

Simulation-Based Assessment to Evaluate Cognitive Performance in an Anesthesiology Residency Program

AVNER SIDI, MD
TEZCAN OZRAZGAT BASLANTI, PhD
NIKOLAUS GRAVENSTEIN, MD
SAMSUN LAMPOTANG, PhD

Abstract

Background Problem solving in a clinical context requires knowledge and experience, and most traditional examinations for learners do not capture skills that are required in some situations where there is uncertainty about the proper course of action.

Objective We sought to evaluate anesthesiology residents for deficiencies in cognitive performance within and across 3 clinical domains (operating room, trauma, and cardiac resuscitation) using simulation-based assessment.

Methods Individual basic knowledge and cognitive performance in each simulation-based scenario were assessed in 47 residents using a 15- to 29-item scenario-specific checklist. For every scenario and item we calculated group error scenario rate (frequency) and individual (resident) item success. For all analyses, alpha was designated as 0.05.

Results Postgraduate year (PGY)-3 and PGY-4 residents' cognitive items *error rates* were higher and *success rates* lower compared to basic and technical performance in each domain tested ($P < .05$). In the trauma and resuscitation scenarios, the cognitive *error rate* by PGY-4 residents was fairly high (0.29–0.5) and their cognitive *success rate* was low (0.5–0.68). The most common cognitive errors were anchoring, availability bias, premature closure, and confirmation bias.

Conclusions Simulation-based assessment can differentiate between higher-order (cognitive) and lower-order (basic and technical) skills expected of relatively experienced (PGY-3 and PGY-4) anesthesiology residents. Simulation-based assessments can also highlight areas of relative strength and weakness in a resident group, and this information can be used to guide curricular modifications to address deficiencies in tasks requiring higher-order processing and cognition.

Editor's Note: The online version of this article contains a table of ranked cognitive errors, details for how the feasibility and acceptability of the simulation assessment method were established, checklist items and clinical scenarios used in the study, and item descriptions and performance grading.

All authors are at Department of Anesthesiology, College of Medicine, University of Florida. **Avner Sidi, MD**, is Associate Professor of Anesthesiology; **Tezcan Ozrazgat Baslanti, PhD**, is Biostatistician; **Nikolaus Gravenstein, MD**, is Professor of Anesthesiology; and **Samsun Lampotang, PhD**, is Professor of Anesthesiology.

Funding: The authors report no external funding source for this study.

This work was previously presented at the University of Florida Celebration of Research, Gainesville, FL, March 11, 2013.

The authors would like to thank Isaac Luria, Center for Safety, Simulation & Advanced Learning Technologies, Department of Anesthesiology, College of Medicine, University of Florida.

Corresponding author: Avner Sidi, MD, Department of Anesthesiology, University of Florida College of Medicine, PO Box 100254 JHMHSC, 1600 SW Archer Road, Gainesville, FL 32610-0254, 352.413.8073, asidi@anest.ufl.edu

Received June 19, 2013; revision received September 1, 2013; accepted September 23, 2013.

DOI: <http://dx.doi.org/10.4300/JGME-D-13-00230.1>

Introduction

Applying learning theories^{1–3} and competency assessment^{4–6} of nontechnical and technical skills^{7,8} cannot be accomplished using only traditional examinations,^{9–12} including objective structured clinical examinations (OSCEs) and multiple-choice questions. These examinations fail to capture the uncertainty that will be encountered in some clinical scenarios. Problem solving in the operating room requires knowledge and experience.¹⁰ Current evaluations (including simulation-based assessments) typically measure basic knowledge and performance rather than competency in the complex tasks of acute care.¹³ This is why it is important to develop better methods to measure acute care clinical performance. Simulation could be used to measure advanced nontechnical cognitive diagnostic and therapeutic management skills and the ability to integrate knowledge, judgment, communication, and teamwork into the simulated practice setting.

In this article, we follow the definition of anesthesia nontechnical skills (ANTS).^{14–18} These ANTS include task management, teamwork, situation awareness, and decision making. Technical skills are those that are not ANTS: basic

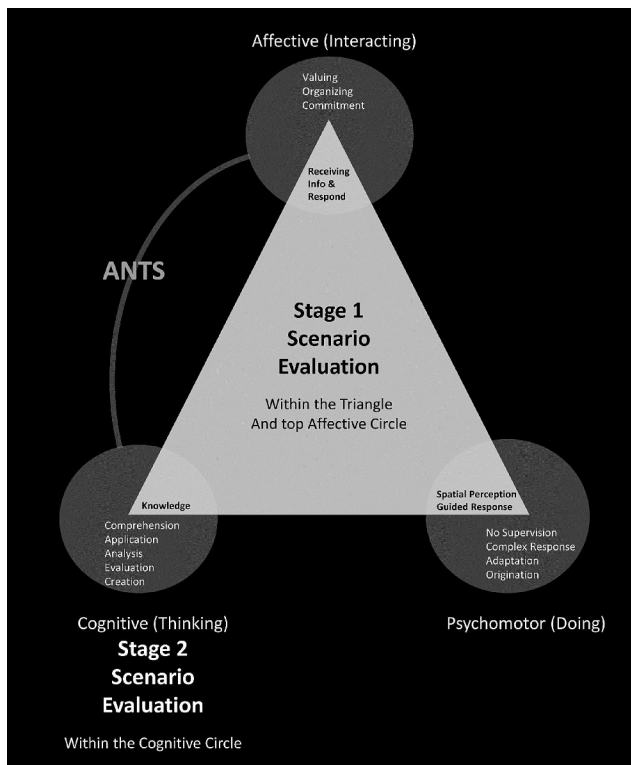


FIGURE 1 | **AFFECTIVE, COGNITIVE, AND PSYCHOMOTOR SKILLS EVALUATED DURING EACH STAGE OF A SCENARIO**

Affective (interacting) skills: (a) receiving information, (b) responding, (c) valuing, (d) organizing, and (e) commitment (see the affective circle); cognitive (thinking) skills: (a) basic knowledge, (b) comprehension, (c) application, (d) analysis, (e) evaluation, and (f) creation (see the cognitive circle); psychomotor (doing) skills: (a) technical information (spatial perception, guided response, eye-hand coordination), (b) no supervision, (c) technical complexity and adaptation, and (d) origination (see the psychomotor circle).

ANTS: combination of anesthesia-nontechnical skills (affective and cognitive skills).

Stage 1: In this stage of each scenario we evaluated basic and technical knowledge, receiving information, and responding (lower skill level is represented by the shaded areas) with affective skills (see the area enclosed in the triangle and the affective circle); stage 2: In this stage of each scenario we evaluated mainly cognitive skills (see the cognitive circle).

and technical knowledge (gathering information, preparation, and working with protocols and checklists)^{11,19–23} and psychomotor skills (spatial perception, eye-hand coordination).⁸ The ANTS can also be divided into cognitive skills (decision making, planning, strategy, risk assessment, situation awareness) and interpersonal affective skills (teamwork, communication, leadership). Both are necessary for safe and effective performance^{24,25} and represent 2 legs in the skills triangle with psychomotor skills being the third leg (FIGURE 1).^{7,8} The ANTS concept was designed using methods of task analysis similar to the model used for pilots.^{15,26} ANTS include the main nontechnical skills

What was known

Traditional assessment is not suited for measuring skills in uncertain situations that require higher-order cognitive processing.

What is new

Simulation-based assessment differentiated between higher-order (cognitive) and lower-order (basic and technical) skills in relatively experienced anesthesiology residents.

Limitations

Single-institution study and limited sample reduce generalizability.

Bottom line

Gaps in anesthesia nontechnical skills (ANTS) can be addressed at the level of the individual trainees or, for recurring themes, should inform curriculum development.

(cognitive and affective) associated with good anesthetic practice^{11,19,27} that should be specifically taught and evaluated in all anesthesiology training programs.^{28–30} Cognitive errors are thought process errors that lead to incorrect diagnoses or treatments. Understanding and correcting cognitive errors³¹ cannot be overemphasized (a table of ranked cognitive errors is provided as online supplemental material). A goal for each resident within an anesthesiology training program should be to explore, define, and pinpoint his or her own cognitive errors, and the program should plan an education strategy designed to decrease these errors.

To accomplish this goal, we used a well-established simulation-based assessment design that integrates OSCE modalities into assessment of anesthesiology residents.^{32–35} This method of testing knowledge and exploring cognition by discussion interrogation during simulated scenarios was developed by the Israeli Board of Anaesthesiology Examination Committee.^{33,34} Using scenarios similar to those developed in Israel provided an opportunity for direct comparison of trainees in American and non-American residencies^{36,37} and exploring the portability of such scenarios.

Our primary aim was to evaluate the deficiencies in cognitive performance according to error rates and performance grades within and across different clinical domains, and between postgraduate year (PGY) levels. Based on our previous work,³² we hypothesized that we would uncover some deficiencies in knowledge and skills, and that there would be fewer higher-order cognitive deficiencies in graduating compared with starting PGY residents.

Methods

Assessment Model

We followed an evaluation model that integrates assessment of lower-level knowledge and basic skill learning with

assessment of higher levels of attitude, skills, behavior, and cognition.²¹ That model integrates Miller's 4 progressive capabilities: understanding (knows), application (knows how), integration (shows how), and practice (does).²³ Knowledge is the base of this framework and action or doing is at the top.

To accomplish this we used an established simulation design that integrated OSCE modalities into the assessment process for anesthesiology residents (FIGURE 1).^{32–35} In the first stage of each scenario we evaluated basic and technical knowledge (lower skill level)²³ with affective skills.⁸ In the second stage we evaluated mainly cognitive skills.⁸

Scenarios

Two similar but not identical scenarios were used in each of 3 clinical domains (cardiac resuscitation, trauma management, intraoperative crisis management) in a simulated environment. These 6 scenarios were originally developed and used by the Israeli Board of Anaesthesiology Examination Committee.^{32–36} Faculty members in the Department of Anesthesiology at the University of Florida, assisted by educational experts (an assistant dean for medical education and an assistant professor of medical education), translated and adapted the material and methods with adherence to scenario protocol, language, and assessment tools.

Assessment Tools

Individual performance in each scenario was assessed using a 15- to 29-item scenario-specific checklist. Applicable items of the checklist differed by scenario, and thus a different number of total items was applied to each scenario. The checklist included evaluation of basic knowledge and cognition. In the first stage (10 to 15 minutes), the examinee handled a simulated case while the examiner followed the preset evaluation checklist and observed passively. In the second stage (10 to 15 minutes), the examinee was asked and scored about the rationale behind actions (or inactions) in the immediately preceding scenario. All questions, such as asking for clarifications, were handled and presented to the examinees in a similar fashion and were dependent on the examinee's performance in the first stage; they were intended to explore the reasoning and logic behind the examinee's ANTS performance. Basic knowledge checklist items were scored as *done* or *not done*. Criteria for a well-performed task were provided to the evaluators to standardize the measure of quality. All checklist items were weighted equally (provided as online supplemental material). The evaluation checklist was in a Yes/No format to minimize bias^{38,39} and included only checklist items scored in a binary format (*done/not done*). This checklist scores performance using the

item-based Angoff method,^{38,39} which relies on experts who examine the content of each question (item) and then predict how many minimally qualified candidates would, in a binary format, answer the item correctly.

Participants

A total of 47 PGY-2 through PGY-4 residents participated 80 times in the simulation assessment. After obtaining informed consent, each resident received oral instruction and printed materials explaining the study objectives (of evaluating teaching or learning errors), and assurance that results were confidential and had no impact on their residency program evaluations. All residents had prior orientation to the high-fidelity Human Patient Simulator as part of their curriculum. Practice and assessment of clinical skills in a simulator environment was not novel to the participants.

Study Protocol

Each resident performed 1 or 2 different scenarios from 1 or 2 clinical domains (provided as online supplemental material). The only reason not to limit the study to 1 scenario per resident was to overcome the difficulty in resident availability.

Assessors

All residents were evaluated in real time (not via video) by the same experienced evaluator (A.S.), who was part of the original Israeli Examination Committee responsible for developing the OSCE component of the examination^{33,34} and had participated as an examiner in Israel. This evaluator had no clinical interaction with the residents during the study period and was blinded to the residents' PGY level. The assessor followed the original Israeli examination instructions.^{33,34}

Scoring

Each item in the assigned scenario was scored as 1 or 0 (ie, correct or incorrect, respectively). Proportions of items performed satisfactorily in each scenario were calculated. Error rates were calculated as the rate at which the examinees did not perform items.

The study received approval by the University of Florida Institutional Review Board. Informed consent was obtained from all examinees.

Calculations

For every item in each scenario, the following parameters were calculated as previously described^{32,34,35} and compared between PGY groups:

- *Group (PGY) error rate*: The ratio of total errors made in a scenario by a certain PGY group to the total number of items answered by all residents in that PGY group for that scenario.

- *Item performance grade*: Ratio of residents who performed an item satisfactorily in a scenario to the total number of residents who managed the same scenario.
- *Individual (resident) success rate*: Ratio of items answered satisfactorily to the total number of items in a scenario.

Group error rate gives information about how much error examinees in a certain PGY level make for a specific scenario whereas item performance grade gives information about the performance of examinees for each item and hence the difficulty level of the item.

Statistical Analysis

Individual success rates are presented as mean $\pm$ SD; grouped error rates are presented as ratios of errors for each scenario within a clinical domain for each PGY level. We tested whether individual success rates were significantly different between scenarios within each domain using *t* tests and Kruskal-Wallis tests. An equivalence test was conducted between scenarios in each domain.⁴⁰ Group error rates for nontechnical and technical items were compared for each scenario within each PGY by using a 2-proportional z-test. Scenarios within each domain and PGY level were similarly compared for error rates.

Linear mixed models were used to compare individual success and error rates between PGY groups. PGY level, domain, and scenario were considered fixed effects and identification of the resident was considered a random effect in order to adjust for correlations among observations from the same subject. The Kenward-Roger method was used to calculate the denominator degrees of freedom due to the unbalanced study design. The Tukey-Kramer method was used to adjust for multiple comparisons. For all analyses, alpha was designated as 0.05. Data were analyzed using SAS 9.3 (SAS Institute Inc, Cary, NC).

Cognitive Errors Analysis

All items tested in each scenario script were evaluated, concentrating on the grouped error rates > 0.7 by the graduating PGY-4 group during the first (nongnitive) stage and the second (cognitive) stage. We then related the deficiencies we observed to a list from a recent publication that identified important cognitive errors in anesthesiology practice.³¹

Results

Eighty examinations were conducted over 4 months with 47 different residents: 18 PGY-2, 13 PGY-3, and 16 PGY-4. Fifteen residents were tested on only 1 scenario, 29 (randomly on the basis of availability) were tested on 2

scenarios from 2 domains, and 3 were tested on 3 scenarios from 3 domains.

The TABLE presents the distribution of residents in each scenario within each domain by their PGY level and number of items tested (15 to 29 items) in each scenario. There were no significant differences in the ratio of cognitive to nongnitive items between or within domains.

FIGURE 2 presents the error rate for each scenario within each domain by PGY level. The error rate for the cognitive items was higher for all residents than the nongnitive within each domain and for most scenarios.

There were no significant differences in error rates between the pair (scenario 1 versus 2) in resuscitation or operating room (OR) domains for all items ($P = .14$ or $P = .44$, respectively) within any of the PGY levels, whereas the difference in error rates was significant for trauma scenario 1 versus 2 ($P = .001$) for PGY-2 versus PGY-4 residents.

When we investigated differences in error rates for nongnitive or cognitive items, a significant difference within pairs of scenarios (error rates in scenario 1 versus 2) was found only in the OR among PGY-3 residents tested for cognitive performance and in trauma for PGY-2 and PGY-4 residents for nongnitive performance.

FIGURE 3 presents the success rate for residents in each scenario within each domain by residents' PGY level. The success rate was lower for the cognitive items than the nongnitive items for at least 1 of the scenarios within resuscitation and trauma among PGY-2 and PGY-4 residents. The cognitive success rates by PGY-4 residents were 0.5–0.68, and significantly lower than the nongnitive success rate for resuscitation and trauma.

Difference in overall (all items) success rates between the 2 scenarios within a clinical domain was significant only for trauma. A difference between the 2 scenarios was significant for cognitive success rates only among PGY-3 residents for the OR ($P = .02$) and for nongnitive performance among PGY-4 residents for trauma ($P = .02$).

A list of items tested with performance grades < 0.6 for each item among the PGY-4 group and the error in each item related to known possible cognitive errors is provided as online supplemental material.³¹ The most common cognitive errors observed in the resuscitation scenarios were availability bias (choosing a diagnosis because it is in the forefront of your mind due to an emotionally charged memory of a bad experience) and premature closure (accepting a diagnosis prematurely, or failure to consider reasonable differential of possibilities); the most common cognitive errors in the trauma scenarios were anchoring (focusing on a single issue at the expense of understanding the whole situation) and premature closure; and the most common cognitive errors in the OR scenarios were anchoring, availability bias, premature closure, and

TABLE	RESIDENTS' DISTRIBUTION IN EACH POSTGRADUATE YEAR (PGY) AND FIELD, INCLUDING SCENARIOS AND NUMBER OF ITEMS TESTED IN EACH SCENARIO ^a									
	Scenario	OR Scenario 1	OR Scenario 2	OR Domain	Trauma Scenario 1	Trauma Scenario 2	Trauma Domain	Resuscitation Scenario 1	Resuscitation Scenario 2	Resuscitation Domain
	PGY-2	5	5	10	5	6	11	4	4	8
	PGY-3	4	4	8	4	4	8	4	5	9
	PGY-4	4	4	8	4	4	8	5	5	10
	No. of items	29	29	58	18	15	33	20	18	38
	No. of basic noncognitive items	15	16	31	7	6	13	6	6	12
	No. of cognitive items	14	13	27	11	9	20	14	12	26
	Ratio of cognitive items in domain	0.52	0.55	0.53	0.39	0.4	0.39	0.3	0.33	0.32

Abbreviation: OR, operating room.

^a Scenario type has noncognitive items in the first stage; cognitive items in the second stage; noncognitive items are technical, basic knowledge/treatment with affective performance.

confirmation bias (seeking or acknowledging only information that confirms the desired or suspected diagnosis).

Discussion

For PGY-3 and PGY-4 residents the error rates were higher and success rates lower for the cognitive items compared with the noncognitive items in each domain tested. The most common cognitive errors in all 3 domains were ranked within top cognitive errors (anchoring, availability bias, premature closure, confirmation bias).³¹

Similar cognitive errors were also found in other US programs,³¹ and our findings are similar to the results of a preliminary study that showed higher error rates in clinical domains with more cognitive or advanced knowledge items.⁴¹ Also, when comparing the results of OSCE scenarios previously used with non-US graduating residents³² to those in our program we observed comparable error rates, performance grades, and pass rates for the US and non-US (Israeli) graduating PGY-4 equivalent residents.

Our findings indicate that even graduating (PGY-4) residents remain challenged by advanced cognitive processes, probably because of training (teaching and learning) challenges. Because cognitive qualities of comprehension, application, analysis, evaluation, and decision making are highly important and at the top of necessary clinical skills or competence or performance (as a higher-order skill rather than knowledge retention),²³ our main findings are important and very different from our hypothesis.

When tested in scenarios designed to evaluate cognitive performance, PGY-4 resident performance was most deficient in higher-order items in resuscitation and trauma but not in the OR scenarios. However, even in better-performed OR scenarios we still found significant cognitive errors; again, the most common higher errors were anchoring, availability bias, premature closure, and confirmation bias. These advanced ANTS items are higher-order decisions, such as choice of actions and interventions or comprehensive differential diagnosis, not basic knowledge of treatment.

We based our examination on testing for Minimal Requirement Task Performance (used in the OSCE^{34,35}), and almost half or more of the tasks are basic knowledge and technical skills (47% in OR, 61% in trauma, and 68% in resuscitation scenarios; TABLE). It appears that although a smaller number of the tasks and items were advanced or applied knowledge and skills, this type of task was more problematic (and had higher error rates) for residents in each PGY.

The definition of performance in anesthesia varies dramatically—from vague (vigilance, data interpretation,

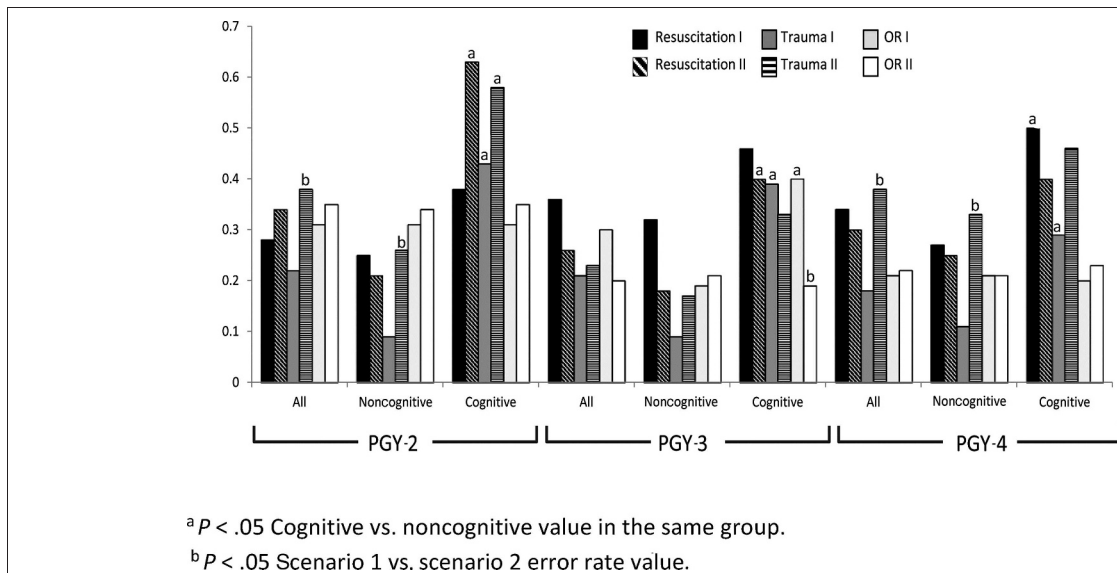


FIGURE 2

ERROR RATE FOR ALL COGNITIVE/NONCOGNITIVE ITEMS ACCORDING TO THE SPECIFIC SCENARIO AND POSTGRADUATE YEAR (PGY)

plan formulation, and implementation⁴²) to very technical, organized, and detailed (gathering information, equipment preuse preparation and check, induction technique, intra-operative checks, postoperative management, airway assessment).^{13,19} Some investigators evaluate performance in anesthesia by separating basic (gathering information) or technical knowledge (initiating and working with protocols, reviewing checklists) from the cognitive and behavioral or affective (decision making and team interaction) aspects.^{43,44} This separation is based on strong analogies to performance during management of critical

events in aviation.⁴⁴ It is important to measure these 2 aspects of skilled performance separately in managing crisis situations: basic technical actions from appropriate crisis solving and management ANTS behaviors. These nontechnical skills are highly important, and should be looked for.²³

To achieve performance at a high level for both noncognitive and cognitive skills, learning objectives and curriculum or teaching should be adjusted to address the deficiencies identified. There are still many questions regarding which errors are most important to address and

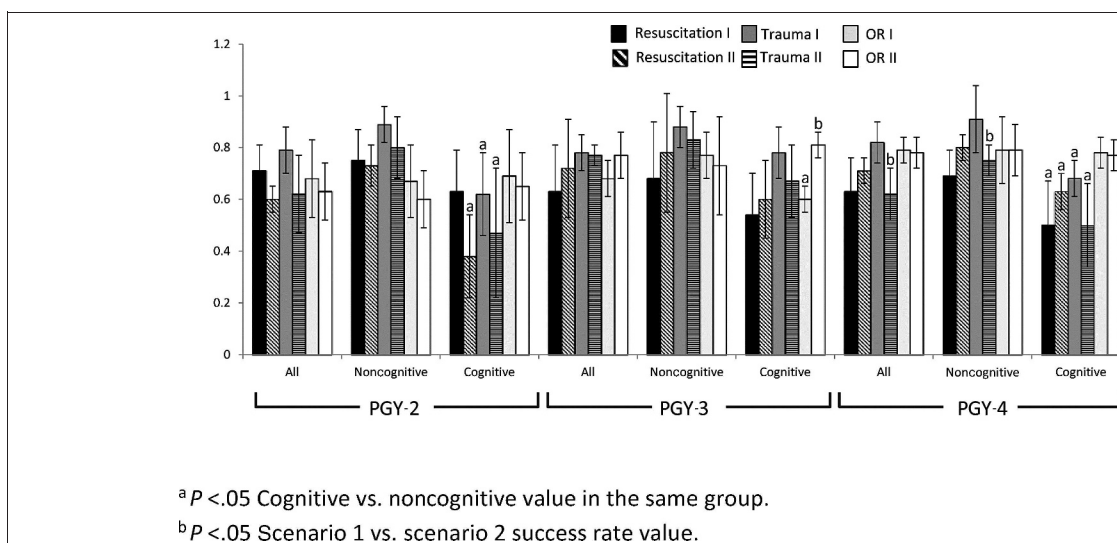


FIGURE 3

GROUPED SUCCESS RATES FOR ALL ITEMS ACCORDING TO THE SPECIFIC SCENARIO AND POSTGRADUATE YEAR (PGY)

which adjustment learning strategies are most appropriate and effective in anesthesiology.^{45–53}

Study Limitations

Our study has several limitations. The study was conducted at a single institution, we did not perform a comprehensive calculation of sample size, and a limited sample of residents was used, limiting generalizability. This study is also limited in its ability to differentiate learning (by the residents) from teaching (according to a systematic curriculum). Despite a consensus that anesthesia acute care skills should be taught systematically and perhaps using simulation,²⁸ these skills are taught sporadically rather than systematically. If they had been taught systematically, it would highlight a very different problem and suggest a problem with the teaching methods.

We did not consider the more advanced clinical year as the sole indicator of competence. The simulation-based format of our practical examination tested the upper level of competence—the “does” stage of Miller’s model of medical competence.²³

We do not have a very good understanding of how cues in the simulated environment affect decision making and problem solving. Thus, what we are witnessing may in part be due to the limitations of using simulation for summative assessment. For example, residents often perform relatively well in resuscitation scenarios because the cues received in the simulated environment are often clear-cut (eg, arrhythmia on monitor) and the treatment follows well-known algorithmic approaches (eg, Advanced Cardiovascular Life Support). Scenarios that are less clear-cut (eg, evolving hypertension or hypotension) may depend on multiple cues from various sources with varying degrees of fidelity.

Conclusion

A cognitive and noncognitive simulation-based skills assessment that included ANTS-identified areas of strength and weakness can be used to guide an anesthesiology residency curriculum. Such ANTS deficiencies need to be addressed in any training program. Gaps identified during debriefing can be used to adjust learning, and gaps that are recurring themes should inform curriculum development.

References

- McGaghie WC, Issenberg SB, Petrusa ER. Simulation—savior or satan? A rebuttal. *Adv Health Sci Educ Theory Pract*. 2003;8(2):97–103.
- Holzinger A, Kickmeier-Rust MD, Wassertheurer S, Hessinger M. Learning performance with interactive simulations in medical education: Lessons learned from results of learning complex physiological models with the HAEMOdynamics SIMulator. *Comput Educ*. 2009;52(2):292–301.
- Hodges BD, Kuper AI. Theory and practice in the design and conduct of graduate medical education. *Acad Med*. 2012;87(1):25–33.
- Accreditation Council for Graduate Medical Education. ACGME Outcomes Project, 2013. <http://www.acgme.org/acgmeweb/Portals/o/PDFs/SlideDecks/SLIDEDECK-FDMilestones2013.pptx>. Accessed January 13, 2014.
- Implementing Milestones and Clinical Competency Committees. ACGME Competencies: Suggested Best Methods for Evaluation. Released April 24, 2013. <http://www.acgme.org/acgmeweb/Portals/o/PDFs/ACGMEMilestones-CCC-AssesmentWebinar.pdf>. Accessed January 13, 2014.
- Journal of Graduate Medical Education Announces March 2014. Release of Milestones for All Phase II Specialties. http://www.acgme.org/acgmeweb/Portals/o/PDFs/newsRel_o6_10_13.pdf. Accessed January 13, 2014.
- Lampotang S, Lizardo D, Rajon D, Luria I, Gravenstein N, Bisht Y, et al. Mixed simulators: augmented physical simulators with virtual underlays. *Proceedings of the IEEE Virtual Reality 2013 meeting* 13: 7–10, 2013.
- Rovai AP, Wighting MJ, Baker JD, Grooms LD. Development of an instrument to measure perceived cognitive, affective, and psychomotor learning in traditional and virtual classroom higher education settings. *Internet Higher Educ*. 2009;12(1):7–13.
- Howard SK, Gaba DM, Fish KJ, Yang G, Sarnquist FH. Anesthesia crisis resource management training: teaching anesthesiologists to handle critical incidents. *Aviat Space Environ Med*. 1992;63(9):763–770.
- Meterissian SH. A novel method of assessing clinical reasoning in surgical residents. *Surg Innov*. 2006;13(2):115–119.
- Reich DL, Uysal S, Bodian CA, Gabriele S, Hibbard M, Gordon W, et al. The relationship of cognitive, personality, and academic measures to anesthesiology resident clinical performance. *Anesth Analg*. 1999;88(5):1092–1100.
- Epstein RM, Hundert EM. Defining and assessing professional competence. *JAMA*. 2002;287(2):226–235.
- Murray DJ, Boulet JR, Kras JF, Woodhouse JA, Cox TC, McAllister JD. Acute care skills in anesthesia practice: a simulation-based resident performance assessment. *Anesthesiology*. 2004;101(5):1084–1095.
- Glavin RJ. Excellence in anesthesiology: the role of nontechnical skills. *Anesthesiology*. 2009;110(2):201–203.
- Flin R, Patey R, Glavin R, Maran N. Anaesthetists’ non-technical skills. *Br J Anaesth*. 2010;105(1):38–44.
- Fletcher G, Flin R, McGeorge P, Glavin R, Maran N, Patey R. Anaesthetists’ Non-Technical Skills (ANTS): evaluation of a behavioural marker system. *Br J Anaesth*. 2003;90(5):580–588.
- Patey R, Flin R, Fletcher G, Maran N, Glavin R. Developing a taxonomy of Anaesthetists’ Nontechnical Skills (ANTS). In: Henriksen K, Battles JB, Marks ES, Lewin DI, eds. *Programs, Tools, and Products*. Rockville, MD: Agency for Healthcare Research and Quality; 2005. *Advances in Patient Safety: From Research to Implementation*, vol 4.
- University of Aberdeen. ANTS website. www.abdn.ac.uk/iprc/ants. Accessed October 24, 2013.
- Forrest FC, Taylor MA, Postlethwaite K, Aspinall R. Use of a high-fidelity simulator to develop testing of the technical performance of novice anaesthetists. *Br J Anaesth*. 2002;88(3):338–344.
- Castanelli DJ. The rise of simulation in technical skills teaching and the implications for training novices in anaesthesia. *Anaesth Intensive Care*. 2009;37(6):903–910.
- Rodriguez-Paz JM, Kennedy M, Salas E, Wu AW, Sexton JB, Hunt EA, et al. Beyond “see one, do one, teach one”: toward a different training paradigm. *Postgrad Med J*. 2009;85(1003):244–249.
- Objective Structured Assessment of Technical Skill (OSATS) in ACGME Outcomes Project Toolbox of Assessment Methods. <http://www.acgme.org/acgmeweb/Portals/o/PFAssets/Nov4NASImpPhasell.pdf> and Reference No. 5 [<http://www.acgme.org/acgmeweb/Portals/o/PDFs/ACGMEMilestones-CCC-AssesmentWebinar.pdf>]. Accessed January 13, 2014.
- Miller GE. The assessment of clinical skills/competence/performance. *Acad Med*. 1990;65(suppl 9):63–67.
- Foerde K, Shohamy D. The role of the basal ganglia in learning and memory: insight from Parkinson’s disease. *Neurobiol Learn Mem*. 2011;96(4):624–636.
- Morgan PJ, Cleave-Hogg DM, Guest CB, Herold J. Validity and reliability of undergraduate performance assessments in an anesthesia simulator. *Can J Anaesth*. 2001;48(3):225–233.
- Flin R, Martin L, Goeters KM, Hormann HJ, Amalberti R, Valot C, et al. The development of the NOTECHS system for evaluating pilots’ CRM skills. *Human Factors Aerospace Saf*. 2003;3(2):95–117.
- Reader T, Flin R, Lauche K, Cuthbertson BH. Non-technical skills in the intensive care unit. *Br J Anaesth*. 2006;96(5):551–559.
- Matveevskii AS, Gravenstein N. Role of simulators, educational programs, and nontechnical skills in anesthesia resident selection, education, and competency assessment. *J Crit Care*. 2008;23(2):167–172.
- Schwid HA, Rooke GA, Carline J, Steadman RH, Murray WB, Olympio M, et al. Evaluation of anesthesia residents using mannequin-based simulation: a multi-institutional study. *Anesthesiology*. 2002;97(6):1434–1444.
- Scavone BM, Sproviero MT, McCarthy RJ, Wong CA, Sullivan JT, Siddall VJ, et al. Development of an objective scoring system for measurement of

- resident performance on the human patient simulator. *Anesthesiology*. 2006;105(2):260–266.
- 31 Stiegler MP, Neelankavil JP, Canales C, Dhillon A. Cognitive errors detected in anaesthesiology. *Br J Anaesth*. 2012;108(2):229–235.
 - 32 Sidi A, Berkenstadt H, Ziv A, Euliano TY, Lampotang S, White CB. Evaluating construct validity of simulation-based OSCE for summative assessment in an anesthesiology teaching program. Abstract A-151. Presented at the American Society of Anesthesiologists Annual Meeting, Chicago, IL, October 15, 2011.
 - 33 Ziv A, Rubin O, Sidi A, Berkenstadt H. Credentialing and certifying with simulation. *Anesthesiol Clin*. 2007;25(2):261–269.
 - 34 Berkenstadt H, Ziv A, Gafni N, Sidi A. Incorporating simulation-based OSCE into the Israeli National Board Examination in Anesthesiology. *Anesth Analg*. 2006;102(3):853–858.
 - 35 Berkenstadt H, Ziv A, Gafni N, Sidi A. The validation process of incorporating simulation based accreditation into the anesthesiology Israeli national board exams. *Isr Med Assoc J*. 2006;8(10):728–733.
 - 36 Berkenstadt H, Kantor GS, Yusim Y, Gafni N, Perel A, Ezri T, et al. The feasibility of sharing simulation-based evaluation scenarios in anesthesiology. *Anesth Analg*. 2005;101(4):1068–1074.
 - 37 Boulet JR, Murray DJ. Simulation-based assessment in anesthesiology: requirements for practical implementation. *Anesthesiology*. 2010;112(4):1041–1052.
 - 38 Angoff WH. Scales, norms, and equivalent scores. In: Thorndike RL, ed. *Educational Measurement*. 2nd ed. Washington, DC: American Council on Education; 1971:508–600.
 - 39 Yudkowsky R, Downing SM, Wirth S. Simpler standards for local performance examinations: the yes/no Angoff and whole-test Ebel. *Teach Learn Med*. 2008;20(3):212–217.
 - 40 Hung HMJ, Wang S-J, Tsong Y, Lawrence J, O'Neill RT. Some fundamental issues with non-inferiority testing in active controlled trials. In: Proceedings from the Annual Meeting of the American Statistical Association; Atlanta, GA; August 5–9, 2001.
 - 41 Sidi A, Ziv A, Euliano TY, Lampotang S, White CB. Using simulation-based education to pinpoint curriculum deficiencies in an anesthesiology teaching program. Abstract A-154. Presented at the American Society of Anesthesiologists Annual Meeting, Chicago, IL, October 15, 2011.
 - 42 Devitt JH, Kurrek MM, Cohen MM, Fish K, Fish P, Noel AG, et al. Testing internal consistency and construct validity during evaluation of performance in a patient simulator. *Anesth Analg*. 1998;86(6):1160–1164.
 - 43 Weller JM, Bloch M, Young S, Maze M, Oyesola S, Wyner J, et al. Evaluation of high fidelity patient simulator in assessment of performance of anaesthetists. *Br J Anaesth*. 2003;90(1):43–47.
 - 44 Gaba DM, Howard SK, Flanagan B, Smith BE, Fish KJ, Botney R. Assessment of clinical performance during simulated crises using both technical and behavioral ratings. *Anesthesiology*. 1998;89(1):8–18.
 - 45 Regehr G. It's NOT rocket science: rethinking our metaphors for research in health professions education. *Med Educ*. 2010;44(1):31–39.
 - 46 Schmidt HG. Problem-based learning: rationale and description. *Med Educ*. 1983;17(1):11–16.
 - 47 van Gessel E, Nendaz MR, Vermeulen B, Junod A, Vu NV. Development of clinical reasoning from the basic sciences to the clerkships: a longitudinal assessment of medical students' needs and self-perception after a transitional learning unit. *Med Educ*. 2003;37(11):966–974.
 - 48 Tallentire VR, Smith SE, Skinner J, Cameron HS. Understanding the behaviour of newly qualified doctors in acute care contexts. *Med Educ*. 2011;45(10):995–1005.
 - 49 Pugh CM, DaRosa S, Santacaterina S, Clark RE. Faculty evaluation of simulation-based modules for assessment of intraoperative decision making. *Surgery*. 2011;149(4):534–542.
 - 50 Chipman SF, Schraagen JM, Shalin VL. Introduction to cognitive task analysis. In: Schraagen JM, Chipman SF, Shute VJ, eds. *Cognitive Task Analysis*. Mahwah, NJ: Lawrence Erlbaum Associates; 2000:3–23.
 - 51 Anderson JR. *Cognitive Psychology and Its Implications*. 4th ed. New York, NY: Freeman; 1995.
 - 52 Bordage G. Conceptual frameworks to illuminate and magnify. *Med Educ*. 2009;43(4):312–319.
 - 53 Kahol K, Vankipuram M, Smith ML. Cognitive simulators for medical education and training. *J Biomed Inform*. 2009;42(4):593–604.