Part Task and Variable Priority Training in First-year Anesthesia Resident Education

A Combined Didactic and Simulation-based Approach to Improve Management of Adverse Airway and Respiratory Events

Ken B. Johnson, M.D.,* Noah D. Syroid, M.S.,† Frank A. Drews, Ph.D.,‡ L. Lazarre Ogden, M.D.,§ David L. Strayer, Ph.D.,‖ Nathan L. Pace, M.D., M.Stat.,# Diane L. Tyler, M.S.,** Julia L. White, R.N.,†† Dwayne R. Westenskow, Ph.D.‡‡

Background: Part task training (PTT) focuses on dividing complex tasks into components followed by intensive concentrated training on individual components. Variable priority training (VPT) focuses on optimal distribution of attention when performing multiple tasks simultaneously with the goal of flexible allocation of attention. This study explored how principles of PTT and VPT adapted to anesthesia training would improve first-year anesthesiology residents' management of simulated adverse airway and respiratory events. The authors hypothesized that participants with PTT and VPT would perform better than those with standard training.

Methods: Twenty-two first-year anesthesia residents were randomly divided into two groups and trained over 12 months. The control group received standard didactic and simulation-based training. The experimental group received similar training but with emphasis on PTT and VPT techniques. Participant ability to manage seven adverse airway and respiratory events were assessed before and after the training period. Performance was measured by the number of correct tasks, making a correct diagnosis, assessment of perceived workload, and an assessment of scenario comprehension.

Results: Participants in both groups exhibited significant improvement in all metrics after a year of training. Participants in the experimental group were able to complete more tasks and answered more comprehension questions correctly. There was no difference in perceived workload or the number of correct diagnoses between groups.

Conclusion: This study in part confirmed the study hypotheses. The results suggest that VPT and PTT are promising adjuncts to didactic and simulation-based training for management of adverse airway and respiratory events.

PART task training (PTT) and variable priority training (VPT) are techniques that have been developed by psychologists to optimize human performance when com-

* Associate Professor and Director, † Technical Director, ** Education Director, Center for Human Simulation, Department of Anesthesiology, ‡ Assistant Professor, ‖ Professor, Department of Psychology, § Associate Professor, # Professor, †† Research Coordinator, Department of Anesthesiology, ‡‡ Professor, Department of Biomedical Engineering, University of Utah.

Address correspondence to Dr. Johnson: Center for Patient Simulation, Department of Anesthesiology, University of Utah, 30 North, 1900 East, Room 3C444, Salt Lake City, Utah 84132-2304. ken.b.johnson@hsc.utah.edu. Information on purchasing reprints may be found at www.anesthesiology.org or on the masthead page at the beginning of this issue. ANESTHESIOLOGY's articles are made freely accessible to all readers, for personal use only, 6 months from the cover date of the issue.

pleting complex tasks. These techniques have been successfully implemented in a number of simulator-based professional training arenas and have led to higher pass rates in settings where students are asked to manage multiple tasks.[1]

Part task training is defined as the decomposition of large multicomponent tasks into a set of component tasks that when trained as individual components either separately or in various combinations can become highly automatized.[2–4] This training reduces processing demands by streamlining effort associated with the individual elements of the task. Focused training also leads to more rapid development of automatic skills that might otherwise not be achieved in the context of the whole task. *Variable priority training* is a method for training people to flexibly distribute attention over multiple aspects of a task. Participants in VPT learn to coordinate and control how attention is allocated to components of a task and assign different processing priorities to the components as they are performed in concert. VPT fosters flexible cognitive style that reduces the likelihood of cognitive tunnel vision.[5]

We have used PTT and VPT techniques as part of the didactic and simulation-based training for first-year anesthesia residents (CA-1s) over a 12-month period. Training was directed toward detection and appropriate treatment of adverse airway and respiratory events reported in the closed anesthesia malpractice claims database.[6–8] These events were made up of unrecognized esophageal intubations as a result of difficult intubations,[8] airway trauma, pneumothorax, airway obstruction, aspiration, and bronchospasm.[7] These adverse events occur with a higher frequency in pediatric patients with more severe consequences (*e.g.*, higher rate of mortality or brain injury).[6] Airway management difficulties, impaired vigilance, inadequate supervision, poor judgment, diversion of attention, and misinterpretation and misuse of data were also noted as potential sources for bad outcomes associated with adverse airway and respiratory events.[6]

The aim of this study was to demonstrate that PTT and VPT would improve CA-1 management of simulated adverse airway and respiratory events. Compared with CA-1s with conventional simulator training, we hypothesized that CA-1s with PTT- and VPT-oriented simulator training would (1) complete more critical tasks essential

831

Table 1. Performance Expectations for Each Scenario

Scenario 1 (adult): unanticipated difficult airway
    After induction of anesthesia, that participants recognize an unanticipated difficult airway, assess ability to mask ventilate, perform a
        second attempt at laryngoscopy, and place a laryngeal mask airway
Scenario 2 (adult): bronchospasm on emergence
    After emergence and extubation, that participants recognize the development of severe bronchospasm unresponsive to conventional
        therapy requiring emergent reintubation
Scenario 3 (adult): development of a tension pneumothorax
    After reintubation, that participants recognize development of a left-sided tension pneumothorax and perform a needle decompression
        of the left thorax
Scenario 4 (adult): aspiration after induction
    After loss of consciousness with significant emesis, that participants recognize the aspiration risk and need for emergent intubation
        and mechanical ventilation to restore oxygenation
Scenario 5 (child): esophageal intubation
    After induction of anesthesia and mechanical ventilation, that participants recognize the absence of exhaled carbon dioxide, hypoxia,
        an esophageal intubation (placed by an actor), and the need to intubate the trachea and restore ventilation
Scenario 6 (child): laryngospasm, bronchospasm, no intravenous access
    After mask induction and inability to attain intravenous access or secure the airway, that participants recognize the presence of a
        laryngospasm and the need for intramuscular succinylcholine and emergent intubation; after intubation, that participants recognize
        and treat bronchospasm
Scenario 7 (child): faulty ventilator circuit
    After mask induction, intubation, and institution of mechanical ventilation (performed by an actor), that participants recognize
        inadequate ventilation due to an anesthesia circuit failure and either replace the circuit or provide mechanical ventilation with an
        Ambu bag or Jackson Reese circuit

to managing an adverse event, (2) reach a correct diagnosis more often, (3) report a decreased perception of workload, and (4) demonstrate an increased level of comprehension when managing simulated adverse airway and respiratory events than CA-1s with conventional training.

Materials and Methods

After University of Utah institutional review board approval (Salt Lake City, Utah), 22 University of Utah CA-1s were consented to participate over a 2-yr period. Training consisted of 12-month-long rotations in general adult anesthesia, pediatric anesthesia, and surgical intensive care practice. Participants were randomly divided into two equal-size groups: control and experimental. The random allocation process was computer generated. To identify potential differences between groups, preliminary in-training examination scores and US Medical Licensing Examination parts I, II, and III were compared between groups with a two-tailed Student *t* test. Operating room supervision, simulation training, and didactic sessions were conducted by board-certified or board-eligible staff anesthesiologists.

Simulation-based assessment of anesthesia provider skill is an emerging method of characterizing performance in managing critical events.[9–11] Previous work has led to the development of checklist, time-based, and participant self-assessment methods of measuring performance. Using these techniques, participants cared for simulated patients using an adult and a pediatric human simulator (HPS version 5.55; METI, Sarasota, FL) to es-

tablish a baseline skill level. Scenario topics and performance expectations are presented in table 1.

A physiologic monitor (Datex AS/3; Helsinki, Finland) displayed the electrocardiogram, pulse oximeter (with tone), and capnogram waveforms and digital values for heart rate, blood pressure, oxygen saturation, end-tidal carbon dioxide, and fraction of inspired oxygen. All standard alarms were set to default limits. Mechanical ventilation was provided by an anesthesia machine (Narcomed 2B; North American Dräger, Telford, PA).

A preanesthetic evaluation, anesthetic record, and scenario introduction describing the recent course of events were prepared for six scenarios: three adult and three pediatric scenarios. The second adult scenario consisted of two parts, each treated as a unique scenario, for a total of seven adverse events. Participants reviewed the preanesthesia evaluation and scenario introduction in a quiet room before entering the simulation laboratory. Once in the simulation laboratory, participants were encouraged to think aloud. Video and audio information were recorded for all simulations. A video image of the physiologic monitor was inset into a video image of the participant caring for the simulated patient. A timer with a resolution to 1 s was superimposed on the video image. The order of scenario presentation was randomized for all participants at two levels: Participants were randomized first as to which set of scenarios they would receive first, pediatric or adult, and second as to the order of the three scenarios in each set. Each scenario lasted 7 min from the start of the adverse event. Testing of all participants was conducted over a 3-day period at the beginning of the academic year and again 12 months later. Two hours were allotted to each participant to complete all scenarios.

Anesthesiology faculty at the authors' institution developed by consensus a list of appropriate diagnostic and therapeutic tasks for each scenario. A series of three pilot studies were conducted with volunteer residents not assigned to a study group. A case report form was created that recorded the start of each adverse event, the time when the appropriate diagnosis was made, and the task lists for each scenario.

Validation of the task list was accomplished by distributing the list among six experts consisting of three board-certified anesthesiologists from the authors' institution (different from those involved in developing the tasks lists) and three from outside institutions. A modified Delphi technique was used to gain a consensus among experts[12,13] and develop a weighted task score list for each scenario (appendix 1).

During each scenario, two investigators, blinded to participant group assignment, watched the video image in an adjacent room. Based on previous work, two observers were considered adequate to properly capture data of this type and achieve adequate interrater reliability.[12,14] They independently checked off tasks from the task list as they were performed and recorded whether participants identified the correct diagnosis. Interrater reliability values for the number of tasks completed and the number of participants making a correct diagnosis were assessed with a κ measure of agreement. Weighted task scores were defined as the sum of all correctly performed items on the weighted task score list. The number of correct diagnoses was defined as the number of correct diagnoses each participant made in the seven scenarios.

After each scenario, participants completed a comprehension questionnaire (appendix 2). The number of correct responses to comprehension questions was defined as the number of correct responses to questions presented in appendix 2 for each participant. After each scenario, participants also completed a self-assessment of their perceived physical and cognitive workload using the National Aeronautic Space Administration Task Load Index (NASA-TLX) questionnaire.[15] The NASA-TLX evaluated six areas: mental demand, physical demand, temporal demand, self-assessment of performance, self-assessment of effort, and self-assessment of frustration. From these six areas, a composite score was derived for each scenario.

Training interventions were of identical duration for both study groups; however, the teaching method differed between groups. The control group received standard didactic and simulation training. The experimental group received PTT-based didactic and VPT-based simulation-based sessions.

All participants in both groups received forty 45-min didactic sessions. The didactic sessions consisted of grand rounds, case conferences, textbook chapter reviews,[16] and visiting professor lectures. For 15 of the 40 didactic sessions, the control group received instruction on airway management, management of a difficult airway, cardiopulmonary physiology, and administration of anesthesia to patients with cardiopulmonary disease, and the experimental group received 15 PTT sessions focused on information from four competence areas: the American Society of Anesthesiologists Difficult Airway Algorithm,[17] a differential diagnosis for hypoxia, treatment options for each item in the differential diagnosis, and knowledge of normal ranges of cardiopulmonary variables and key relations between selected variables. Study sheets§§ were created for each of these areas and were distributed to all participants in both groups at the beginning of the 12-month intervention period.[16,18,19]

As a study aid, computerized flash cards were developed to review study sheet content (Java; Sun Microsystems Inc., Santa Clara, CA). A personal computer was used to automate question presentation and to record response time and accuracy. The flash cards were time limited, randomly reintroduced incorrectly answered questions, presented the correct response when an incorrect answer was entered, and recorded and presented the number of correctly answered questions.

Three types of flash cards were developed. The first type, "fill in the blank," introduced information by having participants look up answers to questions and then fill in blanks. After the first three PTT sessions, participants were asked to fill in the blank from memory. The second type, "qualitative assessment," was designed to solicit a qualitative evaluation of values from the study sheets (e.g., a qualitative responses of high, low, or normal for selected vital signs). The intent was to accelerate interpretation of physiologic monitor and ventilator values. The third type, "patient management," was designed to put into practice information contained from the study sheets. Participants were presented with cardiopulmonary data and were asked to comment on their state (high, low, or normal) and identify the most likely diagnosis and treatment consistent with the variable profile.

During each session, 25 min was allocated to flash card use. Participants were not allowed to use the electronic flash cards outside of the 15 didactic sessions dedicated to PTT. The remaining time was dedicated to small group discussions where participants were asked to recall information from the study sheets in the presence of their peers and discussion proctor. The goal of PTT was to achieve a level of mastery of the study sheet material such that it would be easily recalled during stressful moments.

All participants in both groups received five 90-min simulation sessions covering topics on the difficult airway, hypoxia, hypertension, tachycardia, bradycardia,

---

§§ The study sheets are available at: http://abl.med.utah.edu/ptt_vpt.html. Accessed December 27, 2007.

Table 2. Variable Priority Training Checklist

History *(check if considered)*
- ☐ Findings related to intraoperative patient course
- ☐ Findings related to potential therapeutic interventions

Physical *(check if considered)*

Airway
☐ Suction ETT	
☐ Secretions	(present, absent)
☐ Check ETT cuff pressure	(high, normal, low)
☐ Check ETT depth	(deep, normal, shallow)
☐ Check for ETT disconnect	(connected, disconnected)

Breathing
☐ Evaluate chest excursion	(equal bilaterally, unequal, none)
☐ Auscultate lung sounds	(normal, crackles, wheezes, none)

Circulation
☐ Check pulses	(present, absent)
☐ Evaluate skin color	(normal, abnormal)
☐ Estimated blood loss	(normal, abnormal)

Physiologic data *(check if considered)*
☐ Heart rate	(high, normal, low)
☐ Heart rhythm	(normal, abnormal)
☐ Spo_2	(high, normal, low)
☐ Blood pressure	(high, normal, low)
☐ End-tidal CO_2	(high, normal, low)
☐ Respiratory rate	(high, normal, low)
☐ Temperature	(high, normal, low)
☐ PA pressures	(high, normal, low)
☐ Wedge pressure	(high, normal, low)
☐ CVP	(high, normal, low)

Mechanical data *(check if considered)*

Settings
☐ Fio_2	
☐ Respiratory rate	
☐ Tidal volume	

Measured
☐ Fio_2	(high, normal, low)
☐ Respiratory rate	(high, normal, low)
☐ Tidal volume	(high, normal, low)
☐ Peak inspiratory pressure	(high, normal, low)

Key relations
☐ Relation of peak airway pressure to tidal volume	(high compliance, normal, low compliance)
☐ Relation of Fio_2 to oxygen saturation	(saturations higher, normal, lower than expected with given Fio_2)

CO_2 = carbon dioxide; CVP = central venous pressure; ETT = endotracheal tube; Fio_2 = inspired oxygen content; PA = pulmonary artery; Spo_2 = peripheral hemoglobin oxygen saturation measured by pulse oximetry.

and hypovolemia. Scenarios were preprogrammed into the simulator computer. Teaching objectives were standardized for each session. Participants were divided into smaller subgroups of five or six to facilitate simulation training. Each participant managed at least one adverse cardiopulmonary event per 90-min session while the remaining participants observed. After each scenario, a short debriefing was conducted to review participant performance. In the control group, instruction focused on the teaching objectives; in the experimental group, instruction focused on the teaching objectives using VPT.

During each adverse cardiopulmonary event, VPT consisted of participants reviewing four areas of patient data: relevant segments of the patient history, a targeted physical examination (airway, breathing, and circulation), physiologic data from the monitors, and mechanical ventilation data. Participants were trained to quickly

go through all the items contained within the VPT checklist (table 2). At the same time, participants were asked to synthesize and order a differential diagnosis during data collection and prioritize their therapeutic interventions according to the differential diagnosis. The main purpose of this technique was to ensure that information from all four areas of patient data was reviewed, to allocate additional attention as needed to accurately describe abnormal findings, and to reduce the likelihood of cognitive tunnel vision, all while managing an adverse event.

For the first two simulation sessions, participants were allowed to look at table 2. During the last three simulation sessions, they were asked to use table 2 from memory. After each scenario, participants who observed their peer managing an adverse event provided an item-by-item critique of their peer's performance using table 2. The goal was to ensure that all available data were

considered in a timely and consistent manner and to ensure that participants flexibly distributed attention to abnormal findings.

After the 12-month training period, participants underwent an assessment of their skill in managing seven adverse events as described for the baseline assessment. The vignettes used to introduce the participant to each scenario were altered. The adverse events remained the same as those used in the baseline analysis.

Four metrics were compared for the effect of teaching method: weighted task scores, the number of correct diagnoses, the number of correct responses to comprehension questions, and the NASA-TLX scores using statistical software (Statview, version 5.1; SAS Institute, Cary, NC). Weighted task scores were compared with a repeated-measures multivariate analysis of variance (MANOVA). Response variables were the weighted task scores for each scenario before and after 1 yr of training. Interactions between teaching method, results before and after 1 yr of training, and the seven scenarios were explored. MANOVA statistical tests were performed with the Roy largest root criteria transformed to an F statistic; statistical significance was declared for $\alpha < 0.05$. If MANOVA tests were significant, a *post hoc* analysis between teaching methods by individual scenario was performed using a Bonferroni/Dunn test.

The numbers of correct diagnoses for all scenarios were compared between teaching methods with a Kruskal–Wallis test. If significant, paired comparisons within each group before and after training were made with a Wilcoxon signed rank test, and unpaired comparisons between groups before and after training were made with a Mann–Whitney U test. To account for multiple comparisons, a Bonferroni-corrected P value less than 0.01 was required to declare significance at a nominal α of 0.05. A similar approach was taken with the number of correct responses to comprehension questions. Composite NASA-TLX scores for all scenarios were compared between groups with repeated-measures MANOVA.

Results

Of the 22 participants enrolled, 1 did not complete the study. That participant pursued training in a different specialty. Eleven (3 female, 8 male) and 10 (3 female, 7 male) participants were randomly assigned to the control and experimental groups. No difference in US Medical Licensing Examination parts I, II, and III and preliminary year in-training examination results were observed between groups (P values of 0.712, 0.297, 0.609, and 0.182, respectively). All participants completed the baseline assessment, the simulation sessions, the didactic sessions, and the postintervention assessment.

Using κ as measure of agreement between two observers, there was substantial interobserver agreement (κ between 0.6 and 0.79)[20]; interrater reliability for the task list scoring and correct diagnoses were 0.70 and 0.82 at baseline and 0.72 and 0.86 after 12 months of training.

Weighted task scores for each scenario before and after training are presented in table 3. The repeated-

Table 3. Weighted Task Scores and Number of Correct Diagnoses for Each Scenario before and after Training for Control and Experimental Groups

	Before Training		After Training	
	Control	Experimental	Control	Experimental
Scenario 1 (adult): unanticipated difficult airway				
Weighted task scores	32 ± 3	27 ± 4	31 ± 2	33 ± 2
Diagnostic accuracy	82%	50%	36%	70%
Scenario 2 (adult): bronchospasm on emergence				
Weighted task scores	26 ± 3	29 ± 2	48 ± 3	55 ± 2
Diagnostic accuracy	27%	50%	73%	70%
Scenario 3 (adult): development of a tension pneumothorax				
Weighted task scores	9 ± 2	5 ± 2	24 ± 2	33 ± 2
Diagnostic accuracy	9%	10%	45%	50%
Scenario 4 (adult): aspiration after induction				
Weighted task scores	28 ± 3	31 ± 3	35 ± 2	46 ± 2
Diagnostic accuracy	45%	40%	64%	70%
Scenario 5 (child): esophageal intubation				
Weighted task scores	33 ± 5	34 ± 4	36 ± 3	46 ± 3
Diagnostic accuracy	45%	70%	91%	80%
Scenario 6 (child): laryngospasm, bronchospasm, no intravenous access				
Weighted task scores	29 ± 4	27 ± 3	48 ± 5	53 ± 3
Diagnostic accuracy	9%	20%	64%	90%
Scenario 7 (child): faulty ventilator circuit				
Weighted task scores	28 ± 4	30 ± 5	37 ± 2	41 ± 3
Diagnostic accuracy	27%	30%	55%	80%

Weighted tasks scores are presented as mean $\pm$ SE. The number of participants with a correct diagnosis (diagnostic accuracy) are presented as percentages of the total number of participants in each group (11 in the control group and 10 in the experimental group).
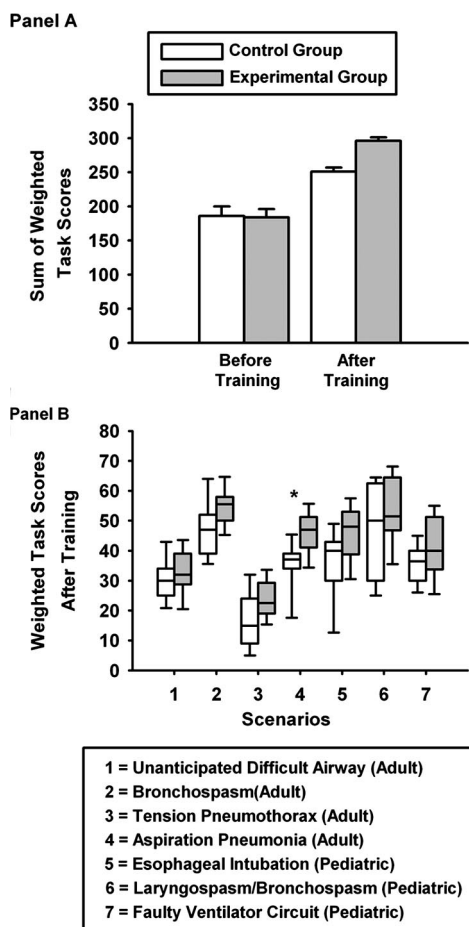
Panel A

Panel B

Fig. 1. (*A*) Weighted task scores by group for all scenarios combined. Data are presented as mean ± SE before and after the 12-month training period. (*B*) Box plot of weighted task scores after 12 months of training by scenario. *$P < 0.005$. $P > 0.05$ for all other results. The *center line* is the median, the *lower* and *upper boundaries* are the 25th and 75th percentiles, and the *error bars* are the 5th and 95th percentiles.

measures MANOVA revealed that both groups demonstrated a significant increase in their weighted task scores after 12 months of training ($P < 0.0001$); there was a significant difference between groups after 12 months of training ($P = 0.014$; fig. 1A). A *post hoc* analysis of the after training weighted task scores by scenario indicated significant difference between groups in scenario 4, aspiration pneumonia ($P = 0.005$; fig. 1B).

The numbers of correct diagnoses for each scenario before and after training are presented in table 3. Participants improved their number of correct diagnoses from baseline to 12 months after training ($P < 0.001$). After training, there was no difference between groups in determining the correct diagnosis (median count of 5 out of 7 and 4 out of 7 for the experimental and control groups, respectively; $P = 0.249$; fig. 2A).

Participants improved their number of correct responses to comprehension questions from baseline to 12
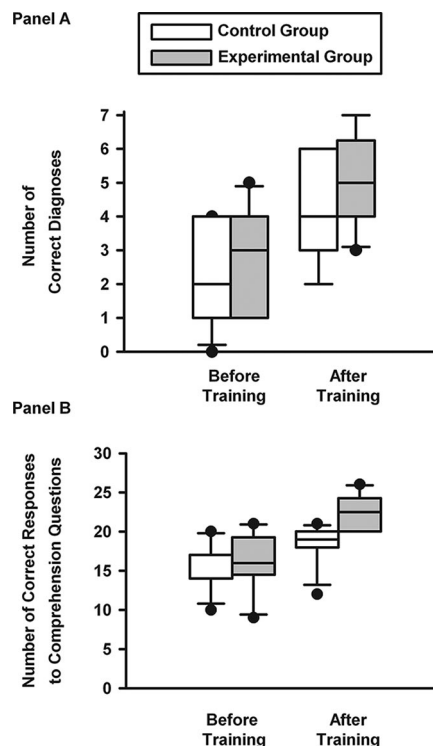
Panel A

Panel B

Fig. 2. (*A*) Box plots of the number of correct diagnoses. (*B*) Number of correct responses to comprehension questions by group before and after the 12-month training period for all scenarios combined. The *center line* is the median, the *lower* and *upper boundaries* are the 25th and 75th percentiles, the *error bars* are the 5th and 95th percentiles, and the *dots* are data outside the 5th and 95th percentiles.

months after training ($P < 0.0001$; fig. 2B). Both groups answered approximately 15 of the questions correctly at baseline. After training, the experimental group completed more questions correctly (median count of 23 out of 30) than participants in the control group (median count of 19 out of 30; $P < 0.001$).

Participants reported a decrease in their composite NASA-TLX scores of perceived workload between baseline and posttraining ($P < 0.001$); however, there was no difference between groups ($P = 0.259$; fig. 3).
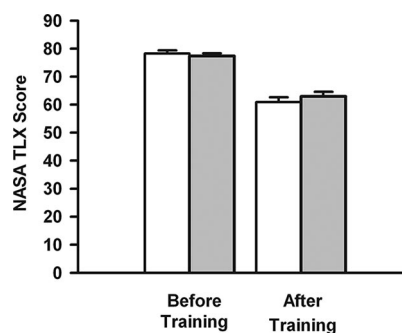

Fig. 3. National Aeronautic Space Administration Task Load Index (NASA-TLX) scores of perceived workload are presented by group at baseline and after 12 months of training for all scenarios combined. Data are presented as mean ± SEM.

Discussion

We explored how PTT and VPT techniques applied over a 12-month period would improve CA-1 management of adverse airway and respiratory events. The results in part confirmed our hypotheses. Participants with PTT and VPT outperformed those in the control group in two of three performance metrics. In the remaining one and the assessment of perceived workload, there was no difference between groups.

After a year of training, we observed an improvement in all performance metrics in both groups. Whether this is a function of the clinical training, didactic training, or simulation-based training is difficult to discern. This finding also suggests that the measurement methods used are a valid construct of assessing resident performance in managing adverse airway and respiratory events.

Participants in the experimental group were able to complete on average 9% more tasks than those in the control group. Those trained with PTT and VPT may have allocated their attention more effectively and were mentally prepared to assess and prioritize the information available to them in an efficient manner. This may be a function of an efficient yet complete survey of all available information and use of easily retrieved mental templates to organize and use information to implement therapeutic interventions. Another potential explanation for this finding is that the PTT and VPT were directly applicable to managing the seven simulated adverse events and may be a function of the focused training time rather than the training techniques.

A large improvement in the number of correct diagnoses (fig. 2) was observed from baseline (35–39% correct) to the end of training (61–73%) independent of group assignment. A subtle trend suggested that PTT and VPT may improve diagnostic performance, but the trend was not significant. Interpretation of this result is difficult because in several instances the correct diagnosis was mentioned amid verbalization of a list of potential diagnoses. Furthermore, participants may have known the correct diagnosis but neglected the "talk aloud" protocol. One nuance of this analysis is the possibility that VPT may have slowed participants in the experimental group in reaching a diagnosis within 7 min. Although we did not ask participants in the experimental group whether they used VPT to reach their diagnosis, it is conceivable that going through the items in table 2 would prolong the time required to reach a diagnosis.

Although there is overlap, clinical task completion and diagnostic accuracy differentiate two important aspects of adverse event management. Clinical task completion reveals what participants do while managing an adverse event under stressful conditions but does not explicitly reveal their diagnostic assumptions. By contrast, diagnostic accuracy reveals the participant's verbal acknowledgment of the diagnosis, whereas it does not describe what the participant does with that information. Our results indicate that PTT and VPT improved task performance under stressful conditions but did not improve diagnostic accuracy.

After 12 months of training, perceived workload decreased by 30% from baseline. This is an expected finding given that participants, with a year of clinical experience, are more likely to have a better subjective belief about their performance than after 1 week of residency. PTT and VPT made no impact on perceived workload. It is interesting to note that additional work generated by participants using cognitive aids developed through PTT and VPT did not contribute to increasing the perceived workload.

Given that participants in the experimental group improved their number of correct responses to comprehension questions from baseline more than the control group suggests that PTT and VPT allowed participants to collect and retain more pertinent information in regard to the adverse event. This may be a direct result of participants in the experimental group completing more diagnostic and therapeutic tasks. In so doing, they gained a better understanding of the adverse events and were able to answer more questions correctly.

In terms of study limitations, general concerns with simulation-based evaluations include the difficulty of creating scenarios that mimic real patients,[21] controlling for atypical participant behavior (*i.e.*, a hypervigilant or cavalier participant),[22] and controlling for scenario recognition from the beginning of the year to the end of the year. In this study, the scenario vignettes were changed, but the adverse events were identical.

An additional confounder to this study is that conventional training may already contain many elements of PTT and VPT, diminishing the usefulness of these techniques. Although we found significant improvements with PTT and VPT, the long-term benefit beyond 1 yr is unknown. Furthermore, the 10 PTT and 5 VPT sessions were spread out over 12 months and led to long time periods between training sessions. Shortening the time between training sessions may have improved use of these techniques during the end-of-year testing phase of this study.

When interpreting results presented in figure 1, participants who performed lifesaving maneuvers without completing all of the tasks for a given scenario listed in table 1 may have been inadvertently penalized. For example, should a participant have immediately recognized an esophageal intubation and immediately intubated the trachea without recognizing low saturation, absence of carbon dioxide, or failing to auscultate the lungs, the participant would have successfully managed the adverse event but would have been penalized for not completing several tasks. This limitation underscores the importance of using the Delphi method to weight absolutely essential tasks higher than other useful, but not

essential, tasks. In addition, this limitation emphasizes the need to use other assessment methods to mitigate this potential limitation.

In summary, we implemented PTT and VPT techniques aimed at improving management of adverse airway and respiratory events as part of CA-1 training. The goal of these training techniques was to develop a flexible cognitive style that would increase vigilance and produce a high level of automaticity during critical events. After a year of didactic and simulation-based training, resident performance significantly improved in both groups. PTT and VPT led to modest improvements in performance when compared with conventional training. Further work is warranted to explore the potential value of using innovative teaching techniques such as PTT and VPT to better prepare anesthesia personnel for their role in managing adverse events.

References

1. Gopher D, Weil M, Bareket T: Transfer of skill from a computer game trainer to flight. Hum Factors 1994; 36:387–405

2. Whightman D, Lintern G: Part-task training for tracking and manual control. Hum Factors 1985; 27:201–312

3. Lintern G, Wickens C: Issues for acquisition and transfer of timesharing and dual-task skills, Multiple Task Performance. Edited by Damos D. London, Wiley, 1991, pp 123–38

4. Damos D: Daul task methodology: some common problems, Multiple Task Performance. Edited by Damos D. London, Wiley, 1991, pp 101–19

5. Kieras DE, Meyer DE, Ballas JA, Lauber EJ: Modern computational perspectives on executive mental processes and cognitive control: Where to from here? Control of Cognitive Processes: Attention and Performance XVIII. Edited by Monsell S, Driver J. Cambridge, MIT Press, 2000, pp 681–712

6. Morray JP, Geiduscheck JM, Caplan RA, Posner KL, Gild WM, Cheney FW: A comparison of pediatric and adult anesthesia closed malpractice claims. ANESTHESIOLOGY 1993; 78:461–7

7. Cheney FW, Posner KL, Caplan RA: Adverse respiratory events infrequently leading to malpractice suits: A closed claim analysis. ANESTHESIOLOGY 1991; 75:932–9

8. Caplan RA, Posner KL, Ward RJ, Cheney FW: Adverse respiratory events in anesthesia: A closed claim analysis. ANESTHESIOLOGY 1990; 72:828–33

9. Gaba D: What makes a "good" anesthesiologist? ANESTHESIOLOGY 2004; 101:1061–3

10. 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

11. Murray DJ, Boulet JR, Kras JF, Woodhouse JA, Cox T, McAllister JD: Acute care skills in anesthesia practice: A simulation-based resident performance assessment. ANESTHESIOLOGY 2004; 101:1084–95

12. Scavone BM, Sproviero MT, McCarthy RJ, Wong CA, Sullivan JT, Siddall VJ, Wade LD: Development of an objective scoring system for measurement of resident performance on the human patient simulator. ANESTHESIOLOGY 2006; 105:260–6

13. Clayton M: Delphi: A technique to harness expert opinion for critical decision making tasks in education. Educ Psychol 1997; 17:373–86

14. Weller JM, Bloch M, Young S, Maze M, Oyesola S, Wyner J, Dob D, Haire K, Durbridge J, Walker T, Newble D: Evaluation of high fidelity patient simulator in assessment of performance of anaesthetists. Br J Anaesth 2003; 90:43–7

15. Hart S, Staveland L: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research, Human Mental Workload. Edited by Hancock P, Meshkati N. Amsterdam, North Holland, Elsevier Science, 1988, pp 139–83

16. Morgan G, Murray M, Mikhail M: Respiratory physiology: The effects of anesthesia, Clinical Anesthesiology, 4th edition. Edited by Morgan G, Murray M, Mikhail M. New York, Lange Medical Books/McGraw-Hill Medical Publishing Division, 2005, pp 537–70

17. Practice guidelines for management of the difficult airway: An updated report by the American Society of Anesthesiologists Task Force on Management of the Difficult Airway. ANESTHESIOLOGY 2003; 98:1269–77

18. Guyton AC: Cardiac output, venous return, and their regulation, Textbook of Medical Physiology, 7th edition. Edited by Guyton AC. Philadelphia, WB Saunders, 1986, pp 272–86

19. Marino PL: Oxygen transport, The ICU Book, 2nd edition. Edited by Marino PL. Philadelphia, Lea & Febiger, 1991, pp 14–24

20. Landis J, Koch G: The measurement of observer agreement for categorical data. Biometrics 1977; 33:159–74

21. Gaba DM: Two examples of how to evaluate the impact of new approaches to teaching. ANESTHESIOLOGY 2002; 96:1–2

22. Wong AK: Full scale computer simulators in anesthesia training and evaluation. Can J Anaesth 2004; 51:455–64

Appendix 1: Modified Delphi Analysis to Develop a Weighted Task Score List for Each Scenario

Anesthesiology faculty at the authors' institution created a list of appropriate diagnostic and therapeutic tasks for each of the seven scenarios. Three board-certified anesthesiologists from the authors' institution different from those who created the lists and three from outside institutions were independently and iteratively asked to rank tasks according to importance using a Likert scale, where 1 = not important and 5 = extremely important. Median and interquartile ranges for each task and suggested additions and eliminations to the task list were redistributed to the experts. Experts were then asked to consider modifying their rankings that deviated from the median or justify their reasons for not changing scores. This process was repeated until an acceptable level of concordance between experts (overall concordance >0.75 using a Kendall W statistic) was observed. The content validity index of the task list was calculated as the percentage of tasks rated by the experts as 4 or 5. The task lists were weighted from 1 to 5 based on median rankings from the final round.

Acceptable concordance among experts was reached within three rounds. The Kendall coefficients of concordance (Kendall W) for rounds 1, 2, and 3 were 0.312, 0.532, and 0.752. The final scoring system consisted of 110 tasks weighted for importance with a total score of 520.5. The content validity of the final weighted task list as a percentage of items ranked greater than 3 was very high (95%) (table 4).

Appendix 2: Comprehension Questions for Each Scenario

Correct answers are presented in **bold**.

Adult Scenario 1: Unanticipated Difficult Airway

1. The room air oxygen hemoglobin saturation was (**low**, normal, high).
2. The arterial partial pressure of oxygen was (**low**, normal, high).
3. The relation between the room air oxygen saturation and arterial partial pressure of oxygen is (normal, **suggestive of an intrapulmonary shunt**, suggestive of an enlarged function residual capacity).
4. Structures or conditions that impaired successful intubation of this patient include all of the following: (**tongue**, **posterior pharynx**, laryngospasm, bronchospasm, nasopharyngeal airway).
5. Transtracheal jet ventilation would have been a logical next step in airway management (true, **false**).

Adult Scenario 2: Bronchospasm on Emergence and Development of a Tension Pneumothorax

1. After extubation, airway resistance was (low, normal, **high**).
2. After reintubation, the patient developed (endobronchial intubation, fat embolism, negative-pressure pulmonary edema, an increase in the vital capacity, **none of these**).
3. After reintubation, the peak airway pressure was (low, normal, **high**) and the measured tidal volume was (**low**, normal, high).

Table 4. Weighted Task Score Lists for Each Scenario from the Final Round of the Modified Delphi Analysis

Adult scenario 1: unanticipated difficult airway	
Detects enlarged tongue	4 (4, 4)
Detects swollen posterior pharynx	4 (4, 4)
Recognizes low Sp_{O_2}	5 (5, 5)
Recognizes inadequate ventilation	5 (5, 5)
Recognizes an unanticipated difficult airway	5 (5, 5)
Calls for help	5 (5, 5)
Ensures F_{IO_2} set to 100%	5 (5, 5)
Attempts laryngoscopy (blade enters mouth)	5 (5, 5)
Mask ventilation after first laryngoscopy attempt	5 (5, 5)
Second laryngoscopy attempt (different blade)	4 (4, 4)
Mask ventilation after second laryngoscopy attempt	5 (5, 5)
Places LMA and inflate cuff	5 (5, 5)
Awakens patient	5 (4.5, 5)
Adult scenario 2A: bronchospasm on emergence	
Auscultates pulmonary fields	5 (5, 5)
Detects decreased breath sounds with wheezing	5 (5, 5)
Recognizes low Sp_{O_2}	5 (5, 5)
Detects high peak airway pressure	5 (5, 5)
Detects low tidal volume	4 (4, 4)
Recognizes severe bronchospasm	5 (5, 5)
Calls for help	5 (5, 5)
Ensures F_{IO_2} set to 100%	5 (5, 5)
Suctions oral airway	3 (3, 3.5)
Attempts manual ventilation	5 (5, 5)
Administers intravenous epinephrine	5 (4.5, 5)
Considers terbutaline, lidocaine, and/or ketamine	3 (3, 4)
Administers a sedative hypnotic and muscle relaxant	5 (3.5, 5)
Laryngoscopy, endotracheal intubation, and inflates cuff	5 (4.5, 5)
Administers β agonists inhaler	5 (5, 5)
Uses inhaler circuit adapter	4 (4, 4)
Administers potent inhaled agent	4 (4, 4)
Adult scenario 2B: development of a tension pneumothorax	
Auscultates pulmonary fields	5 (5, 5)
Detects decreased breaths sounds on left side	5 (5, 5)
Recognizes low Sp_{O_2}	5 (5, 5)
Detects high peak airway pressure	5 (5, 5)
Detects low tidal volume	4 (3.5, 4)
Detects presence of a left tension pneumothorax	5 (5, 5)
Calls for help	5 (5, 5)
Ensures F_{IO_2} set to 100%	5 (5, 5)
Performs a needle decompression of left chest	5 (5, 5)
Provides mechanical ventilation	4 (4, 4)
Adult scenario 3: aspiration after induction	
Auscultates pulmonary fields	5 (5, 5)
Detects course breath sounds with crackles	5 (4.5, 5)
Recognizes low Sp_{O_2}	5 (5, 5)
After intubation, detects high peak airway pressures	4 (4, 4)
After intubation, detects low tidal volume	4 (4, 4)
Calls for help	5 (5, 5)
Ensures F_{IO_2} set to 100%	5 (5, 5)
Places patient in Trendelenburg and turns head	4 (4, 5)
Suctions oral airway	5 (5, 5)
Attempts mask ventilation	4 (4, 4)
Administers a sedative hypnotic and muscle relaxant	4 (4, 4)
Laryngoscopy and intubation to include inflation of the endotracheal tube cuff	5 (5, 5)
Suctions endotracheal tube	5 (5, 5)
Provides mechanical ventilation	5 (5, 5)
Sets positive end-expiratory pressure to 5 cm H_2O	5 (4.5, 5)
Considers intravenous furosemide 5–20 mg	2 (2, 2)

(continued)

Table 4. Continued

Pediatric scenario 1: esophageal intubation	
Checks endotracheal tube depth	>5 (5, 5)
Checks endotracheal tube cuff pressure	3 (3, 3.5)
Laryngoscopy—validates endotracheal tube placement	5 (5, 5)
Auscultates pulmonary fields	5 (5, 5)
Detects absence of breath sounds	5 (5, 5)
Recognizes low Sp_{O_2}	5 (5, 5)
Recognizes minimal or no end-tidal CO_2	5 (5, 5)
Detects no tidal volume with manual ventilation	5 (4.5, 5)
Recognizes esophageal intubation	5 (5, 5)
Calls for help	5 (4.5, 5)
Ensures F_{IO_2} set to 100%	5 (5, 5)
Removes esophageal tube	5 (5, 5)
Mask ventilates to improve oxygen saturation	5 (5, 5)
Laryngoscopy, intubates trachea, and inflate cuff	5 (5, 5)
Suctions stomach	5 (4, 5)
Pediatric scenario 2: laryngospasm, bronchospasm, no intravenous access	
Auscultates pulmonary fields	5 (5, 5)
Detects decreased breath sounds and wheezing	5 (5, 5)
Recognizes low Sp_{O_2}	5 (5, 5)
Recognizes tachycardia	4 (4, 4)
Recognizes minimal or no end-tidal CO_2	5 (5, 5)
Detects high peak airway pressures	5 (5, 5)
Detects no tidal volume with manual ventilation	5 (5, 5)
Recognizes inadequate ventilation	5 (5, 5)
Recognizes laryngospasm	5 (5, 5)
Recognizes bronchospasm	5 (5, 5)
Calls for help	5 (5, 5)
Ensures F_{IO_2} set to 100%	5 (5, 5)
Attempts manual ventilation	5 (5, 5)
Administers subcutaneous epinephrine or terbutaline	3 (3, 4)
Administers intramuscular succinylcholine/atropine	5 (5, 5)
Performs laryngoscopy after succinylcholine	5 (4.5, 5)
Intubates trachea and inflates cuff	5 (4.5, 5)
After intubation, administers β agonist inhaler	5 (5, 5)
Uses circuit adapter for inhaler	4.5 (4, 5)
Suctions endotracheal tube	4 (3.5, 4)
Pediatric scenario 3: faulty ventilator circuit	
Checks endotracheal tube depth	5 (5, 5)
Checks endotracheal tube cuff pressure	5 (5, 5)
Laryngoscopy to validate endotracheal tube placement	5 (5, 5)
Auscultates pulmonary fields	5 (5, 5)
Detects normal breath sounds	5 (5, 5)
Recognizes low Sp_{O_2}	5 (5, 5)
Recognizes minimal or no end-tidal CO_2	5 (5, 5)
Detects low peak airway pressure (<10 cm H_2O)	5 (5, 5)
Detects low tidal volume (<500 ml)	5 (5, 5)
Recognizes inadequate ventilation	5 (5, 5)
Recognizes anesthesia circuit has an air leak	5 (5, 5)
Replaces ventilator circuit	5 (5, 5)
Uses Jackson Reese circuit to ventilate patient	5 (5, 5)
Calls for help	5 (5, 5)
Ensures F_{IO_2} set to 100%	5 (5, 5)
Increases tidal volume	4 (4, 4)
Increases respiratory rate	3 (2.5, 3)
Increases inspiratory flow rate	5 (4, 5)
Increases oxygen fresh gas flow	5 (5, 5)

Data are presented as median (25th, 75th quartiles).

CO_2 = carbon dioxide; F_{IO_2} = fraction of inspired oxygen; LMA = laryngeal mask airway; Sp_{O_2} = percentage of oxygen saturation in peripheral blood.

4. Before extubation, the depth of the endotracheal tube was (deep, **normal**, shallow).
5. Auscultation of lung sounds after reintubation revealed (course breath sounds, **wheezing**, absent breath sounds, normal breath sounds).

Adult Scenario 3: Aspiration after Induction

1. After intubation, the delivered tidal volume was (**low**, normal, high).
2. The breath sounds were (normal, **coarse**, wheezing, absent).
3. The pulmonary compliance was (**low**, normal, high).
4. The airway resistance was (low, **normal**, high).
5. After aspiration, bronchoalveolar lavage is warranted (true, **false**).

Pediatric Scenario 1: Esophageal Intubation

1. The peak airway pressure after intubation by the anesthesia resident you were supervising was (low, **normal**, high).
2. The endotracheal tube cuff was intact (**true**, false).
3. The end-tidal carbon dioxide after intubation by the anesthesia resident you were supervising was (**low**, normal, high).
4. The delivered tidal volume after intubation by the anesthesia resident you were supervising was (**low**, normal, high).

5. The most likely source of hypoxia was (**inadequate ventilation**, bronchospasm, laryngospasm, pulmonary embolism, carboxyhemoglobin).

Pediatric Scenario 2: Laryngospasm, Bronchospasm, No Intravenous Access

1. After intubation, the peak airway pressure was (low, normal, **high**).
2. Bronchospasm is an example of (dead space, **intrapulmonary shunt**, chest wall rigidity, aspiration of foreign body).
3. After intubation, the delivered tidal volume was (**low**, normal, high).
4. The inspiratory flow rate was set at (low, **medium**, high).
5. The patient became hypoxic because of a ventilator malfunction (true, **false**).

Pediatric Scenario 3: Faulty Ventilator Circuit

1. The peak airway pressure was (**low**, normal, high).
2. The delivered tidal volume was (**low**, normal, high).
3. The inspiratory flow rate was set at (low, **medium**, high).
4. The endotracheal tube cuff was intact (**true**, false).
5. The wall oxygen source was low (true, **false**).