Design of the Continuous Knowledge Self-Assessment

Thomas R. O’Neill, Ph.D.
Michael R. Peabody, Ph.D.
Aimee R. Eden, Ph.D.
Disclosure

• All three presenters are salaried employees of ABFM
PURPOSE
Purpose

• The CKSA was designed to be a self-assessment activity that provides physicians with knowledge about their medical knowledge and clinical decision-making ability.
• It is hoped that it will improve their selection of CME activities.
• The CKSA is NOT intended as an activity in which medical information is learned or reinforced, although this may in fact occur.
• As such, it is a Part II activity, not a Part III activity.
• It cannot be “failed”; it is “no-stakes”; it has no security.
Purpose

• Part II (Self-Assessment and Life-Long Learning) and Part III (secure medical specialty test to establish sufficient mastery to warrant certification) are different perspectives on the same specialty.
  – CKSA, ITEs, and the core of the FMCE are made to comparable specifications.

• Part III provides the public with good evidence that the physician does have a sufficient fund of relevant medical knowledge and clinical decision-making ability that meets the standard implied in the certification. [Once every 10 years.]

• Part II is how the physician maintains his/her fund of medical knowledge.
Purpose

• Part II activities
  – “Self-Assessment and Life-Long Learning”
    OR

• Self-Assessment is really a Metacognitive activity.
  – What do you know? What don’t you know? Do you know the difference?
  – This should guide your learning.

• Life-Long Learning is about acquiring information or updating your knowledge-base.
INTEGRATION OF COMPONENTS
ITE PGY1 → ITE PGY2 → ITE PGY3 → FMCE (Once Every Ten Years)
Bayesian Score Predictor

If you need assistance with the Bayesian Score Predictor, please contact Thomas O'Neill, Ph.D. by phone at 888-995-5700 Ext. 1223 or email at tonell@theabfm.org

Video:
- Navigation & Instructions
- Interpretation
CKSA

Year 1
Q1 Q2 Q3 Q4

Year 10
Q1 Q2 Q3 Q4

FMCE
Once Every 10 Years

Old ITEs
Old ITEs
Old ITEs

FMCE
Once Every 10 Years
DATA COLLECTION
Don’t do this! Lots of effort with very little pay-off.
Which one of the following, when confirmed with a repeat test, meets the diagnostic criteria for diabetes mellitus?

- A fasting blood glucose level of 120 mg/dL
- A 2-hour value of 180 mg/dL on an oral glucose tolerance test
- A random glucose level of 180 mg/dL in a patient with symptoms of diabetes mellitus
- A positive urine dipstick for glucose

- A hemoglobin A1c of 7.0%
Registration to this forum is free. We request that you abide by the rules and policies detailed below. If you agree to the terms, please check the "I Agree" checkbox and press the "Complete Registration" button below. Although the administrators and moderators of the ABFM Knowledge Self-Assessment Commenting forum will attempt to keep all objectionable messages off the site, it is impossible to review all the content, which is continuously contributed by the community. All messages express the views of the author, and the ABFM will not be held responsible for the content posted by an individual user. By agreeing to these rules, you warrant that you will not post any messages that are obscene, inappropriate or otherwise violate any laws. The information communicated in this forum is confidential to the community of those commenting on the ABFM Continuous Knowledge Self-Assessment (CKSA) activities. The ABFM reserves the right to remove, edit or delete any users who do not adhere to the listed policies.

***Please only post CKSA related questions in this forum. If you have questions/comments about the certification requirements, the examination; the clinical simulation; or other matters, please submit to the ABFM Support Center at help@theabfm.org.***

I Agree
I was wondering if the first answer is correct.
ABILITY ESTIMATION
Ability Estimation

- Physicians who do not practice broad-spectrum family medicine or do not practice at the top of the specialty may be over-confident because they are confident in their day-to-day patient interactions and believe that these interactions are representative of broad-spectrum family medicine.

- Others may be underconfident because the FMCE is a high-stakes proposition.

- Knowing where you stand relative to the FM passing standard is helpful.

- It can help guide a physician’s continuing development.
Ability Estimation

• The CKSA uses the dichotomous Rasch model to calibrate the questions and estimate physician ability.
• All CKSA questions have been pretest on the FMCE and therefore are already calibrated.
• The ITE, the core of the FMCE, and the CKSA are all built to the same content specifications.
• CKSA results should be “generally” predictive of ITE and CKSA results.
  – Except that CKSA has no security and is a no stakes proposition.
  – There is NO penalty for not taking it seriously.
Predictive Validity

• Does the CKSA measure the same thing as the FMCE?

• Does Q1 performance predict later (Q2) FMCE performance?

• Does (Q2) FMCE performance predict Q2 performance?

• Does (Q2) FMCE performance predict Q3 performance?
  – Q3 had better questions than CKSA Q1 or Q2.
$r = 0.54$

$\rho = 0.83$ disattenuated

$N = 540$
$r = 0.47$

$r = 0.73$ disattenuated

$N = 587$
$r = 0.51$
$r = 0.79$ disattenuated
$N = 491$
The means and SDs presented above are based upon the 321 participants that took the April 2017 certification examination and who participated in CKSA Q1, Q2, Q3.

We expected no one to be in this group because no Part II activities are required in their 10th year of certification. People usually test in year 10.

<table>
<thead>
<tr>
<th></th>
<th>Mean (SD)</th>
<th>Difference from FMCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>FMCE</td>
<td>575 (106)</td>
<td>-</td>
</tr>
<tr>
<td>Q1</td>
<td>619 (175)</td>
<td>44</td>
</tr>
<tr>
<td>Q2</td>
<td>627 (173)</td>
<td>52</td>
</tr>
<tr>
<td>Q3</td>
<td>504 (130)</td>
<td>-71</td>
</tr>
</tbody>
</table>
Issues with Q1 and Q2 Predictive Validity

- All 321 recently tested.
- Probably look at the iPhone Test Prep app
- Q1 and Q2 CKSA items were on the iPhone app.
- The mean for these people on the exam was 575.
METACOGNITIVE ACCURACY INDEX
Metacognition

• Metacognition is knowledge you have about your own cognitive processes
• Not always an accurate assessment of reality.
Availablility Heuristic

• Causes people to make judgements about the likelihood of an event based on how easily an example comes to mind.

• Being able to easily recall a few patients that seemingly represent full-spectrum care would make a physician believe s/he is practicing full-spectrum care when, in fact, s/he is not.
Representativeness Heuristic

- Judging the probability of an event belonging to a certain class based on the degree to which the event resembles the class; however, this neglects the probability of the class occurring in the first place.
Dunning-Kruger Effect

• Mistakenly believing their ability to be much higher than it really is.

• Highly confident in their wrong answers.
Imposter Syndrome

• An inability for successful people to internalize their accomplishments.

• Not confident in their correct answers.
Outcomes Bias in Decision Evaluation

Jonathan Baron
University of Pennsylvania

John C. Hershey
Department of Decision Sciences
University of Pennsylvania

In a study, undergraduate subjects were given descriptions and outcomes of decisions made by
other people. Decisions concerned whether medical fires or monetary
outcomes were consistent with the information. Subjects rated the
quality of thinking, the decision maker’s blame, the decision maker’s
credibility, and their confidence in the decision maker. The
results showed that subjects who were given the
information that the decision maker had received less
information than they had, were more likely to
evaluate the decision maker as incompetent, and were
generally less confident in the decision maker. The
results suggest that the information given about
the decision maker’s knowledge can influence the
subjects’ evaluation of the decision maker.

Baron, J., & Hershey, J. C. (1988). Outcome bias in
decision evaluation. Journal of Personality and
Social Psychology, 54, 569-579.
Outcomes Bias

• Tendency to judge a decision based on the outcome rather than on the quality of the decision at the time it was made.
<table>
<thead>
<tr>
<th>KNOWLEDGE</th>
<th>CONFIDENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>YES</td>
<td>Imposter Syndrome</td>
</tr>
<tr>
<td>NO</td>
<td>Dunning-Krueger Effect</td>
</tr>
</tbody>
</table>

- Imposter Syndrome
- Dunning-Krueger Effect
MetaCognitive Accuracy Index (MCAI)

• To assess the accuracy of judgments, we correlate metacognitive judgments (confidence) and performance outcomes.

• Pearson correlation between correct/incorrect and confidence
  – Extreme scores create null values
  – All correct, all incorrect, all same confidence rating
Distribution of MCAI (Q1)
NULL Value MCAI (Q1)

- 50 all correct
- 0 all incorrect
- 116 all same confidence
  --
  - 25 all correct and same confidence
NULL Value MCAI
Good MCAI (.69)
Bad MCAI (.21)
PARTICIPANT RESPONSE TO CKSA
## Diplomate Feedback—First Quarter

<table>
<thead>
<tr>
<th></th>
<th>Relevant</th>
<th>Current</th>
<th>Useful</th>
<th>Favorable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extremely</td>
<td>32% (1713)</td>
<td>34% (3146)</td>
<td>28% (1525)</td>
<td>38% (2064)</td>
</tr>
<tr>
<td>Very</td>
<td>46% (2513)</td>
<td>58% (1860)</td>
<td>47% (2573)</td>
<td>49% (2647)</td>
</tr>
<tr>
<td><strong>Subtotal</strong></td>
<td><strong>78%</strong></td>
<td><strong>92%</strong></td>
<td><strong>75%</strong></td>
<td><strong>87%</strong></td>
</tr>
<tr>
<td>Somewhat</td>
<td>21% (1131)</td>
<td>8% (408)</td>
<td>23% (1230)</td>
<td>12% (668)</td>
</tr>
<tr>
<td>Not at all</td>
<td>1% (73)</td>
<td>0.3% (16)</td>
<td>2% (102)</td>
<td>1% (51)</td>
</tr>
</tbody>
</table>
Diplomate Feedback—First Quarter

<table>
<thead>
<tr>
<th>How likely to:</th>
<th>Continue CKSA</th>
<th>Recommend to others</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extremely</td>
<td>62% (3368)</td>
<td>45% (2428)</td>
</tr>
<tr>
<td>Very</td>
<td>33% (1812)</td>
<td>40% (2169)</td>
</tr>
<tr>
<td><strong>Subtotal</strong></td>
<td><strong>95%</strong></td>
<td><strong>85%</strong></td>
</tr>
<tr>
<td>Somewhat</td>
<td>4% (235)</td>
<td>12% (675)</td>
</tr>
<tr>
<td>Not at all</td>
<td>0.3% (15)</td>
<td>3% (158)</td>
</tr>
</tbody>
</table>
Please tell us how the CKSA activity could be improved.

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
<th>Example Quote</th>
</tr>
</thead>
</table>
|        | **Explanation** Desire to receive an explanation or discussion of correct and/or incorrect responses | "more extensive explanations"  
"give reason why correct answer is correct for clarification and improve our knowledge." |
|        | **Q. Content** Suggestions to improve the content of the questions: needing more details in the questions; more "bread & butter" FM questions and fewer "zebras"; ensuring info is up to date; ensuring a range of topics; problems with specific questions | "Questions too esoteric"  
"There weren't many pediatrics or gynecology questions"  
"I recommend providing a little more detail on cases."  
"Using questions that more closely follow current practices and guidelines." |
|        | **2 Post Qs** Expressions of annoyance with one or both of the follow-up questions about confidence and looking up info | "Asking after every question if I got help is redundant and annoying." |
|        | **App** Requests for, or looking forward to, mobile phone or ipad format | "make it in mobile format too" |
What did you like best about the CKSA?

<table>
<thead>
<tr>
<th>Code</th>
<th>Example Quote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feedback</td>
<td>“Immediate feedback and opportunity to read the critique”</td>
</tr>
<tr>
<td>Convenience</td>
<td>“Able to answer questions on my schedule, when I had time.”</td>
</tr>
<tr>
<td>Content</td>
<td>“Wide variety of conditions and populations represented.”</td>
</tr>
</tbody>
</table>
Please explain why you may continue.

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
<th>Example Quote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ease</td>
<td>Ease, convenience, short, flexibility</td>
<td>“quick, nice bite of activity able to do in one session”</td>
</tr>
<tr>
<td>Learning</td>
<td>Contributes to learning; good review in general or for exam</td>
<td>“it is appropriate and forces me to confront the complete scope of Family Medicine (whether I am comfortable with that or not, given that I am aging with my practice).”</td>
</tr>
<tr>
<td>Useful</td>
<td>Useful, relevant to practice</td>
<td>“The questions are very clinical and appropriate to every day patient seen and managed.”</td>
</tr>
<tr>
<td>Fun</td>
<td>Fun, enjoyed it</td>
<td>“These were actually FUN to do...”</td>
</tr>
</tbody>
</table>
IDEAS FOR FUTURE IMPROVEMENTS
Ideas for Future Improvements

• Create a bolus of items for new participants, so they can get the initial feedback quickly.
• Automate the feedback process (ability estimates with confidence intervals, predications of passing, MCAI, etc.).
• Add new indices such as an Engagement Index,
• Add more trend information.
• Test for participant-specific content clusters that the participant might want to consider reviewing.
• Administer the items adaptively.
ENGAGEMENT
Engagement

• It is not just being agreeable.
• It is providing value.
• More important to provide value to the majority than appease a few.
• Friendly approach.
  – Be able to articulate the value of what you are doing.
  – Be accessible
  – Be user-friendly
  – Listen to users concerns
• More positive frequent interactions with the board, not just Dec.
Questions?
Thomas O’Neill (toneill@theabfm.org)

Michael Peabody (mpeabody@theabfm.org)

Aimee Eden (aeden@theabfm.org)