

Simulation-based Assessment of Pediatric Anesthesia Skills

James J. Fehr, M.D., F.A.A.P.,* John R. Boulet, Ph.D.,† William B. Waldrop, M.D.,‡
Rebecca Snider, M.H.S., R.N.,§ Megan Brockel, M.D.,|| David J. Murray, M.D.#

ABSTRACT

Background: Assessment of pediatric anesthesia trainees is complicated by the random nature of adverse patient events and the vagaries of clinical exposure. However, assessment is critical to improve patient safety. In previous studies, a multiple scenario assessment provided reliable and valid measures of the abilities of anesthesia residents. The purpose of this study was to develop a set of relevant simulated pediatric perioperative scenarios and to determine their effectiveness in the assessment of anesthesia residents and pediatric anesthesia fellows.

Methods: Ten simulation scenarios were designed to reflect situations encountered in perioperative pediatric anesthesia care. Anesthesiology residents and fellows consented to participate and were debriefed after each scenario. Two pediatric

What We Already Know about This Topic

- Medical simulation using multiple scenarios may enhance training and predict clinical skills in anesthesia residents, but this has not been confirmed in pediatric anesthesia

What This Article Tells Us That Is New

- In a study of residents and pediatric anesthesia fellows, a multisenario simulation assessment was developed with high interrater reliability and correlation with degree of clinical experience

anesthesiologists scored each scenario by key action checklist. The psychometric properties (reliability, validity) of the scores were studied.

Results: Thirty-five anesthesiology residents and pediatric anesthesia fellows participated. The participants with greater experience administering pediatric anesthetics generally outperformed those with less experience. Score variance attributable to raters was low, yielding a high interrater reliability.

Conclusions: A multiple-scenario, simulation-based assessment of pediatric perioperative care was designed and administered to residents and fellows. The scores obtained from the assessment indicated the content was relevant and that raters could reliably score the scenarios. Participants with more training achieved higher scores, but there was a wide range of ability among subjects. This method has the potential to contribute to pediatric anesthesia performance assessment, but additional measures of validity including correlations with more direct measures of clinical performance are needed to establish the utility of this approach.

ANESTHESIOLOGY has pioneered the use of simulation technologies to enhance medical education and advance the evaluation of practitioner knowledge and skills. Simulation has been used to provide education in a number of domains in anesthesia practice that are difficult to assess.

◇ This article is featured in "This Month in Anesthesiology." Please see this issue of ANESTHESIOLOGY, page 9A.

* Associate Professor, Division of Pediatric Anesthesiology, Departments of Anesthesiology and Pediatrics, Director of Saigh Pediatric Simulation Center, Washington University School of Medicine, St. Louis, Missouri. † Associate Vice President for Research and Data Resources, Foundation for Advancement of International Medical Education and Research, Philadelphia, Pennsylvania. ‡ Assistant Professor, Division of Pediatric Anesthesiology, Departments of Anesthesiology and Pediatrics, Washington University School of Medicine. § Assistant Director of Education, Department of Anesthesiology, Washington University School of Medicine. || Instructor, Division of Pediatric Anesthesiology, Department of Anesthesiology, Washington University School of Medicine. # Professor, Division of Pediatric Anesthesiology, Department of Anesthesiology, Director, Wood Simulation Center, Washington University School of Medicine.

Received from the Department of Anesthesiology, Washington University School of Medicine, St. Louis, Missouri. Submitted for publication May 1, 2011. Accepted for publication September 13, 2011. Supported by the Department of Anesthesiology, Washington University School of Medicine; and the Foundation for Anesthesia Education and Research: Research in Education Grant, Rochester, Minnesota, James Fehr 2009: "Skills in Pediatric Anesthesia: A Simulation-Based Assessment Program." Salary support from the Agency for Healthcare Research and Quality, Rockville, Maryland, R01 HS018734-01 (Drs. Murray and Fehr).

Address correspondence to Dr. Fehr: One Children's Place, St. Louis Children's Hospital, 5S31, St. Louis, Missouri 63110. fehrj@anest.wustl.edu. This article may be accessed for personal use at no charge through the Journal Web site, www.anesthesiology.org.

Copyright © 2011, the American Society of Anesthesiologists, Inc. Lippincott Williams & Wilkins. Anesthesiology 2011; 115:1308-15

Some recent studies have used simulation to evaluate whether novice anesthesia residents are able to recognize and manage intraoperative hypoxia or hypotension, to assess team performance during obstetric emergencies, to determine whether residents effectively manage an intraoperative pediatric cardiac arrest, and to provide a method to train residents to wean patients from cardiopulmonary bypass.^{1–4} Simulation has also been used to evaluate and train pediatric residents in the performance of cardiopulmonary resuscitation.⁵ In previous studies, a resident's skill was more reliably estimated when multiple scenarios were included in the assessment.^{6–8} In this investigation, a multiple-scenario assessment was designed to assess the advanced skills required in pediatric anesthesia settings.

Studies that have examined risks to children receiving anesthesia have described younger patient age, higher degree of illness, and provider inexperience as contributing to poor anesthesia outcomes. A physician's skill and experience are essential components of competent clinical care that contribute to safer outcomes in pediatric patients. The scenarios in this study were designed to reflect the broad range of ages and sizes of pediatric patients and to simulate the types of adverse events reported in pediatric anesthesia practice.^{9–10}

Anesthesiology specialty training stipulates that residents should have a minimum of 2 months of pediatric anesthesia experience and provide anesthesia for a variety of pediatric patients.* However, these experiences may not effectively provide residents with the skills needed to manage various pediatric anesthetic conditions. One of the purposes of this study was to determine how experience and training time in pediatric anesthesia affected a resident's ability to care for a variety of pediatric anesthesia critical events. Although the American Board of Anesthesiology sets this as a minimum training bar to achieve board eligibility, there are no studies that compare the performance in simulation scenarios of residents who have completed a 2-month minimum training period of pediatric anesthesia with those who have completed less than 2 months. In this project, we used simulation technology to assess pediatric anesthesia skills in trainees in anesthesiology who are learning to administer anesthesia to children. Our first goal was to design and construct a set of simulated scenarios that a resident who was board-eligible in anesthesiology would be expected to manage capably. A second goal was to use the performances of the anesthesia residents and pediatric anesthesia fellows to determine whether the scenario scores were reproducible and provide an estimate of ability.

* http://www.acgme.org/acWebsite/downloads/RRC_progReq/040_anesthesiology_f07012011.pdf, p 5. Accessed August 1, 2011.

†† <http://www.theaba.org/pdf/ITEContentOutline.pdf>, pp 34–37. Accessed August 1, 2011.

Materials and Methods

Scenarios

Scenarios were designed to incorporate topics from the pediatric anesthesia component of the content outline for the American Board of Anesthesiology's In-Training Examination.†† Scenario design was also guided by the Accreditation Council for Graduate Medical Education's expectations for pediatric anesthesia competence for a graduating resident in anesthesiology. After delineating the scenario content, input about expectations of resident performance was obtained from six faculty from the division of pediatric anesthesia at Washington University in St. Louis, Missouri, and a set of checklist scoring items was established for each scenario. Initial scenarios were pilot-tested with residents to determine the feasibility of the project and the logistics of running the simulations. Upon further revision, the scenarios were further fine-tuned by having three pediatric anesthesiologists complete the scenarios and provide feedback about the scenario design and the scoring points. The goal of the scenarios was to present content reflecting clinical situations that are commonly encountered in pediatric anesthesia settings (laryngospasm, neonatal resuscitation) rather than those typically referred to tertiary pediatric care centers (complex congenital heart disease). The goal was to provide a balanced set of scenarios to broadly assess the practitioner in diverse aspects of pediatric anesthetic management.

Scoring

Ten scenarios were developed to reflect a broad array of potential clinical situations encountered in the perioperative management of pediatric patients (table 1). For each scenario, a checklist was created with 8–10 key diagnostic and therapeutic items. For example, in the neonatal resuscitation scenario, the checklist included the following items: dries baby, stimulates baby, suctions mouth, begins bag-mask ventilation, calls for help, starts chest compressions for bradycardia, intubates baby, requests umbilical venous catheterization, and requests a neonatal intensive care unit bed. The scenarios and checklists were designed, piloted, and revised with the input of actively practicing academic pediatric anesthesiologists.

Participants

The Institutional Review Board at Washington University approved this study, and written informed consent was obtained from all participants. There were 51 anesthesiology residents and 8 pediatric anesthesia fellows during the period when the study was conducted. Sessions generally occurred during the participants' pediatric anesthesiology rotation, and 35 anesthesiology residents and fellows (59%) were available and consented to participate. Participation was not mandatory, although no one refused to participate. The results were kept confidential and not shared with the program director nor included in the participants' evaluation on their

Table 1. Pediatric Anesthesia Simulation Scenarios

| Scenario | Scenario Descriptor | Scoring Items |
|--------------------------|---|--|
| Bronchospasm | 5-yr-old, 20-kg child with asthma is wheezing after extremity fracture repair | Administer oxygen, auscultate lungs, review past medical history, inquire about intraoperative events, recognize wheezing, administer albuterol, confirm chest x-ray has been done, communicate with surgeons |
| Malignant hyperthermia | During inguinal hernia repair in a 6-yr-old, 20-kg child, MH occurs with elevated carbon dioxide, tachycardia | Give 100% oxygen, state diagnosis, discontinue volatile anesthetic, hyperventilate, call for malignant hyperthermia cart, administer dantrolene, administer bicarbonate, cool patient, order laboratory studies, call for ICU bed |
| Venous air embolus | Intraoperative resection of a Wilms tumor in a 6-yr-old, 20-kg child, complicated by sudden desaturation and decrease in end-tidal carbon dioxide | Recognize decreased end-tidal carbon dioxide, administer 100% oxygen, state diagnosis, notify surgeon, ask surgeon to flood surgical field, position patient in left lateral decubitus position, bolus fluid, administer epinephrine, arrange for ICU bed |
| Laryngospasm | At the beginning of strabismus surgery with a laryngeal mask airway in place, a 5-yr-old, 20-kg child loses end-tidal carbon dioxide | Auscultate in <60 s, administer 100% oxygen, remove laryngeal mask airway, attempt bag-mask ventilation, call for help, deepen anesthetic, administer muscle relaxant, intubate patient, confirm end-tidal carbon dioxide, auscultate bilateral breath sounds |
| Appendicitis with sepsis | A 7-yr-old, 30-kg child with appendicitis is ill, appearing with heart rate 136, respiratory rate 38, and blood pressure 72/48 | Check laboratory results, inquire about pain, check blood pressure before induction, bolus fluid, state diagnosis of possible sepsis, prepare suction, preoxygenate, rapid sequence induction with cricoid, decrease dose of induction agent, confirm end-tidal carbon dioxide |
| Airway foreign body | A 4-yr-old, 15-kg child preoperative for myringotomy tubes has a cough. Nurse is concerned. Mannequin with unilateral breath sounds and not on monitor | Ask for vital signs, administer oxygen, obtain past medical history, obtain history of cough, auscultate chest, obtain chest x-ray, state diagnosis, consult otolaryngology, plan for induction with maintenance of spontaneous respiration |
| Newborn resuscitation | Resuscitation of 2-kg newborn who was born less than 1 min ago: no respiratory effort, decreased tone, heart rate <60, and poor perfusion. The special care nurse and pediatrician have not yet arrived | Dry patient, stimulate and warm baby, suction mouth, begin bag-mask ventilation, call for help, start chest compressions for heart rate <60 after 30 s, intubate, request umbilical venous line, request bed in the NICU |
| Infant seizure | Called to evaluate 7-wk-old, 4.8-kg infant found seizing on the general pediatrics ward | Recognize seizure, give oxygen, place intraosseous needle when can't get an intravenous line, give anticonvulsant, recognize apnea, bag-mask ventilation, call for help, intubate patient, check glucose and electrolytes, request ICU bed |
| Postoperative apnea | 2-month-old, 2.8-kg former 32-week premature infant having apnea and bradycardia in PACU after cleft lip repair; had been given an excessive dose of narcotic | Inquire about neonatal history, inquire about intraoperative course, recognize desaturation, apply oxygen, recognize apnea, inquire about narcotic dose, administer naloxone or intubates, consider caffeine, arrange monitored bed and/or alerts surgeon |
| Accidental extubation | 3-month-old, 5-kg infant is brought to the ICU after colostomy takedown has accidental extubation | Recognize desaturation, auscultate and/or check endotracheal tube position, remove endotracheal tube if SpO ₂ <90%, recognize tube out before SpO ₂ <80%, bag-mask ventilate, call for help, reintubate, verify bilateral breath sounds, request end-tidal carbon dioxide confirmation |

ICU = intensive care unit; MH = malignant hyperthermia; NICU = neonatal intensive care unit; PACU = postanesthesia care unit.

pediatric anesthesia rotation. These included residents in the Department of Anesthesiology at Washington University in at least their second month of pediatric anesthesiology training and pediatric anesthesia fellows. The scenarios were per-

formed in a 2-h session in the Saigh Pediatric Simulation Center at St. Louis Children's Hospital using the Laerdal Sim NewB® (Wappingers Falls, NY) and the METI Pediatric SIM HPS® (Sarasota, FL) mannequins. The perfor-

Table 2. Checklist Scores and Discrimination (35 Trainees)

| Scenario | Mean (%) | Min–Max (%) | Discrimination |
|--------------------------|-------------|-------------|----------------|
| Bronchospasm | 66 ± 14.9 | 37.5–93.8 | 0.38 |
| Malignant hyperthermia | 64.5 ± 22.6 | 5–100 | 0.6 |
| Venous air embolus | 48.5 ± 20.5 | 16.7–88.9 | 0.45 |
| Laryngospasm | 58.2 ± 16.4 | 20–100 | 0.71 |
| Appendicitis with sepsis | 45.5 ± 13 | 20–65 | 0.58 |
| Airway foreign body | 64.6 ± 26.9 | 10–100 | 0.49 |
| Newborn resuscitation | 54.7 ± 16.1 | 22.2–88.9 | 0.34 |
| Infant seizure | 65.9 ± 14.1 | 30–90 | 0.37 |
| Postoperative apnea | 62 ± 14 | 33.3–88.9 | 0.51 |
| Accidental extubation | 61.8 ± 14.2 | 33.3–88.9 | 0.29 |

manes were documented using the B-line SimCapture® (Washington, DC) audiovisual recording system. Each participant was given an orientation to the mannequins and the method of the study, and then participated in the scenarios individually. They had 5 min to complete each simulation with a short debriefing between scenarios. The assessment cohort included 11 first-year anesthesia residents with less than 2 months of experience in pediatric anesthesia. Of the 24 trainees with more than 2 months of pediatric anesthesia experience, 7 were second-year residents, 9 were third-year residents, and 8 were pediatric anesthesia fellows. There were 22 males (62.9%) and 13 females (37.1%). The average age of the participants was 32.2 yr (SD = 3.6). Most of the participants (n = 31, 89%) were currently certified in advanced cardiac life support. Most (n = 31, 89%) were not currently pediatric advanced life support certified.

Rater Training/Scoring

Three raters scored the performances, the lead investigator (JF) and two pediatric anesthesiologists (MB, WW). All scenario performances were evaluated independently by at least two raters (one subject was evaluated by all three raters). The lead investigator designed and conducted all of the sessions and scored all of the performances. The other two raters (MB and WW) provided scores for 17 and 19 residents, respectively. One of these raters (MB) had previously participated in the pilot testing as a subject and as a rater. The lead author oriented the raters to the purpose and the design of the study, including the use of the checklist. Any ambiguities in the checklists noted by the raters were addressed before implementation, and the lead author was always available to attend to any concerns. The raters have completed pediatric anesthesia fellowships, are board-certified in anesthesiology, and have active pediatric anesthesia practices. Two independent scores were obtained for each resident and all of their 10 scenarios. Scenario scores, by rater, were calculated as a percentage of checklist items credited. Individual resident scenario scores were the average of the two rater scores. An overall score was calculated as the average of the scenario scores.

Statistical Analysis. Several analyses were performed to investigate the psychometric properties of the scenario scores.

To explore potential sources of measurement error in the checklist scores, generalizability theory was used.¹¹ Estimates of the reliability of the scores, over raters and simulation scenarios, were obtained. Because all scenarios may not be equally effective for measuring pediatric anesthesia skills, scenario-level discrimination (D) statistics were calculated (correlation between individual scenario scores and the overall score obtained by a resident). To examine potential performance differences by pediatric anesthesiology experience, a repeated measures analysis of variance (RM ANOVA) was completed. In performing this analysis, it was hypothesized that residents with more experience would outperform those with less experience. Residents were assigned to one of three groups: those who had completed 2 months or less of pediatric anesthesia training, those who had 3–5 months of pediatric anesthesia training, and those who had completed more than 5 months of pediatric anesthesia training. The first group has less training than the American Board of Anesthesiology's minimal requirement of 2 months of pediatric anesthesiology. The second group represents the typical duration of pediatric anesthesia rotations completed in the anesthesiology residency at Washington University. The third group reflects those who have a particular interest in pediatric anesthesiology or are pursuing fellowship training in pediatric anesthesiology.

Results

Descriptive Statistics

Descriptive statistics (checklist score) are presented in table 2 by scenario. Based on the average performance of all residents, the lowest scores were achieved on the Appendicitis with sepsis scenario (mean = 45.5%) followed by venous air embolus (mean = 48.5%). The highest scores were achieved on the bronchospasm scenario (mean = 66.0%). The D values were all positive, indicating that performance on one scenario was related to overall performance. The laryngospasm scenario had the highest discrimination value (D = 0.71).

Reliability

To estimate the reliability of the assessment scores, a generalizability study was conducted and the estimated variance

Table 3. Generalizability Results

| Variance Components of Checklist Scores | | | Decision Studies | | |
|---|----------|---------------------|------------------|------------------|------------------|
| Variance Component | Estimate | % of Total Variance | t = 10, r = 2 | t = 10, r = 1 | t = 20, r = 1 |
| Person (p) | 37.9 | 9.4 | 37.9 | 37.9 | 37.9 |
| Rater (r) | 3.3 | 0.8 | 4.4 | 4.4 | 2.2 |
| Task (t) | 44.5 | 11.1 | 1.7 | 3.3 | 3.3 |
| Person × task (p × t) | 240.7 | 60.1 | 24.1 | 24.1 | 12 |
| Person × rater (p × r) | 1.3 | 0.3 | 0.6 | 1.3 | 1.3 |
| Rater × task (r × t) | 7.3 | 1.8 | 0.4 | 0.7 | 0.4 |
| Error (p × r × t) | 65.8 | 16.4 | 3.3 | 6.5 | 3.3 |
| Generalizability coefficient (ρ^2) | — | — | 0.57 | 0.54 | 0.7 |

components for the generalizability study are presented in table 3. Variance components were estimated based on a person (P) by rater (R) by task (T) design. That is, each of the residents (person, $n = 35$) was rated by at least two independent raters across the 10 tasks (scenarios). Decision studies were conducted to estimate the generalizability of the checklist scores for different measurement conditions (*i.e.*, number of simulation scenarios, number of raters/per scenario).

The person (resident) variance component is an estimate of the variance across residents in resident-level mean scores. Ideally, most of the variance should manifest in the person component, indicating that individual abilities account for differences in observed scores. In this design, the other “main effect” variance components include task (scenario) and rater. The task component is the estimated variance of scenario mean scores. Because the estimate is much greater than zero, this indicates that the 10 tasks vary in average difficulty. As shown in table 2, mean performance, by simulation scenario, ranged from 45.5% for appendicitis with sepsis to 66.0% for the bronchospasm scenario. The rater component is the variance of rater mean scores. The relatively small value indicates that the raters do not differ much in terms of average stringency. The fact that the task component is much larger than the rater component indicates that the raters differ much less in average stringency than the simulation scenarios differ in average difficulty. The large interaction variance component, person × task, suggests that there are considerably different rank orderings of resident mean scores for each of the simulation scenarios. The relatively small rater by task components indicates that the raters rank order the difficulty of the scenarios similarly. The final variance component, error, is the residual variance that includes the triple-order interactions and all other unexplained sources of variation.

Based on the generalizability study described previously, the variance components were used to estimate the reliability of the resident scores for various measurement designs. This process, referred to as decision studies, allows the most efficient measurement procedures for future operations. For the simulation assessment, the goal is to generalize the residents' scores for a universe of generalization that includes many other tasks (scenarios) and many other raters. For a situation

where each resident is rated in each of the 10 scenarios (tasks; $n_t = 10$) and each of the tasks is rated by two independent raters ($n_r = 2$), the sample sizes for the decision study are the same as those for the generalizability study. Based on the estimated variance components for $n_t = 10$ and $n_r = 2$, the generalizability coefficient (ρ^2) is 0.57. The generalizability coefficient for a design that incorporates $n_t = 10$ tasks (scenarios) and only a single ($n_r = 1$) rater is estimated to be $\rho^2 = 0.54$, only slightly smaller than for a design with two raters per task. The precision of the resident ability estimate is therefore only minimally affected by the number of raters for each scenario. The estimated variance components for a design involving twice as many tasks ($n = 20$ scenarios) and a single rater are also displayed in table 3. Based on these estimates, $\rho^2 = 0.70$. Taken together, the results from the decision studies indicate that gains in measurement precision can best be achieved by increasing the number of tasks (scenarios) and not the number of raters per given task.

Although the reliability of the resident scores is most dependent on the length of the assessment (number of scenarios), it is still important to quantify the association between scores provided by the two independent raters. The standardized coefficient of interrater reliability was 0.80 (average correlation between the scores assigned by the same two raters to resident performances to the same task).

Validity Evidence

In an attempt to gather evidence to support the validity of the assessment scores, the performance of residents with more pediatric anesthesiology experience was compared with that of those with less experience. Residents were grouped based on reported months of pediatric anesthesia (group 1, <2 months, $n = 14$; group 2, 2–5 months, $n = 12$; group 3, >5 months, $n = 9$). RM ANOVA was conducted to test the hypothesis that there was no difference in scores based on experience. The independent variables were resident group (1, 2, 3) and scenario (repeated measure). The dependent variable was the summary checklist scenario score.

Based on the RM ANOVA, there was no significant group × scenario interaction, indicating that potential difference in group mean scores were consistent across scenarios. There was a significant main effect attributable to sce-

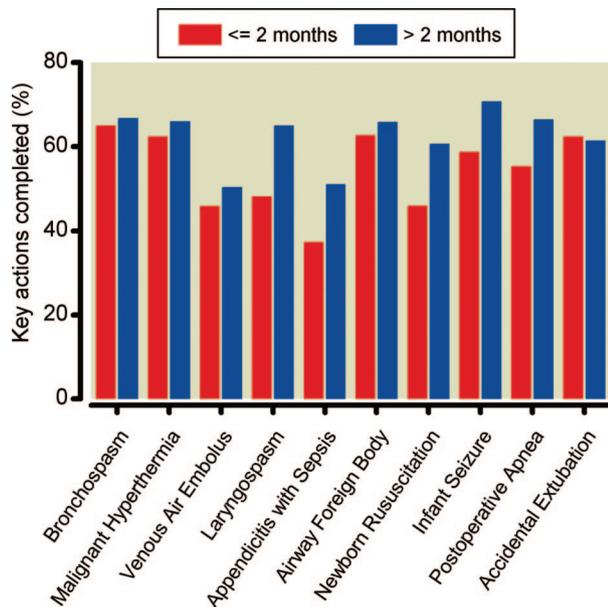


Fig. 1. Mean key action scores by scenario and pediatric anesthesia experience. The total key action scores out of 100% for the simulation scenarios for each group of trainees with either less (*red bars*) or more (*blue bars*) than 2 months of pediatric anesthesia training.

nario ($F_{1,9} = 6.4$, $P < 0.01$), indicating that, averaged over resident groups (experience levels), the scenarios were not of equivalent difficulty. There was also a significant main effect attributable to group ($F_{1,2} = 4.9$, $P = 0.01$). This indicates that, averaged over scenarios, there were significant differences in performance between groups. Based on a *post hoc* analysis of group scores (Scheffé F test), there was a significant difference between group 3 and group 1 and group 2 and group 1, but not between group 2 and group 3. Residents with at least 2 months pediatric anesthesia experience significantly outperformed those with less than 2 months experience.

Because there were no significant differences in scores between participants with 2–5 months of experience (group 2) and those with greater than 5 months of experience (group 3), these two groups were combined. A comparison of group performance (≤ 2 months of experience, > 2 months of experience) is presented in figure 1. Although there was no significant group \times scenario interaction from the RM ANOVA, between-group performance differences were negligible for some scenarios. For example, on the endotracheal tube out scenario, the less experienced residents (≤ 2 months, mean = 62.4%), on aggregate, actually outperformed those with more experience (more than 2 months, mean = 61.4%). Based on 35 participants, there were positive associations between the overall performance scores (aggregated over raters and scenarios) and months of postgraduate training ($r = 0.22$) and months of pediatric anesthesia training ($r = 0.27$). There was considerable variability in the overall performance of the residents, especially within the more experienced group (fig. 2).

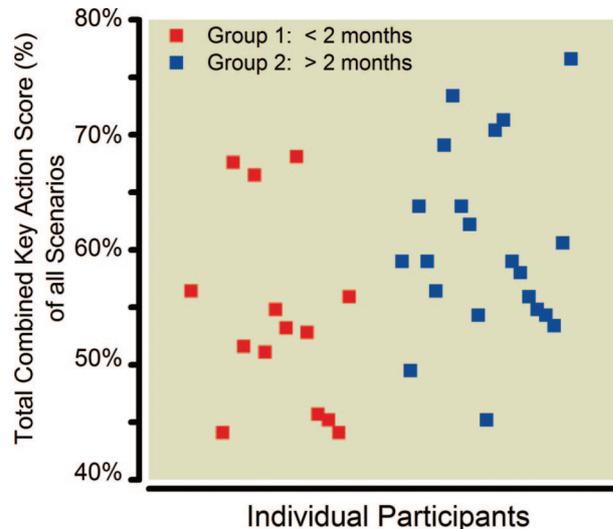


Fig. 2. Total key action score by participant. The total key action scores achieved by each trainee with either less (*red squares*) or more (*blue squares*) than 2 months of pediatric anesthesia training. Each point represents a specific trainee's percent overall score. The scores are scattered to highlight both the overlap and the score variation that occurs among individuals within and between groups.

Discussion

Children, most notably neonates and infants, are more likely than adults to experience morbidity and mortality during the perioperative period.⁹ Closed claims analyses demonstrate that major sources of liability in children include cardiovascular and respiratory events, younger age, and higher American Society of Anesthesiologist physical status.^{12–13} A variety of analyses, including reviews of case reports, investigations of perioperative deaths, and closed insurance claims, indicate that a lack of physician experience and skill are contributing factors to adverse outcomes.^{14–17} In this study, we designed a set of scenarios to reflect the nature of cardiac and respiratory conditions that can lead to adverse outcomes in the perioperative care of children.

Finding methods to assess advanced skills in anesthesiology is essential not only to establish a performance expectation for training, but also to delineate some of the differences in skills required and expected in fellowship training. The recent decision to establish a certification examination for pediatric anesthesiologists beyond the accreditation program for fellowships that began in 1997 suggests that defining differences is of increasing importance. Simulation can be a very effective method to assess those advanced skills and judgments acquired during pediatric anesthesia fellowship training. As part of our study, we developed a set of simulated scenarios, and associated scoring rubrics, to evaluate a trainee's skill in managing acute events in pediatric anesthesia.

In this investigation, residents and pediatric anesthesia fellows were assessed on a set of pediatric scenarios designed to cover a broad range of skills and ability in anesthesia prac-

tice. Subjects with more training and experience received higher scores than residents with less experience, suggesting that the assessment provides construct validity as a measure of skills acquired during residency training. One of the reasons pediatric fellows did not outperform residents could be the decision not to include the more complex scenarios such as managing anesthesia for congenital cardiac disease, solid organ transplantation, or congenital syndromes. Alternatively, the variability in experience among a small cohort of fellows may have affected the results.

The scenarios were designed based on the pediatric anesthesia component of the content outline for the American Board of Anesthesiology's In-Training Examination and the Accreditation Council for Graduate Medical Education's expectations of a graduating resident in anesthesiology for competence in pediatric anesthesia. The scenario score analysis, which ranged from 45.5% (for appendicitis with sepsis) to 66% (for bronchospasm), suggests that a challenging range of content was presented with some scenarios much more difficult than others. For most scenarios, some of the residents were able to achieve high scores in excess of 88%, which suggests scoring expectations were reasonable. However, the lower mean scenario scores indicate that the assessment, at least for some participants, was quite difficult. The scenario difficulty may relate to the relatively inclusive checklist items with less straightforward content.

The scoring analysis offers a method to evaluate whether individual scenario construct and scoring items contribute to the overall assessment. For example, the appendicitis scenario (mean score = 46%) required participants to check the blood pressure before induction, recognize and treat relatively severe hypotension in a 7 yr old, note the potential for sepsis, and anticipate the potential need for intensive care unit admission. Similarly, the venous air embolus (mean score = 48%) required the participant to consider the possibility in a major procedure (nephrectomy for tumor). The discrimination analysis indicated that although these scenarios were relatively difficult, participants who achieved higher overall scores also were able to manage these more difficult scenarios. A detailed checklist item analysis would be useful to identify those items where performance was weak or unrelated to the overall scenario score to modify content or scoring items for use in future assessments. The relative difficulty of the scenarios could be addressed by either eliminating or modifying some of the checklist items that did not contribute to the assessment. This item analysis could also be useful to further increase rater reliability by identifying scoring items that raters were more likely to disagree.

The wide range of scores and their substantial variation may reflect the differences in skill acquisition that exists among resident classes. There are likely multiple factors that lead to the wide range in performance including factors related to the simulation assessment including the design and construct of the scenarios. The preliminary results suggest that this variance in ability increases throughout training and

may lead to even greater differences in ability at the completion of training and entry into clinical practice. Although performance improves as training progresses, the variance in ability suggests that clinical experience and approaches to training must be tailored to ability. A competency-based training approach, with the use of individual assessment techniques such as simulation, could be used to progress training more rapidly for some residents and to provide remedial experiences for others who do not meet a minimal standard.

The results from our generalizability study were similar to those reported elsewhere.^{7,18} As noted in the performance assessment literature, content (or context) specificity can have an appreciable effect on performance. Therefore, although we attempted to measure general skills in pediatric anesthesia, performance on one type of scenario may not predict performance on another type. As a result, to obtain a reproducible measure of resident ability, several scenarios are required. For a 10-scenario assessment with 2 independent raters, the generalizability coefficient was moderate ($\rho^2 = 0.57$). Although this is likely adequate for low stakes, formative assessments, additional measurement precision would be required if the assessment were to be used for summative (e.g., certification) purposes. An even more reliable assessment would be possible by adding scenarios that have similar measurement properties (difficulty, discrimination) to construct an assessment form of reasonable length (e.g., 10–15 scenarios) that yields more precise ability estimates. In this study, the precision of the scores was not greatly affected by the rater or by the number of raters per scenario. This may be due in part to the fact that all of the raters were pediatric anesthesiologists who not only trained as raters for these scenarios, but had participated in a number of previous simulation studies.

In this study, individuals with additional pediatric anesthesia experience outperformed, on average, those with less. Although some evidence to support the validity of the assessment scores was garnered, the process of validation is far from complete. First, although there were significant differences in performance between individuals with less than 2 months pediatric anesthesia experience and those with more experience, this does not say anything about competence. The marked variance among individuals at the same level of training and experience indicates that a more in-depth process of defining individual ability is necessary. As evidenced by the individual overall scores, many residents with 2 months of experience may meet a performance standard; conversely, many participants with considerably more experience performed at a level even lower than the median of the less experienced residents. These findings support the concept that progression through residency could effectively be competence-based rather than time-based. In order to determine an acceptable performance standard, through some form of standard setting process, one could investigate the relationship between competence and experience.^{19,20} Second, with

a 10-scenario assessment in the current format, the generalizability of a resident's performance to another potential set of scenarios is only moderate. It would therefore be worthwhile to investigate, in more detail, the content-specific characteristics of scenarios that underlie individual performance. Information gathered through this process will be valuable in constructing better assessment forms (mixes of scenarios)—ones that yield more reliable and valid estimates of skills in pediatric anesthesia. Third, the consequential effect of introducing simulation-based assessment in pediatric anesthesia needs to be documented. Our cross-sectional analysis of resident performance could, and should, be augmented by longitudinal studies that follow residents over the course of training. This would help determine how skill and ability change among residents and define how to effectively accelerate the acquisition as well as retention of skill and ability during training. Finally, although there is some evidence to suggest that experience in actual patient care settings relates to performance in the simulated environment, specific data related to the practice of pediatric anesthesia need to be gathered.

In summary, we describe a multiple-scenario simulation using high-fidelity pediatric mannequins that provides a means to assess performance in areas that are critical to patient care in pediatric anesthesia. Simulation-based education and assessment offer an additional modality for performance-based credentialing of resident competence in managing anesthetics for children. This method has the potential to contribute to performance-based pediatric anesthesia assessment, but additional measures of validity obtained from comparisons with clinical performance are needed to establish this approach as a method to evaluate competence. Ultimately this approach could be used to assess competence and, as a result, have an effect on how other pediatric specialists assess the skills and performance of physicians who care for children.

The authors thank Margie Hassler, R.N., M.E.D., Nurse Facilitator, Saigh Pediatric Simulation Center, St. Louis Children's Hospital, St. Louis, Missouri, and Anne L. Glowinski, M.D., M.P.E., Associate Professor, Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri, for their assistance.

References

1. Park CS, Rochlen LR, Yaghmour E, Higgins N, Bauchat JR, Wojciechowski KG, Sullivan JT, McCarthy RJ: Acquisition of critical intraoperative event management skills in novice anesthesiology residents using high-fidelity simulation-based training. *ANESTHESIOLOGY* 2010; 112:202-11
2. Morgan PJ, Pittini R, Regehr G, Marrs C, Haley MF: Evaluating teamwork in a simulated obstetric environment. *ANESTHESIOLOGY* 2007; 106:907-15
3. Bruppacher HR, Alam SK, LeBlanc VR, Latter D, Naik VN, Savoldelli GL, Mazer CD, Kurrek MM, Joo HS: Simulation-based training improves physicians' performance in patient care in high-stakes clinical setting of cardiac surgery. *ANESTHESIOLOGY* 2010; 112:985-92
4. Howard-Quijano KJ, Stiegler MA, Huang YM, Canales C, Steadman RH: Anesthesiology residents' performance of pediatric resuscitation during a simulated hyperkalemic cardiac arrest. *ANESTHESIOLOGY* 2010; 112:993-7
5. Hunt EA, Walker AR, Shaffner DH, Miller MR, Pronovost PJ: Simulation of in-hospital pediatric medical emergencies and cardiopulmonary arrests: Highlighting the importance of the first 5 minutes. *Pediatrics* 2008; 121:e34-43
6. Murray DJ, Boulet JR, Kras JF, Woodhouse J, Cox T, McAllister JD: Anesthesia acute care skills: A simulation-based anesthesia skills assessment. *ANESTHESIOLOGY* 2004; 10:1084-95
7. Murray DJ, Boulet JR, Avidan M, Kras JF, Henrichs B, Woodhouse J, Evers AS: Performance of residents and anesthesiologists in a simulation-based skill assessment. *ANESTHESIOLOGY* 2007; 107:705-13
8. Waldrop WB, Murray DJ, Boulet JR, Kras JF: Management of anesthesia equipment failure: A simulation-based resident skill assessment. *Anesth Analg* 2009; 109:426-33
9. Murat I, Constant I, Maud'huy H: Perioperative anaesthetic morbidity in children: A database of 24,165 anaesthetics over a 30-month period. *Paediatr Anaesth* 2004; 14:158-66
10. Collins CE, Everett LL: Challenges in pediatric ambulatory anesthesia: Kids are different. *Anesthesiol Clin* 2010; 28:315-28
11. Boulet JR: Generalizability Theory: Basics, Encyclopedia of Statistics in Behavioral Science. Edited by Everitt BS, Howell DC. Chichester, John Wiley & Sons, Ltd., 2005, pp 704-11
12. Jimenez N, Posner KL, Cheney FW, Caplan RA, Lee LA, Domino KB: An update on pediatric anesthesia liability: A closed claims analysis. *Anesth Analg* 2007; 104:147-53
13. Ramamoorthy C, Haberkern C, Bhananker SM, Domino KB, Posner KL, Campos JS, Morray JP: Anesthesia-related cardiac arrest in children with heart disease: Data from the pediatric perioperative cardiac arrest [POCA] registry. *Anes Analg* 2010; 110:1376-82
14. Cohen MM, Cameron CB, Duncan PG: Pediatric anesthesia morbidity and mortality in the perioperative period. *Anes Analg* 1990; 70:160-7
15. Auroy Y, Ecoffey C, Messiah A, Rouvier B: Relationship between complications of pediatric anesthesia and volume of pediatric anesthetics. *Anesth Analg* 1997; 84:234-5
16. Van der Walt JH: Searching for the holy grail: Measuring risk in paediatric anaesthesia. *Paediatr Anaesth* 2001; 11:637-41
17. Bhananker SM, Ramamoorthy C, Geiduschek JM, Posner KL, Domino KB, Haberkern CM, Campos JS, Morray JP: Anesthesia-related cardiac arrest in children: Update from the Pediatric Perioperative Cardiac Arrest Registry. *Anesth Analg* 2007; 105:344-50
18. Murray DJ, Boulet JR, Kras JF, McAllister JD, Cox TE: A simulation-based acute skills performance assessment for anesthesia training. *Anesth Analg* 2005; 101:1127-34
19. Boulet JR, Murray DJ: Simulation-based assessment in anesthesiology: Requirements for practical implementation. *ANESTHESIOLOGY* 2010; 112:1041-52
20. Boulet JR, Murray DJ, Kras J, Woodhouse J: Setting performance standards for mannequin-based acute-care scenarios: An examinee-centered approach. *Sim Healthcare* 2008; 3:72-81