American Board of Internal Medicine

USING SIMULATION FOR PHYSICIAN ASSESSMENT FROM THE PERSPECTIVE OF A TESTING ORGANIZATION

Rebecca S. Lipner, PhD

Coalition for Physician Enhancement Conference

June 2012, Denver

LEARNING OBJECTIVES

Discuss measurement issues that challenge the use of simulation in competency assessment

Describe the applicability of simulation to the evaluation of the post-licensure physician

Differentiate between the needs and standards appropriate for high-stakes competency assessments versus those for directed remedial education

PHYSICIAN EDUCATION, LICENSURE, AND CERTIFICATION

Medical Societies
- National membership organizations (ACP, ACC)
- Promote education and provide CME
- Develop clinical guidelines & publish medical journals

Licensing Boards
- State governed, non-profit, federated (FSMB/USMLE/NBME)
- Issue and regulate medical licenses— required for practice
- Varying requirements for CME for maintenance of licensure

Certifying Boards (American Board of _____)
- Non-profit "oversight" organizations
- Of the Profession, for the Public"
- Define "the field or disciplines"

AMERICAN BOARD OF MEDICAL SPECIALTIES (ABMS)

Founded in 1917 out of concern for quality care

Largest self-regulatory group of physicians in US

24 ABMS member boards

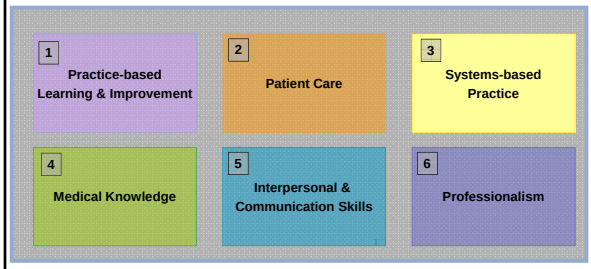
About 725,000 practicing certified physicians

About 230,000 in Internal Medicine

Certification: Immediately after training

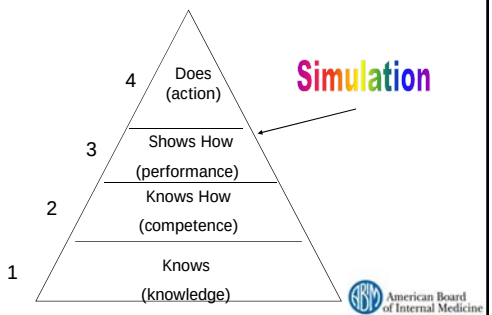
Maintenance of Certification (MOC): For practitioners every 10 years

PARADIGM SHIFT IN MEDICAL EDUCATION: OUTCOMES-BASED EDUCATION/ASSESSMENT



INSTITUTE OF MEDICINE QUALITY REPORTS

THEORY: MILLER'S FRAMEWORK FOR CLINICAL ASSESSMENT (1990)



TESTING ORGANIZATION'S PERSPECTIVE: WHY USE SIMULATION?

Enhance programs beyond attestation of procedural skills or numbers of procedures

Current summative assessments are insufficient at evaluating procedural and technical skills
- Examples: Interventional Cardiology and Gastroenterology

Higher fidelity to practice may yield higher validity of scores

Higher fidelity to practice may ultimately increase patient safety

MEASUREMENT ADVANCES IN SIMULATION: THE GOOD NEWS!

Administration somewhat easier

Skills to be evaluated are being better defined

Outcome measures more abundant and obtainable

Metrics are being captured by the computer

Scoring approaches are more sophisticated

Methods of equating available

Methods of standard-setting growing

Reliability and reproducibility better understood

Validation studies slowly being done

Boulet, Society for Academic Emergency Medicine, 2008

MEASUREMENT ADVANCES IN SIMULATION: THE *NOT* SO GOOD NEWS

Fidelity to practice
- Technology limitations
- Construct irrelevant variance may arise from lack of familiarity with simulator

Minimize false positives/negatives
- # of cases and raters needed for good measurement is significant
- Standardization is difficult

Valid inferences based on scores
- Automated scoring rubrics are complicated
- Adequate sampling of domain is limited due to time constraints
- Cases may be too memorable

Cost and logistics may be prohibitive

SCORING RUBRICS

Checklists - analytic
- Subject to potential bias (differing opinions)
- Timing and sequence are not scored

Ratings - holistic
- Subject to potential bias (unreliability)
- Rater training and calibration and controlled conditions important

Regression-based or rule based
- More than one correct way to manage patient
- Weights of actions and outcomes are important
- Timing and sequence can be scored

RELIABILITY AND REPRODUCIBILITY

Multiple cases or scenarios are needed
- Shorter cases in duration means more of them but lowers the fidelity to practice (tradeoff)

Internal consistency within one case

Generalizability studies and sources of error
- # of cases, # of raters per case, interaction
- Task specificity matters (more cases needed)

Reproducibility of decisions depends on where the cut score (i.e., standard) is set

VALIDITY

Content validity
- Cases modeled after actual practice
- Testing points and scoring rubrics use experts in field

Construct validity
- More experienced do better
- Relationship between real outcomes and simulated outcomes
- Relationship to other measures

Need many studies to build a body of evidence

DIFFERENCES BETWEEN SUMMATIVE AND FORMATIVE ASSESSMENTS

Element	Formative	Summative
Primary Goal	For physician: Feedback for improvement	For public: Physician competent or not
Measurement concepts	Educational impact, feasibility	Reproducibility of decision, validity
Standardization of cases	Important	Essential
Breadth of skills	Important	Essential
Stability of simulator actions	Essential	Essential
Scoring rubric	Detailed feedback	Validated approach
Case fidelity	Important	Important
Metrics automatically captured	Useful	Essential

SIMULATORS AND TECHNOLOGIES

Part-task trainers
- Body parts with functional anatomy (e.g., arm)
- Evaluate specific skills (e.g., suturing)
- Passive (i.e., no response from the simulator)
- Less sophisticated, lower fidelity, less costly, no data capture

Computer-enhanced life-sized mannequins
- Full body
- Evaluate individual skills or procedures, group skills
- Active (e.g., give a drug and BP may change)
- Higher fidelity, more costly, data capture, delivery difficult

Virtual reality
- Computer simulates physical world/sounds/visuals
- High fidelity, costly, data capture, more feasible delivery

Issenberg, Scalese. Perspectives in Biology and Medicine, 2008.

SPECIALTY AREAS CONSIDERED

General skills
- Eye, ear, breast exams, cyst aspiration
- Anatomy – shoulder, elbow, wrist, hand, knee

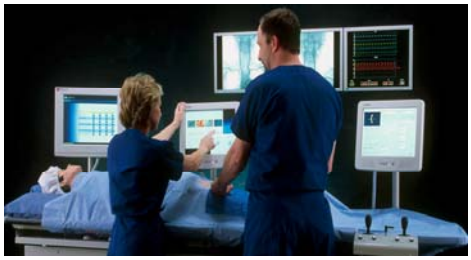
Critical care and Pulmonary disease
- Airway management (head/neck), intubation
- Cardiopulmonary

Internal Medicine
- Lumbar puncture, spinal injection

Gastroenterology
- Colonoscopy

Interventional Cardiology
- Cath lab procedures

COULD SIMULATION CASES BE USED FOR TESTING?



SimSuite® (Medical Simulation Corporation)

Lipner et al., Simulation in Healthcare, 2010

RESEARCH PROTOCOL

Demographic questionnaire
- Questions about training, practice, procedures

Survey about experience with simulator

Practice case on simulator
- Mechanical capabilities
- Decision making and initiation of therapy

Six cases
- Areas not tested via multiple-choice exam
- Use real clinical situations with range of difficulty
- Develop specific testing points

RESEARCH PROTOCOL (cont.)

Six case scenarios for evaluation
- About 20 minutes each
- Counterbalanced to control for order effect
- Varying levels of difficulty

Metrics automatically collected by simulator

Global ratings of procedural skills by clinical specialist

Rule-based scoring by subject matter experts

METHODS: SUBJECTS AND SITES

120 subjects and 10 clinical sites

3 levels of expertise (4 subjects/site at each level)
- Novice
  - » Cardiology fellows in first 3 months of 2nd year of training
- Skilled
  - » Interventional Cardiology fellow or Interventional Cardiologist in practice for 2 years or less, performing 75-100 therapeutic interventional cardiology procedures per year as primary operator (can be certified or not in interventional cardiology)
- Expert
  - » Certified Interventional Cardiologist in practice for 3 or more years, performing minimum of 100 therapeutic interventional cardiology procedures/year as primary operator, with a minimum of 1,000 during entire career

SIMULATOR SURVEY RESULTS

	Positive
Cases histories were clearly stated	93
Visual elements were simulated well	90
Auditory elements were simulated well	93
Tactile forces were simulated well	53
Advancement of catheter was simulated well	67
Stent placement was simulated well	84
Cases covered situations encountered in practice	97
Cases assessed my skills in a consistent manner	83

(86% Response Rate)

SIMULATOR SURVEY RESULTS (CONT.)

	Positive
The simulator was "User-friendly"	95
The case simulations were engaging and simulated well.	96
The orientation adequately prepared me for using the simulator.	95
During the study the clinical specialist provided adequate support as described in the instructions.	99
The study covered a reasonable range of interventional skills.	91

(86% Response Rate)

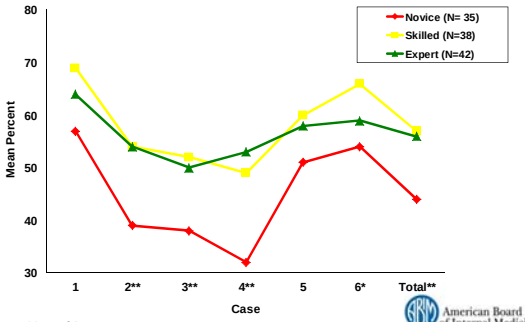
RELIABILITY AND VALIDITY EVIDENCE

Reliability = .69

Validity: Correlation of total score with
- Clinical specialist rating .55**
- Percent of time performing therapeutic procedures .41**
- Mean # of procedures in past year .35**

** $p<.01$

CASE PERFORMANCE BY SKILL LEVEL



** $p<.01$

RESEARCH STUDY CONCLUSIONS

Distinguished between novice and skilled or expert groups

Did *not* distinguish between skilled and expert groups

Reliability satisfactory for low stakes testing

Good evidence of validity

RESEARCH STUDY LIMITATIONS

Fidelity of tactile forces was not adequate

Convenience sample
- Classification of skill level was difficult

Raters were also clinical specialists

Scoring dependent on fields captured

Mentor ("talking head") was used

Cases did not reflect full array of skills

IMPLEMENTATION IN MOC

Contract with MSC to supply testing platform
- Six fixed centers in U.S.
- Related conferences
- Other testing venues: Hospitals, smaller conferences

No additional cost to examinees (currently)

Low stakes testing part of program

Continue to monitor environment for improvements in simulator's technical capabilities

Collaborate with appropriate societies

Provide feedback report

FEEDBACK TO USERS

	# Elements Scored	Percent Correct
I. Overall Performance	50	78%
II. Case Scores		
Case 1: A 75-year-old man has had diagnostic coronary angiography because of unstable angina; you will perform PCI.	8	59%
……		
Case 5: A 65-year-old man had a recent MI, and coronary angiography shows a lesion of the RCA (reference diameter, 3.0 mm).	11	76%
III. Domain Scores (across all cases)		
Angiographic Techniques	9	64%
Interventional Techniques	29	85%
Cognitive Skills/Decision Making	12	96%

FEEDBACK TO USERS (cont.)

Percentiles	Overall Performance	Angiographic Techniques	Interventional Techniques	Cognitive Skills/Decision Making
10	56	48	55	42
20	60	57	60	49
30	62	64	63	53
40	64	64	67	58
50	66	71	69	62
60	69	71	73	64
70	72	79	76	69
80	75	79	79	72
90	79	86	85	78
100	86	100	94	96

NEXT STEPS

Better understanding of state of simulation for Interventional Cardiology
- In conjunction with ACC and SCAI

Survey program directors
- Barriers: Cost, feasibility, technical features (e.g., tactile forces, advancement of the catheter, and stent placement) *not* simulated well

Survey vendors for state of simulators

Update current cases for MOC program

Investigate simulation in other disciplines

Preliminary Survey Findings

Sample

132 Interventional Cardiology Fellowship Programs

45% response rate (N=59)

- 83% were program directors, 5% associate program director

Results

24% - Use simulator to teach catheterization lab procedures

- Most use a computer-enhanced mannequin simulator

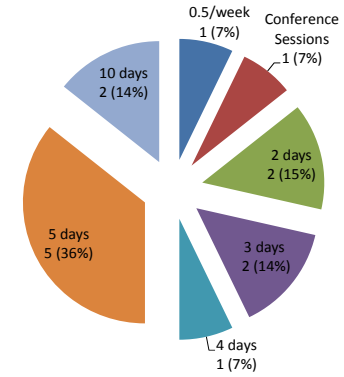
58% - Use simulator to teach in other areas of medicine

- Surgery, internal medicine, anesthesia, cardiology, ER

American Board of Internal Medicine

Question 8

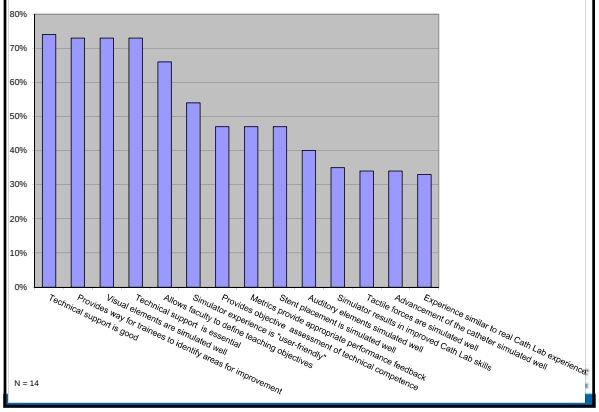
Number of Days Fellows Use the Simulator



N=14

Question 14

Agree / Strongly Agree to Statements about the Primary Simulator Teaching Procedures



N = 14

Question 15

What Respondents Like MOST About the Simulator

It allows an objective assessment of the trainees decision making process in an unsupervised and non life-threatening environment. It helps to teach the fellows to anticipate potential complications. It may have a role very early in their training but not later on.
The individualized experience of it for each fellow, both in enhancing their technical skills and showing them how to feel about their performance.
simple, convenient and available.
The cognitive elements, that is, having to plan each next step.
Uncommon scenarios or complications can be simulated.
The ability to create situations/complications for the fellows to have to address without putting a real patient at risk.
Allows Fellows to experience complications as "primary operator."

American Board of Internal Medicine

Question 16

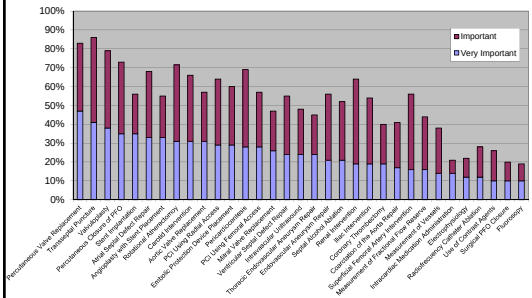
What Respondents Like LEAST About the Simulator

Less realistic than desirable
Does not replicate the real live cardiac cath lab experience
Additional feedback in the future could be via videos from the lab monitors.
The tactile sense to catheter and wire manipulations is not true to reality.
Limited in types of cases (albeit there is a good core set of cases).
Not a very good simulation of true clinical environment.
Financial burden of using simulation and need for central simulation center

American Board of Internal Medicine

Question 17

Ranked Importance of Including Specific Features within Simulation Cases



N = 59

SUMMARY

Measurement issues challenge the use of simulation in competency assessment but they are being addressed

Simulation may be able to be used for summative evaluation of the post-licensure physician to supplement live patient cases if

- Measurement issues are satisfactorily addressed
- Technical and cost barriers are overcome

The needs for high-stakes competency assessments are slightly different and may be more difficult to achieve than the needs for remedial education

American Board of Internal Medicine