Medical education and the maintenance of incompetence

BRIAN HODGES
Wilson Centre for Research in Education, University of Toronto, Canada

ABSTRACT We think of medical education as a process that moves novices from a state of incompetence to one of competence. This paper explores the idea that education may, at times, actually lead to incompetence as a result of over-emphasizing particular discourses that construct what competence is. This paper explores four discourses each with its own terminology and core conceptualizations of competence; each of which creates different roles for students and teachers. No one discourse is ideal and all drive teaching and assessment in particular ways. Sometimes these forms of teaching or assessment may inadvertently foster incompetence. In this paper I argue that, as with medical treatments, medical educators must pay more attention to the side-effects of the discourses that shape medical education.

Introduction

We think of medical education as a process that moves novices from a state of incompetence, to one of competence. This paper, based on a keynote presentation given at the Association for Medical Education in Europe on 16 September 2006 in Genoa, Italy, explores the idea that medical education may, at times, actually lead to incompetence.

When I refer to incompetence, I am not talking here about the uncommon cases of gross incompetence, such as sexual abuse of patients, drug addiction, psychopathy or the rare but high profile serial killers who attract so much media attention. Rather I am focusing on the more grinding and mundane incompetence that harms the quality of patient care and diminishes the contribution of the profession as a whole to societies that have conferred on it so much privilege. These forms of incompetence are harder to characterize, are often hidden, and most importantly are in constant flux.

In this paper, I argue that this type of incompetence is largely a 'side-effect' of medical education and more specifically is the result of overemphasizing particular models of education. We all adhere to these models, which sociologists call 'discourses' because we believe in them, because they help order our world and because they give meaning to our work in such a way that we can communicate it to others. Discourses are ways of seeing the world. They act like lenses or filters, and they make it possible for us to say some things but not others. Discourses also make it possible to act in certain ways, and to have certain jobs.

Michel Foucault wrote about the discourses of madness over the last four centuries. He showed that there was a discourse of 'madness as spiritual possession' that was dominant in the Middle Ages in western countries; a discourse of 'madness as deviancy' that was dominant in the Victorian era and most recently a discourse of 'madness as medical illness' that has been dominant since the beginning of the twentieth century (Foucault, 1965).

The most important implication of these different discourses is that they create very different possibilities for people and for institutions. Thus, a discourse of 'madness as spiritual possession' makes visible 'possessed individuals' and creates a role for spiritual healers and religious institutions. On the other hand, a discourse of 'madness as deviancy' makes visible 'deviant individuals' and creates a role for judges and jailors working in courts and prisons. Finally, a discourse of 'madness as medical illness' makes visible 'mentally ill individuals' and creates a role for psychiatrists and psychologists who work in clinics and hospitals.

As Foucault said, '...we are not dealing with the same madmen' (Foucault, 1969). Clearly, 'possessed' is not the same as 'deviant' which is not the same as 'mentally ill'. Similarly, a spiritual healer is not equivalent to that of judge or jailor, both of which are substantially different from a psychiatrist or psychologist.

A similar analysis can be undertaken with regard to incompetent doctors, because incompetence, like madness, has also been defined in different ways at different times. I will examine some of the older variations first. In the 1700s a competent doctor was a member of a guild who carried a blade for blood letting and emetics for purging with the goal of balancing the humours of the body (Shorter, 1985). In 1850 by contrast, a competent doctor was a gentleman (there were almost no women doctors) with a walking stick who diagnosed patients by looking at their tongue, and smelling their urine (Cathell, 1890; Shorter, 1985). By 1950 a competent doctor, still most likely to be a man, wore not a suit but a white coat, discussed a woman's health with her...
herself: of competence/incompetence that are in current use. I have called these: Hodges, 2005a).

Now I will turn to a discussion of the discourses of competence/incompetence that are in play today. The following is based on an analysis I have been undertaking over the past five years that has involved examining over 600 medical education articles, a set of 25 interviews with key figures in medical education and visits to medical education institutions in the US, UK, Canada, France, Israel, China, Jordan, Ethiopia, Pakistan, Poland and Japan (Hodges, 2005a).

Through this research I have characterized four discourses of competence/incompetence that are in current use. I have called these:

1. Harrison’s Textbook and competence-as-knowledge;
2. Miller’s Pyramid and competence-as-performance;
3. Cronbach’s Alpha and competence-as-reliable test score;

Table 1 summarizes the keywords, roles of teachers and students and measure of competence associated with each of these discourses.

(1) Harrison’s Textbook and competence-as-knowledge

The discourse of ‘competence-as-knowledge’ is characterized by use of words such as: ‘facts, foundational knowledge, basic science, first principles, fund of knowledge, classic text books, classic articles and multiple-choice tests’. In this discourse, the role of teacher is to be the source of wisdom, and the main activity revolves around helping students to receive or integrate knowledge. Core teaching activities are didactic lectures and seminars that aim to transmit and integrate knowledge. The most common measure of competence-as-knowledge is a written test, usually consisting of multiple-choice questions. This combination of teaching and assessment methods creates a role for students that consists to a large degree of memorizing and reproducing knowledge, with much time spent reading. Thus I have chosen the image of Harrison’s classic textbook of medicine (Kasper, 2005) to represent the discourse of competence-as-knowledge. Taken together, this set of ideas, roles and activities associated with a discourse of competence-as-knowledge construct an incompetent individual as one who does not or cannot memorize, reproduce and integrate large amounts of factual data.

After the 1960s George Miller and others argued that too much emphasis on knowledge risked creating knowledge-smart doctors who had poor interpersonal and technical skills (Miller, 1990). This phenomenon is illustrated by an email that I received from a student some years ago. The student wrote, ‘Sitting here studying, I was wondering how important your two lectures are for the exam. I don’t see any questions from your lectures on any old exams and wanted to know if your stuff was “testable” this year’. The lectures that the student referred to addressed how homeless people might be engaged better in the health care system. Along the same lines, Jacques Barzun argued in the New York Times in 1988 that a preoccupation with doing well on recall tests has ‘conditioned the way young people in America think’ and that they have ‘better-developed cognitive abilities to recognize random facts than to construct patterns or think systematically’ (Barzun, 1988, p. A31).

An overemphasis on competence-as-knowledge may lead to ‘hidden incompetence’ such as poor integration of knowledge with performance, a lack of appropriate interpersonal behaviours and poor technical abilities. An example is a student I observed who asked a patient, ‘Madam, do you have higher conjugated or unconjugated bilirubin?’

(2) Miller’s Pyramid and competence-as-performance

As a reaction to a perceived over-emphasis on knowledge, a new discourse of competence-as-performance emerged in the 1960s. The essence of this discourse is summarized by a quote from Ronald Harden, a professor at Dundee University who is credited with inventing the performance-based Objective Structured Clinical Examination (Harden & Gleeson, 1979). He said, ‘In many places they would ask students to write an essay on the origin of the word shoelace, or give them a multiple choice question on the design of shoelaces or even ask them to describe the steps in tying a shoelace. Whereas really the only way of doing it is showing you know how to tie a shoelace’ (Harden, 2005).

The discourse of ‘competence-as-performance’ is associated with a very different set of words including: ‘simulated patient, programmed patient, patient instructors, feedback, performance, skills, OSCE, multiple observations and stations’. In this discourse, the role of the teacher shifts to the demonstration and observation of skills. Competence is measured with performance-based assessments that require students to demonstrate their skills. This discourse has been associated with the development of a wide variety of new testing methods including simulated patients, objective structured clinical examinations (OSCEs) and their many variants, and the incorporation into traditional oral examinations of a period of observing students taking a history and performing a physical examination before making a case presentation. This combination of teaching and assessment methods creates a role for students that consists to a large degree of performing for observers, with much time spent practising performances (Hodges, 2003a). I have chosen the widely cited graphic of Miller’s Pyramid (Miller, 1990) to represent the discourse of competence-as-performance. Miller’s Pyramid is a conceptual taxonomy that places knowledge at the bottom of a pyramid, and a sequence of performance verbs ‘knows how, shows and does’ at increasingly higher levels on the pyramid. This set of ideas, roles...
Table 1. Four discourses of competence and the associated roles and assessments.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Conceptualization of competence</th>
<th>Role of teacher</th>
<th>Role of student</th>
<th>Measure of competence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harrison's Textbook</td>
<td>Knowledge</td>
<td>Provision of facts, and knowledge, explanation of first principles</td>
<td>Memorizing facts for recall</td>
<td>Knowledge tests (multiple choice questions)</td>
</tr>
<tr>
<td>Miller's Pyramid</td>
<td>Performance</td>
<td>Teaching of skills, creation of simulations, observation</td>
<td>Practising and demonstrating skills</td>
<td>Performance test: (Observed real or simulated scenarios)</td>
</tr>
<tr>
<td>Cronbach's Alpha</td>
<td>Reliable test score</td>
<td>Preparation of students for standardized assessments</td>
<td>Maximising data points on standardized measures</td>
<td>Standardized tests</td>
</tr>
<tr>
<td>Schön's Reflective Practitioner</td>
<td>Reflection</td>
<td>Guiding introspection, mentoring, acting as 'confessor'</td>
<td>Demonstrating self-assessment and self-direction</td>
<td>Portfolios, reflective exercises</td>
</tr>
</tbody>
</table>
and activities constructs incompetence as an individual who is unable to demonstrate communication, interpersonal, physical examination, or other skills for observers in structured, often simulated environments.

Although still very prominent today, beginning in the 1990s cognitive psychologists and sociologists began to have worries about too much emphasis on performance. Norman for example argued that ‘cracks started to appear in the pyramid’ as research emerged illustrating that ‘knowledge wasn’t quite so low down and skills quite so high up as one might have thought’ (Norman, 2005, p. 2). Schuwirth and van der Vleuten (2006) and Eva (2003) argued that ‘relevant knowledge is essential for real-life problem solving…knowledge is highly domain-specific, so is problem solving’ (Schuwirth & van der Vleuten, 2006). The strong inter-dependence of skills and knowledge – so-called ‘content specificity’ of knowledge – means that teaching of content-free, generalizable performance skills is probably an unrealistic undertaking.

Approaching the issue from a different angle, sociologists have begun to argue that exclusive reliance on a pedagogical approach of simulation training might encourage students to become ‘simulation doctors’ who act out a good relationship with their patients but have no authentic connection with them (Hanna & Fins, 2006).

Thus an over-emphasis on competence-as-performance may lead to ‘hidden incompetence’ such as poorly integrated knowledge or fake performances. This phenomena became clear to me during a teaching session when I observed a student, who had trained for many weeks in a communication programme say to a real patient on the ward, ‘Oh that student, who had trained for many weeks in a communication programme but was very quiet and did not engage with her patient’. I remember saying how difficult this is? This interview is difficult!’

(3) Cronbach’s Alpha and competence-as-reliable test score

During the 1980s a series of shifts, including a drive for models of ‘accountability’, the growth of large testing institutions and an influx of medical educators with training in measurement and evaluation led to the rise of a new discourse that focused on the psychometric reliability of tests. The essence of this discourse is illustrated by a quote from Howard Barrows, a neurologist who is credited with inventing ‘simulated patients’ as well as propelling standardized performance-based testing (Wallace, 1997). He said ‘The significance of the standardized-patient technique in assessment is that it can produce a valid clinical test item to assess performance that has many of the same advantages of the multiple-choice question. It is a standardized item, can be given in multiples, and can be scored in reliable and valid ways’ (Barrows, 1993).

The discourse of competence-as-reliable test score is characterized by words such as: ‘reliability, validity, generalizability, data, psychometrician, candidate, checklist, item-banking, cut-point, standardization’. In this discourse, there is a major emphasis on the ‘rigour’ of testing. Further the cost and labour implications, accountability imperative and legal ramifications mean that much more attention is given to ‘high stakes’ tests at the end of training. This is not evenly experienced in all western countries, and there has been a much greater role for summative, high stakes examinations in the United States and Canada than in European countries (Seguin & Hodges, 2005) Nevertheless, the psychometric discourse of competence-as-reliable test score has risen to significant prominence at educational conferences, in publications and in common parlance across the world.

The powerful driving effect of psychometrically reliable, summative examinations means the role of teachers often falls in line with examination preparation. In some places, private organizations have sprung up to offer training that paying students hope will improve their chances for high scores. Measures of competence are instruments that can be statistically analysed to show reliability and psychometrically defined validity. In particular, binary or multi-point checklists and ratings are applied to a wide range of tasks such as performance assessments (OSCEs, oral examinations), intraining assessments and also to the assessment of values, attitudes and ‘professionalism’. A drive toward psychometric reliability has also meant removing test items that do not contribute to the overall reliability of a test. This means removing multiple choice questions that ‘do not discriminate between test takers’ or reducing the ‘variance’ of performance-based assessments by standardizing the examiners or the performances and demographic characteristics of patients. The most common statistical formula for test reliability is called Cronbach’s alpha, thus I have used this as a metaphor to represent the discourse of competence-as-reliable test score.

For students, the discourse of competence-as-reliable test score drives them to activities that they perceive will maximize their scores. Interestingly, concerns about ‘security’ of high stakes testing has meant that where this discourse is dominant, feedback is often very limited or impossible. Taken together this set of ideas, roles and activities associated with a discourse of competence-as-reliable test score constructs an incompetent individual as one who cannot score highly on standardized measurement instruments.

While this discourse is currently very prominent, critiques are emerging. Schuwirth and van der Vleuten have recently written, ‘We dismiss variance between observers as error because we start from the assumption that the universe is homogeneous, where in fact the more logical conclusion would have been that the universe is more variant’ (Schuwirth & van der Vleuten, 2006). One of the side-effects of stringent adherence to standardized checklists is an apparent discriminatory effect on individuals at higher levels of expertise, who use pattern recognition and synthesis and who simultaneously gather information and manage problems. This results in low scores on inclusive, detail-oriented checklists. A study conducted in Toronto showed that clinicians with the most experience actually scored lower on OSCE checklists than did residents or medical students (Hodges et al., 1999).

Further, the effect of learning to interview patients in a way that will maximize ‘hits’ on checklists may be inappropriate. Norman recently wrote, ‘I have heard enough anecdotes about the shotgun behaviour induced by checklists to shift the burden of proof onto the advocates of this strategy’ (Norman, 2005). Schuwirth and van der Vleuten similarly question the assumptions that underlie strict
standardization and high stakes testing: ‘Assessment should be fair, honest and defensible… but the strict operationalisation of these values is – in our humble opinion – currently of limited value’ (Schuwirth & van der Vleuten, 2006).

Thus an over-emphasis on competence-as-reliable test score may result in ‘hidden incompetence’ including ‘shotgun interviews’ and discourages the use of pattern recognition, integration and synthesis. This phenomenon was made clear to me when a student preparing for an OSCE said, ‘You keep saying to take time, to be nice, to listen to the patient and to make a synthesis of the problem, but if we don’t ask as many questions as possible we will not pass this examination’.

(4) Schön’s Reflective Practitioner and the discourse of competence-as-reflection

Since the mid 1990s, the work of Donald Schön (1987) has become a popular antidote to a perceived over-emphasis on standardized testing, and has shifted the focus to internal reflection and self-direction. The essence of this discourse is captured in this quote from the Ontario Ministry of Education: ‘The concept of learner as a mere processor of information has been replaced by the image of a self-motivated, self-directed problem solver’ (Ontario Ministry of Education, 1980).

The discourse of ‘competence-as-reflection’ is characterized by words such as: ‘reflection, self-directed learning, insight, learning contracts, portfolios and adult learner’. In this discourse, the teacher takes on a role of guide or mentor, or what might even be called ‘confessor’. Common measures of assessment are portfolios and the use of reflective exercises such as diaries, reflective essays and learning contracts. At a national level, some countries such as Canada now require the submission of a self-directed learning portfolio as evidence of maintenance of competence (RCPSC, 2006). Taken together this set of ideas, roles and measures constructs an incompetent individual as one who cannot engage in what I have called elsewhere the ‘trinity’ of this discourse: self-reflection, self-assessment, self-regulation (Hodges, 2004).

As with other discourses, there are emerging critiques of competence-as-reflection. One thrust can be summarized by the words of Jeremy Taylor, the seventeenth-century cleric who is said to have commented, ‘It is impossible to make people understand their ignorance, for it requires knowledge to perceive it; and therefore, he that can perceive it, hath it not’. In a recent review Davie confirmed this worry with his analysis of 17 studies comparing doctors’ self-assessments against objective, external reviews. He reported that in most studies there was a subset of clinicians ‘who appear, either by training or personality, unable to judge themselves’ (Davie et al., 2006). This work is consistent with the studies of Kruger and Dunning who showed that in a whole host of areas of competence (problem solving, logical reasoning, humour) a significant proportion of individuals demonstrated a wide gap between their own assessments and those of others. This gap persists even when they are given the opportunity to observe other performances and to reassess their own. Our group replicated these findings in a study in which family medicine residents were asked to manage a case of child abuse. Among those who scored very low on management, knowledge and interpersonal skills there were several who rated themselves substantially higher than the observers. When given an opportunity to watch a variety of others handle the same situation and then rescore their own performances, the inflation of self-ratings persisted (Hodges et al., 2001).

Approaching the discourse from a more sociological perspective, there have also been critiques of reflection itself. Nelson argues that in nursing ‘reflective practice provides the mechanism whereby nurses internalize the new professional ethos of self-government…’ and that ‘regulators appear quite unconcerned about the lack of coherence between what is being monitored at a distance and the actual professional knowledge (needed) to function skillfully and competently’ (Nelson & Purkis, 2004).

Thus an overemphasis on competence-as-reflection may result in ‘hidden incompetence’ such as those who can produce portfolios but nevertheless cannot identify significant deficits; those who despite identifying deficits cannot or will not direct own learning; and those for whom the goal becomes reflection itself at the expense of attaining actual knowledge and skills. These phenomena bedevil other reflective processes such as performance reviews and interviews and are illustrated by the resident whom I asked to identify an area in which he needed to do some work. He said, ‘Sometimes I am too dedicated’.

Table 2 summarizes the officially recognized incompetence associated with each discourse and the hidden forms of incompetence that each may conceal.

Conclusions

What teachers and evaluators choose to emphasize in medical education drives behaviour of students and colleagues to such an extent that it can actually create forms of incompetence. Like medical treatments, it is helpful to pay attention to the side-effects of medical education generally, and of discourses specifically. Because medical students are bright and highly motivated, it is likely that only a minority of them will experience these side-effects. Most will be well equipped to function as independent and safe professionals after graduation despite the vagaries of curriculum and assessment. Nevertheless there are some recommendations that can be drawn from the foregoing analysis that may mediate some of the most concerning of these problematic effects. As a general observation, it is wise to avoid unquestioning adherence to any one discourse. All have strengths and weaknesses worth considering carefully.

Conclusions that can be drawn from the literature reviewed in this paper include:

(1) Avoid teaching and testing ‘pure’ knowledge. Knowledge and skills should be integrated early and often, and both should be anchored to clinical, social, cultural and other contexts.

(2) Avoid teaching and testing ‘general skills’. Skills are bound to domain-specific knowledge, the learning of which should be integrated with skills development.

(3) Limit use of highly standardized scenarios and measures. Foster expert forms of thinking and embrace variance in the presentation of situations and cases.
### Table 2. Incompetence constructed by and hidden by discourses of competence.

<table>
<thead>
<tr>
<th>Discourse</th>
<th>Conceptualization of incompetence</th>
<th>Hidden incompetence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harrison's Textbook: competence-as-knowledge</td>
<td>Inability to recall</td>
<td>Stronger ability to recognize random facts rather than to interpret patterns or synthesize contextual information. Poor interpersonal or technical skills.</td>
</tr>
<tr>
<td>Miller’s Pyramid: competence-as-performance</td>
<td>Inability to perform</td>
<td>Lack of integration of relevant knowledge with performance. Inauthentic skills such as superficial displays of pseudo-empathy or cursory examination skills.</td>
</tr>
<tr>
<td>Cronbach’s Alpha: competence-as-reliable test score</td>
<td>Inability to score highly on checklists/standardized measures</td>
<td>Prolongation of novice behaviours (shotgun data collection) and lack of development of expert reasoning and behaviours. Inability to address the vagaries and variance of real contexts.</td>
</tr>
<tr>
<td>Schön’s Reflective Practitioner: competence-as-reflection</td>
<td>Inability to produce convincing self-assessment</td>
<td>Superficial self-assessment that masks inability to detect or address deficiencies. Development of reflective abilities alone in the absence of knowledge and skills.</td>
</tr>
</tbody>
</table>
(4) Implement reflection carefully. Don’t use self-directed learning methods without establishing the capacity for self-assessment. Further, don’t let competence assessment rest on reflection alone, it should remain tied to the development and demonstration of the acquisition of knowledge and skills.

Finally, I should like to say one more thing about discourses. They are not simply autochthonous entities that emerge like shooting stars from a dark night. Discourses emerge because there are important sociological, political, economic and cultural contingencies that make them possible. Discourses are associated with power. The dominance of one discourse over another has significant implications for what is considered legitimate, what positions are made available for individuals, what will get published, what will be funded and which institutions will gain power and influence. These aspects warrant a detailed study that I have only hinted at in this paper. Nevertheless, we can all begin to imagine what such an analysis might yield by asking ourselves the question, ‘What does my discourse make possible?’

References


HODGES, B. (2005a) The Birth and World-wide Adoption of the OSCE: An Advanced Socio-historical Study (Amsterdam, Association for Medical Education in Europe).


