

# 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 design 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” AND “Life-Long Learning”
- 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**

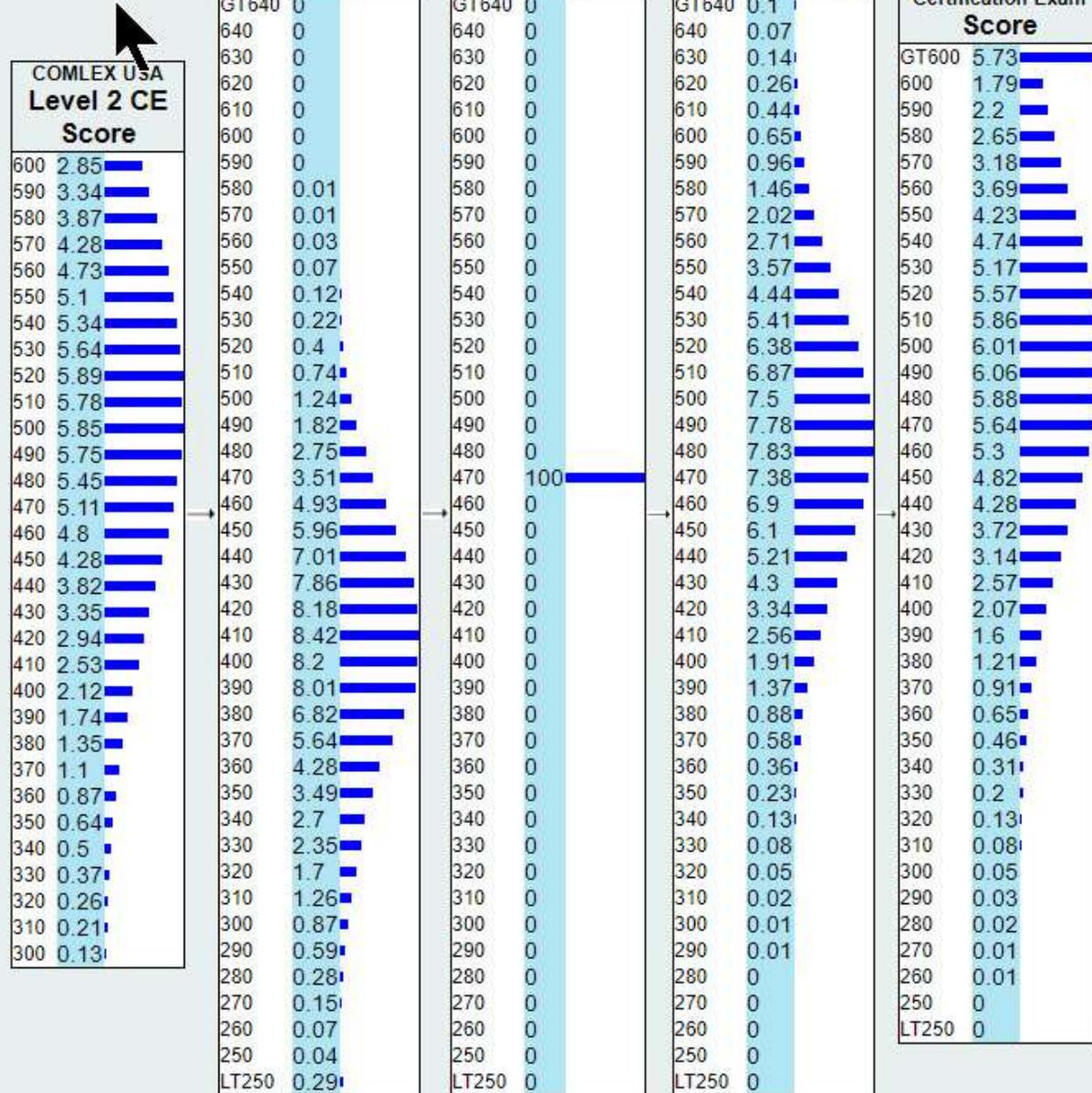


**ITE  
PGY3**



**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. 1225 or email at [toneill@theabfm.org](mailto:toneill@theabfm.org)

Reset

Instructions

Pass Fail Status	
Pass	97
Fail	3

## Video:

Navigation & Instructions

Interpretation



# DATA COLLECTION





- Home >
- Questions >
- Report >
- Settings >
- Withdraw >
- Return to Portfolio >

### Question Rating

How confident are you that your answer was correct?

- Not at all Confident
- Slightly Confident
- Moderately Confident
- Pretty Confident
- Very Confident
- Extremely Confident

Don't do this!  
Lots of effort  
with very little  
pay-off.

A hemoglobin A<sub>1c</sub> of 7.0%

Submit



- Home >
- Questions >
- Report >
- Settings >
- Withdraw >
- Return to Portfolio >

Question Critique Comment

Question 1

Which one of the following, when confirmed with a repeat test, meets the diagnostic criteria for diabetes mellitus?

Incorrect

A fasting blood glucose level of 120 mg/dL

A 2-hour value of 180 mg/dL on an oral glucose tolerance test

**X A random glucose level of 180 mg/dL in a patient with symptoms of diabetes mellitus**

A positive urine dipstick for glucose

**✓ A hemoglobin A<sub>1c</sub> of 7.0%**

← Previous

Next →



Home >

Questions >

Report >

Settings >

Withdraw >

Return to Portfolio >

## User Agreement

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](mailto:help@theabfm.org).\*\*\***

I Agree

Complete Registration

Cancel

Submit

Next ➔





- Home >
- Questions >
- Report >
- Settings >
- Withdraw >
- Return to Portfolio >

Question Critique **Comment**

### User Comments

Add new comment

Submit

I was wondering if the first answer is correct.

Dr. Smith at 11/15/2016, 9:10 AM



← Previous

Next →



# 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.



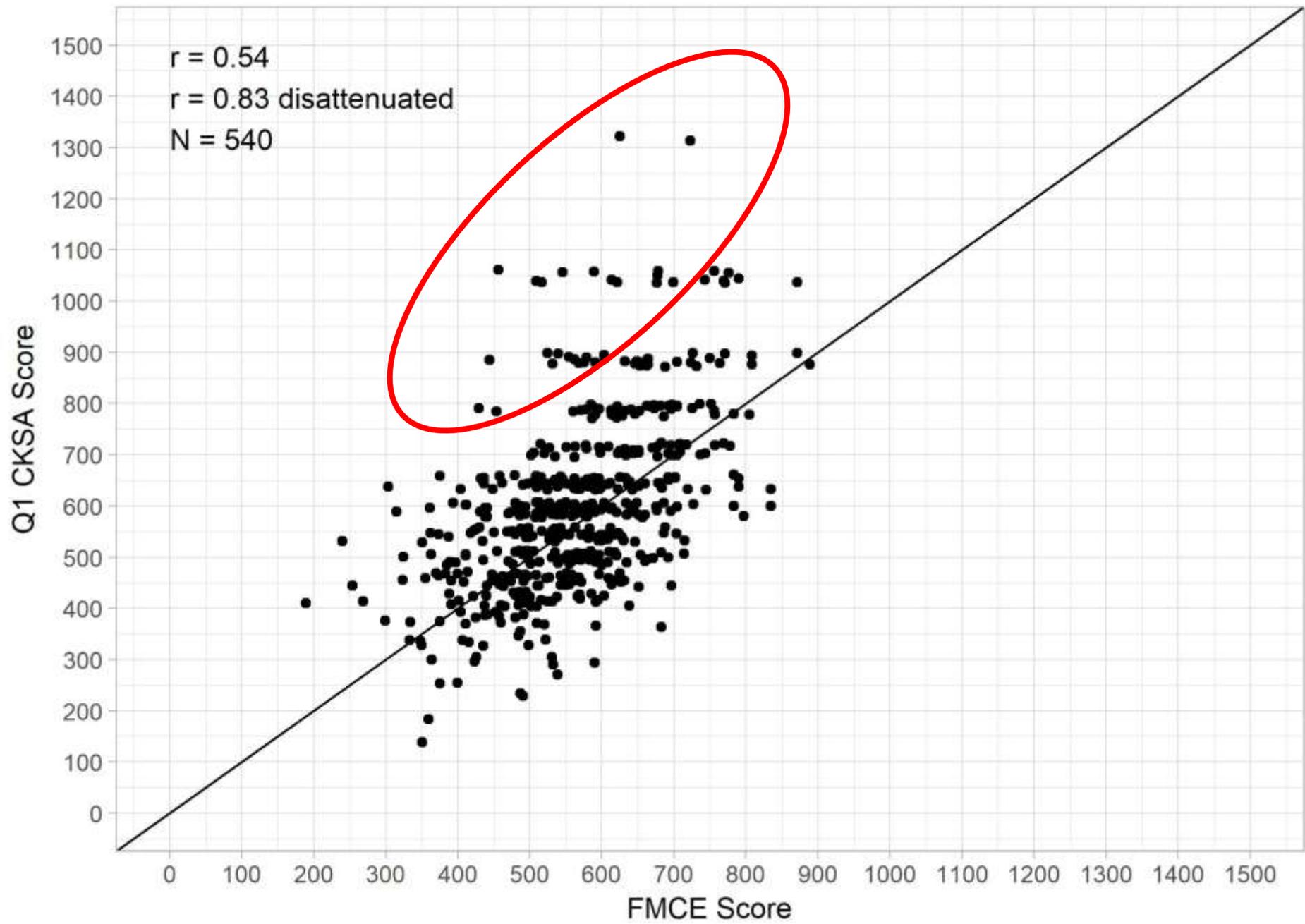
# Ability Estimation

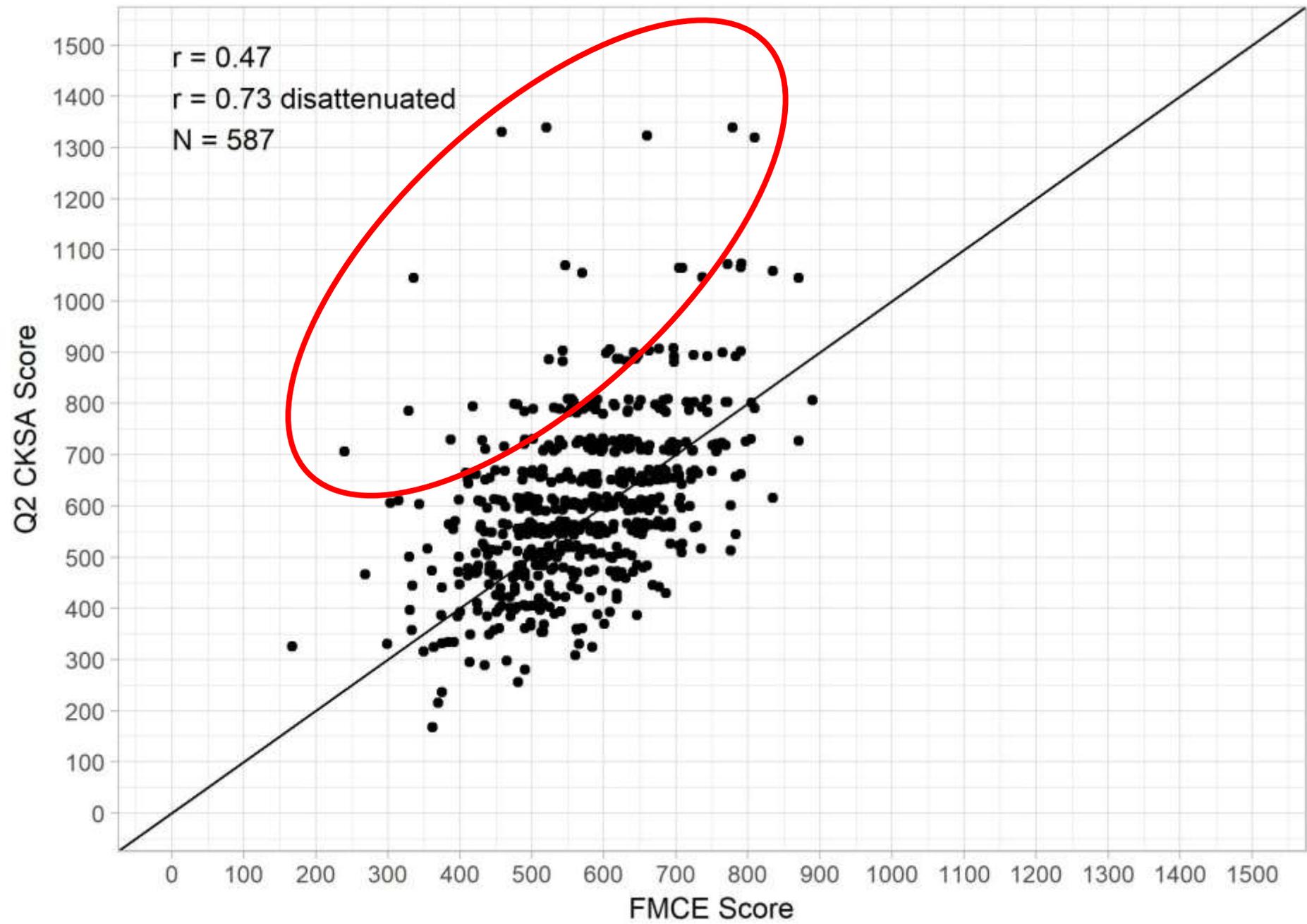
- The CKSA uses the dichotomous Rasch model to calibrate the questions and estimate physician ability.
- All CKSA questions have been pretest on the FMCE and therefore are already calibrated.
- The ITE, the core of the FMCE, and the CKSA are all built to the same content specifications.
- CKSA results 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.**

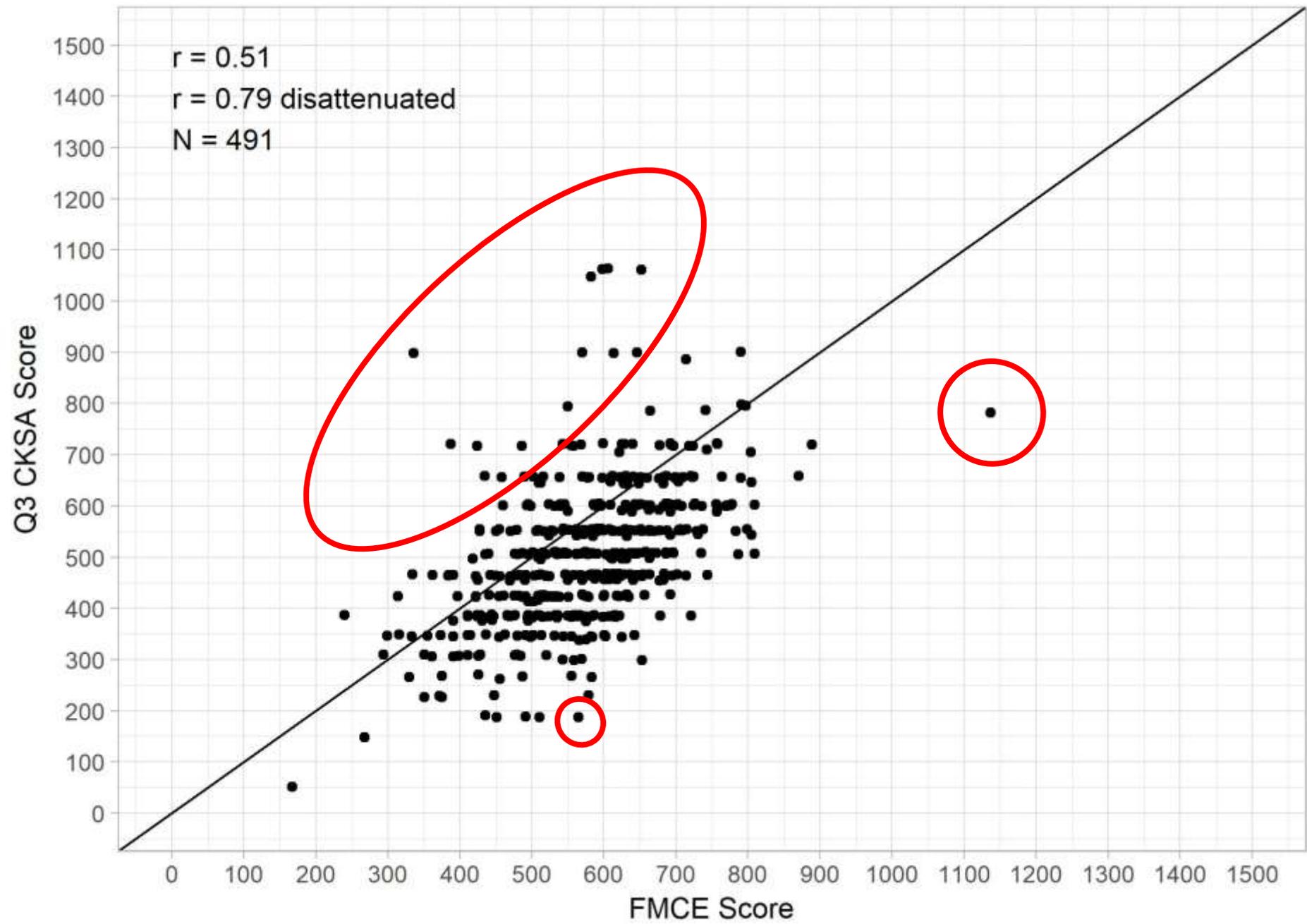


# Predictive Validity

- 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.







	Mean (SD)	Difference from FMCE	
<b>FMCE</b>	575 (106)	-	321
<b>Q1</b>	619 (175)	44	321
<b>Q2</b>	627 (173)	52	321
<b>Q3</b>	504 (130)	-71	321

- 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 10<sup>th</sup> year of certification. People usually test in year 10.

# 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

American Board of Family Medicine Inc.



# Metacognition

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



A,  
COGNITIVE PSYCHOLOGY 5, 207-232 (1973)

## Availability: A Heuristic for Judging Frequency and Probability<sup>1,2</sup>

AMOS TVERSKY AND DANIEL KAHNEMAN  
The Hebrew University of Jerusalem and the Oregon Research Institute

This paper explores a judgmental heuristic in which a person evaluates the frequency of classes or the probability of events by availability, i.e., by the ease with which relevant instances come to mind. In general, availability is correlated with ecological frequency, but it is also affected by other factors. Consequently, the reliance on the availability heuristic leads to systematic biases. Such biases are demonstrated in the judged frequency of combinatorial outcomes, and of repeated events. The correlation is explained as an availability bias. The paper discusses the implications of these findings for the study of subjective probabilities and scenarios on subjective probabilities and consequences.

*Tversky A, Kahneman D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 5(2), 207-232.*

ever, also...  
frequency of...  
We propose that...  
ability or frequency, people...  
reduce these judgments to simple...  
...one such heuristic—represent...  
...population or generating process...  
...in several studies. For exam...  
...believe that the sequence of com...  
...other HHHHTH or HHHHTT, a...  
...of Jerusalem, Jerusalem,  
...from the Central  
National

# Availability 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.

# Representative

# Heuristic

COGNITIVE PSYCHOLOGY 3, 430-454 (1972)

## Subjective Probability: A Judgment of Representativeness

DANIEL KAHNEMAN AND AMOS TVERSKY<sup>1</sup>  
The Hebrew University, Jerusalem

This paper explores a heuristic—representativeness—according to which the subjective probability of an event, or a sample, is determined by the degree to which it: (i) is similar in essential characteristics to its parent population; and (ii) reflects the salient features of the process by which it is generated. This heuristic is explicated in a series of empirical examples illustrating predictable and systematic errors in the evaluation of unpopulations. In particular, it is expected to have little or no effect on decisions and posterior probability judgments are characteristic of the sample (e.g., proportion, size). This prediction is confirmed in studies showing that the present heuristic approach to the

The deci-  
offer

*Kahneman D, Tversky A. (1972). Subjective Probability: A judgement of representativeness. Cognitive Psychology, 3(3), 430-454.*

sions we  
are usually bas  
such as success in a  
of the market. Indeed a  
devoted to the question of how  
abilities of uncertain events  
statistics, and decision mak  
about the psychology of un  
general empirical generalizations  
general conclusion, obtained from

the Research and Development  
6782 to the second author,  
to Oregon Research  
phases of the  
We also  
Sarah

# 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.

Du

fect

Journal of Personality and Social Psychology  
1999, Vol. 77, No. 6, 1121–1134

# Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments

Justin Kruger and David Dunning  
Cornell University

People tend to hold overly favorable views of their abilities in many social and intellectual domains. The authors suggest that this overestimation occurs, in part, because people who are unskilled in these domains suffer a dual burden: Not only do these people reach erroneous conclusions and make unfortunate choices, but their incompetence robs them of the metacognitive ability to realize it. Across 4 studies, the authors found that participants scoring in the bottom quartile on tests of humor, grammar, and logic grossly overestimated their test performance and ability. Although their test scores put them in the 12th percentile in metacognitive skill, or the capacity to distinguish accuracy from error. Paradoxically, believing the limitations of their abilities.

as promoting effective leadership, raising children, constructing a solid logical argument, or designing a rigorous psychological study. Second, people differ widely in the knowledge and strategy they apply in these domains (Dunning, Meyerowitz, & Holzberg, 1989; Kruger & Dunning, 1999). Some of the knowledge and strategy they apply are sound and meet with success. Some of the knowledge and strategy they apply are not sound and meet with failure. Some of the knowledge and strategy they apply are sound and meet with success. Some of the knowledge and strategy they apply are not sound and meet with failure. Some of the knowledge and strategy they apply are sound and meet with success. Some of the knowledge and strategy they apply are not sound and meet with failure.

***Kruger, J. & Dunning, D. (1999). Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments. Journal of Personality and Social Psychology, 77(6), 1121–1134.***

# Dunning-Kruger Effect

- Mistakenly believing their ability to be much higher than it really is.
- Highly confident in their wrong answers.



# Impostor

## The Impostor Phenomenon in High Achieving Women: Dynamics and Therapeutic Intervention

Pauline Rose Clance & Suzanne Imes  
Georgia State University  
University Plaza  
Atlanta, Georgia 30303

**Abstract:** The term impostor phenomenon is used to designate an internal experience of intellectual phonies, which appears to be particularly prevalent and intense among a select sample of high achieving women. Certain early family dynamics and later introjection of societal stereotyping appear to contribute significantly to the development of the impostor despite outstanding academic and professional accomplishments, women who think otherwise. Numerous achievements, which one might expect to contribute to the maintenance of impostor feelings over time, do not appear to affect the phenomenon. Psychotherapy, theme-centered psychotherapy, women – women who work in their fields, or who have earned degrees, and to be effective in helping women change the

*Clance, P.R., & Imes, S.A. (1978). The impostor phenomenon in high achieving women: dynamics and therapeutic intervention. Psychotherapy: Theory, Research and Practice 15(3), 241–247.*

phenomenon. In our clinical work with men and that with women from male-dominated fields, we have found that the impostor phenomenon is a self-unqualified to teach research and practice. In our clinical work with men and that with women from male-dominated fields, we have found that the impostor phenomenon is a self-unqualified to teach research and practice. In our clinical work with men and that with women from male-dominated fields, we have found that the impostor phenomenon is a self-unqualified to teach research and practice.

# Imposter Syndrome

- An inability for successful people to internalize their accomplishments.
- Not confident in their correct answers.



# Outcome Bias in Decision Evaluation

John C. Hershey  
Department of Decision Sciences  
University of Pennsylvania

Jonathan Baron  
University of Pennsylvania

In 5 studies, undergraduate subjects were given descriptions and outcomes of decisions made by others under conditions of uncertainty. Decisions concerned either medical matters or monetary gambles. Subjects rated the quality of thinking of the decision maker, the competence of the decision maker, or their willingness to let the decision maker decide on their behalf. Subjects understood that they had all relevant information available to the decision maker. Subjects rated the thinking as better, rated the decision maker as more competent, or indicated greater willingness to yield the decision when the outcome was favorable than when it was unfavorable. In monetary gambles, subjects rated the thinking as better when the outcome of the option not chosen turned out poorly than when it turned out well. Although subjects who were asked felt that they should not consider outcomes in making these evaluations, they did so. This effect of outcome knowledge on evaluation may be explained partly in terms of its effect on the salience of arguments for each side of the choice. Implications for the theory of rationality and for practical situations are discussed.

think they might have decided otherwise, and that decision makers end up being punished for their bad luck (e.g., Arnauld, 1662/1964; Berlin, 1984; Nichols, 1985). The distinction between a good decision and a good outcome is one to all decision analysts. The quotation from Edwards (1954) is labeled by the author as "a very old idea" (p. 7). In this paper, we explore how decisions and outcomes is recog-

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

decision... (Brown, Kan...  
A good decision cannot...  
sions are made under uncertainty...  
evaluating it as good or not must depend...  
not on the outcome. (Edwards, 1984, p. 7)

Evaluations of decisions are made in our personal lives...  
organizations, in judging the performance of elected officials,  
in certain legal disputes such as malpractice suits, liability  
regulatory decisions. Because evaluations are made  
is often information available to the judge  
decision maker, including informa-  
It has often been sug-  
that reasonable  
back who

to the de... know...  
information, know...  
evaluated, and joint info...  
maker at the time of decision and...  
evaluation. (In some cases, the decision...  
will be the same person, at different times.) In...  
consider, the judge has the outcome information and the...  
does not.

Although outcome information plays no direct role in the...  
of decisions, it may play an appropriate indirect role...  
affect a judge's beliefs about actor informa-  
now the decision maker's probabilit-  
was higher for an outcome...  
and it not occurred...  
nothing

# 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.



**KNOWLEDGE**

YES

Imposter  
Syndrome

NO

Dunning-  
Krueger Effect

NO

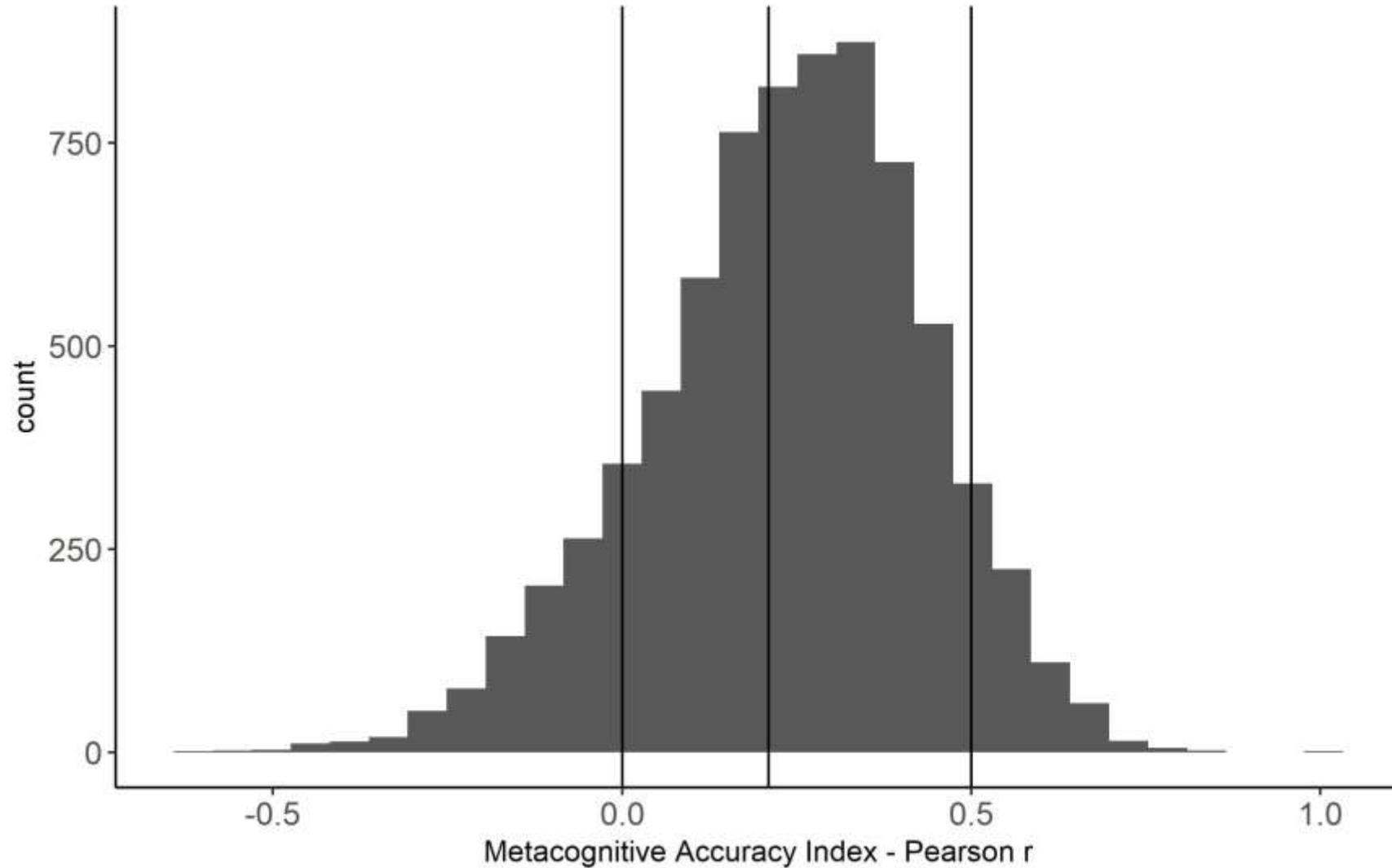
YES

**CONFIDENCE**

# 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

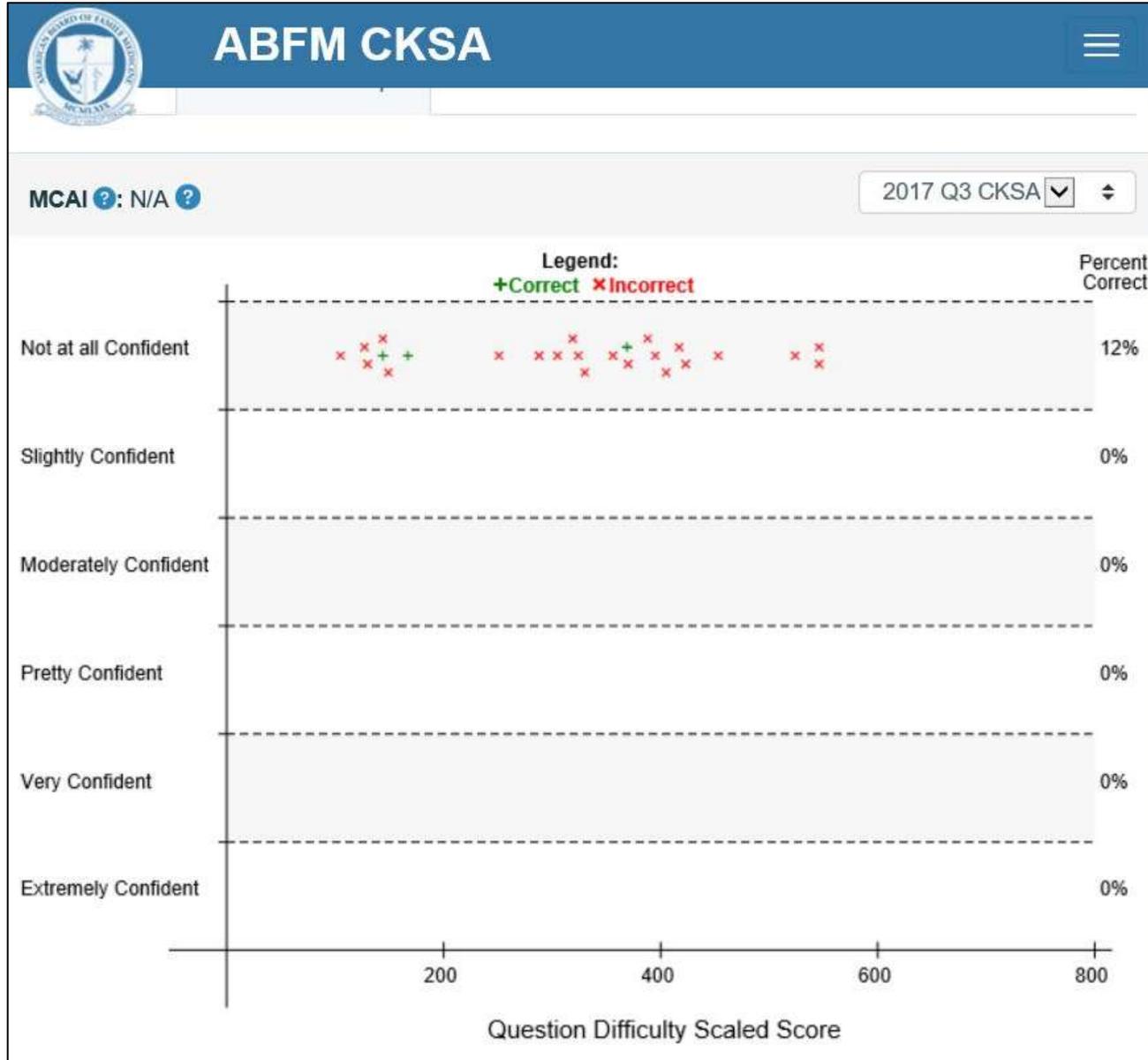
# Distribution of MCAI (Q1)



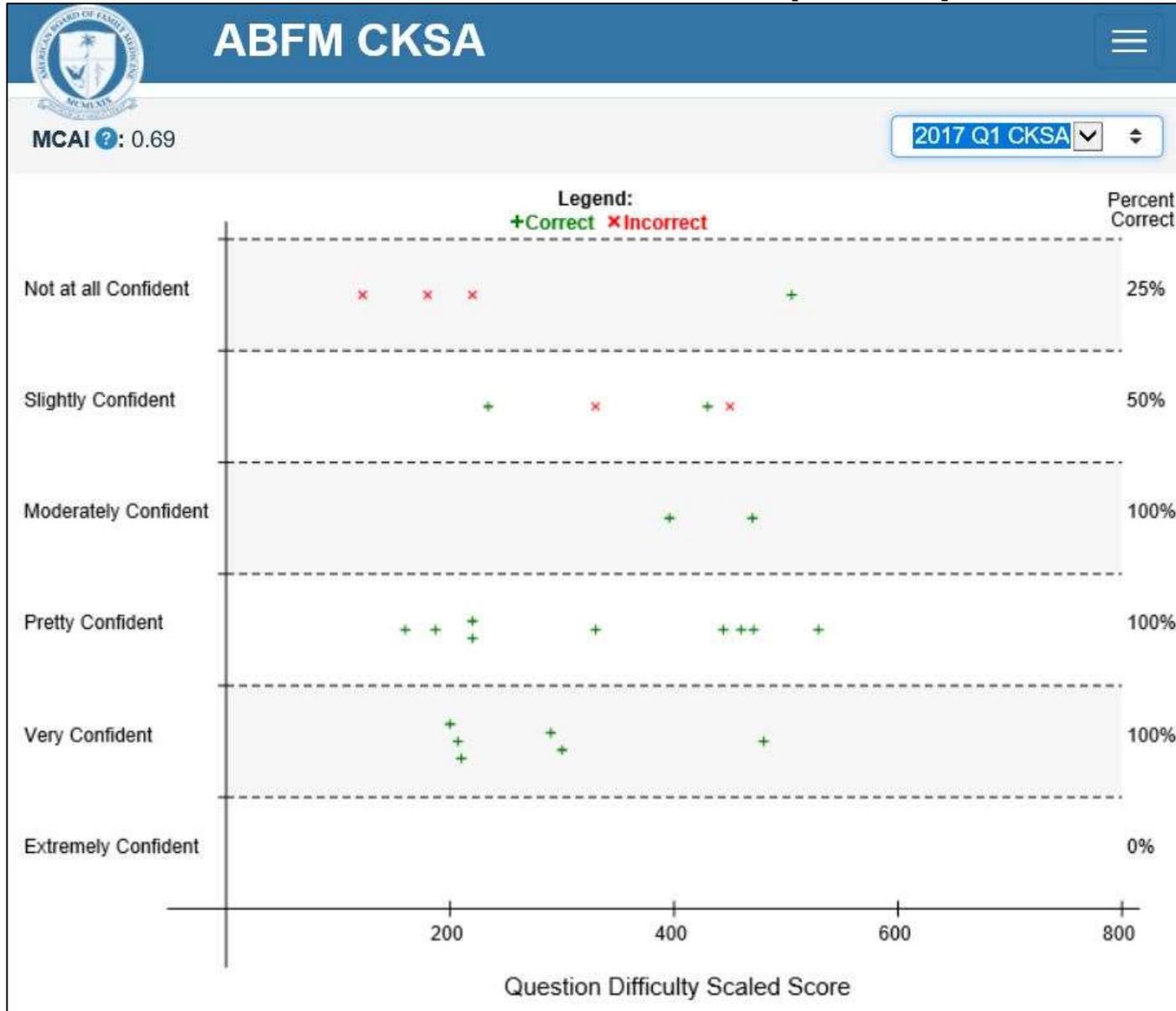
# NULL Value MCAI (Q1)

- 50 all correct
- 0 all incorrect
- 116 all same confidence
  - 
  - 25 all correct and same confidence

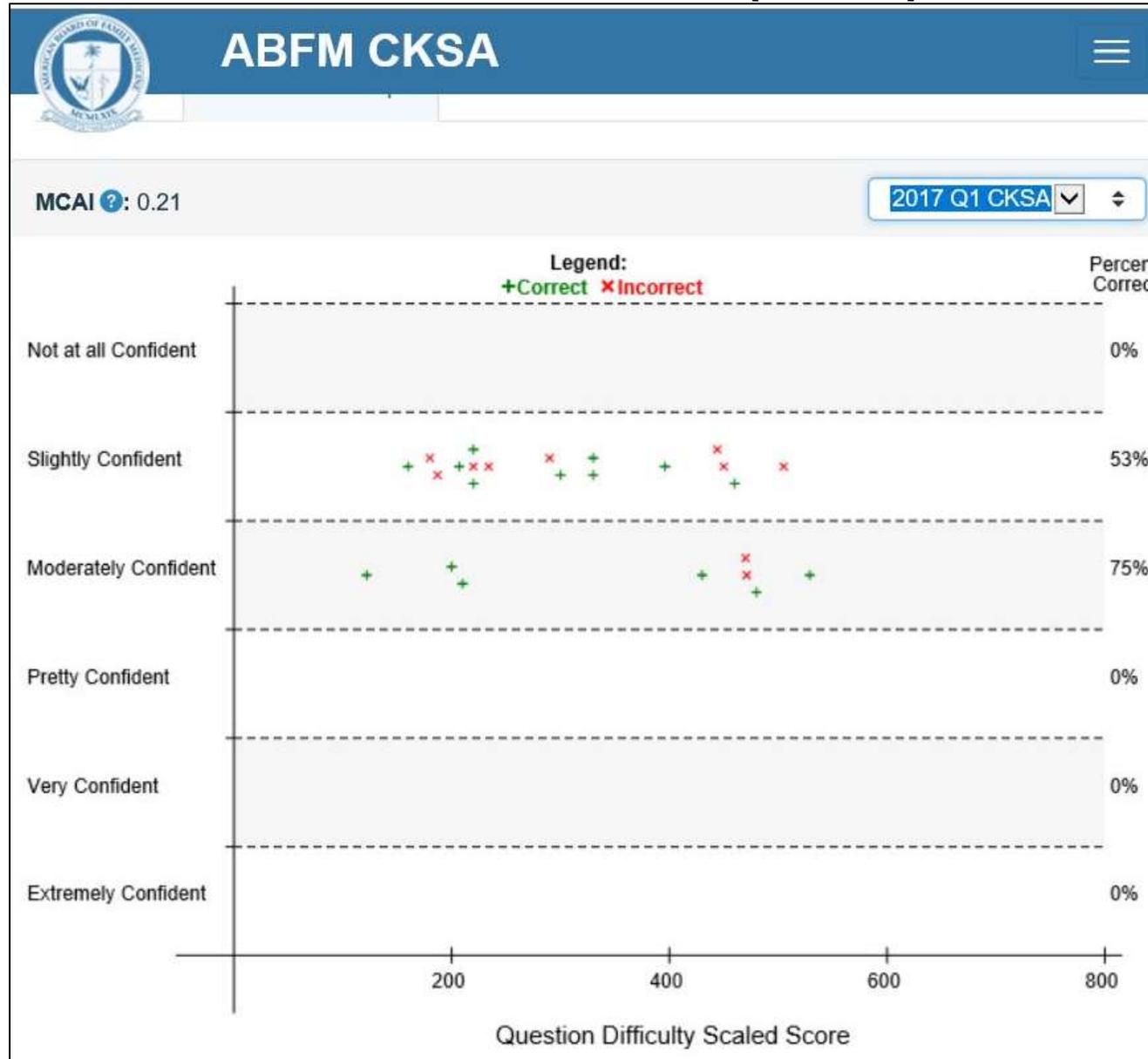
# NULL Value MCAI



# Good MCAI (.69)



# Bad MCAI (.21)



# PARTICIPANT RESPONSE TO CKSA

American Board of Family Medicine Inc.



# Diplomate Feedback—First Quarter

	Relevant	Current	Useful	Favorable
<b>Extremely</b>	32% (1713)	34% (3146)	28% (1525)	38% (2064)
<b>Very</b>	46% (2513)	58% (1860)	47% (2573)	49% (2647)
<b><i>Subtotal</i></b>	<b>78%</b>	<b>92%</b>	<b>75%</b>	<b>87%</b>
<b>Somewhat</b>	21% (1131)	8% (408)	23% (1230)	12% (668)
<b>Not at all</b>	1% (73)	0.3% (16)	2% (102)	1% (51)



# Diplomate Feedback—First Quarter

How likely to:	Continue CKSA	Recommend to others
<b>Extremely</b>	62% (3368)	45% (2428)
<b>Very</b>	33% (1812)	40% (2169)
<b><i>Subtotal</i></b>	<b>95%</b>	<b>85%</b>
<b>Somewhat</b>	4% (235)	12% (675)
<b>Not at all</b>	0.3% (15)	3% (158)

# Please tell us how the CKSA activity could be improved.

Code	Definition	Example Quote
<b>Explanation</b>	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."
<b>Q Content</b>	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 peds or gyn questions" "recommend providing little more detail on cases." "using questions that more closely follow current practices and guidelines."
<b>2 Post Qs</b>	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."
<b>App</b>	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?

Code	Example Quote
Feedback	“Immediate feedback and opportunity to read the critique”
Convenience	“Able to answer questions on my schedule, when I had time.”
Content	“Wide variety of conditions and populations represented.”

# Please explain why you may continue.

Code	Definition	Example Quote
Ease	Ease, convenience, short, flexibility	“quick, nice bite of activity able to do in one session”
Learning	Contributes to learning; good review in general or for exam	“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).”
Useful	Useful, relevant to practice	“The questions are very clinical and appropriate to every day patient seen and managed.”
Fun	Fun, enjoyed it	“These were actually FUN to do...”

# 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](mailto:toneill@theabfm.org))



Michael Peabody ([mpeabody@theabfm.org](mailto:mpeabody@theabfm.org))



Aimee Eden ([aeden@theabfm.org](mailto:aeden@theabfm.org))