Realizing the Promise of CBME and Assessment to Improve Learning and Clinical Outcomes
Disclosures

• Employed by the ACGME
• I receive royalties from Mosby-Elsevier for a textbook on assessment
• I am a member of the board of NBME and Medbiquitous
Outline

• Competency-based medical education
  • What is it?
  • Transformation and the current state of healthcare
    • Nostalgialitis Imperfecta
• Reform and implications of competency-based medical education
• Work-based assessment and programs of assessment
Pair and Share

What does competency-based medical education mean to you?
Early Principles: CBME

• World Health Organization (1978):
  • “The intended output of a competency-based programme is a health professional who can practise medicine at a defined level of proficiency, in accord with local conditions, to meet local needs.”

CBME: Start with System Needs

What Are The Outcomes?

Health of a Population

Experience of Care
- Safe
- Effective
- Patient centered
- Efficient
- Timely
- Equitable

Per Capita Cost

The IHI Triple Aim

Better care for individuals, better health for populations, lower per capita costs
CBME Today

An outcomes-based approach to the design, implementation, assessment and evaluation of a medical education program using an organizing framework of competencies¹

Fundamental Characteristics of CBME

• Graduate outcomes in the form of achievement of predefined desired competencies are the goal.

• Competencies are derived from the needs of patients, organized into a coherent guiding framework.

• Time is a resource for learning, not the basis of progression of competence.

• Teaching and learning experiences are sequenced to facilitate an explicitly defined progression of ability in stages.
Fundamental Characteristics of CBME

- Learning is **tailored** to the learner's individual progression in some manner.

- Numerous **direct observations** and focused **feedback** contribute to effective learner development of expertise.

- **Assessment** is planned, systematic, systemic, and integrative.
Current Realities: Health System to Training System Performance
# EXHIBIT ES-1. OVERALL RANKING

<table>
<thead>
<tr>
<th>Country</th>
<th>AUS</th>
<th>CAN</th>
<th>FRA</th>
<th>GER</th>
<th>NETH</th>
<th>NZ</th>
<th>NOR</th>
<th>SWE</th>
<th>SWIZ</th>
<th>UK</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall (2013)</td>
<td>4</td>
<td>10</td>
<td>9</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Quality Care</td>
<td>2</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>5</td>
<td>4</td>
<td>11</td>
<td>10</td>
<td>3</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Effective Care</td>
<td>4</td>
<td>7</td>
<td>9</td>
<td>6</td>
<td>5</td>
<td>2</td>
<td>11</td>
<td>10</td>
<td>8</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Safe Care</td>
<td>3</td>
<td>10</td>
<td>2</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>11</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Coordinated Care</td>
<td>4</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>11</td>
<td>3</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Patient-Centered Care</td>
<td>5</td>
<td>8</td>
<td>10</td>
<td>7</td>
<td>3</td>
<td>6</td>
<td>11</td>
<td>9</td>
<td>2</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Access</td>
<td>8</td>
<td>9</td>
<td>11</td>
<td>2</td>
<td>4</td>
<td>7</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Cost-Related Problem</td>
<td>9</td>
<td>5</td>
<td>10</td>
<td>4</td>
<td>8</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Timeliness of Care</td>
<td>6</td>
<td>11</td>
<td>10</td>
<td>4</td>
<td>2</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>1</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Efficiency</td>
<td>4</td>
<td>10</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Equity</td>
<td>5</td>
<td>9</td>
<td>7</td>
<td>4</td>
<td>8</td>
<td>10</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>Healthy Lives</td>
<td>4</td>
<td>8</td>
<td>1</td>
<td>7</td>
<td>5</td>
<td>9</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>Health Expenditures/Capita, 2011**</td>
<td>$3,800</td>
<td>$4,522</td>
<td>$4,118</td>
<td>$4,495</td>
<td>$5,099</td>
<td>$3,182</td>
<td>$5,669</td>
<td>$3,925</td>
<td>$5,643</td>
<td>$3,405</td>
<td>$8,508</td>
</tr>
</tbody>
</table>

Notes: * Includes ties. ** Expenditures shown in $US PPP (purchasing power parity); Australian $ data are from 2010.
## AHRQ Quality Report

<table>
<thead>
<tr>
<th>Measure Focus</th>
<th>Measure Name/Description</th>
<th>Baseline Rate</th>
<th>Most Recent Rate</th>
<th>Aspirational Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspirin Use</td>
<td>Outpatient visits at which adults with cardiovascular disease are prescribed/maintained on aspirin</td>
<td>47%&lt;sup&gt;13&lt;/sup&gt;</td>
<td>53%&lt;sup&gt;14&lt;/sup&gt;</td>
<td>Increase to 65% by 2017</td>
</tr>
<tr>
<td>Blood Pressure Control</td>
<td>Adults with hypertension who have adequately controlled blood pressure</td>
<td>46%&lt;sup&gt;15&lt;/sup&gt;</td>
<td>53%&lt;sup&gt;16&lt;/sup&gt;</td>
<td>Increase to 65% by 2017</td>
</tr>
<tr>
<td>Cholesterol Management</td>
<td>Adults with high cholesterol who have adequate control</td>
<td>33%&lt;sup&gt;17&lt;/sup&gt;</td>
<td>32%&lt;sup&gt;18&lt;/sup&gt;</td>
<td>Increase to 65% by 2017</td>
</tr>
<tr>
<td>Smoking Cessation</td>
<td>Outpatient visits at which current tobacco users received tobacco cessation counseling or cessation medications</td>
<td>23%&lt;sup&gt;19&lt;/sup&gt;</td>
<td>22%&lt;sup&gt;20&lt;/sup&gt;</td>
<td>Increase to 65% by 2017</td>
</tr>
</tbody>
</table>
Since physician graduates of American medical education organizations typically lead or heavily influence US health care delivery, one source of indirect, broad, outcome-based evidence [of the effectiveness of the medical education enterprise] is the overall performance of the US health care system. The width of the performance gaps on the aims of effectiveness, safety and efficiency understandably reduces society’s confidence that physicians are adequately honoring their Hippocratic promises.

Milstein A. Trailing Winds and Personal Risk Tolerance: An External Perspective on the Opportunity for Medical Educators to Fulfill Their Social Contract Permanently. Presented at ABIMF Summer Forum, August 2010
Nostalgialitis Imperfecta

- Syndrome characterized by the following signs and symptoms:
  - “When I was an intern…<insert superlative>”
  - “Medicine was so much better 25 years ago”
    - Reality: Not really…
  - “Younger physicians today are less professional, skilled, etc. because of <insert favorite complaint>”
Faculty and Clinical Skills

“Evidently it is not deemed necessary to assay students’ and residents’ clinical performance once they have entered the clinical years. Nor do clinical instructors more than occasionally show how they themselves elicit and check the reliability of the clinical data…”
Faculty and Clinical Skills

To a degree that is often at variance with their own professed scientific standards, attending staff all too often accept and use as the basis for discussion, if not recommendations, findings reported by students and residents without ever evaluating the reporter’s mastery of the clinical methods utilized or the reliability of the data obtained.”
Faculty and Clinical Skills

From George Engel

1976 editorial on JAMA study highlighting deficiencies of student and resident’s basic clinical skills
Diagnostic Errors

IOM Report
Released September 2015

- At least 5 percent of U.S. adults who seek outpatient care each year experience a diagnostic error.
- Postmortem examination research shows diagnostic errors consistently contribute to ~ 10 percent of patient deaths.
- Diagnostic errors account for 6 to 17 percent of hospital adverse events.
The “Miracle” of Medical Education

Residency

“Then a miracle occurs.”

“I Think You Should Be More Explicit Here In Step Two.”

© 2015 Accreditation Council for Graduate Medical Education
Professional Self-Regulation: Assessment

Assessments within Program:
- Direct observations
- Audit and performance data
- Multi-source FB
- Simulation
- ITExam

Qual/Quant "Data" Synthesis: Committee

Residents

Faculty, PDs and others

Milestones and EPAs as Guiding Framework and Blueprint

Unit of Analysis: Program

Accreditation

Certification and Credentialing

Unit of Analysis: Individual
Assessing for the Desired Outcome

Knows (knowledge)

Knows How (competence)

Shows How (performance)

Does (action)

Performance in Practice/Multi-source feedback/
Direct Observation

Standardized Patients/Simulation

Diagnostic Reasoning using clinical vignettes or CSR

Multiple choice Questions

Work-based assessment is mostly accomplished through the observations and questions of faculty, team members, patients, peers and other co-workers

© 2015 Accreditation Council for Graduate Medical Education
Cambridge Model: “Righting” the Pyramid

Work-based assessment has to be the primary focus of our assessment systems

Rethans, Norcini, et al, 2002
Model For Programmatic Assessment
(With permission from CPM van der Vleuten)

- ○ = learning task
- △ = learning artifact
- ▲ = single assessment data-point
- ▲ = single certification data point for mastery tasks
- — = learner reflection and planning
- --- = social interaction around reflection (supervision)
- = learning task being an assessment task also

Committee

Time

Training Activities
Assessment Activities
Supporting Activities

Intermediate Eval
Intermediate Eval
Final Evaluation
Dreyfus & Dreyfus Development Model

MILESTONES and EPAs

<table>
<thead>
<tr>
<th>Novice</th>
<th>Advanced Beginner</th>
<th>Competent</th>
<th>Proficient</th>
<th>Expert/Master</th>
</tr>
</thead>
</table>

Curriculum ↔ Assessment

Time, Practice, Experience

Dreyfus SE and Dreyfus HL. 1980
Carraccio CL et al. Acad Med 2008;83:761-7

© 2015 Accreditation Council for Graduate Medical Education
Importance of Effective Observation and Work-based Assessment
Direct Observation: Assess Core Skills

Legitimizes the subject
Sends message skills are important
Ensures assessment of essential skills
Expert Performance vs. Everyday Skills

Ericsson KA. Acad Med. 2004
Design and Sequencing of Training Activities

* Monitor students’ development
* Design and select training tasks for individual students

From Anders Ericsson: Used by Permission
The Role of the Coach

- “They observe, they judge, and they guide”

- “That one twenty-minute discussion gave me more to consider and work on than I’d had in the past five years”

- “Medical practice is largely unseen by anyone who might raise one’s sights. I’d had no outside ears and eyes.”

Atul Gawande, New Yorker 10/3/2011
Do We Need More Assessment Methods?
Common Assessment Methods

- Descriptive evaluation by teachers
- Records of clinical encounters
- External/ internal evaluations
  - MCQ
  - Key features/script concordance
  - Short answer questions/essays
- Simulation/OSCES
- Checklists
- Rating scales
- Oral examinations
- Chart (record) reviews
- Standardized patients
- A-V reviews
- Educational prescription contracts
- Portfolios
- 360 evaluation
- Patient logs
How to Choose?

I’m a little overwhelmed, guys.
Assessment Utility
Utility Elements of Assessment

- Validity
- Reliability
- Educational impact
- Acceptability
- Cost effectiveness

- Identify gap between resident performance and desired outcome

Trainee Learning & Assessment


Ensure high quality patient
Inform supervision
Educational Impact

**Educational Effect**
“The assessment motivates those who take it to prepare in a fashion that has educational benefit.”

**Catalytic Effect**
“The assessment provides results and feedback in a fashion that creates, enhances, and supports education; it drives future learning forward.”

Norcini J et al. Med Teach 2011;33:206-14
Frame of Reference
Construct Alignment
<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unsatisfactory</td>
<td>Satisfactory</td>
<td>Superior</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Below Expectation</td>
<td>At Expectation</td>
<td>Exceeds Expectation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>NORMATIVE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not What I Do</td>
<td>Close to What I Do</td>
<td>What I Do (or better)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SELF</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Kogan JR et al. Med Educ. 2011;45:1048-60
<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>???</td>
<td>???</td>
<td>???</td>
<td>???</td>
<td>???</td>
<td>???</td>
<td>???</td>
<td>???</td>
<td>???</td>
</tr>
</tbody>
</table>

**GESTALT**

- Missing evidence based elements
- Most evidence based elements
- All evidence based elements

**BEST PRACTICE**

Kogan JR et al. Med Educ. 2011;45:1048-60
Rethinking Our Frame of Reference

- Importance of appropriate supervision
- Entrustment

Trainee performance* X Appropriate level of supervision**

Must = Safe, effective patient-centered care

* a function of level of competence in context
** a function of attending competence in context
Construct Aligned Scales

**Good questions, good answers: construct alignment improves the performance of workplace-based assessment scales**

Jim Crossley, Gavin Johnson, Joe Booth & Winnie Wade

Cognitive Load
Cognitive Load

- There is a limit as to how much you can ask faculty to observe and capture
  - Clinical units: complex environment
  - Selective attention
- Byrne et. al. (Med Educ 2014)
  - Average cognitive load for faculty judging OSCE stations was higher than anesthesia trainees during induction for routine surgery
    - OSCE had 21-22 items in an 8 minute station
Cognitive Load

Figure 3 Comparison of NASA–Task Load Index (NASA-TLX) scores in the study subjects (grey boxes) and trainee anaesthetists (white boxes).

Demand categories:
1 = mental demand
2 = physical demand
3 = temporal demand
4 = performance/success
5 = effort
6 = frustration

Cognitive Load and Milestones

- With few exceptions the reporting Milestones should not be used as a faculty evaluation form
- Reporting Milestones best used as a framework to guide synthetic, aggregate judgments
Creating Assessment Programs

- Competence is specific, not generic. Sample across contexts, assessors, time
- Use multiple assessment methods
- Quantitative not necessarily better than qualitative
- Move assessment back to workplace
- Use credible standards
- Validity resides in instrument user

How Much Does the Institutional Environment and Structure of the Training Program Really Matter?
Evaluating Residency Programs Using Patient Outcomes


Rate of Major Obstetric Complications by Graduates (%)

- Residency Program of Origin, Ranked (Quintile) by Program Complication Rate

Difference remains after correction for USMLE performance

Excess Risk \( \Delta 32\% \) Q1 vs Q5

\( \Delta Q1-Q5 \)
Choosing a Residency

Average # of physician visits in last six months of life (teaching hospitals in red)

From:
What Kind of Physician Will You Be?
Variation in Health Care and Its Importance for Residency Training
Dartmouth Institute for Health Policy & Clinical Practice 2012

Figure 2. Average number of physician per chronically ill Medicare patient during the last six months of life among patients receiving most of their care at teaching hospitals (2010 deaths)
Environment and Conservative Practice

### Table 3. Adjusted Patient Expenditures for Primary Care

<table>
<thead>
<tr>
<th></th>
<th>All Physicians</th>
<th>1-7 Years</th>
<th>8-15 Years</th>
<th>16-19 Years</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>β (95% CI)</strong>^a^</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physicians, No.</td>
<td>2851</td>
<td>480</td>
<td>1694</td>
<td>677</td>
</tr>
<tr>
<td>Medicare beneficiaries, No.</td>
<td>491,948</td>
<td>60,996</td>
<td>302,869</td>
<td>128,083</td>
</tr>
<tr>
<td>Training HRR spending[^C]</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Low</td>
<td>0.05 (0.00 to 0.09)</td>
<td>.04</td>
<td>0.22 (0.01 to 0.44)</td>
<td>.04</td>
</tr>
<tr>
<td>High</td>
<td>0.07 (0.02 to 0.12)</td>
<td>.007</td>
<td>0.29 (0.13 to 0.45)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>


© 2015 Accreditation Council for Graduate Medical Education
The Reality of Assessment Methods

There is NO HOLY GRAIL of Assessment:

George Box: “All models are wrong, some are useful”
Milestone Journey:
Revised Conceptual Model of Rapid Cycle Change


© 2015 Accreditation Council for Graduate Medical Education
Thank You and Questions

eholmboe@acgme.org
Selected References


Selected References


• Norman G, Norcini J, Bordage G. Competency-Based Education: Milestones or Millstones? J Graduate Medical Education. 2014; DOI: http://dx.doi.org/10.4300/JGME-D-13-00445.1 [epub ahead of print]