to move away from that structure once they are good at reflecting; eventual summative assessment; and the availability of experiences or other material on which to reflect.[149] Overall, the focus remains on process, rather than on content, and summative assessment must be accompanied by extensive formative feedback as the portfolio develops over time. The provision of selected literary texts by mentors when needed can help with the issue of availability of material in the absence of appropriate real experiences. Students can also continue to reflect briefly on assigned works of literature, particularly if a series of short works is chosen to accompany their rotations.[‡] Other related items which could be integrated into a portfolio include parallel charts, in which students write 'about aspects of the care of their patients that don't belong in the clinical chart but must be written somewhere',[150] or

[‡] Although the technicalities of their assessment is beyond the scope of this review, other branches of the arts have also been advocated as possible sources and means of reflection. For example, work has been done using 'film, art and drama'[48] as well as literature to provide students with patient, family, and physician perspectives of cancer. Gordon describes students' inclusion of both 'literary and art works (both their own and others)'[112] in portfolios used to assess personal and professional development.

copies of real clinical notes and letters, with patients names removed, to assess written communication skills.

Conclusion

The assessment tools appropriate for a literature and medicine curriculum do not meet the psychometric discourse's traditional evaluation criteria and, if they are to preserve their authenticity, are not likely to conform to them in the future. However, there are rigorous, usable criteria taken from another discourse, that of qualitative research, which are slowly being introduced into medical education research. From a theoretical perspective, using evaluation strategies that shun the positivist notion of truth is consistent with the objectives of a literature and medicine curriculum; given that literature is 'concerned above all with qualitative distinctions',[106] qualitative measures are apposite. Reflective of the need to create assessment congruent with the learning that it drives, the emphasis on individuality and engagement within both the constructivist and hermeneutic approaches to evaluation further highlight their suitability for the context of literature and medicine.

Rather than ignore the development of literature and medicine, the medical education research community should take on the challenges that it presents. Otherwise we run the risk that our immediate leap to objectified, reductionist evaluation systems, and our scorn of any discipline that does not comply with them, will prevent important constructs from being effectively taught and evaluated within medical education. Our students need to be taught empathy, self-reflection, and comfort with uncertainty, and our curricula need to remain open to other subjects that teach the competencies of physician as professional and of physician as person. Given the ongoing public concern with cynicism, lack of

professionalism, and burnout in the medical profession, the accountability agenda demands no less.

Acknowledgements

Support for this research was provided by the Clinician Educator Training Program, Department of Medicine, University of Toronto and by a Fellowship for Studies in Medical Education from the Royal College of Physicians and Surgeons of Canada. The author is grateful to many colleagues at the Wilson Centre for Research in Education for their helpful comments about this paper.

Correspondence: Ayelet Kuper, MD, DPhil, Wilson Centre for Research in Education, University Health Network, 200 Elizabeth Street, Eaton South 1-565, Toronto, Ontario, Canada M5G 2C4; e-mail <ayelet94@post.harvard.edu>.

References

1. Charon R. Narrative and medicine. N Engl J Med 2004; 350(9):862-4.
2. Verghese A. The calling. N Engl J Med 2005; 352(18):1844-7.
3. Jones AH. Narrative based medicine: narrative in medical ethics. BMJ 1999; 318(7178):253-6.
4. Greenhalgh T, Hurwitz B. Narrative based medicine: why study narrative? BMJ 1999; 318(7175):48-50.
5. Charon R. Narrative medicine. A model for empathy, reflection, profession, and trust. JAMA 2001; 286(15):1897-902.
6. Charon R, Montello M. Literature and medicine: an on-line guide. Ann Intern Med 1998; 128(11):959-62.

7. Charon R, Trautmann Banks J, Connelly JE, et al. Literature and medicine: contributions to clinical practice. Ann Intern Med 1995; 122(8):599-606.

8. Schneiderman LJ. Empathy and the literary imagination. Ann Intern Med 2002; 137(7):627-9.

9. Charon R. Narrative medicine: form, function, and ethics. Ann Intern Med 2001; 134(1):83-7.

10. Verghese A. The physician as storyteller. Ann Intern Med 2001; 135(11):1012-6.

11. Bamforth I. Literature, medicine, and the culture wars. The Lancet 2001; 358(9290):1361-4.

12. Skelton JR, Macleod JAA, Thomas CP. Teaching literature and medicine to medical students, part II: why literature and medicine? The Lancet 2000; 356(9246):2001-3.

13. Skelton JR, Thomas CP, Macleod JAA. Teaching literature and medicine to medical students, part I: the beginning. The Lancet 2000; 356(9245):1920-2.

14. Jones AH. Literature and medicine: narratives of mental illness. The Lancet 1997; 350(9074):359-61.

15. McLellan MF. Literature and medicine: narratives of physical illness. The Lancet 1997; 349(9065):1618-20.

16. Jones AH. Literature and medicine: narrative ethics. The Lancet 1997; 349(9060):1243-6.

17. McLellan MF. Literature and medicine: physician-writers. The Lancet 1997; 349(9051):564-7.

18. Jones AH. Literature and medicine: physician-poets. The Lancet 1997; 349(9047):275-8.

19. McLellan MF. Literature and medicine: the patient, the physician, and the poem. The Lancet 1996; 348(9042):1640.

20. Jones AH. Literature and medicine: an evolving canon. The Lancet 1996; 348(9038):1360.

21. McLellan MF. Literature and medicine: some major works. The Lancet 1996; 348(9033):1014.

22. McLellan MF, Jones AH. Why literature and medicine? The Lancet 1996; 348(9020):109.

23. Jones AH. Images of physicians in literature: medical Bildungsromans. The Lancet 1996; 348(9029):734.

24. McLellan MF. Images of physicians in literature: from quacks to heroes. The Lancet 1996; 348(9025):458.

25. Abse D. More than a green placebo. The Lancet 1998; 351(9099):362.

26. Calman KC. Literature in the education of the doctor. The Lancet 1997; 350(9091):1622.

27. Hudson Jones A. Literary perspectives on ageing. The Lancet 1999; 354(Supplement 3):S1.

28. Bolton G. Medicine, the arts, and the humanities. The Lancet 2003; 362(9378):93-4.

29. Evans HM, Greaves DA. Looking for emerging themes in medical humanities - some invitations to our readers. J Med Ethics: Medical Humanities 2003; 29:1-3.

30. Macnaughton J. 'Arts and humanities': a new section in *Medical Education*. Med Educ 2002; 36(2):106-7.

31. Dittrich LR. Preface. Acad Med 2003; 78(10):951-2.

32. Hawkins AH, McEntyre MC, eds. Teaching literature and medicine. New York: The Modern Language Association of America, 2000.

33. Association of American Medical Colleges. Curriculum directory. Washington, DC: Association of American Medical Colleges, 1998.

34. Number of U.S. medical schools teaching selected topics, 2003-2004. <http://services.aamc.org/currdir/section2/03-04hottopics.pdf >. Accessed 16 September. Curriculum Directory, Association of American Medical Colleges, n.d.

35. Meakin R, Kirklin D. Humanities special studies modules: making better doctors or just happier ones? J Med Ethics: Medical Humanities 2000; 26:49-50.

36. Hood K, Jacobson L, Houston H. Medicine and self-image in literature (Correspondence). The Lancet 2002; 359(9310):981.

37. Lazarus PA, Rosslyn FM. The Arts in Medicine: setting up and evaluating a new special study module at Leicester Warwick Medical School. Med Educ 2003; 37(6):553-9.

38. Kirklin D. The Centre for Medical Humanities, Royal Free and University College Medical School, London, England. Acad Med 2003; 78(10):1048-53.

39. General Medical Council. Tomorrow's doctors. Recommendations on undergraduate medical education. London: General Medical Council; 1993.

40. General Medical Council. Tomorrow's doctors. Recommendations on undergraduate medical education. London: General Medical Council; 2003.

41. Lancaster T, Hart R, Gardner S. Literature and medicine: evaluating a special study module using the nominal group technique. Med Educ 2002; 36(11):1071-6.

42. Dittrich LR, Farmakidis AL, eds. Academic Medicine Special Theme Issue: Humanities Education. Acad Med 2003; 78(10):951-1075.

43. Shapiro J, Duke A, Boker J, Ahearn CS. Just a spoonful of humanities makes the medicine go down: introducing literature into a family medicine clerkship. Med Educ 2005; 39(6):605-12.

44. Shapiro J, Longenecker R. Country doctors in literature: helping medical students understand what rural practice is all about. Acad Med 2005; 80(8):724-7.

45. Grant VJ, Jackson A, Suk T. Courses, content, and a student essay in medical humanities. J Med Ethics: Medical Humanities 2002; 28(1):49-52.

46. Jacobson L, Grant A, Hood K, et al. A literature and medicine special study module run by academics in general practice: two evaluations and lessons learned. J Med Ethics: Medical Humanities 2004; 30(2):98-100.

47. Glasser B. From Kafka to Casualty: doctors and medicine in popular culture and the arts - a special studies module. J Med Ethics: Medical Humanities 2001; 27(2):99-101.

48. Kirklin D, Meakin R, Singh S, Lloyd M. Living with and dying from cancer: a humanities special study module. J Med Ethics: Medical Humanities 2000; 26(1):51-4.

49. Anderson R, Schiedermayer D. The Art of Medicine through the Humanities: an overview of a one-month humanities elective for fourth-year students. Med Educ 2003; 37(6):560-2.

50. Wetzel P, Hinchey J, Verghese A. The teaching of medical humanities. Clinical Teacher 2005; 2(2):91-6.

51. Medical humanities syllabi. <http://endeavor.med.nyu.edu/lit-med/Syllabi.html >. Accessed 31 October 2005. Literature, Arts, & Medicine Database, New York University School of Medicine.

52. Syllabus database. <http://www.mhrd.ucl.ac.uk/cgi-bin/c-browse-t >. Accessed 1 November 2005. Medical Humanities Resource Database, University College London.

53. Olthuis G, Dekkers W. Medical education, palliative care, and moral attitude: some objectives and future perspectives. Med Educ 2003; 37(10):928-33.

54. Donohoe M, Danielson S. A community-based approach to the medical humanities. Med Educ 2004; 38(2):204-17.

55. Wear D, Aultman JM. The limits of narrative: medical student resistance to confronting inequality and oppression in literature and beyond. Med Educ 2005; 39(10):1056-65.

56. Jackson M. Back to the future: history and humanism in medical education. Med Educ 2002; 36(6):506-7.

57. Evans M. Reflections on the humanities in medical education. Med Educ 2002; 36(6):508-13.

58. Kneebone R. Total internal reflection: an essay on paradigms. Med Educ 2002; 36(6):514-8.

59. Burns CR. In search of wisdom: William Osler and the humanities. Med Educ 2003; 37(2):165-7.

60. Gull S. Embedding the humanities into medical education. Med Educ 2005; 39(2):235-6.

61. Gordon J. Not everything that counts can be counted. Med Educ 2005; 39(6):551-4.

62. Friedman LD. The precarious position of the medical humanities in the medical school curriculum. Acad Med 2002; 77(4):320-2.

63. Hunter KM, Charon R, Coulehan J. The study of literature in medical education. Acad Med 1995; 70(9):787-94.

64. Charon R. Literature and medicine: origins and destinies. Acad Med 2000; 75(1):23-7.

65. Kirklin D. Medical humanities: a practical introduction. London: Royal College of Physicians, 2001.

66. Evans M, ed. Medical humanities. London: BMJ Books, 2001.

67. Greenhalgh T, Hurwitz B, eds. Narrative based medicine: dialogue and discourse in clinical practice. London: BMJ Books, 1998.

68. Hunter KM. Doctors' stories: the narrative structure of medical knowledge. Princeton, NJ: Princeton University Press, 1991.

69. Chambers T. The fiction of bioethics: cases as literary texts. New York and London: Routledge, 1999.

70. Charon R, Montello M, eds. Stories matter: the role of narrative in medical ethics. New York and London: Routledge, 2002.

71. Regehr G. Trends in medical education research. Acad Med 2004; 79(10):939-47.

72. Frequently asked questions. <http://www.acgme.org/outcome/about/faq.asp >. Accessed 28 December. ACGME Outcome Project, Accreditation Council for Graduate Medical Education, n.d.

73. Learning objectives for medical student education - guidelines for medical schools. <http://www.aamc.org/meded/msop/msop1.pdf >. Accessed 28 December. Medical School Objectives Project, Association of American Medical Colleges.

74. Frank JR, Jabbour M, Tugwell P, et al. Skills for the new millenium: report of the societal needs working group, CanMEDs 2000 project. Annals RCPSC 1996; 29:206-16.

75. Frank JR, ed. The CanMEDS 2005 physician competency framework. Better standards. Better physicians. Better care. Ottawa: The Royal College of Physicians and Surgeons of Canada, 2005.

76. Neufeld VR, Maudsley RF, Pickering RJ, et al. Educating future physicians for Ontario. Acad Med 1998; 73(11):1133-48.

77. Chen FM, Bauchner H, Burstin H. A call for outcomes research in medical education. Acad Med 2004; 79(10):955-60.

78. Swanson DB, Case SM. Assessment in basic science instruction: directions for practice and research. Adv Health Sci Educ 1997; 2:71-84.

79. Eraut M. A wider perspective on assessment. Med Educ 2004; 38(8):800-4.

80. van der Vleuten CPM, Schuwirth LWT. Assessing professional competence: from methods to programmes. Med Educ 2005; 39(3):309-17.

81. BEME collaboration - individuals or institutions who are committed to the promotion of best evidence medical education. <http://www.bemecollaboration.org/index.html >. Accessed 28 December. BEME Collaboration.

82. Dauphinée WD, Wood-Dauphinée S. The need for evidence in medical education: the development of best evidence medical education as an opportunity to inform, guide, and sustain medical education research. Acad Med 2004; 79(10):925-30.

83. Torgerson CJ. Educational research and randomised trials. Med Educ 2002; 36(11):1002-3.

84. Murray E. Challenges in Educational Research. Med Educ 2002; 36(2):110-2.

85. Neufeld VR, Norman GR, eds. Assessing clinical competence. New York: Springer Publishing Company, 1985.

86. Jolly B, Spencer J. The metric of medical education. Med Educ 2002; 36(9):798-9.

87. Schuwirth LWT, van der Vleuten CPM. Merging views on assessment. Med Educ 2004; 38(12):1208-10.

88. Kubiszyn T, Borich G. Educational testing and measurement: classroom application and practice. 6th ed. New York: John Wiley & Sons, 2000.

89. Wass V, van der Vleuten CPM, Shatzer J, Jones R. Assessment of clinical competence. The Lancet 2001; 357(9260):945-9.

90. Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. Acad Med 1998; 73(9):993-7.

91. Norman GR, van der Vleuten CPM, de Graaff E. Pitfalls in the pursuit of objectivity: issues of validity, efficiency and acceptability. Med Educ 1991; 25(2):119-26.

92. Schuwirth LWT, van der Vleuten CPM. Changing education, changing assessment, changing research? Med Educ 2004; 38(8):805-12.

93. Cox R. Value of objective examinations. Nature 1972; 237(5357):489-92.

94. van der Vleuten CPM, Norman GR, de Graaff E. Pitfalls in the pursuit of objectivity: issues of reliability. Med Educ 1991; 25(2):110-8.

95. Meakin R. Medical humanities in undergraduate medical education - moving on. J Med Ethics: Medical Humanities 2002; 28:32.

96. Fleischman S. Call for paper for special issue of the Journal of Medical Humanities: 'Medicine and the humanities: the pedagogical landscape'. J Med Humanit 1999; 20(4):293-4.

97. Schwandt TA. Three epistemological stances for qualitative inquiry: interpretivism, hermeneutics, and social constructivism. In: Denzin NK, Lincoln YS, eds. Handbook of qualitative research. 2nd ed. Thousand Oaks, CA: Sage Publications, 2000:189-213.

98. Singer PA. Medical ethics. BMJ 2000; 321(7256):282-5.

99. Hurwitz B. Narrative and the practice of medicine. The Lancet 2000; 356(9247):2086.

100. Squier HA. Teaching humanities in the undergraduate medical curriculum. In: Greenhalgh T, Hurwitz B, eds. Narrative based medicine: dialogue and discourse in clinical practice. London: BMJ Books, 1998:128-39.

101. Coulehan J. Today's professionalism: engaging the mind but not the heart. Acad Med 2005; 80(10):892-8.

102. Wear D, Nixon LL. Literary inquiry and professional development in medicine: against abstractions. Perspectives in Biology & Medicine 2002; 45(1):104-24.

103. Rosenthal JM, Okie S. White coat, mood indigo -- depression in medical school. N Engl J Med 2005; 353(11):1085-8.

104. Dahlin M, Joneborg N, Runeson B. Stress and depression among medical students: a cross-sectional study. Med Educ 2005; 39(6):594-604.

105. Rachman S. Literature in medicine. In: Greenhalgh T, Hurwitz B, eds. Narrative based medicine: dialogue and discourse in clinical practice. London: BMJ Books, 1998:123-7.

106. Downie R. Medical humanities: means, ends, and evaluation. In: Evans M, ed. Medical humanities. London: BMJ Books, 2001:204-16.

107. Schön DA. The reflective practitioner: how professionals think in action. New York: Basic Books, 1983.

108. Anderson LR, Beer J, Coyle MJ, et al. English. Gloucester, UK: Quality Assurance Agency for Higher Education; 2000.

109. Aizlewood R, Davie M, Griffiths C, et al. Languages and related studies. Gloucester, UK: Quality Assurance Agency for Higher Education; 2002.

110. Altham JEJ, Bowie AS, Cameron JR, et al. Philosophy. Gloucester, UK: Quality Assurance Agency for Higher Education; 2000.

111. Arnot M, Bates D, Clark C, et al. History. Gloucester, UK: Quality Assurance Agency for Higher Education; 2000.

112. Gordon J. Assessing students' personal and professional development using portfolios and interviews. Med Educ 2003; 37(4):335-40.

113. Klenowski V. Developing portfolios for learning and assessment: processes and principles. New York: RoutledgeFalmer, 2002.

114. Cole G. The definition of 'portfolio' (Correspondence). Med Educ 2005; 39(11):1141.

115. Rees C. The use (and abuse) of the term 'portfolio' (Correspondence). Med Educ 2005; 39(4):436.

116. Rees C. 'Portfolio' definitions: do we need a wider debate? (Correspondence). Med Educ 2005; 39(11):1142.

117. Snadden D, Thomas M. The use of portfolio learning in medical education. Med Teach 1998; 20(3):192-9.

118. Challis M. AMEE medical education guide no. 11 (revised): portfolio-based learning and assessment in medical education. Med Teach 1999; 21(4):370-86.

119. Friedman Ben David M, Davis MH, Harden RM, Howie PW, Ker J, J PM. AMEE medical education guide no. 24: portfolios as a method of student assessment. Med Teach 2001; 23(6):535-51.

120. Driessen E, van Tartwijk J, Vermunt J, van der Vleuten CPM. Use of portfolios in early undergraduate medical training. Med Teach 2003; 25(1):18-23.

121. Schuwirth LWT, van der Vleuten CPM. Different written assessment methods: what can be said about their strengths and weaknesses? Med Educ 2004; 38(9):974-9.

122. Schuwirth LWT, van der Vleuten CPM. ABC of learning and teaching in medicine: written assessment. BMJ 2003; 326(7390):643-5.

123. Day SC, Norcini JJ, Diserens D, et al. The validity of an essay test of clinical judgment. Acad Med 1990; 65(9 Supplement):S39-40.

124. Neufeld VR. Written examinations. In: Neufeld VR, Norman GR, eds. Assessing clinical competence. New York: Springer Publishing Company, 1985:94-118.

125. Yang JC. Reliability of grading essay papers in a baccalaureate nursing programme. Nurse Educ Today 1987; 7(3):120-5.

126. Williams R, Sanford J, Stratford PW, Newman A. Grading written essays: a reliability study. Phys Ther 1991; 71(9):679-86.

127. Norcini JJ, Diserens D, Day SC, et al. The scoring and reproducibility of an essay test of clinical judgment. Acad Med 1990; 65(9 Supplement):S41-2.

128. Frijns PHAM, van der Vleuten CPM, Verwijnen GM, van Leeuwen YD, Wijnen WHFW. The effect of structure in scoring methods on the reproducibility of scores of tests using open-ended questions. In: Bender W, Hiemstra RJ, Scherpbier AJJA, Zwierstra RP, eds. Teaching and Assessing Clinical Competence (Proceedings of the Third International Conference on Teaching and Assessing Clinical Competence, Groningen, May 22-24, 1989). Groningen: BoekWerk Publications, 1990:466-71.

129. Fullerton JT, Greener DL, Gross LJ. Scoring and setting pass/fail standards for an essay certification examination in nurse-midwifery. Midwifery 1992; 8(1):31-9.

130. Cannings R, Hawthorne K, Hood K, Houston H. Putting double marking to the test: a framework to assess if it is worth the trouble. Med Educ 2005; 39(3):299-308.

131. Wass V, McGibbon D, van der Vleuten CPM. Composite undergraduate clinical examinations: how should the components be combined to maximize reliability? Med Educ 2001; 35(4):326-30.

132. Roberts C, Newble D, O'Rourke AJ. Portfolio-based assessments in medical education: are they valid and reliable for summative purposes? Med Educ 2002; 36(10):899-900.

133. Carraccio C, Englander R. Evaluating competence using a portfolio: a literature review and web-based application to the ACGME competencies. Teach Learn Med 2004; 16(4):381-7.

134. Davis MH, Friedman Ben David M, Harden RM, et al. Portfolio assessment in medical students' final examinations. Med Teach 2001; 23(4):357-66.

135. O'Sullivan PS, Reckase MD, McClain T, Savidge MA, Clardy JA. Demonstration of portfolios to assess competency in residents. Adv Health Sci Educ 2004; 9(4):309-23.

136. Rees C, Sheard C. The reliability of assessment criteria for undergraduate medical students' communication skills portfolios: the Nottingham experience. Med Educ 2004; 38(2):138-44.

137. Melville C, Rees M, Brookfield D, Anderson J. Portfolios for assessment of paediatric specialist registrars. Med Educ 2004; 38(10):1117-25.

138. Pitts J, Coles C, Thomas P. Educational portfolios in the assessment of general practice trainers: reliability of assessors. Med Educ 1999; 33(7):515-20.

139. Pitts J, Coles C, Thomas P, Smith F. Enhancing reliability in portfolio assessment: discussion between assessors. Med Teach 2002; 24(2):197-201.

140. Pitts J, Coles C, Thomas P. Enhancing reliability in portfolio assessment: 'shaping' the portfolio. Med Teach 2001; 23(4):351-5.

141. Snadden D. Portfolios - attempting to measure the unmeasurable. Med Educ 1999; 33(7):478-9.

142. O'Donovan N. There are no wrong answers: an investigation into the assessment of candidates' responses to essay-based examinations. Oxford Review of Education 2005; 31(3):395-422.

143. McMullan M, Endacott R, Gray MA, et al. Portfolios and assessment of competence: a review of the literature. J Adv Nurs 2003; 41(3):283-94.

144. Webb C, Endacott R, Gray MA, Jasper M, McMullan M, Scholes J. Evaluating portfolio assessment systems: what are the appropriate criteria? Nurse Educ Today 2003; 23(8):600-9.

145. Driessen E, van der Vleuten C, Schuwirth L, van Tartwijk J, Vermunt J. The use of qualitative research criteria for portfolio assessment as an alternative to reliability evaluation: a case study. Med Educ 2005; 39(2):214-20.

146. Seale C. The quality of qualitative research. London: Sage Publications, 1999.

147. Archibald DA, Newmann FM. Assessing authentic academic achievement in the secondary school. Reston, VA: National Association of Secondary School Principals, 1988. As cited in: Friedman Ben David M, Davis MH, Harden RM, Howie PW, Ker J, J PM. AMEE medical education guide no. 24: portfolios as a method of student assessment. Med Teach 2001; 23(6):535-51.

148. Moss PA. Can there be validity without reliability? Educational Researcher 1994; 23(2):5-12.

149. Driessen E, van Tartwijk J, Overeem K, Vermunt J, van der Vleuten CPM. Conditions for successful reflective use of portfolios in undergraduate medical education. Med Educ 2005; 39(12):1230-5.

150. Charon R. Rita Charon. The Lancet 2004; 363(9406):404.