Simulation-Based Assessment Identifies Longitudinal Changes in Cognitive Skills in an Anesthesiology Residency Training Program

Avner Sidi, MD,* Nikolaus Gravenstein, MD,† Terrie Vasilopoulos, PhD, * and Samsun Lampotang, PhD* †‡§

Objectives: We describe observed improvements in nontechnical or “higher-order” deficiencies and cognitive performance skills in an anesthesi/residency cohort for a 1-year time interval. Our main objectives were to evaluate higher-order, cognitive performance and to demonstrate that simulation can effectively serve as an assessment of cognitive skills and can help detect “higher-order” deficiencies, which are not as well identified through more traditional assessment tools. We hypothesized that simulation can identify longitudinal changes in cognitive skills and that cognitive performance deficiencies can thus be remediated over time.

Methods: We used 50 scenarios evaluating 35 residents during 2 subsequent years, and 18 of those 35 residents were evaluated in both years (post graduate years 3 then 4) in the same or similar scenarios. Individual basic knowledge and cognitive performance during simulation-based scenarios were assessed using a 20- to 27-item scenario-specific checklist. Items were labeled as basic knowledge/technical (lower-order cognition) or advanced cognitive/nontechnical (higher-order cognition). Identical or similar scenarios were repeated annually by a subset of 18 residents during 2 successive academic years. For every scenario and item, we calculated group error scenario rate (frequency) and individual (resident) item success. Grouped individuals' success rates are calculated as mean (SD), and item success grade and group error rates are calculated and presented as proportions. For all analyses, α level is 0.05.

Results: Overall PGY4 residents' error rates were lower and success rates higher for the cognitive items compared with technical items in the operating room and resuscitation domains. In all 3 clinical domains, the cognitive error rate by PGY4 residents was fairly low (0.00–0.22) and the cognitive success rate by PGY4 residents was high (0.83–1.00) and significantly better compared with previous annual assessments (P < 0.05). Overall, there was an annual decrease in error rates for 2 years, primarily driven by decreases in cognitive errors. The most commonly observed cognitive error types remained anchoring, availability bias, premature closure, and confirmation bias.

Conclusions: Simulation-based assessments can highlight cognitive performance areas of relative strength, weakness, and progress in a resident or resident cohort. We believe that they can therefore be used to inform curriculum development including activities that require higher-level cognitive processing.

Key Words: simulation, cognitive performance, simulation-based assessment, curriculum development

The Accreditation Council for Graduate Medical Education (ACGME) provides detailed standards for teaching and evaluating medical competency. Using its “Outcomes” project, the ACGME created an initiative that makes it mandatory for every resident to be assessed for competency in several domains of medical practice. This initiative was known as the “Milestones” project.2,3 The “Milestones” project and evaluation toolbox recommended by the ACGME are meant to elevate the role of simulation for medical performance (cognitive and technical) and procedures.2,3 It is not possible to accomplish competency assessment of nontechnical (i.e., cognitive and affective) and technical (e.g., psychomotor) skills using only the traditional written, oral, and objective structured clinical exams (OSCEs).6–9 These traditional assessments do not effectively capture the nuance encountered in important clinical scenarios seen in practice. To reliably solve problems in the operating room (OR) requires that a practitioner has both knowledge and experience.7 Therefore, it is incumbent on us to create better methods to measure actual realistic acute clinical performance and the complex task competency that is required in actual clinical practice. Simulation is an attractive approach to assess advanced cognitive skills, knowledge integration, clinical judgment, communication, and teamwork.10 Life-sized patient simulators have been used to assess complex skills.11 Anesthesia nontechnical skills (ANTS) should be specifically taught, emphasized, and evaluated in training programs.8,12–17 These skills encompass 2 subgroups: cognitive skills (decision making, planning, strategy, risk assessment, situational awareness) and affective skills (teamwork, communication, leadership).6,8,18,19 Anesthesia basic/technical knowledge (“lower-order” cognition) is very different from ANTS, which is composed of advanced cognitive skills (“higher-order” cognition).12,15–17 Both are necessary for safe and effective performance in the OR and constitute 2 of 3 sides of the skills triangle (with psychomotor/technical being the third leg in that triangle).4,5,10 Anesthesia nontechnical skills can be evaluated in a simulated environment,21 and that was the primary aim of the present study: to evaluate higher-order, cognitive performance. That approach of separating and testing higher-order performance has the same basis for evaluating 4 progressive capabilities, each of which builds upon the previous understanding (knows, application (knows how), integration (shows how), and practice (does)).10,22 Cognitive errors are thought-process errors that lead to incorrect conclusions or actions. Cognitive error science has evolved to the extent that there is now a cognitive error catalog specific to anesthesiology practice22,23 (see Appendix, Supplemental Digital Content 1, http://links.lww.com/JPS/A95).23 Understanding that the key types of cognitive errors that anesthesiology practitioners are vulnerable to is the first step toward using debiasing strategies in training to further improve patient safety. Anesthesiology training programs should explore, define, and pinpoint their local cognitive error vulnerabilities.
Simulation exercises that integrate OSCE modalities into the assessment process have been used to accomplish this. In the first stage of each OSCE scenario, basic and technical knowledge along with affective skills was evaluated, and in the second stage, ANTS were evaluated. Thus, a simulation-based assessment that included ANTS can be used to modify learning, identify common or recurring gaps, and inform curriculum development.

In one study, deficiencies in cognitive performance were identified using simulation error rates and performance grades within different clinical domains and between postgraduate year (PGY) levels. The findings in that study indicated that even graduating (PGY4) residents remain challenged by advanced cognitive processes, such as choice of actions and interventions or making comprehensive differential diagnoses. During scenarios designed to evaluate cognitive performance, PGY4 resident performance was most deficient (0.29–0.5 error rate) in higher-order items in resuscitation and trauma scenarios. Even in the better-managed (0.20–0.23 error rate) OR scenarios, significant cognitive errors were still noted.

If the goal is high-level ANTS and technical skills, we should expect residents to perform on the same high level for both. To achieve this, “adjustments”—such as education in cognitive errors, meta-cognition, and debiasing strategies—are needed. However, these kinds of “adjustments” are not easy to achieve.In addition, the most important errors and the best “adjustment” strategies remain to be determined.

In any case, we did not intend to prove that any specific intervention or measurement we applied was responsible for the change, and we considered it to be outside the scope of this article. Thus, we see the main value of using simulation in helping to highlight and assess deficiencies as well as inform curricular development.

Our primary aims were to evaluate cognitive performance and to detect “higher-order” deficiencies according to error rates and performance grades within 3 different clinical domains (OR, trauma, and cardiac resuscitation) and between PGY levels, comparing 2 successive academic years. Our main objective was to demonstrate that simulation can effectively serve as an assessment of cognitive skills and can help detect “higher-order” deficiencies, which are not as well-identified through more traditional assessment tools. We hypothesized that simulation can identify longitudinal changes in cognitive skills and that cognitive performance deficiencies remediated over time.

METHODS

After institutional review board approval, the study was completed at the University of Florida College of Medicine (Gainesville, Florida).

Scenarios

Two similar but not identical scenarios (scenarios 1 and 2) were used in the following 3 domains: cardiac resuscitation, trauma management, and intraoperative crisis management, in a simulated environment. We duplicated a previously described scenario approach (first stage: basic knowledge; second stage: exploring advanced cognition by discussion/debriefing) (see Appendices B–D and figure in our previous publication).

Assessment Tools

Individual performance in each scenario was assessed using a 20- to 27-item scenario-specific checklist. The checklist included evaluation of basic knowledge and cognition (see also Appendices B–D in our previous publication). In the second stage of every scenario, the examinee was asked and scored about the rationale behind actions taken (or not taken) in the first stage. All questions were preset and concentrated on the ANTS performance. Basic knowledge checklist items were scored in a binary format as “done” or “not done.” A set of criteria for a well-performed task was provided to the evaluator to standardize the measure of quality. All checklist items were weighted equally. The exam script was designed as a nonsubjective binary “yes/no” format to minimize bias and to include only checklist items scored in a binary format (“done”/“not done”) (see also Appendices B–D and “flowchart” figure of the scenario in our previous publication).

Participants

After informed consent, 35 PGY3 and PGY4 residents (of 50 in the residency program) were recruited by the chief residents in the Department of Anesthesiology to participate in the study. We used 50 scenarios evaluating 35 residents during 2 subsequent years, and 18 of those 35 residents were evaluated in both years (first as PGY3 then as PGY4) using the same scenarios. Thus, we studied 35 residents (18 of these 35 residents had the same scenarios) across 2 successive years as they graduated to the next level the following year. The exam was administered 50 times for a period of 3 to 4 months to 35 residents; 23 were PGY3 and 12 were PGY4. Whereas 20 residents were tested on 1 scenario only, 15 residents (7 PGY3 and 8 PGY4) were tested on 2 scenarios from different fields (those residents were picked randomly on the basis of same-day availability and were tested on a second scenario on the same day). Eighteen examinees were evaluated in the same domain and in identical or similar scenarios during 2 consecutive years (2011–2012 and 2012–2013). Three residents were tested in 4 scenarios (those residents were picked randomly) during those 2 years.

METHODS

After institutional review board approval, the study was completed at the University of Florida College of Medicine (Gainesville, Florida).

Scenarios

Two similar but not identical scenarios (scenarios 1 and 2) were used in the following 3 domains: cardiac resuscitation, trauma management, and intraoperative crisis management, in a simulated environment. We duplicated a previously described scenario approach (first stage: basic knowledge; second stage: exploring advanced cognition by discussion/debriefing) (see Appendices B–D and figure in our previous publication).

Assessment Tools

Individual performance in each scenario was assessed using a 20- to 27-item scenario-specific checklist. The checklist included evaluation of basic knowledge and cognition (see also Appendices B–D in our previous publication). In the second
TABLE 1. Residents Distribution in Each Postgraduate Year and Field, Including Scenarios and Number of All/Cognitive Items Tested in Each Scenario

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Resuscitation Domain</th>
<th>Trauma Domain</th>
<th>OR Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGY3 and PGY4 residents</td>
<td>10</td>
<td>18</td>
<td>22</td>
</tr>
<tr>
<td>No. all items</td>
<td>38</td>
<td>50</td>
<td>54</td>
</tr>
<tr>
<td>No. cognitive items</td>
<td>36</td>
<td>18</td>
<td>28</td>
</tr>
</tbody>
</table>

academic years 2012–2013 compared with the 2011–2012 evaluation.¹⁰ Eighteen almost identical or very similar scenarios were repeated annually by the same 18 examinees (9 OR, 5 resuscitation, and 4 trauma scenarios). The trauma and resuscitation scenarios were almost identical in academic years 2011–2012 and 2012–2013; the OR scenarios were similar but not identical.

Scoring

Each item in a scenario was scored as 0 or 1 depending on whether it was answered correctly. Proportions of items performed satisfactorily out of all possible items (noncognitive and cognitive) in a scenario were calculated.

Data Collection

Checklist results were manually entered into an Excel (Microsoft, Redmond, Wash) spreadsheet.

Calculations

For every item in each of the scenarios, the following parameters were calculated as previously described²⁶,²⁸,²⁹ and compared between PGY groups: group (PGY) error rate, item performance grade, scenario performance grade (grouped), and individual (resident) success rate.

Group (PGY) Error Rate

The ratio of the total number of errors made in a given scenario by a certain PGY group to the total number of items answered in a given scenario in that PGY group.

Item Success Grade

The ratio of residents who performed an item satisfactorily in the scenario.

Scenario Success Grade (Grouped items)

The average of item success grade using all items in the scenario.

Individual (Resident) Success Rate

The ratio of items answered satisfactorily by the resident versus the total number of items answered in the scenario.

Group error rate determines how many errors examinees in a certain PGY level make for a specific scenario, whereas the item success grade indicates the performance of all examinees for each item and hence the difficulty level of the item.

Statistical Analysis

Grouped individuals success rates are presented as mean (SD) (Table 3); item success grade and group error rates are presented separately as proportions of success or errors for each item or scenario within a field for each PGY level. The t and Kruskal-Wallis tests were used to determine whether individual success rates were significantly different between 2 scenarios within each field. Group error rates for cognitive and technical items were compared for each scenario within each PGY using a 2-proportion 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. The Tukey-Kramer method was used to adjust for multiple comparisons. For all analyses, α level was designated as 0.05. Data were analyzed using SAS 9.3 (SAS, Cary, NC).

Cognitive Errors Analysis

All items tested in each scenario script were examined and evaluated, concentrating on item success grade of less than 0.7 for the graduating PGY4 group during the first (noncognitive) stage and the second (cognitive) stage. We then related the deficiencies we observed to a list from a recent publication that identified the cognitive errors most important in anesthesiology practice.²⁵

RESULTS

Table 1 presents the distribution of the number of residents in each scenario within each field by their PGY level and the number of items tested (20–27 items) in each scenario within each field. There were no significant differences in the ratio of cognitive to noncognitive items between and within fields between the 2 consecutive years.

Table 2 presents group scenario error rates (for all items and for cognitive items), according to the specific scenario within each field and the resident's PGY level. The grouped (individuals) success rates are not presented in Table 2 but still reported here. The PGY4 residents error rates were lower for the cognitive items compared with basic and technical item performance in

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Resuscitation</th>
<th>Trauma</th>
<th>OR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>PGY3 All</td>
<td>0.18</td>
<td>0.17</td>
<td>0.22</td>
</tr>
<tr>
<td>Cognitive</td>
<td>0.26*</td>
<td>0.22</td>
<td>0.19*</td>
</tr>
<tr>
<td>PGY4 All</td>
<td>0.15</td>
<td>0.06*</td>
<td>0.15</td>
</tr>
<tr>
<td>Cognitive</td>
<td>0.11</td>
<td>0.00*</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>*</td>
<td>†</td>
<td>‡</td>
</tr>
</tbody>
</table>

*Value different from the previous annual assessment, same PGY level.
†P< 0.05 scenario 1 versus scenario 2 error rate value.
‡P< 0.05 technical versus nontechnical value in the same group.
We isolated 16 major cognitive errors in 4 fields: knowledge and skills and hypothesized that there would also be error rates showed greater decreases (comparing with OR error change) in 2012–2013. Eventually the error rates within all 3 fields became similar in 2012–2013 (Fig. 1). The differences between cognitive and technical error rates across years in trauma and resuscitation are significantly higher ($P < 0.05$ and $P < 0.003$, respectively) in 2011–2012 compared with OR, but both trauma and resuscitation show significant decreases, and eventually, the differences between cognitive and technical error rates are similar in 2012–2013.

Table 3 presents the grouped success rates for 18 examinees who were evaluated in the same domain or the same scenario in the years 2011 and 2012, according to specific scenarios and PGY level. In all 3 clinical domains, the cognitive success rate by PGY3 and PGY4 residents in 2012–2013 was higher (range = 0.74–1.00) than the previous year’s value (range = 0.39–0.87), and significantly better ($P < 0.05$) compared with the previous annual assessments. Cognitive error rate values within the same examinees in trauma and resuscitation were initially higher compared with OR, but they both showed greater decreases in 2012–2013; thus, the error rates within all 3 fields were similar in 2012–2013.

Presented in Table 4 are the following: item success grades, major (with the success grade $<0.7$ for the PGY4 group) cognitive errors within the list of items tested during the second (cognitive) stage of each scenario, and the possible error type contributing to each deficient item. We isolated 16 major cognitive errors in 4 scenarios (n = 88 items; 17% error rate). Such major cognitive errors were decreased in the trauma scenarios during 2 subsequent years. The amount and quality of cognitive errors in the resuscitation and OR domains were minimally changed compared with the previous (2011–2012) annual evaluation. The cognitive errors ranked as 1 and 3 “cognitive errors” (anchoring, premature closure; see Appendix, Supplemental Digital Content 1, http://links.lww.com/JPS/A95) in the trauma domain were decreased compared with the previous (2011–2012) annual evaluation. The most common cognitive errors in the trauma scenarios remained ranked as 2 and 3 “cognitive errors” (availability bias, premature closure); the most common cognitive errors in the trauma scenarios were still ranked as 1 and 3 “cognitive errors” (anchoring, premature closure); and the most common cognitive errors in the OR scenarios remained ranked as 1, 2, 3, and 5 “cognitive errors” (anchoring, availability bias, premature closure, confirmation bias). Thus, the most common cognitive errors observed remained anchoring, availability bias, premature closure, and confirmation bias.

**DISCUSSION**

We expected to see improvement in some deficiencies in knowledge and skills and hypothesized that there would also be significant decreases between years, with this effect more predominant in resuscitation. Error rates in trauma and resuscitation were higher in 2011 compared with OR. Both of these error rates show greater decreases, so the error rates across all 3 fields are similar in 2012.

![Graph showing change in overall mean cognitive error rates across years](image-url)

**FIGURE 1.** Change in overall mean nontechnical (cognitive) error rates across years, stratified by field. *Value significantly different ($P < 0.05$ and $P < 0.003$, respectively, for trauma and resuscitation) from the previous year’s assessment. Error rates in trauma and resuscitation were higher in 2011 compared with OR. Both of these error rates show greater decreases, so the error rates across all 3 fields are similar in 2012.

**TABLE 3.** Grouped Individuals Success Rate of PGY3 and PGY4 Residents, for Cognitive Items According to the Specific Scenario and Domain and the Year Evaluated

<table>
<thead>
<tr>
<th>Domain</th>
<th>Pairs of 2012 Scenarios Versus Identical or Similar 2011 Scenarios</th>
<th>All Items</th>
<th>Cognitive Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resuscitation</td>
<td>5</td>
<td>0.78 (0.12)*</td>
<td>0.62 (0.12)</td>
</tr>
<tr>
<td>Trauma</td>
<td>4</td>
<td>0.85 (0.07)</td>
<td>0.70 (0.08)</td>
</tr>
<tr>
<td>OR</td>
<td>9</td>
<td>0.81 (0.07)</td>
<td>0.72 (0.10)</td>
</tr>
</tbody>
</table>

*Data are presented as mean (SD).

* $P < 0.05$; success rate different from the previous year's assessment, in the same examinees.
fewer higher-order cognitive deficiencies for residents in the subsequent academic year from a learning effect. This learning effect is known as “construct validity,” or progression of scores within different levels of training.26,30,43 We expected that progression in scores would also be evident for the whole group of graduating residents evaluated in other fields and different scenarios.

The most interesting finding is that our simulations’ cognitive deficiencies mirrored the top anesthesiology cognitive errors. We demonstrated that simulation is effective at identifying these errors and that simulation may be a valuable way to teach and combat these errors. We describe observed improvements in cognitive or “higher-order” deficiencies and cognitive performance skills as discerned from an item and scenario analysis within OR, trauma, and resuscitation domains in an anesthesia residency cohort over a 1-year time interval.

The definition of skilled performance in anesthesia varies from vague (vigilance, data interpretation, plan formulation, and implementation45) to very technical, organized, and detailed (gathering information for preoperative evaluation, equipment preuse preparation and check, induction technique, intraoperative checks, postoperative management, and airway assessment).11,18 Some investigators evaluate performance in anesthesia by completely separating basic knowledge (gathering information) or technical (initiating and working with protocols, reviewing checklists), and psychomotor (intubating, perception, detecting cues to act, guided response)5 skills from the cognitive and behavioral or affective (decision making and team interaction) aspects.44,45 This separation is based on strong analogies to performance during management of critical events in other complex, dynamic domains such as aviation.45 We also believe that it is important to measure 2 separate aspects of skilled performance in managing crisis situations, that is, implementing appropriate technical actions (technical performance) and manifesting appropriate crisis-solving and management behaviors (ANTS).

In a preliminary study where the performance of noncognitive and cognitive items was evaluated, differences between

| TABLE 4. Item Success Grade and Cognitive Error Catalog Ranking for PGY4 Examinees in Cognitive Stage of Different Scenarios |
|---|---|
| **Resuscitation scenario 1 (wide complex tachycardia with hypotension)** |
| Principals of deciding how to treat the dysrhythmia |
| β-Blockers—wrong decision (failing heart) |
| Understanding the new CPR 2010 |
| C-A-B concept (rather than A-B-C) |
| Higher compressions frequency |
| Higher depth for compressions |
| Same ratio of ventilation: compression |
| 3 trials cardio versions attempts |
| Immediately after cardioversion |
| Chest compressions |
| Pulse check/cycle interpretation |
| **Resuscitation scenario 2 (narrow complex irregular tachycardia and hypotension)** |
| Principals of deciding how to treat the dysrhythmia |
| Cardioversion—presents a problem (patient needs sedation) |
| Understanding the new CPR 2010 |
| C-A-B concept (rather than A-B-C) |
| 3 Trials cardio versions attempts |
| **Trauma scenario 1 (chest and neck)** |
| Alternatives (plan B) for chest drain location (not in the wound) |
| Higher insertion |
| Highest location (axilla) |
| Chest film x-ray systematic interpretation |
| Heart size and borders |
| **Trauma scenario 2 (head and neck)** |
| Concerns regarding: re-intubation |
| Full stomach/cricoid pressure |
| Airway evaluation during complication |
| Neck veins |
| **OR scenario 1 (hemodynamic instability and myocardial ischemia)** |
| Differential diagnosis and reasons for hypotension |
| Septic/metabolic/allergic |

*Item performance grade PGY4 2012 > 0.7; 2011 < 0.7.
†Item performance grade PGY4 2012 < 0.7; 2011 > 0.7.
The errors catalog ranking is defined according to reference.25
error rates for the PGY4 residents could be explained by the fact that the OR scenarios had a much higher number of cognitive or advanced/applied knowledge items than the trauma and resuscitation scenarios.\textsuperscript{30,46}

Findings in another study\textsuperscript{10} strengthened those preliminary findings\textsuperscript{30,46} and indicated that even graduating (PGY4) residents remain challenged by advanced cognitive processes such as higher-order decisions, choice of actions and interventions or comprehensive differential diagnosis, and higher-level knowledge of treatment. Because cognitive qualities of comprehension, application, analysis, evaluation, and decision making are highly important and necessary clinical skills,\textsuperscript{22} our (Sidi et al\textsuperscript{10}) initial findings were alarming and important enough for a revisit and re-evaluation, even though we did not aim for a specific intervention to improve the cognitive performance.

Our evaluations in this study demonstrated significant improvements in each domain compared with a previous annual assessment in certain scenarios where learning gaps had been previously identified. There was an overall average decrease in error rates, mostly driven by decreases in cognitive errors.

It is notable that the most common cognitive errors observed by our group (Sidi et al\textsuperscript{10}) and others (Stiegl et al\textsuperscript{25}) remained anchoring, availability bias, premature closure, and confirmation bias. In addition, when comparing the results of identical OSCE scenarios previously used with residents graduating from Israeli programs,\textsuperscript{26} we observed comparable error rates and performance grades.\textsuperscript{30} Some items that scored as critical in the “top 10” catalog specific to anesthesiology practice\textsuperscript{25} were observed relatively infrequently in simulation, in the published catalog study, and in our study. One possible explanation is that the catalog was built from the most serious errors rather than the most frequent.

It remains to be seen if specific training in cognitive errors, meta-cognition, and debiasing strategies can improve this performance further or can be specifically applied to those demonstrating the most opportunity for improvement.\textsuperscript{25} Many questions outside of the scope of this study remain regarding which errors are the most important to address and which of the “adjustment” learning strategies are the most effective.\textsuperscript{7,31–39} There are multiple recommendations for “adjustment” learning strategies to improve cognitive/higher-order learning or performance, including the following: problem-based learning,\textsuperscript{22,23} focus groups,\textsuperscript{38} other learning modules,\textsuperscript{35–37} or frameworks.\textsuperscript{38} Using a simulation-based assessment that highlights cognitive mistakes while providing feedback via debriefing can serve also as an “adjustment” learning strategy, and it may partially explain the improved performance we observed.\textsuperscript{10} It can also serve as curriculum development.\textsuperscript{10,21}

Study Limitations

This study is 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,\textsuperscript{47} these skills are taught sporadically rather than systematically. Thus, it is not surprising that residents do not perform consistently well.

Simulations are not always ideal evaluators of actual performance because of many variables that are difficult to quantify, and sometimes, we are not able to simulate all necessary variables. We do not yet completely understand how cues in the simulated environment affect decision making and problem-solving. For example, residents often perform well in resuscitation scenarios because they receive relatively unambiguous cues (e.g., arrhythmia on monitor) and the treatment then follows a familiar algorithm (e.g., ACLS). Less clear-cut scenarios (e.g., evolving blood pressure changes) may depend on multiple cues from various sources with varying degrees of clinical authenticity.

As we did not intend to prove that anything specific was responsible for the observed improvements and our primary aim was to evaluate for deficiencies in cognitive performance and detect “higher-order” deficiencies, we did not use a control group. Only 18 of 35 were tested with the same/similar scenarios, and their progression of scores could be in part attributed to the repeated “learning effect” factor related to a specific scenario.

Because a single evaluator (A.S.) evaluated all residents, the generalizability of the findings may be restricted. However, the generalizability of these scenarios has already been proven.\textsuperscript{30} A detailed description of the psychometric properties of all items tested appears in our previous publication appendixes.\textsuperscript{10,20} The major error items from the same list tested appear in Table 4 of this article.

It is not completely clear why there is a significant difference in the performance of PGY4 residents between the 2 resuscitation scenarios and why the difference was not observed in our previous study.\textsuperscript{10} These findings probably suggest (as other investigators found as well)\textsuperscript{31} that a range of exposure to events may be required for a generalized performance to other domains, and the reason for the inconsistencies may be related to variable exposure of the residents to a certain subject in a certain field during the curriculum and study period.

CONCLUSIONS

Our study demonstrates that there is a measurable improvement in demonstration of higher-order knowledge and skills over a 1-year period in a residency training program, as one would hope to find. The most interesting finding is that our simulations’ cognitive deficiencies mirrored the top anesthesiology cognitive errors. We demonstrated that simulation is effective at identifying these errors, and that simulation may be a valuable way to teach and combat these errors. Simulation-based assessments can highlight areas of relative strength, and weakness in a resident cohort and could be used to guide modifications in the curriculum with regard to identified deficiencies in tasks requiring higher-order processing.

REFERENCES


© 2017 Wolters Kluwer Health, Inc. All rights reserved. www.journalpatientsafety.com