Tests of Knowledge: How Can They Contribute to Maintenance of Certification (MoC) and Revalidation?

David B Swanson, PhD
American Board of Medical Specialties
Chicago, USA

THE EFFECTIVENESS OF REVALIDATION SYSTEMS AND TOOLS: ARE WE ACHIEVING OUR MANDATE?
2015 IAMRA Revalidation Symposium
October 29-30, 2015 - Montréal, Québec
Overview of Presentation

• Challenges in using tests of knowledge for Maintenance of Certification (MoC) and revalidation
• Assessment of Learning: Using knowledge tests summatively for doctors having performance problems
• Assessment for Learning: A new approach to Maintenance of Certification Assessment (MOCA)
• Need for institutional cooperation in knowledge test development and use
• Discussion
Challenges in Using Tests of Knowledge for MoC and Revalidation

• The primary purpose of the test may be unclear
Using Assessment of Learning to Eliminate the Poor Performers
Using Assessment for Learning to Shift the Performance Curve to the Right

Pass/Fail Standard

55  60  65  70  75  80  85  90  95
Challenges in Using Tests of Knowledge for MoC and Revalidation

- The primary purpose of the test may be unclear
- It is difficult to determine what content areas to assess: practices evolve (mostly narrow) over the course of a career
<table>
<thead>
<tr>
<th>Category</th>
<th>Health Maintenance</th>
<th>Mechanisms of Disease</th>
<th>Diagnosis</th>
<th>Management</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>04 Immunologic Disorders</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>05 Blood and Blood Forming Organs</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>06 Mental Disorders</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>07 Nervous System &amp; Special Senses</td>
<td>2</td>
<td>7</td>
<td>5</td>
<td>4</td>
<td>18</td>
</tr>
<tr>
<td>08 Cardiovascular Disorders</td>
<td>3</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>24</td>
</tr>
<tr>
<td>09 Diseases of Respiratory System</td>
<td>2</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>22</td>
</tr>
<tr>
<td>10 Nutritional and Digestive Disorders</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>7</td>
<td>18</td>
</tr>
<tr>
<td>11 Gynecologic Disorders</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>12 Renal, Urinary, Male Reproductive</td>
<td>0</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>13 Pregnancy, Childbirth, Puerperium</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>14 Skin and Subcutaneous Tissue</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>15 Musculoskeletal, Connective Tissue</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>16 Endocrine, Metabolic Disorders</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>TOTAL</td>
<td>20</td>
<td>50</td>
<td>70</td>
<td>60</td>
<td>200</td>
</tr>
</tbody>
</table>
## Sample Blueprint for a 200-item Licensing Exam

<table>
<thead>
<tr>
<th>Classification</th>
<th>Category</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Site of Care</strong></td>
<td>Ambulatory</td>
<td>106</td>
<td>114</td>
</tr>
<tr>
<td></td>
<td>Emergency Department</td>
<td>46</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>Hospital</td>
<td>36</td>
<td>44</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td>Male</td>
<td>90</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>100</td>
<td>110</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>Newborn to 12 months</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>13 months to 5 years</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>6 to 12 years</td>
<td>16</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>13 to 17 years</td>
<td>22</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>18 to 64 years</td>
<td>64</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>65 to 75 years</td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>75+ years</td>
<td>18</td>
<td>26</td>
</tr>
</tbody>
</table>
Practice Narrows over Time

Broad Specialty Practice (as represented on an initial certifying exam)
Practice Narrows over Time

Broad Specialty Practice (as represented on an initial certifying exam)

Actual Practice
Practice Narrows over Time

- Actual Practice
- Potential Practice
- Broad Specialty Practice (as represented on an initial certifying exam)
Challenges in Using Tests of Knowledge for MoC and Revalidation

• The primary purpose of the test may be unclear
• It is difficult to determine what content areas to assess: practices evolve (mostly narrow) over the course of a career
• Doctors often view tests of knowledge as assessing recall of facts that are of marginal relevance to patient care
Recall Version
Which of the following is the most appropriate initial management of suspected optic neuritis in patients with multiple sclerosis?
A. Intravenous methylprednisone
B. Oral methylprednisolone
C. CT scan of the orbit
D. MRI scan of the brain

Application Version
A 24-year-old woman experiences sudden loss of vision in her right eye with a visual acuity of 20/400. She has a right afferent pupillary defect, optic disc swelling, and a central scotoma. Her left eye is normal. A complete neurological history and physical examination reveal no abnormalities. Which of the following is the most appropriate next step?
(same option list)
A 52-year-old man with alcoholism comes to the ED because of a 2-day history of nausea, vomiting, and increasingly severe abdominal pain that radiates to his left shoulder and back. He appears extremely dehydrated and is short of breath. Vital signs are: temp 37.8°C, pulse 120/min, resp 18/min, and BP 80/60 mm Hg. Abdominal examination shows distention with epigastric tenderness. Bowel sounds are decreased. Rectal exam shows no abnormalities. CT scan of the abdomen is shown.

Which of the following is the most likely diagnosis?

A. Acute portal vein thrombosis
B. Hemorrhagic pancreatitis
C. Perforated posterior gastric ulcer
D. Ruptured gastric varices
E. Splenic artery rupture
Sample Patient Management Item

A 45-year-old man hospitalized for chemotherapy for treatment of acute myelogeneous leukemia develops shaking chills and fever. He appears pale. Vital signs show a temperature of 38.8°C, blood pressure of 118/76 mmHg, pulse of 112/min, and respirations of 34/min. Examination shows petechiae over the inner forearm. Laboratory studies show a hematocrit of 24%, leukocyte count of 440/mm³, and platelet count of 15,000/mm³. A blood culture is pending. The most appropriate next step is intravenous administration of which of the following?

A. Cefazolin and gentamicin  
B. Cefazolin and ticarcillin  
C. Cefazolin and tobramycin  
D. Gentamicin and ticarcillin  
E. Gentamicin and tobramycin  
F. Tobramycin and ticarcillin
Challenges in Using Tests of Knowledge for MoC and Revalidation

- The primary purpose of the test may be unclear.
- It is difficult to determine what content areas to assess: practices evolve (mostly narrow) over the course of a career.
- Doctors often view tests of knowledge as assessing recall of facts that are of marginal relevance to patient care.
- A high-stakes test taken every 10 years may lack credibility from public perspective, particularly toward end of a career.
The “Ballistic Model” of Competence: The Skilled Trainee is “Launched into Practice”
The “Ballistic Model” of Competence: Competence Changes at Different Rates

- Medical School
- Postgraduate Training
- Practice
- Retirement

Performance vs. Time (across a practice lifetime)

Acceptable performance
The “Ballistic Model” of Competence: Competence Changes at Different Rates

- Medical School
- Postgraduate Training
- Practice
- Retirement

Time (across a practice lifetime)
Challenges in Using Tests of Knowledge for MoC and Revalidation

• The primary purpose of the test may be unclear
• It is difficult to determine what content areas to assess: practices evolve (mostly narrow) over the course of a career
• Doctors often view tests of knowledge as assessing recall of facts that are of marginal relevance to patient care
• A high-stakes test taken every 10 years may lack credibility from public perspective, particularly toward end of a career
• Development of high-quality tests of knowledge customized to practice requires a lot of effort and test material
  — Need for normative information to aid in setting standards
Using Knowledge Tests Summatively for Doctors Having Performance Problems

• Going back a decade, good examples found in the UK: the Competence component of GMC’s Performance Procedures
• Given to doctors identified through complaints from a variety of sources
• Performance component included practice visit with interviews and chart audit (Case-based Discussion)
• Competence (MCQ-based) tests customized to reflect specialty area and practice content assembled as needed
  – Test material contributed by multiple institutions
  – Tests also taken by doctors in good standing to obtain normative information and set pass/fail standards
Overview of Presentation

- Challenges in using tests of knowledge for Maintenance of Certification (MoC) and revalidation
- Assessment of Learning: Using knowledge tests summatively for doctors having performance problems
- Assessment for Learning: A new approach to Maintenance of Certification Assessment (MOCA)
  - Initially developed by American Board of Anesthesiology
  - Longitudinal knowledge tests used formatively to identify knowledge gaps and assist doctors in keeping up to date
- Need for institutional cooperation in knowledge test development and use
- Discussion
A More Reasonable Model of Competence: Performance Managing a Single Condition
Performance Falling below Minimum Personal Acceptable Level.

- Aha! Surprise! Critical incident
- Diabetes
- Minimum personal acceptable level

Performance vs. Time Graph
Variation in Minimum Personal Acceptable Levels of Performance

![Graph showing variation in performance over time with minimum acceptable levels and a condition labeled as "Diabetes".]
Depending upon Degree, This Variation Can Come to the Attention of Regulators
Performance across Different Conditions

- Cervical smears
- Optimal professional performance
- Hypertension
- Breaking bad news
- Diabetes
- Minimum personal acceptable level
- Minimum professionally acceptable level

Time → Performance
Point-in-Time Summative Assessment

- Cervical smears
- Hypertension
- Breaking bad news
- Diabetes

Performance vs. Time

Summative Assessment
Criteria for Summative Assessment of Learning (from Norcini et al, 2011)

• **Validity**: There is a coherent body of evidence supporting *use of assessment results for a specific purpose.*

• **Reproducibility (reliability)**: Results of the assessment would be roughly the same *if a similar assessment were repeated under similar circumstances*

• **Equivalence**: The assessment *yields equivalent scores* or decisions when administered at different locations or points in time
  - *If the test is long enough with appropriate content coverage, it can provide a good basis for making overall summative decisions*
  - *It does not provide a good basis for providing detailed feedback regarding strengths and weaknesses in individual content areas*
Additional Criteria for Assessment for Learning (from Norcini et al, 2011)

- **Educational effect:** The assessment *motivates* those who take it to prepare in an educationally beneficial way.

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

- **Specific, actionable feedback** is provided on an ongoing basis.

*Case/Problem-based Learning for Practitioners?*
Principles of Assessment for Learning

1. Assessment is viewed as an *instructional design problem*

2. Assessment is seen as having large *direct and indirect* effects on learning and retention; achieving the desired impact (and avoiding undesirable ones) is emphasized in assessment design
   - Growing body of research evidence that this is the case

3. Assessment delivery is designed to *promote learning and retention, providing feedback* about areas of strength/weakness

4. Multiple assessment methods are used; *testing is frequent/ongoing*

5. Assessments incorporate *authentic, integrated clinical tasks* that require multiple competencies for successful completion

6. Across longitudinally administered assessments, contexts and clinical problems are *broadly sampled*
Research on Test-Enhanced Learning Shows

Direct Effects of Testing:
• Material is better remembered when it is tested than when it is not
• In well-controlled studies, being tested boosts retention substantially more than studying for equal amounts of time

Indirect Effects of Testing:
• Study time increases and study strategies improve with frequent, longitudinal assessments
• Benefits for retention greater when testing is spaced out over time
• Frequent, longitudinal assessments encourage keeping up to date

Direct and indirect effects of testing occur in tandem;
Suggests a very different approach to MoC exams
Principles of Assessment for Learning

1. Assessment is viewed as an instructional design problem
2. Assessment is seen as having large direct and indirect effects on learning and retention; achieving the desired impact (and avoiding undesirable ones) is emphasized in assessment design
   - Growing body of research evidence that this is the case
3. Assessment delivery is designed to promote learning and retention, providing feedback about areas of strength/weakness
4. Multiple assessment methods are used; testing is frequent/ongoing
5. Assessments incorporate authentic, integrated clinical tasks that require multiple competencies for successful completion
6. Across longitudinally administered assessments, contexts and clinical problems are broadly sampled
Maintenance of Certification Assessments (MOCA): A New Approach to MoC Exams Pioneered by ABA

- Design emphasizes assessment for learning: Identifying knowledge gaps and assisting diplomates in keeping up to date
- Practice-relevant, with diplomates able to customize content coverage and control assessment timing/length
- Frequent, longitudinal administrations with spaced repetition of assessment content to promote learning and retention
Longitudinal Formative Assessments

- Cervical smears
- Breaking bad news
- Diabetes
- Hypertension
Maintenance of Certification Assessments (MOCA): A New Approach to MoC Exams Pioneered by ABA

- Design emphasizes assessment for learning: Identifying knowledge gaps and assisting diplomates in keeping up to date
- Practice-relevant, with diplomates able to customize content coverage and control assessment timing/length
- Frequent, longitudinal administrations with spaced repetition of assessment content to promote learning and retention
- Immediate feedback on performance – right/wrong answers and explanation of correct responses ("critiques")
- Optionally, results accumulated over time* can provide a basis for making summative decisions regarding continuing certification

*eg, instead of 200 items once every 10 years, 100 items/year = 1000 items over 10 years with ongoing feedback on performance
Options for MOCA Design to Meet Differences in Needs and Emphasis

Assessment Organization

• Question-based assessments (similar to current MoC exams)
• Article-based assessments using recent publications to assist doctors in keeping up to date clinically
• Problem/Topic-based assessments (eg, items grouped into themes like management of asthma, CHF, diabetes, hypertension)
• Combinations of the above

Item Formats

• Single-best-answer items presented in clinical vignette format to increase relevance
• “Unfolding cases” (multi-item sets addressing patient care over an episode of illness)
• Support for use of images and multimedia
• Access to reference material during assessments

Alternate administration/security modes, including remote proctoring and support for mobile devices to increase user convenience
Need for Institutional Cooperation in Knowledge Test Development and Use

- Both summative assessments of learning and formative assessments for learning can be useful, but they
  - Serve different purposes
  - Require different assessment designs
  - Are handled by different groups (regulatory authorities vs Royal Colleges/Specialty Boards or Societies)
  - Involve different sources of information for “triage” of doctors into the assessments

- Coordination of activities (and, perhaps, *international* sharing of resources and test materials) are desirable to maximize effectiveness and reduce costs
Questions and Some References


