Credentialing with Simulation

PP Chen

COS, Department Anaesthesiology & Operating Services

North District Hospital & Alice Ho Miu Ling Nethersole Hospital

Director, NTE Simulation & Training Centre

Outline

- Why do we need simulation-based assessment for credentialing?
- Who are already using simulation-based assessment?
- Is this method valid and reliable?
- Limitations and barriers

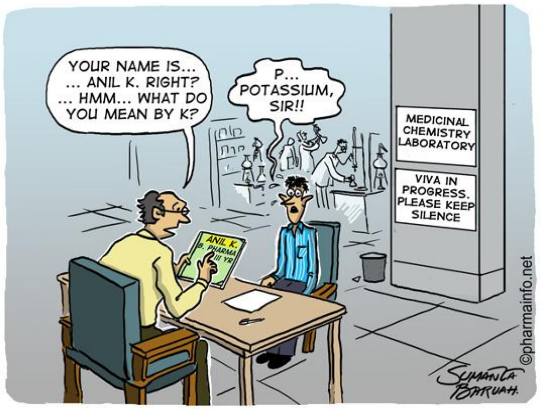
Definitions

Credentialing = Certification & Licensure

Indicates a healthcare professional is competent and able to practise independently

How we are judged to be competent?

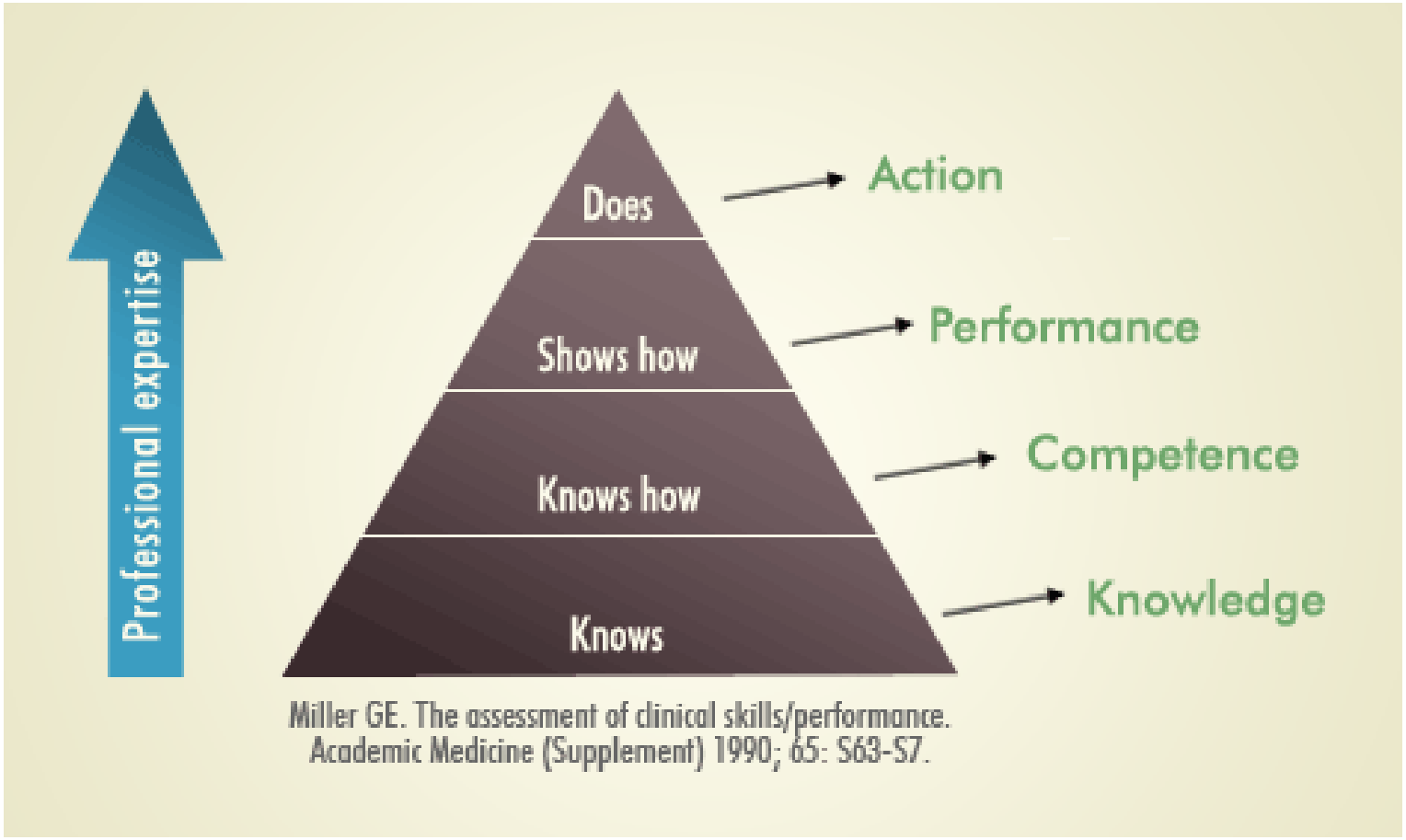
- Written examination
- Oral examination
- Clinical case evaluation
- Logged cases



Knowledge-based

Focused on clinical skills & management on real patient

Miller's Pyramid



Miller GE. The assessment of clinical skills/performance. Academic Medicine (Supplement) 1990; 65: S63-S7.

Anaesth Intensive Care 2010; 38: 876-882

A prospective comparison between written examination and either simulation-based or oral viva examination of intensive care trainees' procedural skills

L. NUNNINK*, B. VENKATESH†, A. KRISHNAN‡, K. VIDHANI§, A. UDY**

Intensive Care Unit, Princess Alexandra Hospital, Brisbane, Queensland, Australia

SUMMARY

We compared results of written assessment of intensive care trainees' procedural skills with results obtained from one of two live assessment formats for the purposes of assessing the concurrent validity of the different test methods. Forty-five Australasian senior trainees in intensive care medicine completed a written test relating to a procedural skill, as well as either a simulation format or oral viva assessment on the same procedural skill.

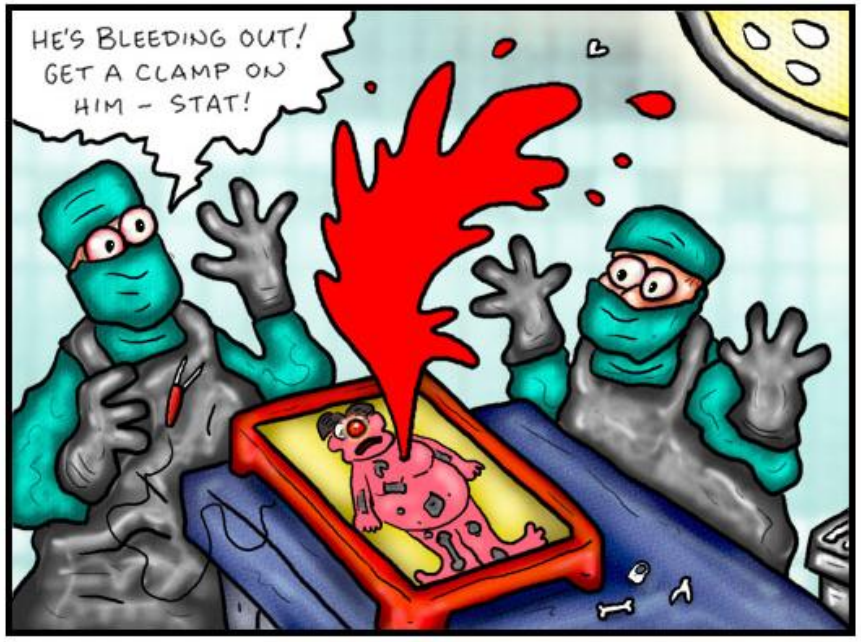
We analysed correlation between written exam results and results obtained from simulation format or oral viva assessment. For those who completed the simulation format examination, we also maintained a narrative of actions and identified critical errors.

There was limited correlation between written exam results and live (simulation or viva) procedure station results ($r=0.31$). Correlation with written exam results was very low for simulation format assessments ($r=0.08$) but moderate for oral viva format assessment ($r=0.58$). Participants who passed a written exam based on management of a blocked tracheostomy scenario performed a number of dangerous errors when managing a simulated patient in that scenario.

The lack of correlation between exam formats supports multi-modal assessment, as currently it is not known which format best represents workplace performance. Correlation between written and oral viva results may indicate redundancy between those test formats, whereas limited correlation between simulation and written exams may support the use of both formats as part of an integrated assessment strategy. We hypothesise that identification of critical candidate errors in a simulation format exam that were not exposed in a written exam may indicate better predictive validity for simulation format examination of procedural skills.

Key Words: intensive care medicine, examination, oral viva, written, simulation

Can simulation enhance the assessment for clinical competence?



What's new in medical school simulation?

Simulation

- Offers the potential to teach & assess complex skills
- Skills can be practised repeatedly until competence is achieved
- Learn from error without harm to patient
- Uncommon procedures and events can be practiced
- Ability to vary difficulty & complexity
- Allows training in teamwork, communication skills, leadership, decision making, resource management

2011; 33: 364–369 MEDICAL TEACHER

Technology-enabled assessment of health professions education: Consensus statement and recommendations from the Ottawa 2010 conference

ZUBAIR AMIN[1], JOHN R. BOULET[2], DAVID A. COOK[3], RACHEL ELLAWAY[4], AHMAD FAHAL[5], ROGER KNEEBONE[6], MOIRA MALEY[7], DORIS OSTERGAARD[8], GOMINDA PONNAMPERUMA[9], ANDY WEARN[10] & AMITAI ZIV[11]

[1]National University of Singapore, Singapore, [2]Education Commission for Foreign Medical Graduates, USA, [3]Mayo Clinic College of Medicine, USA, [4]Northern Ontario School of Medicine, Canada, [5]University of Khartoum, Sudan, [6]Imperial College, UK, [7]The University of Western Australia, Australia, [8]Copenhagen University, Denmark, [9]University of Colombo, Sri Lanka, [10]University of Auckland, New Zealand, [11]Israel Centre for Medical Simulation, Israel

Abstract

The uptake of information and communication technologies (ICTs) in health professions education can have far-reaching consequences on assessment. The medical education community still needs to develop a deeper understanding of how technology can underpin and extend assessment practices. This article was developed by the 2010 Ottawa Conference Consensus Group on technology-enabled assessment to guide practitioners and researchers working in this area. This article highlights the changing nature of ICTs in assessment, the importance of aligning technology-enabled assessment with local context and needs, the need for better evidence to support use of technologies in health profession education assessment, and a number of challenges, particularly validity threats, that need to be addressed while incorporating technology in assessment. Our recommendations are intended for all practitioners across health professional education. Recommendations include adhering to principles of good assessment, the need for developing coherent institutional policy, using technologies to broaden the competencies to be assessed, linking patient-outcome data to assessment of practitioner performance, and capitalizing on technologies for the management of the entire life-cycle of assessment.

Formative vs Summative

- Formative assessment
  - Primarily provide feedback of strength and weakness

- Summative assessment
  - Part of the process to determine competence and readiness to practice independently
  - High-stake assessment
  - Credentialing

Boulet JR. Acad Emerg Med 2008, Boulet JR et al. Simul Healthc 2009

Some examples of simulation-based high stake assessment

- USMLE Step 2 and Step 3 assessments
- ECFMG's SP-based clinical skills assessment for assessment of foreign doctors
- Fundamental of Laparoscopic surgery (ABS)
- Israeli National Board Examination in Anaesthesiology
- Nursing licensure exam in Canada, Israel
- MOC Part-4 ABA, ABIM, ABFM
- ANZCA, RCA, HKCA for CPD required for recertification
- HKCA Simulation-based OSCE station in Final exam

Holmboe E, et al. Sim Healthcare 2011, Steadman RH, et al. Best Pract & Res Clin Anaesth 2012, Ziv A, et al. Anesthesiology Clinic 2007

How resilient is clinical simulation in high stake assessment?

Is this method valid, reliable, feasible?

Challenges

- Face validity – does the tool measure the intended skill or behaviour?
- Construct validity – is the tool able to differentiate among different levels of skills and competence?
- Transferability – does the tool measure skills & performance that translate to actual clinical setting?
- Is the assessment reliable and consistent between raters, across different scenarios and testing centres?
- Is this type of assessment feasible large scale?

Boulet JR. Acad Emerg Med 2008, Boulet JR, Murray DJ. Anesthesiology 2010

ADULT CARDIAC

The Joint Council on Thoracic Surgery Education Coronary Artery Assessment Tool Has High Interrater Reliability

Richard Lee, MD, MBA, Daniel Enter, MD, Xiaoying Lou, BS, Richard H. Feins, MD, George L. Hicks, MD, Mario Gasparri, MD, Hiroo Takayama, MD, PhD, J. Nilas Young, MD, John H. Calhoon, MD, Fred A. Crawford, MD, Nahush A. Mokadam, MD, and James I. Fann, MD

Saint Louis University, St. Louis, Missouri; Northwestern University, Chicago, Illinois; University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; University of Rochester, Rochester, New York; Medical College of Wisconsin Affiliated Hospitals, Milwaukee, Wisconsin; New York Presbyterian Hospital-Columbia, New York, New York; University of California-Davis, Davis, California; University of Texas Health Science Center at San Antonio, San Antonio, Texas; Medical University of South Carolina, Charleston, South Carolina; University of Washington, Seattle, Washington; and Stanford University, Stanford, California

Background. Barriers to incorporation of simulation in cardiothoracic surgery training include lack of standardized, validated objective assessment tools. Our aim was to measure interrater reliability and internal consistency reliability of a coronary anastomosis assessment tool created by the Joint Council on Thoracic Surgery Education.

Methods. Ten attending surgeons from different cardiothoracic residency programs evaluated nine video recordings of 5 individuals (1 medical student, 1 resident, 1 fellow, 2 attendings) performing coronary anastomoses on two simulation models, including synthetic graft task station (low fidelity) and porcine explant (high fidelity), as well as in the operative setting. All raters, blinded to operator identity, scored 13 assessment items on a 1 to 5 (low to high) scale. Each performance also received an overall pass/fail determination. Interrater reliability and internal consistency were assessed as intraclass correlation coefficients and Cronbach's α, respectively.

Results. Both interrater reliability and internal consistency were high for all three models (intraclass correlation coefficients = 0.98, 0.99, and 0.94, and Cronbach's α = 0.99, 0.98, and 0.97 for low fidelity, high fidelity, and operative setting, respectively). Interrater reliability for overall pass/fail determination using κ were 0.54, 0.86, 0.15 for low fidelity, high fidelity, and operative setting, respectively.

Conclusions. Even without instruction on the assessment tool, experienced surgeons achieved high interrater reliability. Future resident training and evaluation may benefit from utilization of this tool for formative feedback in the simulated and operative environments. However, summative assessment in the operative setting will require further standardization and anchoring.

(Ann Thorac Surg 2013;95:2064–70)

Joint Council on Thoracic Surgery Education Tool

- 10 attending surgeons evaluated nine videos of 5 individuals (1 med stu, 1 resident, 1 fellow, 2 attendings)
- Performing coronary anastomoses on a synthetic graft task station (low fidelity), a porcine explant (high fidelity) and in the operative setting
- JCTSE tool for coronary anastomosis consists of 13 items
  - Arteriotomy, graft orientation, bite space, needle holder use, use of forceps, needle angles, needle transfer, suture mgt, know tying, hand mechanics, use of both hands, economy of time and motion

Lee R, et al. Annals of Thoracic Surgery 2013

Internal Consistency & Interrater reliability

Variable	Low Fidelity	High Fidelity	OR
Interrater reliability composite score	0.98	0.99	0.94
Individual item			
Arteriotomy	0.93	0.67	0.51
Graft orientation	0.93	0.84	0.56
Bite	0.84	0.93	0.88
Spacing	0.85	0.76	0.74
Needle holder use	0.96	0.93	0.60
Use of forceps	0.80	0.89	0.70
Needle angles	0.95	0.75	0.63
Needle transfer	0.98	0.85	0.70
Suture management	0.85	0.85	0.68
Knot tying	0.91	1.00	0.53
Hand mechanics	0.98	0.88	0.74
Use of both hands	0.93	0.88	0.58
Economy of time	0.96	0.86	0.62
Overall pass/fail	0.54	0.86	0.15
Internal consistency	0.99	0.98	0.97

OR = operating room.

Lee R, et al. Annals of Thoracic Surgery 2013

British Journal of Anaesthesia **90** (**1**): 43–7 (2003)
DOI: 10.1093/bja/aeg002

BJA

Evaluation of high fidelity patient simulator in assessment of performance of anaesthetists

J. M. Weller[1]*, M. Bloch[2], S. Young[3], M. Maze[2], S. Oyesola[2], J. Wyner[2], D. Dob[2], K. Haire[2], J. Durbridge[2], T. Walker[2] and D. Newble[4]

[1]Department of Surgery, Wellington School of Medicine, Otago University, Private Bag 7343, Wellington South, New Zealand. [2]Magill Department of Anaesthetics, Chelsea and Westminster Hospital, London, UK. [3]Department of Clinical Engineering, Charing Cross Hospital, London, UK. [4]Department of Medical Education, Northern General Hospital, Sheffield, UK

**Corresponding author. E-mail: jennifer.weller@wnmeds.ac.nz*

Background. There is increasing emphasis on performance-based assessment of clinical competence. The High Fidelity Patient Simulator (HPS) may be useful for assessment of clinical practice in anaesthesia, but needs formal evaluation of validity, reliability, feasibility and effect on learning. We set out to assess the reliability of a global rating scale for scoring simulator performance in crisis management.

Methods. Using a global rating scale, three judges independently rated videotapes of anaesthetists in simulated crises in the operating theatre. Five anaesthetists then independently rated subsets of these videotapes.

Results. There was good agreement between raters for medical management, behavioural attributes and overall performance. Agreement was high for both the initial judges and the five additional raters.

Conclusions. Using a global scale to assess simulator performance, we found good inter-rater reliability for scoring performance in a crisis. We estimate that two judges should provide a reliable assessment. High fidelity simulation should be studied further for assessing clinical performance.

Br J Anaesth 2003; **90**: 43–7

Keywords: anaesthetists, clinical competence; computers, computer simulation; education, educational measurement

Accepted for publication: June 25, 2002

Evaluation of high-fidelity HPS in the performance of anaesthetists

- 3 primary raters rated 28 ACRM video recordings
  - anaphylaxis, cardiac arrest, O2 fail, MH
- 5 additional raters rated btw 5-17 of the tapes
- Global scoring for 3 categories – knowledge (5 items), behaviours (10), overall scores

Weller JM, et al. BJA 2003

High reliability of assessment of overall performance, behaviour and knowledge when 2 or 3 raters scored

Intraclass Correlation Coeff	**Overall performance**	**Knowledge**	**Behaviour**
ICC for one rater	0.65	0.60	0.54
ICC for two raters	0.79	0.75	0.70
ICC for three raters	0.85	0.82	0.78

High tendency for raters to score similarly, n=28

Qualities	**Spearman's Rank Correlation Coeff** ρ
Overall performance *vs* behaviour	0.89
Overall performance *vs* knowledge	0.95
Behaviour *vs* knowledge	0.88

CHEST

Original Research

SIMULATION-BASED TRAINING

Simulation-Based Objective Assessment Discerns Clinical Proficiency in Central Line Placement

A Construct Validation

Yue Dong, MD; Harpreet S. Suri, MBBS; David A. Cook, MD, MHPE; Kianoush B. Kashani, MD; John J. Mullon, MD; Felicity T. Enders, PhD; Orit Rubin, PhD; Amitai Ziv, MD; and William F. Dunn, MD, FCCP

Background: **Central venous catheterization (CVC) is associated with patient risks known to be inversely related to clinician experience. We developed and evaluated a performance assessment tool for use in a simulation-based central line workshop. We hypothesized that instrument scores would discriminate between less experienced and more experienced clinicians.**

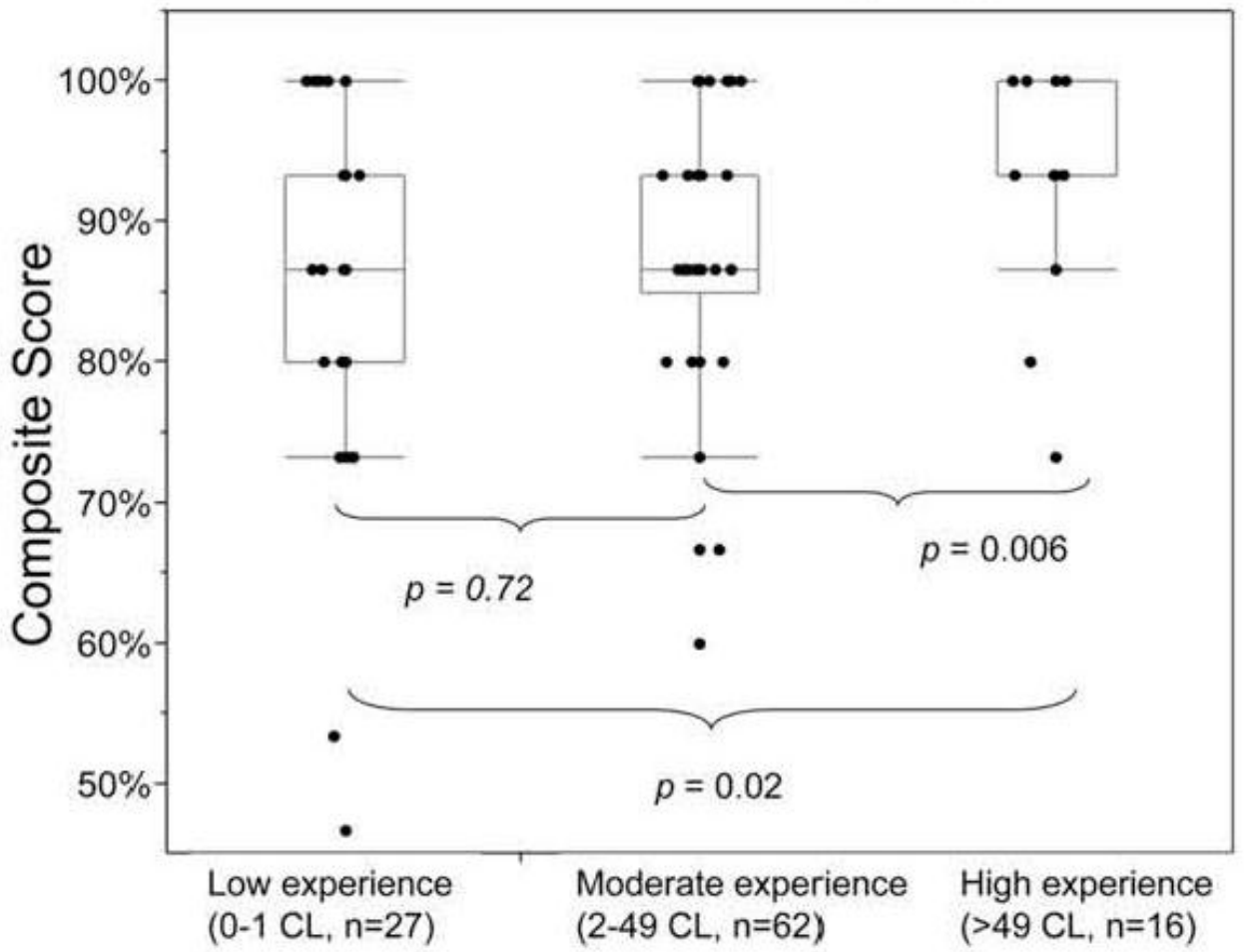
Methods: **Participants included trainees enrolled in an institutionally mandated CVC workshop and a convenience sample of faculty attending physicians. The workshop integrated several experiential learning techniques, including practice on cadavers and part-task trainers. A group of clinical and education experts developed a 15-point CVC Proficiency Scale using national and institutional guidelines. After the workshop, participants completed a certification exercise in which they independently performed a CVC in a part-task trainer. Two authors reviewed videotapes of the certification exercise to rate performance using the CVC Proficiency Scale. Participants were grouped by self-reported CVC experience.**

Results: **One hundred and five participants (92 trainees and 13 attending physicians) participated. Interrater reliability on a subset of 40 videos was 0.71, and Cronbach α was 0.81. The CVC Proficiency Scale Composite score varied significantly by experience: mean of 85%, median of 87% (range 47%-100%) for low experience (0-1 CVCs in the last 2 years, n = 27); mean of 88%, median of 87% (range 60%-100%) for moderate experience (2-49 CVCs, n = 62); and mean of 94%, median of 93% (range 73%-100%) for high experience (>49 CVCs, n = 16) (*P* = .02, comparing low and high experience).**

Conclusions: **Evidence from multiple sources, including appropriate content, high interrater and internal consistency reliability, and confirmation of hypothesized relations to other variables, supports the validity of using scores from this 15-item scale for assessing trainee proficiency following a central line workshop.**

CHEST 2010; 137(5):1050–1056

Abbreviations: CVC = central venous catheterization; IJ = internal jugular; SC = subclavian



Dong Y et al. Chest 2010

A Technical and Cognitive Skills Evaluation of Performance in Interventional Cardiology Procedures Using Medical Simulation

Rebecca S. Lipner, PhD;
John C. Messenger, MD, FACC;
Roberta Kangilaski, BA;
Donald S. Baim, MD†;
David R. Holmes, Jr., MD;
David O. Williams, MD;
Spencer B. King, III, MD

Introduction: Interventional cardiology, with large numbers of complex procedures and potentially serious complications, stands out as an obvious discipline in which to apply simulation to help prevent medical errors. The objective of the study was to determine whether it is feasible to develop a valid and reliable evaluation approach using medical simulation to assess technical and cognitive skills of physicians performing coronary interventions.

Methods: Clinical case scenarios were developed by a committee of subject matter experts, who defined key decision nodes, such as stent positioning, and introduced unanticipated complications, such as coronary perforation. Subjects were 115 physicians from 10 U.S. healthcare institutions at three levels of expertise: novice, skilled, or expert. Subjects completed a questionnaire, one practice case and six test cases on a SimSuite simulator (Medical Simulation Corporation, Denver, CO), and an opinion survey. Clinical specialists rated subjects' procedural skills.

Results: A technical and cognitive skills evaluation of performance in interventional cardiology procedures using medical simulation yielded results that distinguished between a novice group and skilled or expert groups ($P < 0.001$) and scores correlated moderately with clinical specialist ratings of subjects' procedural skills and with number and complexity of procedures performed in practice during the previous year. Approximately 90% of subjects generally thought that the cases were well simulated and presented situations encountered in practice.

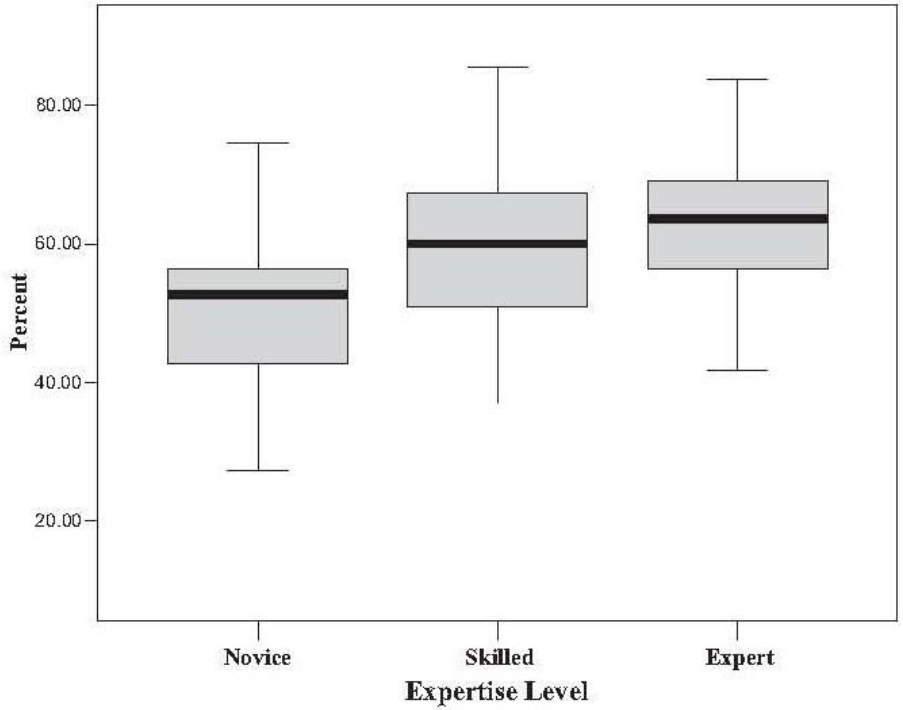
Conclusions: This study suggests that an evaluation approach using high-fidelity medical simulation to assess technical and cognitive skills of physicians performing interventional cardiology procedures can be used to identify physicians who are extremely poor performers and not likely to be providing appropriate patient care. We believe that use of a high-fidelity simulator incorporating situations with multiple events, immediate feedback, and high sensory load would complement the results of traditional written examinations of medical knowledge to provide a more comprehensive assessment of physician ability in interventional cardiology.

(*Sim Healthcare* 5:65–74, 2010)

Key Words: Medical simulation, Medical assessment, Medical evaluation, High-fidelity simulation, Medical education, Educational testing and measurement.

Novice – first 3 months of 2nd year of cardiology training
Skilled – in practice < 2 years doing 75-100 procedures a year
Experts – in practice > 3 years doing > 100 procedures a year, and a min career of >1000 procedures

- Novice (35), skilled (38) or experts (42) performed one practice and 6 test cases on a SimSuite Endovascular simulator



Lipner RS, et al. *Simul Healthc* 2010

Validity of generation Olympus colonscopy simulator

- 10 novices, 13 intermediate trainees, 11 experienced (>1000 procedures) endoscopists
- Complete 3 cases, 15 min each
  - Sigmoid N-loop + mod transverse loop, with low pain threshold, sigmoid alpha loop + mod transverse loop
- Completion rates (37 vs 79 vs 88%)
- Both novices and trainees took significantly longer to reach all landmarks
- Some technical aspects discriminatory

Haycock AV et al. Endoscopy 2009

Validity of the Voxelman TempoSurg Virtual Reality (VR) temporal bone simulator

- 3 groups – novices, intermediate, experienced surg
- Experts and intermediates outperformed novices
  - with respect to the total time taken, total volume of bone removed, efficiency of bone, time spent with the drill tip obscured and number of injuries to the sigmoid sinus
- Simulator-generated objective metrics can be used to differentiate individuals of differing levels of experience using a standardized temporal bone task

Khemani S, et al Otology & Neurotology 2012

Limitations

- Psychometrics
  - predictive validity
  - overreliance on psychometric criteria that can lead to measures (eg, checklists) that may fail to capture the complexities involved in healthcare
  - Lacks validity - especially in maintenance of licensure and certification where little evidence exists

Holmboe E et al. Sim Healthcare 2011

Limitations

- Costs and logistics
- Patient Simulator
  - sweating, skin color, response to pain not modeled well
  - interrelationships between physiology variables imperfect
- Scenario
  - improperly scripted or modeled scenario
  - complex scenario
  - how many is optimal
- Environment
  - realism

Quality Assurance

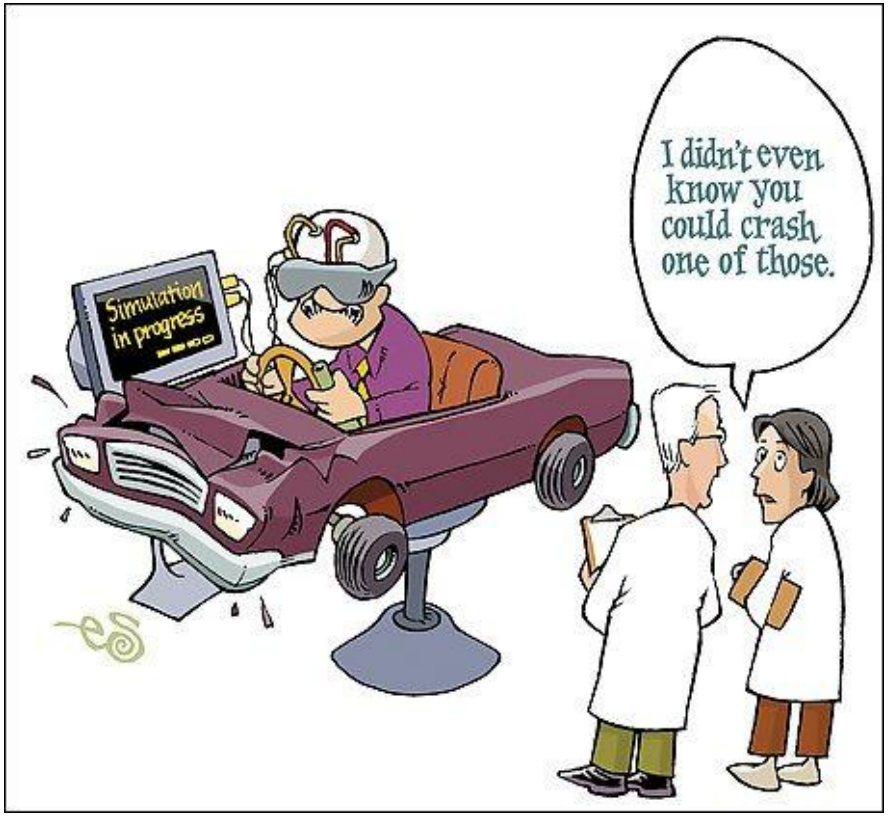
- Lack of standards in simulation in healthcare
  - Simulators
    - Reproducibility
    - Consistency
    - Regularly reviewed and updated
  - Authenticity of the environment & equipment
  - SP training
  - Curriculum
  - Faculty development
  - Methodology of assessment

Boulet JR, et al. Adv Health Sci Educ Theory Pract 2003,
Cumin D, et al. BJA 2010

The way forward

- Simulation holds great promise for high-stakes assessment esp with advancing technology [1,2]
- A part of a multimodal assessment programme inclu assessment activities involving direct patient contact
- Greater amount of experience and research will enhance its validity, reliability and feasibility for credentialing & other high-stakes assessments
- Future direction in setting standards in simulation in healthcare will allow benchmarking, high reliability & confidence in equipment, personnel & processes

1. Gaba DM. Simul Healthc 2009, 2. Boulet JR. Acad Emerg Med 2008



Thank you!

Email: chenpp@ha.org.hk