

"Far Horizons: Extending the Landscape of Assessment"

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

the
Wilson
centre



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:

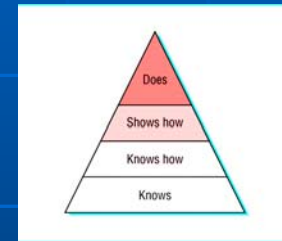
- US, UK, Canada, France, Israel, China, Jordan, Ethiopia, Pakistan, Poland, Japan

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

$$\alpha = \frac{N \cdot \bar{r}}{1 + (N - 1) \cdot \bar{r}}$$

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

Resistance



- After 1960 George Miller and other said that too much emphasis on knowledge created knowledge-smart doctors who had poor inter-personal skills

Side-effects



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

Side-effects



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.

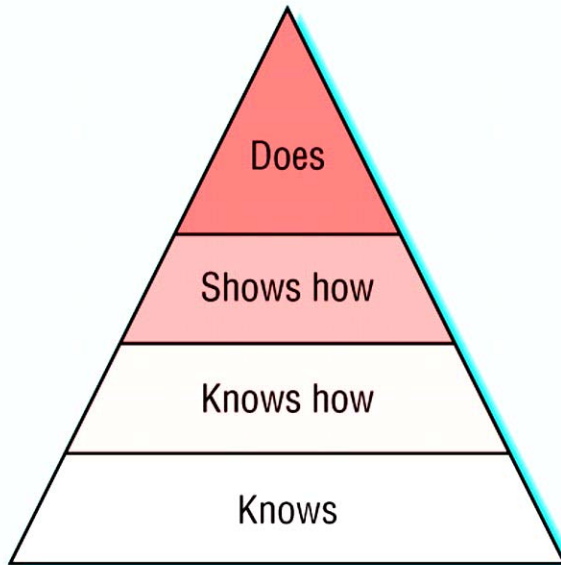
Jacques Barzun, New York Times, 1988

The Hidden Incompetent

- Poor interpersonal behaviours
- Poor technical abilities

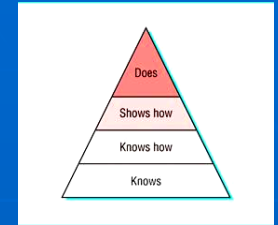
Student: *Madam – do you have higher conjugated or unconjugated bilirubin?*

2. Miller's Pyramid and competence-as-performance



van der Vleuten, C. *BMJ* 2000;321:1217-1219

Competence-as-performance



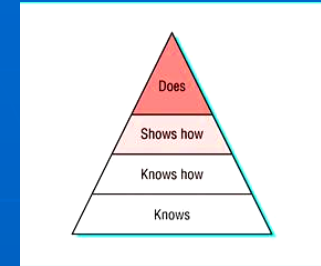
In 1960s the idea of competence-as-performance emerged

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

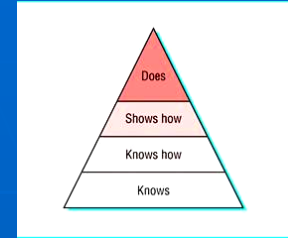


- Teaching involves clinical performances, real and simulated cases
- Performance tests replaced written exams

The official incompetent:

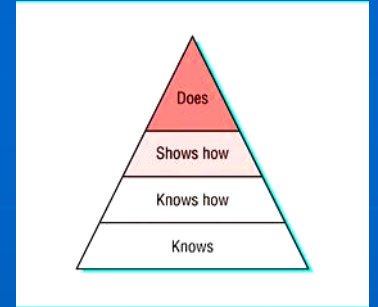
- An individual unable to demonstrate communications, interpersonal, physical examination, or other skills

Resistance



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

Geoff Norman 2005

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

Schurwirth and van der Vleuten 2006

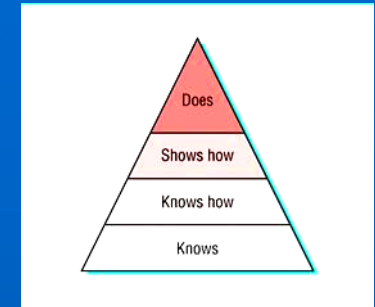
Eva 2005

Side-effects

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

$$\alpha = \frac{N \cdot \bar{r}}{1 + (N-1) \cdot \bar{r}}$$

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

$$\alpha = \frac{N \cdot \bar{r}}{1 + (N-1) \cdot \bar{r}}$$

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

Resistance

$$\alpha = \frac{N \cdot \bar{r}}{1 + (N-1) \cdot \bar{r}}$$

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.

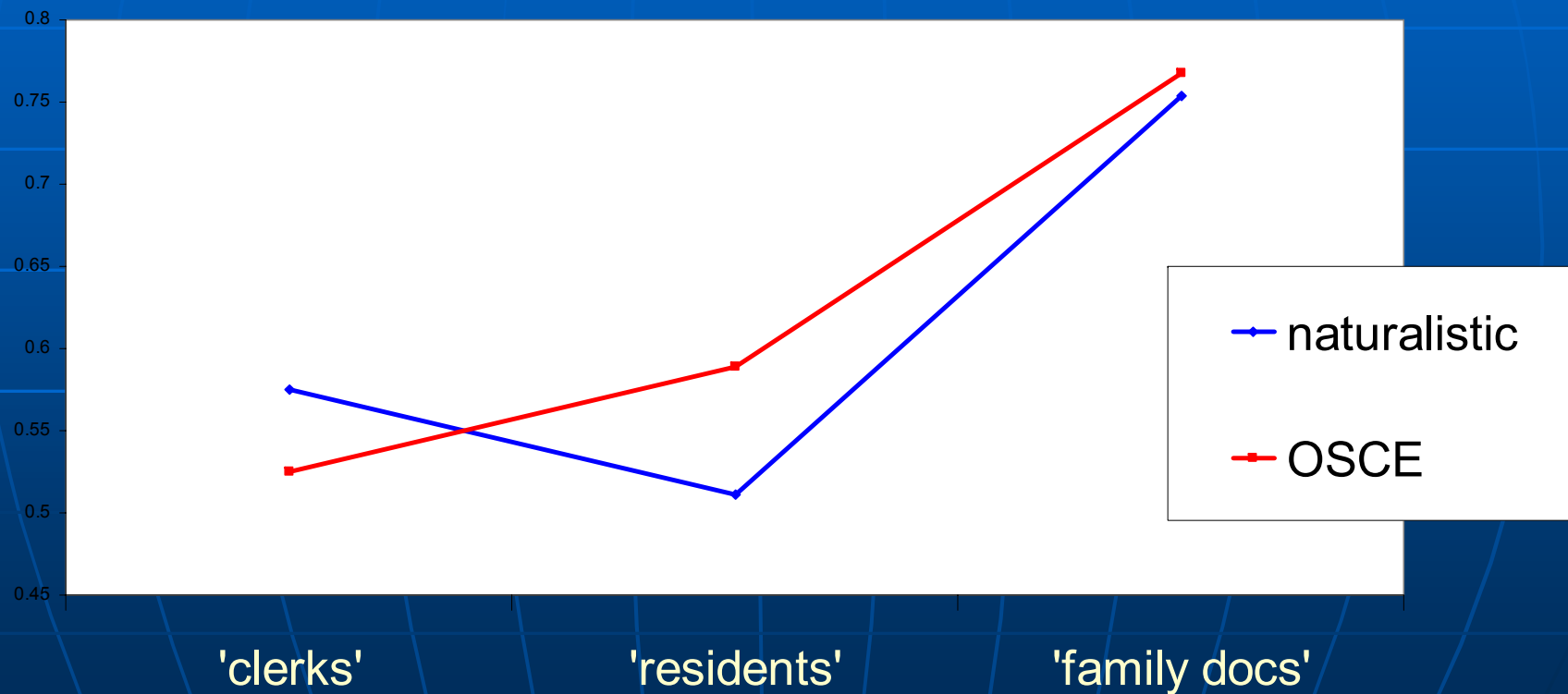
Side-effects: OSCE checklists do not capture increasing levels of expertise



Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M. (1999)
Academic Medicine

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 ($F_{1,55}=5.5, p<0.05$)
 - checklist oriented group had higher checklist scores
 - Process oriented group had higher global scores

Herold-McIlroy et al 2002

Side-effects

$$\alpha = \frac{N \cdot \bar{r}}{1 + (N-1) \cdot \bar{r}}$$

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

The Hidden Incompetent

$$\alpha = \frac{N \cdot \bar{r}}{1 + (N-1) \cdot \bar{r}}$$

- 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

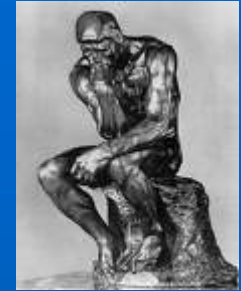


Since the mid 1990s, the work of Donald Schon has become an antidote to standardized testing, emphasizing the idea that competence requires internal reflection and self-direction

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

Resistance



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

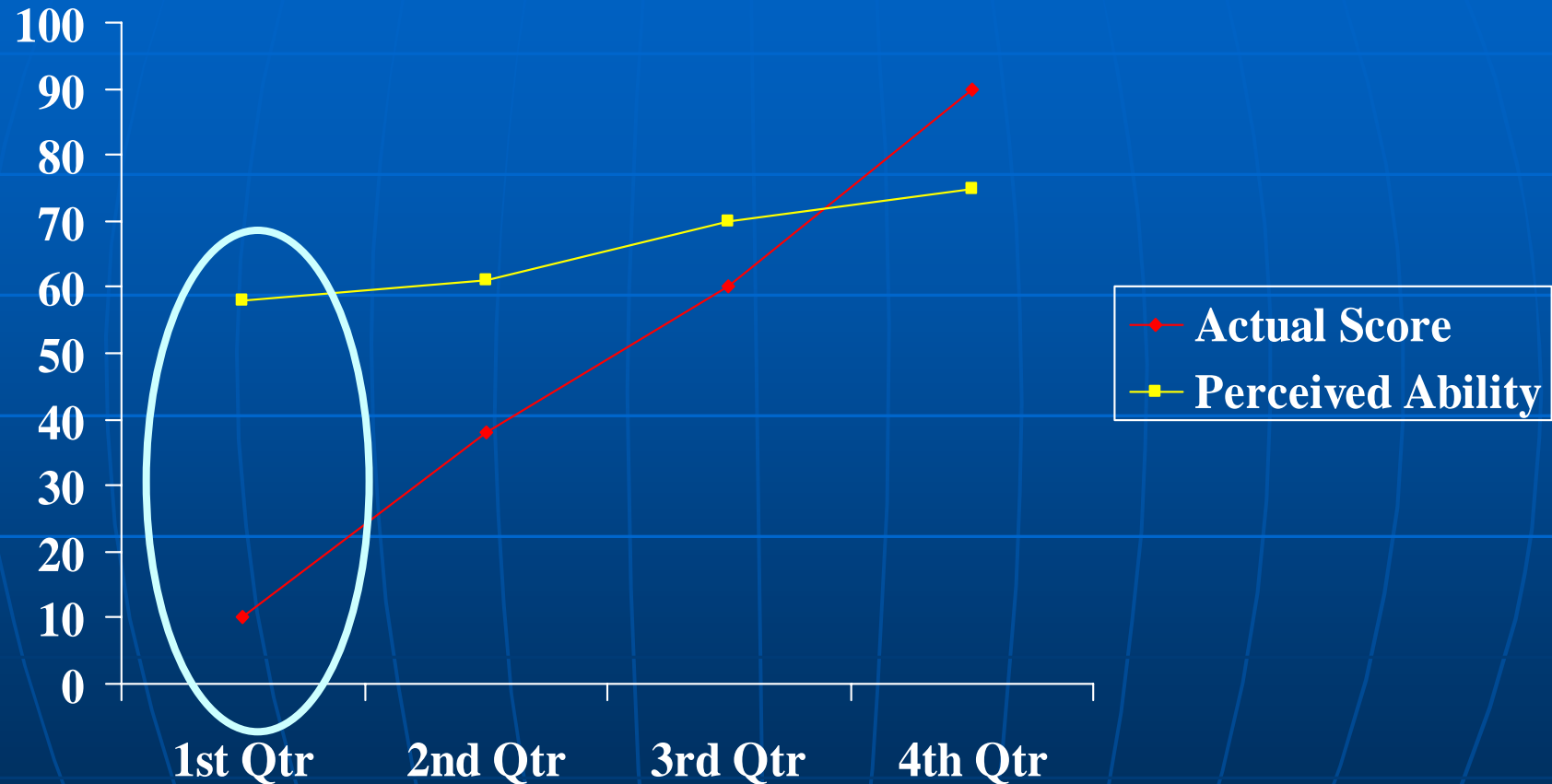
Side-effects



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

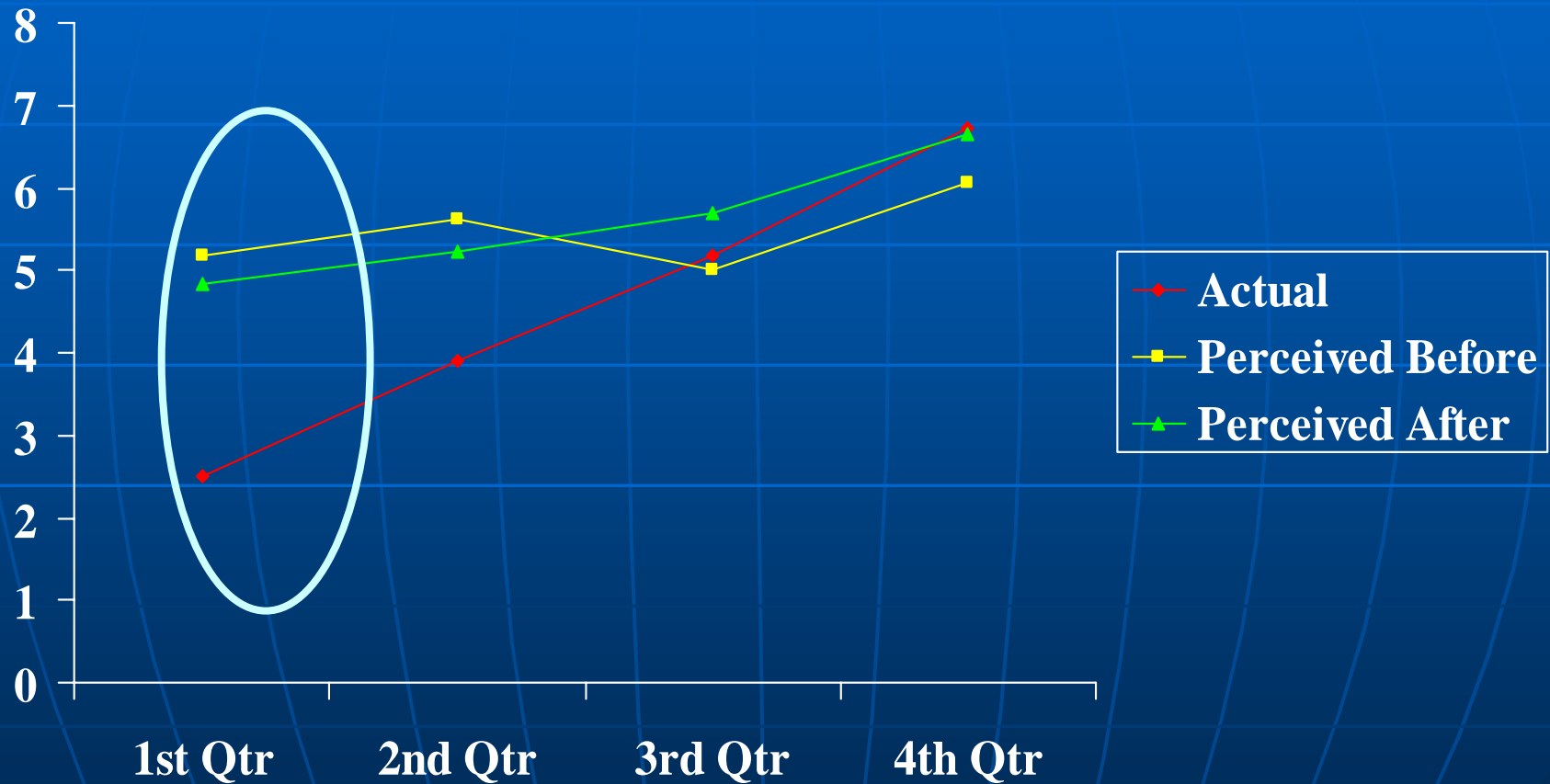
Davis JAMA 2006

Side-effects: Unskilled and Unaware



Kruger and Dunning. *Journal of Personality and Social Psychology*, 1999

Handling a case of child abuse



Hodges, Martin, Regehr *Academic Medicine* 2001

Side-effects

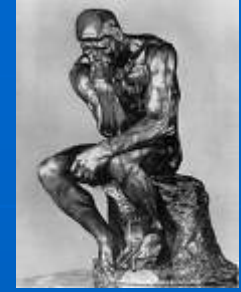


Reflective practice provides the mechanism whereby nurses internalize the new professional ethos of self-government...

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, particular 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 chose 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

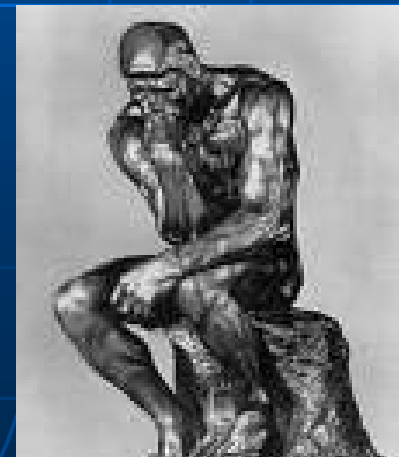
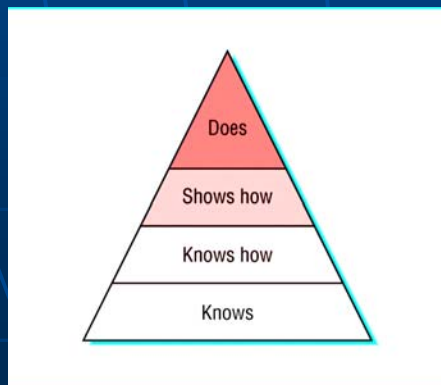
- Explore the "ecological" validity of assessments through studies of impact on student behaviour and thinking
- Consider the negative, "deforming" side-effects as well as the positive outcomes
- Use qualitative methods as well as traditional validity statistics to study validity
- Reflect on how the discourses and models we chose are linked to important historical, political and economic factors

Most importantly each of use 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



Bibliography

Barzun J. (1988) Multiple Choice flunks out. New York Times, October 11.

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

Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. (2006) Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. JAMA. 296(9):1094-102.

Eva KW. (2003) On the generality of specificity. Medical Education. 37, 587-588.

Eva KW, Cunnington 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.

Foucault M. (1969) The Archaeology of Knowledge and the Discourse on Language. Pantheon Books: New York.

Foucault M. (1965) Madness and Civilization; A History of Insanity in the Age of Reason. Random House: New York, USA.

Hanna M and Fins JJ. (2006) Why simulation training ought to be complemented by experiential and humanist learning. Academic Medicine. 81(3):265-270.

Harden RM and Gleason FA. (1979) Assessment of clinical competence using an objective structured clinical examination. *Medical Education*. 13:41-47.

Harden R. (2005) Interview. Recorded at Dundee Scotland. January 2005. Unpublished.

Herold-McIlroy J, Hodges B, McNaughton N, Regehr G. (2002) The effect of candidates' perception of evaluation methods on reliability of checklist and global rating scores in an objective structured clinical examination. *Academic Medicine* 77: 725-728.

Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M. (1999) OSCE checklists do not capture increasing levels of expertise. *Academic Medicine*, 74(10):64-69

Hodges B, Regehr G, Martin D. (2001) Difficulties in recognizing one's own incompetence: Doctors who are unskilled and unaware of it. *Academic Medicine*. 76(10 Suppl):S87-9.

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

Hodges B. (2003b) Validity and the OSCE. *Medical Teacher*. 25(3), 250-254.

Hodges B. (2004) Medical student bodies and the pedagogy of self-reflection, self-assessment and self-regulation. *Journal of Curriculum Theory*. 20(2), 41-51.

Hodges B. (2005a) The birth and world-wide adoption of the OSCE: An advanced socio-historical study. Association for Medical Education in Europe, Amsterdam.

Hodges B. (2005b) The many and conflicting histories of medical education in Canada and the United States: An introduction to the paradigm wars. *Medical Education*. 39:613-621.

Hunter DM, Jones RM and Randhawa BS. The use of holistic versus analytic scoring for large-scale assessment of writing. *The Canadian Journal of Program Evaluation*. 11 (2): 61-85, 1996.

Kasper DL (Ed). (2005) *Harrison's principles of internal medicine*. 16th ed. New York: McGraw-Hill.

Miller G. (1990) The Assessment of Clinical Skills/Competence/Performance. *Academic Medicine*. 65(9):S63-S67.

Nelson S and Purkis ME. (2004) Mandatory reflection: the Canadian reconstitution of the competence nurse. *Nursing Inquiry*. 11(4):247-257.

Norman GR, van der Vleuten CP, De Graaff E. (1991) Pitfalls in the pursuit of objectivity: issues of validity, efficiency and acceptability. *Medical Education*. 25 (2):119-26.

Norman G. (2005) Editorial--checklists vs. ratings, the illusion of objectivity, the demise of skills and the debasement of evidence. *Adv Health Sci Educ Theory Pract*. 10 (1):1-3

Ontario Ministry of Education. (1980) *Issues and Directions*. Government of Ontario, Toronto.

Royal College of Physicians and Surgeons of Canada (RCPSC). (2006) *Maintenance of Competence Program Guide*. http://rcpsc.medical.org/main_e.php. Accessed September 1, 2006.

Schon DA. (1987). Educating the reflective practitioner: Toward a new design for teaching and learning in the professions. San Francisco, CA: Jossey-Bass.

Segouin C and Hodges B. (2005) Educating Physicians in France and Canada: Are the differences based on evidence or history? Medical Education 39:1205-1212.

Shorter, E. (1985). Doctors and their patients: A social history. New York: Simon and Schuster.

Schurwirth LWT and van der Vleuten CPM. (2006) A plea for new psychometric models in educational assessment. Medical Education. 40:296-300

van de Vleuten CPM. (2000) Validity of final examinations in undergraduate medical training. British Medical Journal. 321:1`217-1219.

van der Vleuten CPM, Norman GR, De Graaff E. (1991) Pitfalls in the pursuit of objectivity: issues of reliability. Medical Education. 25(2):110-8.

Wallace P. (1997) Following the treads of innovation: The history of standardized patients in medical education. Caduceus. Vol. 13(2), 5-28.