

PERFORMANCE ASSESSMENT: REQUIREMENTS AND EVIDENCE TO SUPPORT THE USE OF SIMULATION:

David J Murray MD  
 Carol B. and Jerome T. Loeb  
 Professor of Medicine  
 Director, Howard and Joyce  
 Wood Simulation Center  
 Washington University School  
 of Medicine  
 Anesthesiologist-in-Chief  
 St Louis Children's Hospital,  
 St Louis, MO



The Simulation Centers  
 at Washington University School of Medicine

**Towards Safer Patient Care: Measuring the Efficacy of Simulation Training**

- Simulation Training → Professional skill
- Professional skill → Clinical Performance
- Clinical Performance → Practice Improvement
- Practice Improvement → Practice Standards
- Practice Standards → Patient Safety

The Simulation Centers  
 at Washington University School of Medicine

**Simulation assessment: current and future approaches to evaluating performance**

- Psychomotor Skills
- Clinical Judgments (Recognition, Diagnosis and Treatment)
- Communication and Teamwork Skills



**Performance Assessment: Requirements and Evidence to Support the Use of Simulation**

- Evidence of Simulation's Efficacy in Patient Safety
  - Psychomotor Skills, Simulation Drills
- Simulation Assessment: Cognitive, Judgment, Communication and Teamwork
  - Practical Simulation Assessment Design
    - Scenario, Scenario Content, Design and Number
    - Scoring and Performance Assessment
    - Reliability and Validity of Scenarios as Measure of Ability
    - Setting Performance Standards

The Simulation Centers  
 at Washington University School of Medicine

**Are Bad Outcomes From Questionable Clinical Decisions Preventable Medical Errors? A Case of Cascade Iatrogenesis**

Hofer TP, Hayward RA: Ann Intern Med 2002; 137: 327-333

- Priorities for enhancing safety:
  - Recognize that not all errors are "system errors"
  - More emphasis on research rather than initiatives and isolated case analysis
  - Investigate cognitive and procedural errors
  - Foster individual skills and expertise
  - Increase physician responsibility and accountability

The Simulation Centers  
 at Washington University School of Medicine

**Psychomotor Skill Acquisition**

A central venous line training program reduced retained guide wires, arterial puncture and pneumothorax.

Duncan JR, Henderson, K, Street, M, Richmond A, Klingensmith, M, Vanucci, A, Murray DJ. Creating and Evaluating A Data-Driven Curriculum for Central Venous Access Journal of Graduate Medical Education JGME 2: 389-397, 2010.

## Patient Safety Implications: Failure Mode and Effects Analysis (FMEA)

- ◆ Failure Mode and Effects Analysis (FMEA) to identify high-priority failure modes

- ◆ Recognizes that it is impossible to eliminate all failures
- ◆ Training directed to reduce the frequency, decreasing the severity, and improves the detection of failures before failure leads to harm.



## Performance in Practice: Simulation and Task Training

What preparation is appropriate prior to conducting a supervised procedure?

What are the steps required to effectively accomplish each procedure?

Are there technical milestones that could be measured for each step?

Are there common misadventures in performing this procedure?

Can these errors be programmed in a simulated environment?

What is the range of resident skill prior to and following simulation training?

Can the 'learning curve' be predicted from simulation training?

How many 'mentored' or training procedures are needed before a trainee possesses the expertise to function independently?

How can proficiency be measured for the procedure?

In order to meet minimum qualifications, how many procedures are needed?  
How many procedures are needed to maintain requisite skill?

## Simulation-Based Assessment : Why Score? Curriculum Quality

- Scores require faculty to define expectations and commit to the required and essential clinical management
- Participants' scores used to analyze quality of scenarios (difficulty, content, construct and scoring)
- Ability measure provide curriculum affirmation (or reproof) (Validation)
- Potential to direct curriculum effort to performance deficits, the need for an adaptive curriculum
- Individual score assures feedback directed to strengths and weaknesses
- Motivation to acquire and exceed performance expectations (the secondary learning outcome)

## Simulation-Based Assessment : Why Score? Goal-Directed Learning

- Secondary Outcomes
  - Motivate Learning Goals
    - USMLE Clinical Skills Examination and Medical Student Learning
  - Assessment drives learning
    - Teaching to test or testing to Teach? Boulet JR: Med Educ 2008; 42:952-3 4.
    - Test- enhanced learning in medical education. Larsen DP, Butler AC, Roediger HL III: Med Educ 2008; 42:959 - 66

## Simulation-based Assessment: Scenarios in Acute Care (Clinical Judgment, Decision-making, Communication)

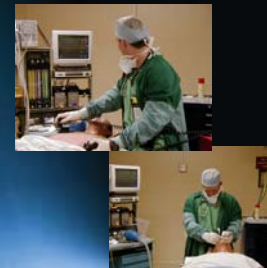
- Acute-care events are often associated with an adverse patient outcome, so improving this practice domain will result in increased patient safety
- 'Borderline' trainees frequently manifest skill deficits in performing a logical, sequential and timely patient evaluation
  - (Winkler TJ, Harris P. The transition out of medical schools: qualitative study of descriptions of borderline trainees interns. Medical Education 2002; 36: 466-72)
- Compressed timeline of a 'crisis' offers an opportunity to study skills and behaviors



The Simulation Centers  
at Washington University School of Medicine

## Simulation-based Assessment: Acute Events

- A specialist's advanced skills in patient care management (judgment, communication, teamwork) might not be evident in routine practice.
- Deficits in these essential skill sets are often difficult to recognize and more difficult to remediate
- Simulation offers a method to assess these skills



The Simulation Centers  
at Washington University School of Medicine

## Simulation-based Curriculum: Problem-based (scenario or events)

### Common Skill Deficits:

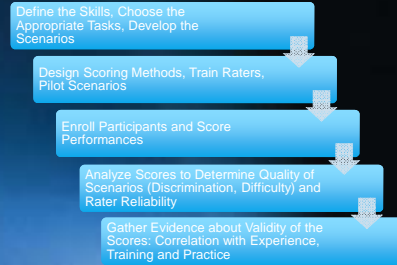
- *setting priorities,*
- *generating hypothesis,*
- *processing knowledge,*
- *assigning probabilities,*
- *recognizing important, from unimportant information,*
- *integrating competing issues,*
- *recognizing limits,*
- *learning when to call for assistance*

*The transition out of medical school—a qualitative study of descriptions of borderline trainee interns.*

*Wilkinson TJ, Harris P. Medical Education 2002; 36: 466-72*

The Simulation Centers  
at Washington University School of Medicine

## Building the Simulation Assessment



The Simulation Centers  
at Washington University School of Medicine

## Defining the Skills and Choosing the Appropriate Simulation Tasks

Scenario is the fundamental building block of the simulation-based assessment.

- Selecting competence domains that are amenable to a simulation environment,
- Defining the expected skills that are needed to diagnose and manage the crisis
- Designing a scenario that has the skills required embedded into the framework
- Establish instruments and metrics to conduct the assessment



The Simulation Centers  
at Washington University School of Medicine

## Malignant Hyperthermia

- *Heart rate increases, unresponsive to intervention,*
- *Arrhythmia, T wave changes ( $K^+$  6.8 meq),*
- *Progressive  $\downarrow$  CO2 and  $\downarrow$  tachycardia throughout (e.g. scenario 3 minutes)*
- *Key Scoring: Recognition of MH and Declaration of Emergency*



The Simulation Centers  
at Washington University School of Medicine

## Hyperkalemia

58 year old, 55 kg, diabetic, for fem-pop bypass [Retinopathy and Nephropathy (Cr=2.3 mg)]

Cross-Clamp removed 20 minutes ago

BP and HR elevated; Arrhythmia, peaked T wave,

$K^+$  to 7.4 meq, BP declines

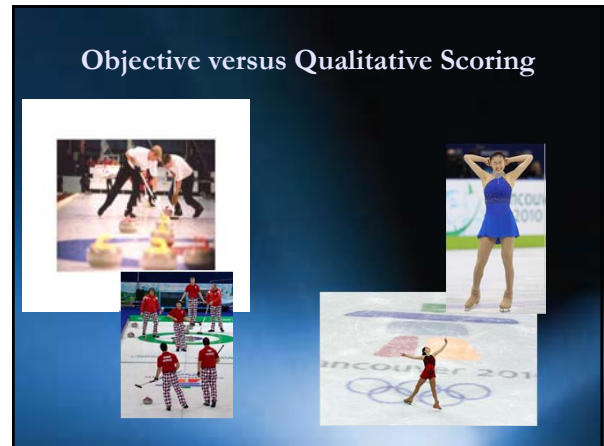
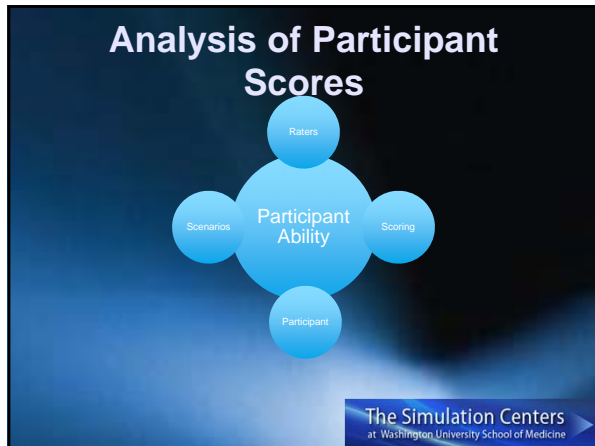


## Scoring Issues in Performance Assessment

- Choice of method
- Choice of person to score
  - Cost, efficiency, availability
- Experts and Consensus on Scoring
- What to score
  - Discreet actions
  - Speed, Accuracy, Skill, Sequence, Degree of Difficulty
- When to score
  - Real-time, audio-video recording

The Simulation Centers  
at Washington University School of Medicine





- ## Scoring is not random ... we hope....
- Can vary as a function of
    - Interpretation
    - Portrayal
    - Favoritism
    - Memory
    - Confusion
    - *Unethical conduct*
    - Etc.
- 
- The Simulation Centers  
at Washington University School of Medicine

- ## Participant Scores: Analysis of Scenarios, Raters and Scoring Systems
- Reliability of Participant Scores (*Participant Ability*)
  - Factors that influence the ability measure
    - Raters (*Definition and Interpretation, Bias, Attention*)
    - Content of the exercise (*Subject, Design, Construct*)
    - Scoring system used to measure skill
  - An Analysis of Sources of Variance Results in a Reliability Measure
- The Simulation Centers  
at Washington University School of Medicine

- ## Analysis of Participant Scores: Raters
- For each participant/scenario combination, what is the range of scores among the raters?
  - Do raters agree on the rank order of participants?
  - Do raters differ in overall stringency?
  - Are raters consistent on the overall score?
  - Do raters define success and failure similarly specific scoring items?
  - Are there specific action items that alter the overall score? ('the lethal action/omission/commission')
- The Simulation Centers  
at Washington University School of Medicine

- ## Scoring Analysis: Quality of Assessment
- Analysis of participant scores can reveal test flaws:
    - Scenario ineffective model of clinical event
    - Scoring system does not captures key elements of performance
    - Raters criteria or interpretation bias
- 
- The Simulation Centers  
at Washington University School of Medicine

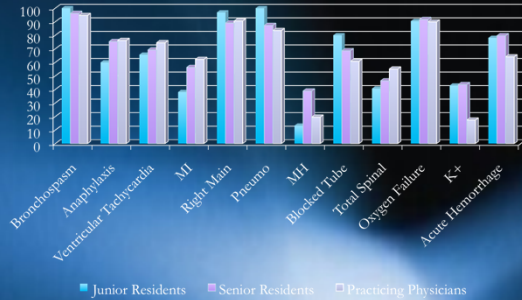
## Analysis of Participant Scores: Scenarios

- What are the range of scores for each scenario?
- Are some scenarios too easy? Are some scenarios too difficult?
- Are there any scenarios, that negatively correlate with performance on other scenarios?
- Is there a scenario that less experienced trainees do better than more experienced trainees?
- What is the correlation of performance across the scenarios? Is it a positive correlation?
- Is there a specific scoring item that correlates with the overall participants scenario score?
- Is there scoring items that do not differentiate final score?

The Simulation Centers  
at Washington University School of Medicine

## Scenario Score Variance: Rationale for the Multiple Scenario Assessment

Performance of Residents and Anesthesiologists in a Simulation-based Skill Assessment  
David J Murray M.D., John R. Boulet Ph.D., Michael Andam M.D., Joseph F. Kras, M.D., Bernadette Henrichs, CRNA, Julie Woodhouse RN  
Anesthesiology 2007; 67:76-83



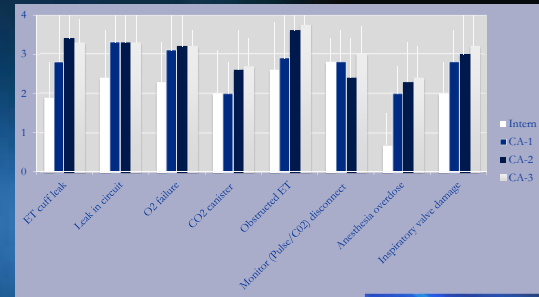
'The most effective method to improve the assessment is to increase the number of scenarios/stations/tasks that the participant manages'

Reliability and Validity of a Simulation-based Acute Care Skills Assessment for Medical Students and Residents  
Boulet JR, Murray DJ, Kras J, Woodhouse J, McAllister JD, Ziv A  
Anesthesiology 99:1270-1280, 2003.

The Simulation Centers  
at Washington University School of Medicine

## Simulation-based Assessment for Anesthesia Equipment Failure

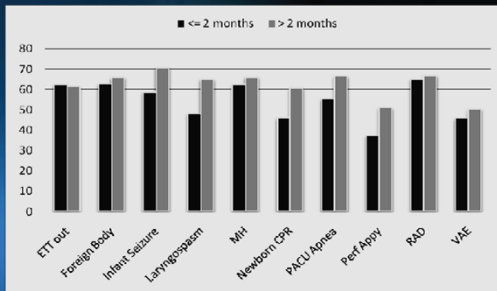
William B. Waldrop MD, David J. Murray MD, John R. Boulet, Joseph F. Kras MD, Anesth Analg (2009)



The Simulation Centers  
at Washington University School of Medicine

## Simulation-Based Assessment of Pediatric Anesthesia Skills

James J. Fehr, MD, FAAP<sup>1,2</sup>, John R. Boulet PhD<sup>3</sup>, William B. Waldrop MD<sup>3,4</sup>, Rebecca Snider, MHS, RN<sup>1</sup>, Megan Brockel, MD<sup>1</sup>, David J. Murray MD<sup>1</sup>  
Anesthesiology 2011; 115: 1308-15.



The Simulation Centers  
at Washington University School of Medicine

Scenario Score Analyses: Difficulty and Discrimination

Scenario	% Key Actions	Discrimination**
Bronchospasm	96.0 ± 10.7	0.26
Anaphylaxis	73.2 ± 21.7	0.3
Ventricular Tachycardia	71.4 ± 17.7	0.51
Myocardial Ischemia	56.5 ± 22.3	0.5
Right Main	91.1 ± 14.1	0.21
Pneumothorax	87.3 ± 19.1	0.34
MH	22.0 ± 34.0	0.39
Tube Obstruction	66.3 ± 25.5	0.33
Total Spinal	49.6 ± 20.0	0.07
Oxygen Failure	90.5 ± 18.5	0.32
Hyperkalemia	29.5 ± 40.0	0.6
Acute Hemorrhage	71.8 ± 35.6	0.55

## Simulation as a method to develop prospective practice guidelines

- Henriksen K, Kaplan H. Hindsight bias, outcome knowledge and adaptive learning Qual Saf Health Care 2003;12(Suppl II):ii46–ii50
  - 'knowledge of the outcome of an adverse event influences their thinking and assessment of the event.'
  - 'fondness for retrospective techniques (such as morbidity and mortality conferences, malpractice claims analysis, error reporting systems, chart reviews) for investigating adverse events that harm patients.'

The Simulation Centers  
at Washington University School of Medicine

## Quality and Safety Indicators in Anesthesia

Anesthesiology 2009; 110:1199-75.  
Guy Haller, MD, M.Sc., Ph.D.,<sup>1</sup> Johannes Stoelwinder, M.B.B.S., M.D., F.R.C.A.M.A., F.A.C.H.S.E., F.F.P.H.M.,<sup>1</sup>  
Paul S. Myles, M.B.B.S., M.P.H., M.D., F.C.A.R.C.S.I., F.A.N.Z.C.A.,<sup>1</sup> John McNeil, M.B.B.S., Ph.D., F.R.A.C.P.

The majority of the clinical indicators identified in order to be directed to physician performance relied on follow-up expert opinion.

## Quality of Care, Performance Assessment and Expert Opinion

- Caplan RA, Posner KL, Cheney FW, et al. Effect of outcome of physician judgments of appropriateness of care. JAMA 1991; 265:1957–60.
  - 112 anesthesiologists were presented with identical clinical scenarios but differing outcomes.
    - Reviewers were more likely to judge anesthesia care as appropriate if the injury was temporary; conversely, reviewers were more likely to judge anesthesia care as substandard or impossible to judge if the injury was permanent.

## Quality of Care, Performance Assessment and Expert Opinion

- Posner KL, Caplan RA, Cheney FW. Variation in expert opinion in medical malpractice review. Anesthesiology 1996; 85:1049-1054.
  - Was the anesthesia care appropriate, less than appropriate, or impossible to judge?
  - Thirty anesthesiologists reviewed 103 claims. Reviewers agreed on 62% of claims and disagreed on 38%.
  - They agreed that care was appropriate in 27% and less than appropriate in 32%.

## Simulation-derived algorithm: A better method to achieve a performance consensus

- Expectations based on overall goals and considerations rather than avoiding a single retrospective adverse outcome
- A prospective simulation approach develops an algorithm with attention to timing and sequence of each step in a 'real time' simulation environment.
- If experts are required to participate in managing the algorithm
  - Inordinately high performance expectations can be modulated
  - Scoring based on what expert actually 'does'

The Simulation Centers  
at Washington University School of Medicine

## The relationship between competence and performance: implications for assessing practice performance.

Rethans JJ, Norcini JJ, Baro n-Maldonado M, Blackmore D, Jolly BC, LaDuca T, Lew S, Page GG, Southgate LH. Med Educ 2002;36:901–9

Performance (what we do in clinical practice) is a product of competence (what we demonstrate under controlled conditions), systems factors (resources, organizational factors), and individual (fatigue, mood) factors.

The Simulation Centers  
at Washington University School of Medicine

### Performance Assessment: Requirements and Evidence to Support the Use of Simulation

- Evidence of Simulation's Efficacy in Patient Safety
  - Psychomotor Skills, Simulation Drills
- Simulation: Patient Safety and Medical Misadventure (Error)
  - Cognitive Skills, Judgment, Communication and Teamwork Skill
- Practical Simulation Assessment
  - Scenario, Scenario Content, Design and Number
  - Scoring and Performance Assessment
  - Reliability and Validity
  - Setting Performance Standards

The Simulation Centers  
at Washington University School of Medicine