Conceptual challenges in tailoring physician performance assessment to individual practice

D E Melnick1, D A Asch2, D E Blackmore3, D J Klass4 & J J Norcini5

Practice inevitably narrows over time. Therefore, testing of established doctors requires that their assessment be tailored to a far narrower practice than is appropriate for testing of new doctors who have not yet differentiated. In this paper, we address the conceptual challenges of tailoring physician assessment to individual practice. Testing of established doctors needs to reflect that physicians specialise, often in idiosyncratic ways; otherwise, the testing will not be credible among established doctors and will not reflect the realities of their practice. Despite the importance of these goals, the conceptual and methodological challenges of creating tailored assessments remain daunting.

Keywords clinical competence/ *standards; physicians, family/ *standards; education, medical, continuing/ *standards; quality of health care/ standards.

Medical Education 2002;36:931–935

Introduction

The knowledge, abilities or potential of doctors recently completing undergraduate medical education can be assessed against the broad standard of what is taught and expected in medical schools. These assessments should cover a wide range of clinical domains (e.g. psychiatry, obstetrics, surgery) as well as professional norms of behaviour, judgement and other general skills and qualities. Indeed, initial licensing, certifying, and registration examinations are typically broad, reflecting in part that new doctors have not yet defined a narrower practice, even if they intend to do so in the future.

The appropriate method of assessment of a doctor a dozen years beyond initial training, however, is not quite as clear. As doctors mature in the workforce, their practice typically narrows. Some treat only children; some treat only women. Some spend most of their time performing surgery; others spend most of their time giving behavioural therapy. To assess the performance of established doctors, it would be inappropriate to assess them solely against the broad standards that were applied when they graduated from medical school.

Instead, it would be more appropriate to tailor the assessment, at least in part, to reflect their actual practice. Such tailoring presents great challenges, however. Although the domains represented by traditional certifying organisations (e.g. in specialties such as internal medicine or subspecialties such as cardiology) are structured and definable, the actual work of many practising clinicians may differ substantially from those structured domains or may disproportionately emphasize certain elements within them.

Most American specialty certifying boards now require periodic assessment for recertification1 and there are calls for periodic re-examination for licensure in the US.2 The General Medical Council (GMC) in the UK has announced its intent to implement a periodic revalidation system, including several approaches to appraisal and assessment.3,4 Canadian provincial licensing authorities are planning or have implemented periodic assessment for some or all doctors.5–7 Similar efforts are underway in Australia.8 All of these efforts are premised on the growing belief that some form of continued assessment beyond that provided at professional entry is required to protect the public and maintain professional standards. In this paper, we identify the conceptual and procedural challenges presented by tailored assessment of practice in support of periodic evaluation.

Multiple audiences

One challenge for tailored assessments concerns the fact that there are multiple audiences for certification,
Key learning points

Doctors differentiate as they practise. Assessment of competency in practice demands recognition of the differentiation.

Tailoring assessment to practice characteristics requires the systematic collection of information about the doctor’s practice.

Practice characteristics may influence both assessment content and assessment method or format.

The purpose of the assessment will determine the mixture of practice-related and domain-related content.

The purpose of the assessment may demand the inclusion of core content in addition to practice-tailored content.

registration and licensure. For example, when assessment is used as part of a system of public accountability – to satisfy public trust in the profession – the assessment must in part reflect the public’s expectations of what doctors ought to know and be able to do. The assessment system must justify the trust that patients put in the profession. The average patient assumes that licensure, registration and/or certification attest to the competence of the doctor. Although doctors recognize the limits of licensure as a reflection of specialty ability, patients may expect more. Some of these expectations may be reasonable. For example, patients may assume that all doctors can either perform cardiopulmonary resuscitation or are aware of their own limitations and know when to refer elsewhere. Some expectations might be unreasonable – for example, that any licensed or registered doctor is skilled at cosmetic surgery. An assessment system accountable to such an unreasonable expectation could never be tailored because the expectations themselves are not bounded.

The profession represents another audience or stakeholder in the assessment process. When assessment leads to certification, the profession expects the certificate to represent the breadth of the discipline in which it is granted. Thus, the assessment and the profession in turn define and support each other.

Individual doctors represent yet another audience. When the individual doctor uses assessment to guide continued professional development, the central requirement for validity is the assessment’s ability to represent the important domains of individual practice. Indeed, the acceptance of assessment by practising doctors will increase if they perceive the assessment to be valid and fair, a perception more likely to be related to a sense of relevance to everyday tasks than to excellent psychometric characteristics.

Overlapping domains

A second challenge to tailored assessments concerns the fact that doctors’ practice domains vary considerably and may be extremely difficult to determine – even conceptually. Figure 1 demonstrates several conceptual domains that might be relevant from different perspectives. If one were to catalogue or sample the activities of an individual doctor, one could develop a picture of the practice domain of that doctor. This observed practice might be the appropriate domain for a tailored assessment, because it reflects what that doctor actually does. However, using that narrow domain falls short for several reasons. Firstly, the doctor’s practice might change over time and assessments tailored to current practice are not necessarily relevant to future practice. Secondly, the observable practice exists within a larger domain that might be called the ‘potential practice’. This represents things a doctor does not ordinarily do, but which might be reasonable expectations. For example, a doctor who limits practice to the care of women might occasionally treat the male partner of a patient with a sexually transmitted disease. The same doctor might do very little office gynaecology and an observation of a sample of practice experiences might

Figure 1 Relationships of observable practice to domains of interest.
reveal no examples. However, that doctor should be prepared to engage in office gynaecology as necessary or might claim to be competent in that area. Moreover, given that patients are an important audience for licensure and certification, such a doctor might have patients who naturally expect proficiency in that area.

Thirdly, the potential practice generally falls within a larger domain, which may represent a subspecialty or specialty discipline, or encompass all of medicine. Nevertheless, it is important to distinguish specialty-defined domains from observable and potential practice domains. A board-certified internist might practise only in a centre devoted to women’s health issues. If one were to observe the practice, one would find that all the patients were women, and nowhere within the observable practice domain would one find the management of erectile dysfunction. Nor might this problem exist in the potential practice, because the doctor never expects to treat men. An assessment tailored only to the actual or potential practice would therefore not include measures of ability to manage erectile dysfunction.

However, management of erectile dysfunction clearly falls within the domain of an internist. To the extent that assessment is meant to help the doctor represent himself more broadly as an internist, the assessment ought to conform to the broader domain defined by that specialty – to what might be called the professional field. Professional fields are defined by organisations rather than by individual practices: they are socially rather than individually constructed. In general, because professional fields are constructed by organisations, they can be applied consistently across individual doctors. They do not, however, exhaust the space within which each individual doctor might practise. Indeed, an individual doctor’s actual or potential practice may fall outside the accepted boundaries of a professional field: a doctor may, for example, sometimes use herbal remedies with patients.

Finally, patient expectations will cut across actual and potential practice. They will include some domain characteristics not represented in the practice and they may extend beyond the professionally defined field. Because patients are legitimate stakeholders in doctor assessment, their expectations are relevant, although not necessarily authoritative, in determining the scope of tailored domains.

**Information required to tailor assessment to practice**

Tailoring assessments to actual practice (observable or potential) will probably have the most relevance and credibility for both patients and doctors. But even if there is agreement that these are appropriate domains for a tailored assessment, a further challenge lies in identifying effective methods of characterising the observable or potential practices of individual clinicians.

In the future, characterisations of observable practice may be derived from comprehensive, integrated data systems that track patient encounters in individual practices. In the absence of such automated data systems, the available options include self-reporting by the doctor and observational methods. The latter would be costly and intrusive.

The simplest approach to self-reporting is to ask the doctor to select assessment modules that appear to be relevant to their practice. Unfortunately, in the absence of objective data, doctors may misperceive the real characteristics of their practice. Alternatively, the doctor can be asked to compile a diary, documenting relevant characteristics of the practice. Some of those characteristics might include a condition profile (complaint, diagnosis and acuity), procedure and treatment profile, patient demographics (age, gender, socio-economic status, race/ethnicity, etc.), settings of care, patient financing, and doctor characteristics (such as professional standing, health status, etc.). The many dimensions for characterising a practice reveal the conceptual and practical complexity of such a task.

Structured self-reporting will suffice for guiding continued professional development. However, as the stakes of the assessment increase, as in recertification, revalidation or relicensure, the potential for manipulating the assessment and its outcome rises. Self-reporting might require external audits to assure both accuracy and reasonable sampling of the practice or methods that add a greater degree of objectivity to the descriptive process.

Historical assessments of practice character, no matter how they are performed, may fail against two other standards. Firstly, if assessments need to cater to public perceptions of doctors’ practices, then they may need to extend beyond actual or potential practice and into areas where the public might reasonably expect doctors to have competence. These judgements are not likely to be made easily, and doctors might justifiably complain that they should not be held to standards they cannot control.

Secondly, historical assessments of practice do not necessarily reflect future practice patterns. Assessments that are tailored to those past practices may fail as enduring reflections of ability – not just because doctors’ abilities may change or because new knowledge develops, but because doctors may substantially change their scope of practice over time. In the USA during the 1990s, some anaesthesiologists who had...
been trained in internal medicine and anaesthesia abandoned their anaesthesia practices and re-established themselves as internists. An assessment tailored to practice before that shift would be misleading if represented as assuring capability for the new practice profile.

Once the doctor’s practice is objectively characterized, the assessment can be tailored to the characteristics of the practice. Most obviously, this tailoring will affect the content of the assessment. The doctor caring for patients only in an ambulatory setting would not be asked about patients in critical care, inpatient settings. The doctor whose practice is limited to diseases of children (whether or not certified as a paediatrician) will not be asked about the management of elderly patients.

Information from the practice might also be used to tailor the assessment methods. Different assessment tools are more appropriate for different purposes. A multiple-choice question or essay test may effectively assess knowledge about diagnosis, but is unlikely to effectively assess skill at performing a procedure. A practice dominated by procedures might call for assessment tools other than multiple-choice or essay examinations. Tailoring the method to match the measurement demands of the doctor’s actual practice is more likely to produce a valid measure and is certain to result in assessments that are better accepted by doctors, but it also presents considerable logistical challenges.

Finally, basing assessment on objective practice information allows a sequential evaluation process in which efficient screening assessment tools can be augmented by more in-depth assessment tools in those areas most critical to practice.

Is there a ‘core’ domain?

A vexing dilemma in practice-tailored assessment concerns the question of the existence of a ‘core’ domain comprising the knowledge, skills and other attributes that should be required regardless of practice characteristics. The definition of the ‘core’ may vary according to the purpose of the assessment.

For example, if a doctor whose observable practice is limited to women’s health should wish to retain her credentials as an internist, the specialty certifying authority for internal medicine would nevertheless require that she also demonstrate competency in conditions unique to men. That authority is likely to define the domain of the internist in a way that is largely independent of the many practice choices that individual internists may make. In the context of specialty certifying assessments, the domain is likely to be the same as ‘core.’ This is a major reason why reassessment for specialty certification is conceptually easier than reassessment for licensure.

For a relicensure or revalidating examination, the core may relate to some subset of the domain encompassing all of medicine – a subset of competencies presumed of all doctors granted the privilege of practice, regardless of specialty or practice profile. But what might those core competencies be? In most instances, the legal authority to practise medicine is granted without restrictions regarding the scope of practice, and the renewal of the license or revalidation of the doctor continues to grant that broad right despite the doctor’s own limitation of the scope of practice. If doctors are assessed later in their careers, are there core elements that all should be expected to satisfy? The notion that the core domain applies to all doctors implies that its definition is not based on data about an individual practice. In some instances, elements of the core domain will relate to criticality – low frequency but high impact knowledge or skills that are unlikely to emerge from a data-based review of practice characteristics. Some aspects of the core domain may represent responsiveness to public or professional concerns, although these may not be directly related to the practice characteristics of every doctor. Others might reflect general professional behaviour. Whatever the elements, they are likely to be so narrow as to provide no useful guide to an assessment tailored to actual practice.

These relationships highlight the need to consider the appropriate balance of both practice-tailored and core domains in constructing assessments relevant to the practising doctor. As demonstrated in Fig. 1, reference to the observed practice alone will ignore both the potential practice and the domains reflecting patient expectations. On the other hand, restricting assessment to core domains will diminish relevance and will miss aspects of both the practice and patient expectations.

Development needs

This paper provides the rationale for the development of practice-tailored assessments for relicensure, revalidation, recertification and continued professional development. However, the tools that would allow this tailoring to occur are rudimentary and much of the conceptual framework that might support such tailored assessment is undeveloped.

Further study is needed to help define the parameters that would be useful in objectively characterising a practice. How accurate is doctor-provided practice data
and what attributes of the data collection system enhance accuracy? What elements are valuable in tailoring assessment? In particular, how can data collection instruments gather information on unanticipated (outside the domain) aspects of the practice? How can information aggregated in other contexts be utilised to tailor assessment?

The notion of tailored assessment raises questions about the meaning of licensure, registration and certification and the audiences these credentials reach. What are the implicit guarantees in a licence or registration to practise medicine and to whom are they provided? What role should public expectations play in licensure, revalidation and certification, and who determines the public will? What do certification and recertification imply about domain mastery? These issues are critical to determining the balance between the core and practice-tailored components of an assessment system.

A number of technical measurement considerations remain after these conceptual questions have been addressed. How can assessments that differ in content and method be equated so as to provide fairness across disparate doctor practices? With many, often nearly unique assessments in use, how are the traditional measurement concerns of reliability and validity of the assessments to be addressed?

As numerous and as daunting as these questions are, they remain essential as we move beyond large-scale, one-size-fits-all assessment systems toward those that reflect what doctors really do, and what patients really need and expect.

**Conclusion**

Doctors naturally narrow their practice over time, and so the assessment tools appropriate for new, undifferentiated doctors can never reflect the realities of practice for more established practitioners. Instead, assessments of established doctors should be tailored to their actual or potential practices. We have described several conceptual and practical challenges to structuring these assessments – difficulties in defining the information required to customise evaluations, the methods for collecting it, and how it should be applied to the tailoring of examinations. Until these challenges are addressed, tailored assessments of doctors’ abilities cannot be achieved.

**Contributors**

All authors contributed equally to the discussions undertaken during the 10th Cambridge Conference that led to the writing of this article. DEM undertook main responsibility for preparing the draft, co-ordinating input from the other authors and writing the final version of the paper.

**Acknowledgements**

Grateful acknowledgement is made to the sponsors of the 10th Cambridge Conference: the Medical Council of Canada, the Smith & Nephew Foundation, the American Board of Internal Medicine, the National Board of Medical Examiners (USA) and the Royal College of Physicians (UK).

**Funding**

Support for the development of this paper was provided by the employing organization of each author and the 10th Cambridge Conference sponsors acknowledged above.

**References**


Received 21 March 2002; editorial comments to authors 14 May 2002; accepted for publication 26 June 2002