

Components of Critical Decision Making and ABSITE Assessment: Toward a More Comprehensive Evaluation

SATISH KRISHNAMURTHY, MD, MCh
 USHA SATISH, PhD
 TINA FOSTER, MD, MPH, MS
 SIEGFRIED STREUFERT, PhD
 MANTOSH DEWAN, MD
 THOMAS KRUMMEL, MD

Abstract

Rationale Accurate assessment of resident competency is a fundamental requisite to assure the training of physicians is adequate. In surgical disciplines, structured tests as well as ongoing evaluation by faculty are used for evaluating resident competency. Although structured tests evaluate content knowledge, faculty ratings are a better measure of how that knowledge is applied to real-world problems. In this study, we sought to explore the performance of surgical residents in a simulation exercise (strategic management simulations [SMS]) as an objective surrogate of real-world performance.

Methods Forty surgical residents participated in the SMS simulation that entailed decision making in a real-world—oriented task situation. The task requirements enable the assessment of decision making along several parameters of thinking under both crisis and noncrisis situations. Performance attributes include “simpler” measures of competency (activity level), intermediate categories (information management and emergency responses) to complex measures (breadth of approach

and strategy). Scores obtained in the SMS were compared with the scores obtained on the American Board of Surgery In-Training Examination (ABSITE).

Results The data were intercorrelated and subjected to a multiple regression analysis with ABSITE as the dependent variable and simulation scores as independent variables. Using a 1-tail test analysis, only 3 simulation variables correlated with performance on ABSITE at the .01 level (ie, basic activity, focused activity, task orientation). Other simulation variables showed no meaningful relationships to ABSITE scores at all.

Conclusions The more complex real-world—oriented decision-making parameters on SMS did not correlate with ABSITE scores. We believe that techniques such as the SMS, which focus on critical thinking, complement assessment of medical knowledge using ABSITE. The SMS technique provides an accurate measure of real-world performance and provides objective validation of faculty ratings.

Introduction

Accurate assessment of resident competency is a fundamental requisite to assure the public that the training of physicians is adequate. Knowledge about how to best develop and assess physician competency continues to grow. Procedural specialties require the assessment of procedural as well as cognitive skills. In surgical disciplines, structured written tests evaluate content knowledge (eg, the American Board of Surgery In-training Examination [ABSITE]).^{1,2}

Satish Krishnamurthy, MD, MCh, is Director, Minimally Invasive Neurosurgery, Department of Neurosurgery, SUNY Upstate Medical University; **Usha Satish, PhD**, and **Siegfried Streufert, PhD**, are Professors, and **Mantosh Dewan, MD**, is Professor and Chair, all in the Department of Psychiatry and Behavioral Sciences, SUNY Upstate Medical University; **Tina Foster, MD, MPH, MS**, is Assistant Professor of Obstetrics and Gynecology and Community and Family Medicine, Dartmouth-Hitchcock Medical Center, and Associate Program Director of the Dartmouth-Hitchcock Leadership Preventive Medicine Residency. **Thomas Krummel, MD**, is Professor and Chair, Department of Surgery, Stanford School of Medicine.

Corresponding author: Usha Satish, PhD, SUNY Upstate Medical University, 750 East Adams Street, Syracuse, NY 13210, 315.464.3114, satishu@upstate.edu

DOI: 10.4300/JGME-D-09-00034.1

Ongoing faculty ratings of resident behaviors during patient care provide documentation on how residents combine content knowledge, skills, and attitudes during real-world challenges. Structured tests measure content knowledge and have limited ability to measure how well the resident uses this knowledge. Faculty ratings, on the other hand, are a better measure of the residents' ability to apply knowledge in a timely fashion, yet they are often subject to bias (positive or negative).³ Additional techniques are beginning to emphasize different variables of performance, including the need for focus on organizational factors,⁴ emergency management,⁵ and improvement of performance through structured observations.⁶ Tools such as the Oxford Non-Technical Skills (NOTECHS) scale have emerged as a reliable and valid instrument for teamwork analysis.⁷

Meaningful assessments of clinical performance and the potential for clinical performance are important to resident selection and evaluation. Brothers and Wetherhalt⁸ found that faculty evaluations of personal characteristics and letters of reference at the time of interview were likely to predict subsequent clinical performance. United States

Medical Licensing Examination scores and academic grade performance were predictive of subsequent formalized testing (ABSITE) but they were poor predictors of clinical performance. Superior faculty rating of resident performance was correlated with better rating on communication and professionalism competencies (residents who are able to communicate better and work in a team),³ but ratings on communication and professionalism competencies did not correlate with ABSITE scores.³ ABSITE scores also had a low correlation with technical skills or operative performance in surgical residents.⁹

Faculty ratings appear to judge the resident performance (in and out of the operating room) better than the ABSITE scores. Robust approaches for global evaluation tools by faculty can use a standard set of questions addressing specific competency-based behaviors and/or proficiencies, with standardization of results between those being rated through the use of a Likert scale.¹⁰ Still, a major concern in the use of faculty evaluation is the potential for assessment bias. A “halo effect” may result from evaluators being unduly influenced by resident affability and availability. This concern is echoed in the tradition of the “3 As of private practice”: affability, availability, and ability, which are the factors most important to a referring physician when selecting a consultant in descending order of priority.³ Physicians are frequently unaware of their own bias toward other physicians, and it has been shown that referral patterns appear to be based on relationships, not physician skill.¹¹ On the other hand, 1 negative experience may establish the perception on the part of faculty that a resident is below average. All these concerns with faculty ratings could be minimized with an objective surrogate for a real-world performance of the surgical resident. This can potentially be achieved using techniques such as the strategic management simulation (SMS) described in this article.

Simulation has become increasingly important in graduate and undergraduate medical education in order to avoid observer bias. A variety of techniques using simulation for both needs assessment and evaluation have been developed. Procedural simulations (both low and high-tech), use of standardized patients and Objective Structured Clinical Examinations, and simulations of team performance and emergency situations are a few of the more commonly described techniques.^{12–16} However, capturing performance and providing useful, structured feedback can be challenging in some simulated situations. Just as we sometimes struggle to teach leadership and decision-making skills in the clinical setting, we face the same challenges in a simulated environment.

Patient care at the frontlines occurs under “VUCAD” conditions: volatility, uncertainty, complexity, ambiguity, and delayed feedback.^{17,18} Although many commonly used procedural simulations can recreate such conditions, they do not necessarily offer defined ways to evaluate

performance in the areas of strategic thinking, prioritizing, decision making, and leadership. Cognitive simulation can do that because it offers another avenue for assessing and promoting the development of competent physicians, especially with regard to areas of decision making and leadership. These simulations are designed to replicate complex realities and can be used to evaluate decision-making and leadership skills in resident physicians.^{19,20}

The technique of strategic management simulation has been widely used in many industries to assess cognitive and behavioral responses to task demands as well as the cognitive and behavioral components of “executive functions.”²¹ Strategic management simulation is a tool that explores *how we think* without assessing the content of the subject matter; in this case, surgery. We have previously reported the use of SMS in the evaluation of surgical residents.^{22,23} The simulation technique was able to accurately assess performance of the surgical residents in an objective manner in a brief period of time. A number of measures obtained on simulation performance (such as activity level, response speed, initiative, adequate usage of and appropriate search for relevant information) were highly correlated with comprehensive faculty assessment (eg, measures of crisis management, team interactions, flexibility of approach). In other words, simulation data were highly similar to faculty ratings that were based on at least 2 years of experience with the resident.

The SMS simulations are designed to assess multiple critical dimensions of decision making under dynamic environmental conditions. The SMS measurement system emphasizes the underlying parameters of thinking that are critical to communication, teamwork, utilization of knowledge, breadth of approach, integration of knowledge with incoming information, use of planning, and strategy.²⁴ The SMS simulations have been used across the world to assess and train professional decision makers.²⁵

In this article, we sought to explore how standardized test scores on ABSITE correlated with performance in SMS in surgery residents. In other words, we were interested in determining the real-world commonalities between ABSITE scores and real-world performance indicators of cognitive simulation (SMS) performance.

Methods

Forty surgical residents from Stanford School of Medicine participated in the SMS simulation. Scores obtained were compared with ABSITE, a test of the general level of knowledge attained by residents.

Strategic Management Simulation

Strategic management simulations assess both cognitive and behavioral responses to task demands. The method provides more than 80 computer-gathered and computer-calculated

TABLE 1 PARAMETERS OF DECISION MAKING

Basic activity level	Overall level of activity (this measure pertains to both relevant goal-directed activity as well as distractions and other activity patterns)
Focused activity level	Level of activity that pertains to a specific tasks at hand
Task orientation	Ability to focus on a task at hand and also focus on larger goals
Basic initiative	Ability to generate activity without an overt external stimulus
Information management	Ability to seek and use information efficaciously
Emergency responses	Ability to think critically and strategically under conditions of emergency and stress
Breadth of Approach	Ability to think along multiple dimensions and find different solutions to problems
Strategy	Ability to form systematic plans and actions that are optimally sequenced and goal-directed in the long term

measures of functioning, loading on 12 reliable factors (based on factor analytic varimax rotation for more than 2000 prior subjects). High levels of predictive validity, reliability, and applicability of the SMS simulations to real-world settings have been repeatedly demonstrated across multiple professions, cultures, and continents (predicting an individual's achievement and future success level on indicators such as "job level at age," "income at age," "promotions," "number of persons supervised," and so forth).²⁶ Overall validity coefficients consistently exceed $r = +.60$. Reliability values range between $r = +.7$ and $r = +.94$.²⁶

During the simulation, participants make decisions during a 90-minute task period that includes responding to scenarios that are specifically designed to assess decision making along multiple parameters.²⁷ The real-world atmosphere of the task and setting, involving multiple potentially interactive components of task demands as well as multiple and interactive options to engage in various aspects of behavior, allows for a more realistic (ecologically relevant) assessment of competency.²⁸ Assessed performance attributes on several validated performance indicators vary from "simpler" measures of competency in categories such as "activity" and "timeliness of response," through intermediate categories such as "information orientation," "information utilization," and "emergency management" to increasingly complex measures in such areas of functioning as "initiative," "breadth of approach to challenges," "planning," "strategy," and so forth (TABLE 1). During the simulation, the task requirements enable the assessment of decision making under both crisis and noncrisis situations along several parameters of thinking. These include levels of activity, task orientation, information management, and strategy among others.

American Board of Surgery In-Training Examination

The ABSITE is a test in clinical and basic science that is designed to serve as an objective measure of resident knowledge to determine competency for progression in the

program. The data were intercorrelated and subjected to a multiple regression analysis with ABSITE as the dependent variable and simulation scores as independent variables.

Results

The intercorrelations of ABSITE scores with simulation scores and their common variance values are presented in TABLE 2. Because we are concerned with potential parallel predictions of ABSITE and SMS simulation scores, a 1-tailed test analysis was used. Only 3 simulation variables correlated with performance on ABSITE at the .01 level (ie, basic activity, focused activity, task orientation). Other simulation variables showed no meaningful relationships to ABSITE scores at all.

To help define the primary contributors to any relationship between ABSITE performance and performance on the simulation, a multiple regression analysis was performed to single out those simulation variables that meaningfully predict the ABSITE score. Using a stepwise model of $F \leq 0.50$ to enter and $\geq .10$ to remove, an overall R^2 (adjusted) value of .827 was obtained for 2 of the simulation variables, basic activity and focused activity, generating an F value of 94.495 ($P < .001$) for the 2-variable model. In other words, the primary simulation predictors for performance on the ABSITE are measures of activity. Other simulation-based variables are less or not at all predictive of ABSITE performance.

Discussion

Assessment of residents' competency in surgery is a complex process. Competence encompasses several components: knowledge of diagnosing and treating disease, appropriate judgment in treating a patient's condition in elective and emergent conditions, knowledge of surgical technique and performing the procedures with technical finesse, empathy and caring for patients, ability to deliver health care within a framework of rules and regulations (local hospital, state, and federal), and the ability to learn from experience

TABLE 2 INTERCORRELATIONS OF ABSITE AND SIMULATION SCORES

Simulation Measure	Correlation	Common Variance, %
Basic activity	.899 ($P < .01$)	80.8
Focused activity	.825 ($P < .01$)	68.1
Task orientation	.837 ($P < .01$)	70.1
Basic initiative	.135 (NS)	1.8
Information management	.038 (NS)	0.1
Emergency responses	.080 (NS)	0.6
Breadth of approach	.136 (NS)	1.8
Strategy	.111 (NS)	1.2

Abbreviation: NS, nonsignificant.

(lifelong learning). Most residency programs use a combination of test scores and faculty evaluations based on performance at the bedside or in the clinic and in the operating room.

In this study, we used SMS to evaluate the surgical resident performance and compared it with ABSITE scores. ABSITE scores correlated significantly with simple measures such as activity level, focused activity, and task orientation. The more complex, real-world–oriented, decision-making parameters did not correlate with the ABSITE scores. These measures included basic initiative, information management, strategy, emergency responses, and breadth of approach. Translating these SMS measures to real-world situations is important to understand the implications of these findings. Initiative to look for information when evaluating or managing a patient's condition is critical in surgery. Similarly, appropriate use of information gathered (information management), considering different surgical and nonsurgical conditions during diagnosis, and considering different options in managing a surgical problem (and nonsurgical such as radiation therapy, medical management including no treatment) are important in optimizing the risk-to-benefit ratio for the patient. Considering various options comes under the rubric of breadth of approach and strategy. Finally, surgical problems can evolve rapidly, and the ability to gather critical pieces of information, come up with a plan of action, and execute the plan in a timely fashion, especially in an emergency.

In conclusion, we believe that techniques such as the SMS, which focus on critical thinking, provide information about resident competency that is different from, and adds to, information generated by ABSITE (and potentially other objective measures across medical specialties). Because it has been repeatedly shown that the SMS technique provides an accurate measure of underlying performance indicators, data obtained by this method would be somewhat similar to

(but because they are unbiased, they can improve on) faculty ratings.

References

- 1 American Board of Surgery In-Training Examination website. Available at: <http://www.absite.org>. Accessed June 21, 2009.
- 2 American Board of Surgery website. Available at: <http://home.absurgery.org/default.jsp?certabsite>. Accessed November 18, 2009.
- 3 Haurani MJ, Rubinfeld I, Rao S, et al. Are the communication and professionalism competencies the new critical values in a resident's global evaluation process? *J Surg Educ*. 2007;64(6):351–356.
- 4 Salas E, Almeida SA, Salisbury M, et al. What are the critical success factors for team training in health care? *Jt Comm J Qual Patient Saf*. 2009;35(8):398–405.
- 5 Solis-Trapala IL, Carthey J, Farewell VT, de Leval MR. Dynamic modeling in a study of surgical error management. *Stat Med*. 2007;26(28):5189–5202.
- 6 Catchpole KR, Giddings AE, Wilkinson M, Hirst G, Dale T, de Leval MR. Improving patient safety by identifying latent failures in successful operations. *Surgery*. 2007;142(1):102–110.
- 7 Mishra A, Catchpole K, McCulloch P. The Oxford NOTECHS system: reliability and validity of a tool for measuring teamwork behavior in the operating theater. *Qual Saf Health Care*. 2009;18(2):104–108.
- 8 Brothers TE, Wetherhalt S. Importance of faculty interview during the resident application process. *J Surg Educ*. 2007;64(6):378–385.
- 9 Scott DJ, Valentine RJ, Bergen PC, et al. Evaluating surgical competency with the American Board of Surgery In-Training Examination, skill testing, and intraoperative assessment. *Surgery*. 2000;6:613–622.
- 10 Likert R. A technique for the measurement of attitudes. *Arch Psych*. 1932;140:5–53.
- 11 Shortell SM. Patterns of referral among internists in private practice: a social exchange model. *J Health Soc Behav*. 1973;14(4):335–348.
- 12 Mehrabi A, Gluckstein C, Benner A, et al. A new way for surgical education: development and evaluation of a computer-based training module. *Comput Biol Med*. 2000;30:97–109.
- 13 Rosser J. CD-ROM multimedia: the step before virtual reality. *Surg Endosc*. 1996;10:1033–1035.
- 14 Trelease RB. The virtual anatomy practical: a stereoscopic 3D interactive multimedia computer examination program. *Clin Anat*. 1998;11:89–94.
- 15 Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. *BMJ*. 1975;1:447–451.
- 16 Reznick M, Harter P, Krummel T. Virtual reality and simulation: training the future emergency physician. *Acad Emerg Med*. 2002; 9(1):78–87.
- 17 Satish U, Streufert S. Value of a cognitive simulation in medicine: towards optimizing decision making performance of health care personnel. *Qual Saf Health Care*. 2002;11:163–167.
- 18 Satish U, Krummel TM, Foster TC, Krishnamurthy S. Using strategic management simulations to evaluate physician competence: a challenge and a vision. *ACGME Bulletin*. 2005;19–21.
- 19 Committee on optimizing graduate medical trainee (resident) hours and work schedules to improve patient safety: improving the resident learning environment. In: Ulmer C, Wolman DM, Johns MME. *Resident Duty Hours*:

- Enhancing Sleep, Supervision, and Safety*. Washington DC: The National Academies Press, 2008.
- 20 Salas E, Wilson KA, Burke CS, Priest HA. Using simulation based training to improve patient safety: what does it take? *Jt Comm J Qual Patient Saf*. 2005;31(7):363–371.
 - 21 Satish U, Streufert S. The measurement of behavioral complexity. *J Appl Soc Psychol*. 1997;27(23):2117–2121.
 - 22 Satish U, Streufert S, Marshall R, et al. Strategic management simulations is a novel way to measure resident competencies. *Am J Surg*. 2001;181(6):557–561.
 - 23 Satish U, Streufert S. Improving strategic performance on multiple dimensions: an analysis via complexity theory. In: *Winning Strategies in a Deconstructing World*. Berlin: Strategic Management Society; 1999.
 - 24 Satish U, Krishnamurthy S. Role of cognitive simulations in healthcare. *Man Simul Healthc*. 2008;477–488.
 - 25 Breuer K, Streufert S. Computergestützte eignungsdiagnostik mit komplexen dynamischen scenarios [Computer-based assessment using complex scenarios]. *Zeitschrift für Arbeits- und Organisationspsychologie*. 1995;39:34–36.
 - 26 Streufert S, Pogash R, Piasecki M. Simulation-based assessment of managerial competence: reliability and validity. *Pers Psychol*. 1988;41: 537–557.
 - 27 Streufert S, Satish U, Barach P. Improving medical care: the use of simulation technology. *Simul Gaming*. 2001;2:164–175.
 - 28 Satish U, Streufert S. The measurement of complexity. *J Health Soc Behav*. 1997;27:2117–2121.