"Far Horizons: Extending the Landscape of Assessment"

Brian Hodges MD, MEd, FRCPC
Director, Wilson Centre for Research in Education, University of Toronto
In appreciation

- Michigan State University's Office of Medical Education Research and Development (OMERAD)
- Dr. Jack Maatsch
  - Professor OMERAD from 1971 to 1990
  - OMERAD Director from 1980 to 1989
- “With interest in both theory and application, Jack Maatsch enjoyed and stimulated spirited debate around important issues in the education and assessment of physician competence”
This presentation is based on a paper that will appear in Medical Teacher in 2007 entitled:

Medical Education and the Maintenance of Incompetence

You may access copies of the slides at:

www.thewilsoncentre.ca
We think of medical education as a process that moves novices from a state of incompetence, to one of competence.

This talk explores the idea that education, and in particular assessment processes may actually lead to incompetence.
What kind of incompetence?

I am not talking about the rare cases of gross incompetence – sexual abuse, drug addiction, serial killers... but the more grinding and mundane incompetence that harms the quality of patient care and the reputation of the profession itself.
Incompetence as a side-effect

This kind of incompetence is a “side-effect” of medical education.

The particular side-effect that occurs is a result of overemphasizing particular models of education and assessment.

We all adhere to these models because we believe in them – sociologists call them “discourses”.
What are “discourses”?

- Discourses are ways of seeing the world
- They act like lenses or filters
- They make it possible for us to say some things but not others
- They make it possible to act in certain ways, and to have certain jobs
For example, there are discourses about madness:

- Madness as Spiritual Possession
- Madness as Deviancy
- Madness as Medical Illness
If you use the discourse of...

Spiritual Possession you make visible:
- Possessed individuals, and create a role for Spiritual Healers working in religious institutions

Deviancy you makes visible:
- Deviant individuals and create a role for Judges/Jailors working in prisons

Medical Illness you make visible:
- Mentally Ill individuals and create a role for Psychiatrists/Psychologists working in hospitals
But as Foucault said, “... we are not dealing with the same madmen”

- Possessed ≠ Deviant ≠ Mentally Ill
- Spiritual Healer ≠ Jailor ≠ Psychiatrist
- Church ≠ Jail ≠ Hospital

Foucault 1969, The Archaeology of Knowledge
What about incompetent doctors?

- Incompetence, like madness, can also been defined in different ways

- Let’s look first at some older variations
A competent doctors in...

- **1700**
  - Member of a guild
  - Carried blade for blood letting
  - Emetics for purging to balance humours

- **1900**
  - Gentleman with a walking stick
  - Diagnosed by looking at the tongue, and
  - Smelling urine

- **1950**
  - Man in a white coat
  - Talked to husbands about their wife’s illness
  - Withheld diagnoses from dying patients so they wouldn’t worry

- **2007**

??
We are not dealing with the same competent behaviours!

- Blood letting, smelling urine, withholding diagnoses – are clearly incompetent today

- How did these changes occur?

- They occurred because our discourses changed

- What then, are the discourses of competence/incompetence we use today?
I have been studying our discourses

Over 600 medical education articles:
  - Coded for key words, metaphors, shifts in paradigms

25 interviews with key figures in education:

Medical education institutions around the world:
  - US, UK, Canada, France, Israel, China, Jordan, Ethiopia, Pakistan, Poland, Japan
We use at least 4 discourses of competence/incompetence

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

4. Donald Schon and competence-as-reflection
1. Harrison’s Textbook and competence-as-knowledge
Competence-as-knowledge

Key words:

- Facts, foundational knowledge, basic science, first principles, fund of knowledge, classic text books, classic articles, multiple-choice tests

- Teacher role: Lecturer, source of wisdom
- The measure: Knowledge test (MCQ)
- Student role: Memorize, reproduce
Implications

- Teaching consists of didactic lectures
- Studying involves reading
- Testing involves recall

The official incompetent

- An individual who does not or cannot memorize or reproduce large amounts of factual data
After 1960 George Miller and others said that too much emphasis on knowledge created smart doctors who had poor interpersonal skills.
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.

University of Toronto Medical Student 2000
The preoccupation with doing well on standardized tests has literally conditioned the way young people in America think.

They have better-developed cognitive abilities to recognize random facts than to construct patterns or think systematically.

The Hidden Incompetent

- Poor interpersonal behaviours
- Poor technical abilities

*Student:* Madam – do you have higher conjugated or unconjugated bilirubin?
2. Miller’s Pyramid and competence-as-performance

van der Vleuten, C. BMJ 2000;321:1217-1219
In 1960s the idea of competence-as-performance emerged. 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.

Ronald Harden 2005
Competence-as-performance

Key words:
- simulated patient, programmed patient, patient instructors, feedback, performance, skills, OSCE, multiple observations, stations

- Teacher role: Teach skills
- The measure: Performance-based test
- Students role: Perform for observers
Implications

- Teaching involves clinical performances, real and simulated cases
- Performance tests replaced written exams

The official incompetent:
- An individual unable to demonstrate communications, interpersonal, physical examination, or other skills
In the 1990s, cognitive psychologists and sociologists alike began to have worries about too much emphasis on performance.
Side-effects

Cracks started to appear in the pyramid, it seems that knowledge wasn’t quite so low down and skills quite so high up as one might have thought.

Relevant knowledge is essential for real-life problems solving...knowledge is highly domain-specific, so is problem solving.

Geoff Norman 2005

Schurwirth and van der Vleuten 2006
Eva 2005
Exclusive reliance on a pedagogical approach of simulation training may be encouraging students to become “simulation doctors” who act out a good relationship with their patients but have no authentic connection with them.

Hanna and Fins 2006
The Hidden Incompetent

- Poorly integrates knowledge
- Fakes performances

**Student:** Oh that must be hard for you... wow that must be hard for you... oh, yes that must be really hard for you

**Patient:** Can you stop saying that?
3. Cronbach’s Alpha and competence-as-reliable test score

$$\alpha' = \frac{N \cdot \bar{r}}{1 + (N - 1) \cdot \bar{r}}$$
Competence-as-reliable test score

After 1980 psychometric reliability of tests became very important

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.

Howard Barrows 1993
Competence-as-reliable test score

Key words
- reliability, validity, generalizability, data, psychometrician, candidate, checklist, item-banking, cut-point, standardization

Teacher role: Exam preparation
The Measure: Standardized checklists
Student role: Maximizing data-points
Implications

Asks about:

- Onset of pain  ✓
- Site  ✓
- Nature  ✓
- Duration  ✓
- Exacerbation  ✓
- Relieving  ✓
- Nausea  ✓
- Vomiting  ✓
- Shortness of breath  ✓
- Diarrhea  ✓
- Blood  ✓
- Stool colour  ✓

- Teaching shifted to examination preparation and standardized scenarios
- Goal of testing was to reduce all sources of variance and maximize reliability of scores
- Feedback often vanished for reasons of examination security

The official incompetent
- The individual who could not score highly on checklists in standardized simulations
We dismiss variance between observers as error because we start from the assumption that the universe is homogenous, where in fact the more logical conclusion would have been that the universe is more variant.

Assessment should be fair, honest and defensible... but the strict operationalisation of these values is – in our humble opinion – currently of limited value.

Schurwirth and van der Vleuten 2006
Side-effects: OSCE checklists do not capture increasing levels of expertise

Global ratings can capture expertise

Figure 1: Global Ratings
Students adapt their behaviour to the system of evaluation

- 57 clinical clerks randomly assigned to 2 groups in a 10 station OSCE
- Group 1 told that scores were based on checklists
- Group 2 told that performance would be rated using global ratings assessing overall competence
- All candidates scored by blinded MD raters using both checklists and global ratings
- Significant interaction: rating form by orientation (F1,55=5.5, p < 0.05)
  - Checklist oriented group had higher checklist scores
  - Process oriented group had higher global scores

Herold-McIlroy et al 2002
Side-effects

I have heard enough anecdotes about the shotgun behaviour induced by checklists to shift the burden of proof onto the advocates of this strategy.

Geoff Norman 2005

\[ \alpha' = \frac{N \cdot \bar{r}}{1 + (N-1) \cdot \bar{r}} \]
The Hidden Incompetent

- Shot gun interviews
- Lacks use of pattern recognition, integration, synthesis

**Student:** You keep saying to take time to be nice, listen to the patient and make a synthesis of the problem, but if we don’t ask as many questions as possible we will not pass the examination
4. Schon and the discourse of reflection
Competence-as-reflection

Since the mid 1990s, the work of Donald Schon has become an antidote to standardized testing, emphasizing the idea that competence requires internal reflection and self-direction.

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
Competence-as-reflection

Key words:
- reflection, self-directed learning, learning contracts, portfolios, adult learner

- Teacher role: Guide/Mentor/Confessor
- The measure: Portfolio
- Student role: Reflector
Implications

- At the school level: use of dossiers, portfolios, reflective/confessional activities
- At national levels: required self-assessments and submission of learning portfolios

The official incompetent
- Individual who cannot produce a convincing analysis of his/her strengths and weaknesses
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”

Jeremy Taylor
Side-effects

- Analysis of 17 studies comparing doctors’ self-assessments against objective, external review
- There is a subset of clinicians who appear, either by training or personality, unable to judge themselves

Davis JAMA 2006
Side-effects: Unskilled and Unaware

Handling a case of child abuse

Hodges, Martin, Regehr Academic Medicine 2001
Reflective practice provides the mechanism whereby nurses internalize the new professional ethos of self-government...

Meanwhile, 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 skilfully and competently

Nelson and Purkis Nursing Inquiry 2004
The Hidden Incompetent

- Individual who cannot identify deficits
- Individual who cannot direct own learning
- Individual who appears to reflect but doesn’t have adequate knowledge or skills

Teacher: Can you reflect on your weaknesses?

Student: Sometimes I am too committed
Summary

- What we choose to emphasize and to assess in medical education drives behaviour to such an extent that it can actually create forms of incompetence.
- Like medical treatments, we must pay more attention to the side-effects of medical education and assessment methods.
- Probably only a minority of our students will have these side-effects. Many will be quite competent professionals after graduation, but for the rest...
Balance the positive and negative effects of educational discourses

Don’t teach / test pure knowledge
- Integrate knowledge with skills early and often

Don’t teach / test ‘general skills’
- Integrate skills with their contextual knowledge

Limit use of standardized scenarios / measures
- Foster expert forms of thinking and embrace variance

Implement reflection carefully
- Don’t use self-directed learning without establishing the capacity for self-assessment
- Don’t let competence assessment rest on reflection alone
Chose measurement instruments carefully

Don’t use psychometrically rigid measures only
- Performance-based assessments should include global ratings, particularly at higher levels of expertise

“Global ratings” are not all the same
- Use a rubric like Hunters’ classification (5 levels from “atomistic” to “holistic”) to choose the kind of global rating you want

Triangulate multiple perspectives
- Gather ratings from SPs and MD, but also other professionals and peers
- Consider the type of rating most appropriate for each

Don’t let test “security” trump feedback
Push the limits of education research

Don’t be confined by traditional psychometric concepts of validity

- Explore the “ecological” validity of assessments through studies of impact on student behaviour and thinking
- Consider the negative, “deforming” side-effects as well as the positive outcomes
- Use qualitative methods as well as traditional validity statistics to study validity
- Reflect on how the discourses and models we chose are linked to important historical, political and economic factors
Most importantly each of us should ask: What does my discourse make possible?

\[ \alpha' = \frac{N \cdot \bar{r}}{1 + (N - 1) \cdot \bar{r}} \]

Thank you!

www.thewilsoncentre.ca
Cathell DW. (1890) Book on the physician himself. FA Davis, Philadelphia and London.


Eva KW, Cunningham JP, Reiter HI, Keane DR and Norman GR. (2004) How can I know what I don’t know? Advances in Health Sciences Education. 9, 211-24.


Hodges B. (2003a) OSCE! Variations on a theme by Harden. Medical Education. 37 (12):1134-1140.


