Five myths and the case against a European or national licensing examination

RONALD M. HARDEN
University of Dundee, UK

Abstract
The introduction of a European licensing examination or national examinations, where these do not already exist, offers significant advantages. These are more than offset, however, by the disadvantages and the collateral damage incurred. Five myths about centralizing examinations are explored. Myth 1: The claim that a central examination will ensure that candidates are assessed in important areas of medical practice is unfounded. What tends to be assessed are learning outcomes that can be easily assessed. These are often not the important outcomes related to the overall competence of a doctor. Myth 2: It is claimed that a central examination will lead to improvements in assessment practice. The evidence is that this is not the case and that, in fact, a central examination stifles change and inhibits innovation. Myth 3: A central examination, it is suggested, will meet a need for greater uniformity. There is also an important need to recognize diversity. Myth 4: Central examinations are seen as an indicator that will track the performance of the system. The limitations of the data, however, are usually not recognized and there maybe unfortunate and unintended consequences if the results are used in this way. Myth 5: Finally, a major argument proposed for a European or national examination is that it will lead to safer medical practice and that this will protect the patient from substandard practitioners. There is, in fact, no evidence to support this argument. There is a need for further work and new initiatives on standards and quality improvement in assessment approaches. This can be achieved in a number of ways including monitoring the assessment process and sharing tools and assessment approaches between schools.

The case for a European or national examination is a seductive one, particularly at a time when on the agenda are the implications of globalization, the need for standards, the Bologna process and a demand from the public for greater monitoring of the work of the medical profession. Few would disagree about the need for a robust system to assess the competence of a doctor to practice. A European or national examination, where this does not already exist, have been proposed as the answer to the problem. The advantages of a more centralized approach to examinations with the potential of greater uniformity are obvious and have been well articulated. What have not been discussed sufficiently, however, are the associated disadvantages or side effects – the collateral damage. A centralized assessment system may not be the wonder drug it appears to be on first inspection, because of the almost inevitable drawbacks that come with it. Some of these are described in this article. Five myths about a centralized approach to licensing examinations are presented.

Myth 1: A European or national examination will ensure that the candidate is assessed in important areas of medical practice
No! A centralized examination will emphasize not necessarily the most important learning outcomes but instead those which can be readily assessed at a national or European level. The focus will be on written examinations, often with multiple choice questions, where knowledge rather than clinical skills or attitudes are assessed. As Schuwirth (2007, p. 1023) argued ‘National tests typically focus on “low” level skills, whereas currently there is growing interest in higher order skills and competences, such as “professionalism”, “scholarship”, the ability to think as a “scientist”, etc.’. With the move to competence assessment programmes, we need a combination of different assessment methods including both traditional and new forms of assessment with the criteria used to evaluate their quality derived from both psychometrics and edumetrics (Baartman et al. 2007). This combination is unlikely to be found with a central examination.

Burke (2008) noted that 'The implementation of a national qualifying exam risks overemphasizing academic achievement at the expense of the balanced range of skills required to be a good doctor and enjoy a long and prosperous future career’. As judged by their performance in a national qualifying examination, the glittering undergraduate star may be as much use as a chocolate fireguard at postgraduate level (Neilson 2008).

A problem is that in the past we have emphasized in the measurement debate, the concept of reliability. We need now to focus more on Fredericksen and Collins’ (1989) concept of ‘systemic validity’ which they define in terms of inducing in the education system, curricular and instructional changes. A centralized examination, assessing as it does less important...
learning outcomes, is likely to influence and distort a teacher’s teaching and the students’ study patterns. There has been a growing recognition that teaching practice can easily become dominated by what is referred to as an unwelcome ‘backwash’ from national testing programmes (Winter 2003). Current moves to increase the emphasis in the curriculum on ‘soft skills’ such as the ability to generate fresh and original ideas, to work in teams, and to be empathetic are undercut when assessment regimes largely test the recall and manipulation of facts (Cole 2007). In the USA National Board of Medical Examiners’ (NBME) USMLE assessment, a clinical component with simulated patients, was added recently to the examination, but the emphasis remains on the written elements. ‘First Aid for the USMLE Step 1 remains one of the most popular textbooks for students. No one can blame teachers and students from thinking that what is important is what is tested in the examinations.

Myth 2: A European or national examination is likely to encourage change and lead to improvements in assessment practice

No! The evidence is that where there is a system of centralized examinations, innovation in assessment is hindered. It could be argued that as assessment is a highly specialized area of medical education and experts in the field may not be evenly distributed across medical schools and may work in relative isolation from others with a similar interest, bringing together a minimum critical mass of experts in one central location and charging them with the delivery of a central examination will lead to improvements in assessment practice. The NBME is probably the best example of a central team of assessment experts and it has made significant contributions internationally to advances in the practice of assessment and to our understanding of the field. Because, by necessity its priorities have to be different, however, such a body is not well placed to work at the cutting edge of assessment practice and to lead new initiatives in the field. Peter Scoles highlighted, at a session on the NBME position on the Gateway Examination in the USA, that 6–10 years was required to change the process for the USMLE. It was 30 years after the objective structured examination (OSCE) was first introduced in the final examination at the medical school in Dundee (Harden et al. 1975) and many years after it was introduced by a number of other schools that it was adopted as an approach to be used in a centrally organized examination. New approaches such as the OSCE or the use of portfolios for assessment purposes (Friedman et al. 2001) are more readily introduced in schools where assessment decisions can be taken at a local level by the local curriculum committee working with educationists in a medical school where they are not constrained by central dictates or requirements.

There is a fundamental concern about a central examination and its impact on progress with assessment practice. A central examination by its very nature perpetuates an outmoded view of assessment and encourages the misconception that students are best evaluated at one point in time and that assessment is something separate from the curriculum and the teaching programme. As Schuwirth (2007, p. 1022) put it ‘in terms of competences, this (single-shot assessment) constitutes a misalignment between assessment and education’. Broadfoot and Black (2004) have emphasized the limitations of ‘one-shot’ testing in a central examination and have advocated the need for greater use to be made in assessment of multiple sources of data when inferences are drawn about performance. Gipps (1995) has argued that the unintended bias between different students and schools can be reduced only with multimodal tasks encountered in non-threatening settings.

Looking to the future, Schuwirth and Van der Vleuten (2006, p. 18) suggest that ‘Assessment will be less viewed as an external measurement of the results of the educational process but more as an integral part of the process’. They argue that ‘Assessment is no longer seen exclusively as a psychometric measurement problem, but more as an educational design problem. This implies that the purpose of assessment is not merely to determine whether a candidate is up to standard, but more how the information about the candidate’s competence can best be used to tailor the teaching or the courses to individual needs’. This is most likely to be achieved when assessment is integrated into the curriculum and not a single-shot end of course in national or international test.

Myth 3: In an age of standards and globalization, a European or national examination will meet our needs by encouraging uniformity

No! Undoubtedly there are core competencies that are likely to be common to all schools and where a common assessment might be appropriate, but there are also important and significant differences. This is an important downside to the uniformity imposed by a central examination. The ‘Scottish Doctor’ (Scottish Deans’ Medical Education Group (SDMEG) 2008), for example, identifies the learning outcomes that are common to all five Scottish medical schools, while at the same time recognizing that some learning outcomes may vary from school to school. This diversity needs to be recognized in any assessment programme. Indeed in the UK, the General Medical Council has been promoting equality and diversity among the aims of its Quality Assurance Programme. The problem, as defined by Carney (2003), is that the use of tests promoted by international donor nations has resulted in a global emphasis on cognitive achievement without a consideration of what local values might define as good education.

Myth 4: A national examination is of value as an indicator that helps to track the performance of the system

No! While the use of students’ results in a central examination, and the statistics derived from these, is appealing to governments, funding authorities, employers and even potential students as an indicator of success or failure, the limitations of the data may not be recognized and there may be unintended consequences. Smith (1995) identified a ‘huge number of instances of unintended behavioural consequences of the publication of performance data’ and described problems such as ‘tunnel vision, sub-optimization, myopia,
measure fixation, gaming, ossification, misinterpretation and misrepresentation’. In a useful review of ethical issues and indicator systems, Fitz-Gibbon and Tynan (2002) suggested that ‘A major ethical imperative is to do good rather than harm. At the very least we must try to observe the Hippocratic oath and “at least do no harm”’. They argue that an indicator may not reflect the complexity of the situation and that the system will be undermined as gaming takes hold with the efforts of the institution focussed on the indicator (student performance in the national examination, for example) at the expense of other issues. They express concern that ‘because schools freely choose to buy into indicator systems, this is proof that they find indicators useful. However, people buy snake-oil, and the commercial argument is never adequate’.

The study of Hecker and Violato (2007, p. 112) demonstrates the problems that can arise when performance in a national examination is used as an indicator for curriculum evaluation. Based on USMLE scores, they made the sweeping conclusion that ‘changing curricula in medical education reform is not likely to have much impact on improvement in student achievement’. They did recognize, however, that in their study the dependent variables are licensing exam scores, which may not fully access important physician characteristics such as empathy, problem solving, clinical reasoning and so on, that are learning objectives of some medical schools.

This problem with standardized tests was highlighted in ‘Engines for Education’ (Engines for Education 2008). ‘The problem with standardized tests and the fixed curricula they engender is their tendency to kill off the kind of education that matters most. But who can blame a teacher or school for orienting the lesson towards helping students pass those tests with high marks? The temptation to teach students to do well on standardized tests is almost unavoidable when performance on such tests is how entire school systems are evaluated.’

Myth 5: A European or national examination will lead to safer medical practice and protect the public from substandard practitioners

No! There is no evidence that the absence of a centralized examination leads to the graduation of substandard doctors (Noble 2008) nor that the introduction of such an examination will lead to an improvement in the practice of medicine. If one compares medical practice in terms of the number of errors, the quality of care perceived by patients and the overall competence of doctors, in a country without a national examination, such as the UK, with a country where there is a national examination, such as the USA, little difference can be found. There is no evidence that the graduate of a UK school where there is no national examination is in some way inferior in terms of their competence to practice medicine than the graduate of an American school who has passed the relevant national examinations. A national examination almost certainly would not have avoided the high profile examples of unacceptable practice as such as the case of Dr Shipman, who most probably would have passed such an examination.

Indeed, counterintuitively, by neglecting the aspects of competence which are most often associated with problems and errors in practice, a centralized examination may have a negative rather than a positive effect on medical practice. In a case-control study, Papadakis et al. (2005) found that disciplinary action taken against practicing doctors by medical boards in the USA was more strongly associated with records of unprofessional behaviour in medical school, than with their performance as measured by a national examination. Studies of newly qualified doctors too, have identified the deficiencies of graduates to be in areas such as prioritizing work, dealing with acutely ill patients and practical prescribing, not with their knowledge base.

In any consideration of the doctor’s future competence to practice safely, it also has to be noted that on graduation students have to undergo a further period of supervised training with additional examinations at the end of it.

The evidence and the facts relating to quality of care do not support the case for an European or national examination.

Conclusions

There are undoubtedly significant advantages to be gained from the implementation of a common European examination or a national examination in countries where one does not already exist. The case presented in favour of adopting such an approach, however, is seriously flawed and ignores the collateral damage and the major disadvantages associated. These are likely to outweigh any benefits gained. Indeed the result of such an initiative may be to hold back rather than to improve assessment practice. There are significant educational, psychometric, logistical, financial, fiscal, cultural and diversity implications associated with the introduction of a centralized examination. There is the risk too that such a development may be interpreted as downgrading the status of schools by suggesting they are not competent to make important decisions about their students.

Given the lack of evidence as to the overall benefits, one has to ask whether the significant resources required to implement a European or national examination might be used in other ways to improve the assessment practice. It needs to be emphasized that abandoning the idea of a centrally organized examination does not mean that we should ignore the need for further work on standardization and quality control in the area of assessment. This can be considered as a continuum with stages in the continuum:

(1) Assessment procedures in the medical school are completely independent with no external standards.

(2) The school’s assessment procedures are subject to external inspection and approval in line with specified standards (such as those offered by the WFME).

(3) Examinations in a school are monitored by external examiners including student pass/fail decisions (the current practice in the UK, with the process quality assured by the General Medical Council).

(4) Assessment instruments and some questions are shared between schools and common questions are introduced in examinations.

(5) Students are required to take and pass both an internal school examination and an external European or national examination.
(6) Only a single European or national examination is taken by the student.

It is argued in this article that both ends of the continuum (stages 1 and 6 and probably stages 2 and 5) are untenable and that we should position ourselves with regard to assessment some where in the middle (stages 3 or 4). The exchange of information between schools about good assessment practice and the sharing of expertise and technologies should be encouraged along with further work on faculty development, assessment guidelines and standards. This approach is much less likely to be threatening to schools and more likely to result in improvements not only in the assessment procedures but also in the overall curriculum. It is this approach that has been adopted by the five Scottish medical schools building on their earlier collaborative work on learning outcomes as set out in the publication ‘The Scottish Doctor’ (SDMEG 2008).

Declaration of interest: The author reports no conflict of interest. The author alone is responsible for the content and writing of the article.

Notes on contributor

PROFESSOR RONALD M HARDEN was formally Professor of Medical Education and Director of the Centre for Medical Education, Teaching Dean and Postgraduate Dean at the University of Dundee, Scotland. He has wide interests in medical education including curriculum planning, assessment and the use of new technologies in teaching and learning.

References


Schuwirth LWT. van der Vleuten CPM. 2006. How to design a useful test: The principles of assessment. UK: ASME.
