Published in final edited form as:

Simul Healthc. 2018 February ; 13(1): 16–26. doi:10.1097/SIH.0000000000000271.

A Validation Argument for a Simulation-based Training Course centered on Assessment, Recognition and Early Management of Pediatric Sepsis

Gary L. Geis, MD¹ [Associate Professor], Derek S. Wheeler, MD¹ [Chief of Staff], Amy Bunker, PhD² [Assistant Professor], Laura G. Militello, MA³, Regina G. Taylor, CRC¹, Jerome P. Bauer, RN¹, Terri L. Byczkowski, PhD¹ [Associate Professor], Benjamin T. Kerrey, MD, MS¹ [Associate Professor], and Mary D. Patterson, MD, MEd⁴ [Professor]

¹Cincinnati Children's Hospital Medical Center

²University of Cincinnati Medical Center

³Applied Decision Science, LLC

⁴Children's National Medical Center

Abstract

Introduction—Early recognition of sepsis remains one of the greatest challenges in medicine. Novice clinicians are often responsible for the recognition of sepsis and the initiation of urgent management. The objective of this study is to create a validity argument for the use of a simulation-based training course centered on assessment, recognition, and early management of sepsis in a lab-based setting.

Methods—Five unique simulation scenarios were developed integrating critical sepsis cues identified through qualitative interviewing. Scenarios were piloted with groups of novice, intermediate and expert pediatric physicians. The primary outcome was physician recognition of sepsis, measured with an adapted situation awareness global assessment tool. Secondary outcomes were physician compliance with pediatric advanced life support (PALS) guidelines and early sepsis management (ESM) recommendations, measured by two internally derived tools. Analysis compared recognition of sepsis by levels of expertise and measured association of sepsis recognition with the secondary outcomes.

Results—Eighteen physicians were recruited, 6 per study group. Each physician completed 3 sepsis simulations. Sepsis was recognized in 19 (35%) of 54 simulations. The odds that experts recognized sepsis was 2.6 (95% CI 0.5, 13.8) times greater than novices. Adjusted for severity, for every point increase in the PALS global performance score, the odds that sepsis was recognized increased by 11.3 (95% CI: 3.1, 41.4). Similarly, the odds ratio for the PALS checklist score was

Corresponding Author Information: Gary L. Geis, MD, Associate Professor, Department of Pediatrics, University of Cincinnati College of Medicine, Medical Director, Center for Simulation and Research, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, MLC 12000, Cincinnati, OH 45229-3039, Phone (513) 636-9869; fax (513) 636-1173; gary.geis@cchmc.org.

Financial Disclosure Summary

None of the authors have financial disclosures or conflicts of interest, except Mary Patterson who performs occasional consulting for SimHealth Group. This research was funded by the Agency for Healthcare Research and Quality, grant #1R18HS020455-01.

1.5 (95% CI: 0.8, 2.6). Adjusted for severity and level of expertise, the odds of recognizing sepsis was associated with an increase in the ESM checklist score of 1.8 (95% CI: 0.9, 3.6) and an increase in ESM global performance score of 4.1 (95% CI: 1.7, 10.0).

Conclusions—Although incomplete, evidence from initial testing suggests that the simulations of pediatric sepsis were sufficiently valid to justify their use in training novice pediatric physicians in the assessment, recognition and management of pediatric sepsis.

Keywords

expertise; novice; pediatric; resident; sepsis; simulation

INTRODUCTION

The early recognition of sepsis remains one of the greatest challenges in medicine today.¹ The incidence of sepsis has been climbing at a rate of 1.5% annually, with the cost of care approaching \$17 billion annually in the United States.^{2–4} In a recent international study of critically ill children, the prevalence of severe sepsis was comparable to adults at 8.2%.⁵ Failure to recognize the signs and symptoms of sepsis and institute timely and appropriate care leads to higher mortality.^{6–10} Several studies in children with sepsis suggest that early recognition and treatment with antibiotics and fluid resuscitation improves outcomes.^{11–16}

Sepsis is primarily diagnosed clinically based on a constellation of signs and symptoms. In early stages, when treatment is most effective, sepsis is virtually indistinguishable from more benign febrile illnesses. Early recognition of sepsis requires a heightened awareness, increased index of suspicion, and traditionally a substantial amount of clinical experience.^{17–19} In the academic ward setting, like our institution, the novice clinician is typically responsible for the recognition of often subtle signs of sepsis and the initiation of urgent management. Failure to recognize clinical deterioration results in unacceptable delays, culminating in the transfer of a moribund patient to the ICU.^{20,21}

Simulation-based training (SBT) appears to help prevent participants from making clinical mistakes by providing a safe setting in which it is permissible to make mistakes and learn from them.²² Simulation, when utilized for sepsis education, has been shown to improve sepsis- and shock-related knowledge and improved septic shock diagnosis and management in the lab setting.^{22–28} Additionally, in a hospital utilizing in situ simulation, a checklist showed increased compliance with first-hour sepsis related care tasks in comparison with two hospitals where simulation was not utilized (61.7% vs. 23.1%, $p < 0.01$).²⁹

In the first phase of a planned three-phase investigation, we utilized the Critical Decision Method interviewing technique and qualitative analysis to identify and classify behaviors that characterized and differentiated the expert (faculty physicians) from the novice (intern physicians) in the recognition of sepsis at the bedside.³⁰ The Critical Decision Method (CDM) technique utilizes cognitive probes in semi-structured interviews to elicit information about how experts formulate their decision-making strategies. The objective of the current study (second phase) was to create a validity argument for the use of a simulation-based training course centered on assessment, recognition, and early

management of sepsis in a lab-based setting. The planned third stage will be to explore and assess the training course in a larger cohort of novice physicians.

METHODS

This study was reviewed and approved by our institutional review board. To create a validity argument for simulation as a strategy, we needed to create a series of processes (scenario development and testing, outcome measure development and implementation, etc.) to make judgements. To make these judgements, we needed to understand the strengths and limitations of the processes upon which decisions were based. Stated differently, we required evidence to support the validity of our decisions.³¹ We chose to use a validation theory, articulated by Kane, which prioritizes evidence by highlighting key phases or inferences in planning. As described by Cook et al, Kane's validity argument guides the collection and interpretation of validity evidence.³¹ First, an educator must consider the decision at hand and a proposed interpretation that would support that decision. Next, with the desired decision in mind, the educator identifies the key claims, assumptions and inferences associated with this interpretation and use; Kane labels this the 'interpretation/use argument'. The educator then develops a plan to test these assumptions and inferences. Finally, guided by this plan, he or she collects empiric evidence from multiple sources and evaluates this evidence into a validity argument.

Using data from our initial study,³⁰ we developed five pediatric sepsis scenarios incorporating perceptual cues utilized by expert clinicians. We designed scenarios to be accessible to novices, but also challenging and requiring the detection and interpretation of subtle sepsis cues and clusters of cues. We developed one "garden path" simulation, two scenarios of compensated sepsis, and two scenarios of uncompensated septic shock. Garden path scenarios are ones in which the initial presentation suggests a straightforward case. As it unfolds, additional information is presented which is not consistent with the initial diagnosis. Generally, experts are able to recognize the incongruity and adapt to new information, in the process revising their mental models and plans. Novice clinicians are more constrained in their initial hypothesis and do not recognize or reconcile the atypical elements. A sixth simulation, a supraventricular tachycardia (SVT) scenario unrelated to sepsis, was developed to limit the bias of participant anticipation of sepsis. The six scenarios (see Figure, Supplemental Digital Content 1, which shows sepsis scenario content and flow) were reviewed by local experts in emergency medicine and critical care medicine and revised based on their feedback.

The simulation facility is a 400 square foot, on-campus simulation lab, including 2 simulation rooms (one emergency department setting, one inpatient ward setting), a control room separated from the simulation rooms by one-way mirrors, and a debriefing room. Simulation rooms were set up to replicate the intended clinical setting. For example, the emergency department (ED) room included an ED stretcher, bedside monitor, IV pole with intravenous (IV) infusion pump, bedside table, and head wall with air, oxygen and suction capabilities. Non-invasive airway and oxygen equipment were available on a separate wall, arranged as they would be in the ED. Additional resuscitation equipment (e.g. IV access supplies, laryngoscopes, defibrillator) was available in the hall outside the room, as in the

ED setting that equipment is not standard for a non-trauma/resuscitation room. As another example, when IV access was achieved and fluids or medications were ordered, those were obtained from outside the room and delivered in real-time into a fluid collection system (IV catheter, IV tubing and foley catheter bag). Participants could not just verbalize “I have given 20 mL/kg of normal saline,” but instead had to instruct the bedside nurse to or themselves deliver the fluid through the IV catheter before any “response” was achieved by the simulated patient.

The simulation lab represents an environment in which we are able to place residents in a role in which they need to seek and interpret available cues, rather than rely on a senior, more experienced physician. We included critical cues and cue clusters frequently mentioned in interviews, during our initial study, as important indicators of sepsis.³⁰ It is common in SBT to provide participants with findings from radiographs, ECGs, or distal perfusion by simply relaying the results. In contrast, we presented the learner with raw data to facilitate the development of perceptual skills and pattern recognition in the context of a challenging incident, e.g., we presented the physician with the radiographic image so that s/he could interpret it first-hand. Classic indicators of sepsis were relatively easy to incorporate by a single value or a trend over time, and the monitors used to measure and display these cues are generally part of simulations. However, some cues required innovative strategies, such as wrapping the hands and feet of the mannequin in ice prior to training to simulate cold extremities. To improve the fidelity of mental status cues, we used a voice modulator that allowed us to speak through the mannequin in the voice of an infant, a toddler, or school-aged child, depending on the scenario. Videos and screen shots were used to depict changes in skin appearance and distal perfusion; requiring the physician to interpret the image and independently draw his/her own conclusions. Another requirement was authentic timing and urgency, as several CDM incidents included situations in which a patient’s condition started at an early point in sepsis and slowly deteriorated, making it easier to miss subtle cues. Thus, we designed compensated sepsis scenarios with timelines more representative of early sepsis.

To provide consistency, and limit the number of actors needed per session, Simulation Center staff executed scripted roles and provided relevant cues to participants. The scripting of each role, which included verbatim comments identified in the initial phase of this project, was built into the scenarios to increase authenticity. At minimum, a bedside nurse and a charge nurse were made available to the participant. Participants could consult a respiratory therapist for specific tasks (e.g., oxygenation, blood gas sampling), a subspecialist (e.g., orthopedic resident, nephrology fellow) if applicable to the scenario, or a parent. Simulation Center staff practiced these roles prior to course implementation and were trained to only provide supporting information/results when asked, or perform tasks when instructed, by the participant. Staff carried laminated cards (their script) with them to ensure consistency across scenarios. Also, staff were not allowed to provide clinical insight or make suggestions to the participant.

Each simulation was designed to take 15 minutes to complete. The Situation Awareness Global Assessment Technique (SAGAT) (Appendix 1) was applied by a trained facilitator three times during the 15-minute simulation (at 5–8 minutes, 8–11 minutes, and 11–14

minutes). The participant was asked to step out of the simulation room, where standardized SAGAT questions were asked.

After completion of each scenario, a 2-member facilitation team, aware of the participant experience level, debriefed the participant in a separate room. The lead facilitator was a board-certified pediatric emergency medicine or pediatric critical care faculty physician with at least 10 years of clinical and simulation-based experience. The second facilitator was a nurse or research coordinator, present to ensure that a standardized debriefing took place. All facilitators were trained in the CDM interviewing technique.³⁰ Each debriefing session lasted 20–30 minutes. The debriefing format was learner-focused, through open ended questions requiring self-reflection on performance and centered on the participant's assessment of illness severity, identification of potential causes, and early stabilization of critical illness. Table 1 was available as a wall chart in the debriefing room and was used by the facilitator to focus the initial discussion, and prompt the participant to classify the rapid cardiopulmonary assessment. The second portion of the debriefing entailed asking the participants to describe and “weigh” the information they used in the scenario, questions that both paralleled the SAGAT and forced self-reflection on assessment data (e.g., vital signs, exam findings, lab results) and how the participant used this data to prioritize management strategy and interventions. The third portion of the debriefing allowed the participant to ask clinical questions to the lead facilitator regarding assessment and management skills. During the third part of the debriefing, facilitators intentionally did not use the words sepsis, septic or septic shock, and avoided giving the participants “the answer” to the scenario. Upon completion of their 4th scenario, each participant was made aware that three scenarios were sepsis and one was SVT; and a 20-minute sepsis-specific didactic talk was given by the lead facilitator to augment the SBT. Finally, three recent publications on pediatric sepsis were sent by email to each participant.

Study Population and Scenario Assignment

A cohort of pediatric residents and faculty (unique from the CDM interviews) were recruited as a convenience sample for the simulations. Novices were defined as first-year residents (interns), intermediate-level physicians were defined as third-year residents, and experts were defined as faculty physicians >5 years from fellowship training in critical care or emergency medicine. All physicians participated in 4 scenarios: SVT, the sepsis “garden path”, and two of the other four sepsis scenarios. There were six possible combinations of the two non-“garden path” sepsis scenarios; each participant was assigned a 2-scenario combination using a randomized block design (Table 2).

Outcomes of interest

The primary outcome was recognition of sepsis by the physician at the end of the scenario. Recognition of sepsis was determined by the response to the question “What is most likely the cause of the patient's illness” on the final (3rd) SAGAT. All SAGAT interviews were audio-recorded and transcribed by a medical transcriptionist. Transcripts were coded by two of four investigators. Three coders were qualitative researchers (AB, LM, and RT) and one was an ED physician (MP). The interviews were randomly assigned to each pair of coders. Using the same themes and categories developed for the SAGAT, the coders independently

coded each interview. After two interviews were coded, the coders met in person or by phone to discuss each item and come to consensus. This process resulted in a set of master codes (e.g. vital signs or exam findings being used, diagnostic possibilities). Discussions and consensus occurred after every two interviews to prevent coding drift and allow further revision of the coding sheet, if needed. This iterative process continued until each transcribed SAGAT was coded and consensus had been reached. The frequency of each cue type was tabulated across all SAGATs and summarized by participant class (novice, intermediate trainee, and expert). Frequency was calculated based on a single mention of cue rather than the number of times a specific cue was mentioned during the SAGAT.

Secondary outcomes were physician performance in the assessment and early management of simulated patients. Physician performance was assessed by internally derived tools adapted from Pediatric Advanced Life Support (PALS) recommendations and early sepsis management (ESM) literature. Each tool contained two parts: (1) a checklist of behaviors to be assessed as performed or not, with 15 behaviors for assessment and 6 behaviors for sepsis management; and (2) a 1–7 (poor performance to ideal performance) Likert Scale to rate the participant's global (overall) performance in the domains of assessment and sepsis management. The PALS and ESM tools (Appendix 2) were applied during video review of the simulations.

Measures of Outcome

SAGAT is a method that employs a “freeze technique” to assess levels of situation awareness. A simulation of the system of interest is frozen at random points in time and the participant(s) quickly responds to questions concerning their perceptions of the situation at that moment. Participants are turned away from the simulation while answering questions on elements related to all three levels of situation awareness.³² This technique reduces recall bias because the participants are queried immediately regarding their detection of simulation elements rather than questioned following the incident. The nature of the inquiries is factual, allowing for objective scoring. Though potentially disruptive, studies have demonstrated that freezes as long as two minutes, the threat of freezes, and the number of freezes did not affect performance for pilots and air traffic controllers.^{33–36} SAGAT queries are randomly administered and require discrete, well-defined responses.³² Because we were interested in whether participants were noticing appropriate cues, correctly interpreting those cues, and anticipating the participant's future state, our adapted SAGAT used the same four queries at each pause (Appendix 1).

Checklists of correct performance behaviors were developed by a pediatric emergency medicine (GG) and a pediatric critical care physician (DW). The checklists were based on PALS recommendations and ESM literature, and were reviewed by the entire investigative team. Checklists were constructed to reflect observable behaviors only, and were piloted to ensure reliable data abstraction from video recordings of simulations.

Video-reviewer training and scoring

Four members of the Simulation Center, including one investigator (GG), were trained in PALS and ESM checklist application. All four reviewers reviewed a pilot simulation and

independently applied the checklist; discrepancies were discussed as a group and consensus achieved. This method was repeated for separate videos until all items were scored identically by reviewers. Participant videos were randomized and assigned to two reviewers per video.

When the two reviewers disagreed for checklist items, reviewers met to discuss the scores and achieve consensus. The consensus score was used in analysis. When an item was “not applicable” (e.g. patient in compensated sepsis who was never hypotensive for age), it was counted as a “yes” for analysis. When the two reviewers disagreed on a global assessment score, a mean of the two values was entered into the database.

Statistical Analysis

Descriptive statistics were generated for PALS and ESM checklist and global performance scores. Frequency distributions were generated for recognition of sepsis. Additional analyses were conducted using a mixed models approach due to correlated data from two sources. First, all models accounted for repeated measures within participant. Second, each participant was randomized to 2 of 4 scenarios while all were exposed to the 5th scenario. Thus, scenario was included as a random effect. Odds ratios were calculated to compare recognition of sepsis on the 1st, 2nd and 3rd SAGAT among levels of expertise. The dependent variable was whether or not sepsis was recognized. The independent variable of interest was the categorical level of experience. Severity (i.e., compensated or uncompensated) was included as a covariate since participants (the unit of analysis) were exposed to only 3 of the 5 possible scenarios. Models were developed to measure the association between each of the ESM and PALS checklist and global assessment performance scores and level of experience. The dependent variable was the score and the independent variable was the level of experience. Similarly models were developed to measure the association with severity (compensated and uncompensated), but adjusting for level of experience. Finally, four models were developed to determine if the recognition of sepsis was associated with increased early sepsis management and compliance with pediatric advanced life support. The dependent variable was whether or not sepsis was recognized. The independent variables of interest were the ESM and PALS checklist and global assessment performance. The PALS score models adjusted for severity, while the ESM models adjusted for severity and level of experience. SAS version 9.3, including PROC MIXED and PROC GLIMMIX, was used to conduct these analyses.

As internally derived tools, we assessed interrater reliability for the video review. Because 2 out of a possible 4 reviewers determined the presence or absence of each component, intraclass correlation, assuming a one-way random model, was calculated for each component of the PALS and ESM checklists, as well as the percent agreement. Reviewer agreement on the ordinal PALS and ESM global performance scores was measured using the nonparametric Spearman’s rank order correlation. IBM SPSS Statistics 24 was used to conduct these analyses.

RESULTS

Eighteen providers were enrolled (6 faculty, 6 third-year residents and 6 interns). Each participant completed 4 simulations and data collection was complete in all cases, resulting in 54 sepsis simulations (18 participants $\times$ 3 sepsis simulations/participant) available for analysis. Physicians recognized sepsis in 19 (35%) of 54 sessions. Interns (novices) recognized sepsis in 4 (22%) of 18 scenarios, senior residents (intermediate-level) in 8 (44%), and faculty (expert) physicians in 7 (39%). Table 3 shows recognition of sepsis (at 3rd SAGAT) by experience and scenario. Adjusted for scenario severity, the odds that faculty recognized sepsis at 3rd SAGAT was 2.6 (95% CI: 0.5, 13.8) times that of interns (Table 4). Compared to compensated sepsis simulations (lesser severity of illness), a higher percentage of physicians recognized sepsis in uncompensated septic shock simulations (Table 5). Physicians recognized sepsis in 8 (22%) of 36 compensated simulations compared with 11 (61%) of 18 uncompensated shock simulations ($p = 0.005$).

Video-based review of the simulations resulted in PALS-based assessment and ESM checklist scores for each of the 54 sepsis simulations. Intraclass correlation and percent agreement were low for initiation of pressor support, correction of electrolyte disturbance, and recognition of hypotension for age. The following exhibited a high (>90%) agreement, but low intraclass correlation: serum glucose obtained, mental status, oxygen saturation, and airway patency assessment (see Table, Supplemental Digital Content 2, which shows reliability of video-based application of PALS and ESM checklist tools). For the PALS global performance scoring, the correlation coefficient was a moderate 0.534; for the ESM global performance score, the correlation coefficient was strong at 0.622.³⁷

Table 6 shows the unadjusted mean assessment and management scores by experience, scenario and severity. Overall, the 18 participants achieved a mean of 12.89 (SD 1.30) for assessment on the 15-item PALS checklist and a mean of 4.07 (SD 1.16) for management on the 6-item ESM checklist. Global performance score results demonstrated means of 5.23 (SD 1.11) for assessment and 4.76 (SD 1.31) for management. Accounting for correlated measures within subject and scenario and using p -value < 0.05, faculty (experts) scored statistically significantly higher than interns (novices) on the PALS global performance score (5.9 vs. 4.3), ESM global performance score (5.5 vs. 4.2) and ESM checklist scores (4.5 vs. 3.7), but not on the PALS assessment checklist. Faculty scored higher than residents on the ESM global performance score (5.5 vs. 4.6). Residents scored higher than interns on the PALS global performance score (5.4 vs. 4.3). There were no significant differences in scores by severity (compensated vs. uncompensated) adjusted for experience.

Adjusted for severity, for every point increase in the PALS global performance score, the odds that sepsis was recognized increased by 11.3 (95% CI: 3.1, 41.4). Similarly, the odds ratio for the PALS checklist score was 1.5 (95% CI: 0.8, 2.6). Adjusted for severity and level of expertise, the odds of recognizing sepsis was associated with an increase in the ESM checklist score of 1.8 (95% CI: 0.9, 3.6) and an increase in ESM global performance score of 4.1 (95% CI: 1.7, 10.0).

The qualitative analysis of the SAGAT revealed that sepsis was the most commonly mentioned diagnosis for “what’s most likely the cause?” across all scenarios and all levels of expertise. Thirty-one different diagnoses were offered, as well as “I don’t know.” Frequently mentioned diagnoses included pneumonia, infection, hypovolemia, dehydration and shock. For “what else are you considering for causes?” sixty-five different diagnoses were offered, with sepsis, cardiac issue, dehydration, hypovolemia, infection, ingestions/toxins, and pneumonia being the most common. When examining the cues utilized, there was a high level of agreement across providers and all seemed to focus on the classic indicators of sepsis (oxygenation, heart rate, respiratory status and blood pressure), as well as mental status and distal extremity perfusion.

DISCUSSION

Multiple examples of effective use of cognitive task analysis, and specifically CDM, to identify novice and expert differences in a variety of domains exist. In healthcare, this method has been used successfully with acute care nurses, neonatal nurses, and residents rotating in an ICU.^{17,38–40} Despite recognition of differences in their approach to complex clinical problems, there has been little research on cues and cue utilization strategies for experts and novices in healthcare. There has also not been any organized attempt to exploit these differences to accelerate the development of clinical expertise.⁴¹

In our first phase, we identified cues that physicians use to assess and make decisions in the context of real-world incidents involving sepsis.³⁰ The resulting critical cue inventory, detailed examples, and contextual elements provided the foundation for creating SBT scenarios in this second phase aimed at increasing the novice clinician’s exposure to sepsis, developed to portray variety in timing and presentation of sepsis, and focused on the critical strategies practiced by experts. Ideally, training scenarios would allow hypothesis testing and force time dependent decision-making and management in order to accelerate recognition and early management of sepsis.

Thus, we needed to determine whether the combination of our SBT scenarios and outcome measures (SAGAT, PALS assessment tool and ESM performance tool) created a valid construct for improving our intended population’s performance (on sepsis recognition and early management), within our institution, during their first year of pediatric residency training. Creating this validity argument, as framed by Kane, guides the collection and interpretation of validity evidence.³¹ As described by Cook et al, the initial step is to consider the decision at hand and a proposed interpretation that would support that decision.³¹ For our study, the decision at hand was whether pediatric physicians benefit from simulation-based training in their recognition of sepsis and early management directed at sepsis. More specifically does their performance over a series of simulations demonstrate their ability to make correct clinical decisions regarding sepsis, i.e. do they “pass” or do they need more training? Our interpretation of those results, or as Kane would state as our (instructors’) decision about the learners’ abilities based on application of the outcome measures, was that we categorized “correct decisions” as recognition of sepsis by the 3rd SAGAT and achievement of high scores on the ESM checklist and global performance scale. Next, one must identify key claims, assumptions and inferences associated with this

interpretation and use.³¹ Our hypothesis or interpretation/use argument was development of a 5-scenario simulation-based training course using cues embedded from CDM interviews will allow differentiation between pediatric interns (novices) and faculty (experts) in the recognition and early management of sepsis, and thus form the basis for future training of novices. Then, one must develop a plan to test these assumptions/inferences.³¹ This would be our methodology described above, including convenience sampling of different experience cohorts (interns, 3rd years and faculty), event-based scenario design augmented by critical cues from CDM interviews, novel strategies to relay these critical cues to participants and scripting of confederate roles, outcome measure development (SAGAT, PALS, ESM) leveraging local expertise and published literature, and standardized debriefing sessions. Finally, one must collect empiric evidence from multiple sources and organize this evidence into a validity argument.³¹ To test these assumptions, Kane's validity framework "traces an assessment from the scoring of a single observation (Scoring), to using the observation score(s) to generate an overall test score representing performance in the test setting (Generalization), to drawing an inference regarding what the test score might imply for real life performance (Extrapolation), and then to interpreting this information and making a decision (Implications)" [p 563].³¹

In this second phase of study, we have collected empiric evidence at the Scoring and Generalization levels. Over 54 sessions where single observations were scored dichotomously at the 3rd SAGAT, physicians recognized sepsis in 35%. One would hope this frequency would be higher, especially among experts. However, this overall test score in the simulation lab (generalization) likely represents actual, clinical difficulties in recognizing sepsis in the pediatric population. In our limited cohort experts were able to identify sepsis more readily than novices, although the difference was not statistically significant and we would feel more secure in this statement if we could have increased sampling. Also, although comparison of expert-novice groups weakens our argument (see further discussion below), the differences in performance provide limited evidence towards use of these scenarios in future training.

In addition to sepsis recognition, our other foci were assessment of a critically ill child and application of early sepsis management principles. Single observations of PALS assessment and ESM performance (Scoring) demonstrated moderate to high compliance with published recommendations, with mean assessment checklist scores of 12.89 out of 15 and mean management checklist scores of 4.07 out of 6. Generalizing performance to the test (simulation lab) environment there were significant differences between experts and novices in three of four performance measures, as experts scored higher than novices within management (checklist and global performance) measures but not on the assessment checklist. An explanation may be the interns' participation in PALS at the beginning of their training year where assessment strategies and performance are key foci. Also, an intern may have greater comfort with assessment compared to higher level management decisions. A major focus of the debriefing sessions was recognition and weighting of information used to make decisions and discussing how and why that information could be used for rapid cardiopulmonary assessment, independent of diagnosis. This may have promoted interns becoming more efficient in their assessment skills. Interestingly, all physicians performed worse in early sepsis management during less severely ill (compensated sepsis) scenarios.

An explanation might be that in the undifferentiated ill child, sepsis is only one of many potential diagnoses and physicians are less likely to commit to early sepsis management (aggressive fluid resuscitation and early antibiotics). Importantly, we showed that when PALS assessment fundamentals were utilized, physicians were more likely to identify sepsis. When this occurred, and sepsis was identified, physicians were more likely to adhere to early management recommendations.

However, we are limited at this point of study by our inability to infer implications on novice assessment for, recognition of, and management of pediatric sepsis at the bedside (Extrapolation). Additionally, the use of expert-novice, or known-group, weakens our validity argument as association does not imply causation.³¹ To that end, the simulation scenarios were implemented as a 4-session course (third phase) in the following academic year for the cohort of pediatric interns and clinical outcomes for patients were tracked, including need for rapid transfer of septic patients to the intensive care units, need for invasive procedures, volume resuscitation and vasopressor support. Although the analysis of the training year's (third phase) outcomes is pending, the feedback received from the intern class was overwhelmingly positive. Residency leadership made this course mandatory for all residents, and currently we offer a weekly 2-scenario session to residents using the scenarios and debriefing format developed.

Limitations

The first limitation was the small number of participants and sessions per participant. This likely prohibited us from demonstrating a significant difference between novices and experts in the recognition of sepsis. As noted above, increased sampling would increase our ability estimate and improve our validity argument. Second, the video reviewers were not blinded to participant experience. This may have influenced their scoring, especially when behaviors, such as mental status, were difficult to assess by video. A computerized manikin cannot be stimulated verbally and physically. We attempted to minimize this limitation by using a voice modulator to allow a verbal response from the "patient" and use the eye opening function of some of the simulators. Third, non-blinding may have influenced the reviewers assignment of global scores, e.g., a reviewer could have weighted assessment and management decisions made by "an expert" to be better than ones made by a novice. Fourth, physicians were scheduled into sessions by convenience of their schedules, introducing the potential bias of self-reflection between sessions. Participants may have specifically questioned whether all the scenarios involved sepsis, leading to improved performance in subsequent simulations. We included the SVT scenario for each participant to limit this source of bias. During the debriefing sessions, we intentionally did not use the words sepsis, septic or septic shock. Instead, we debriefed around assessment and early stabilization of critical illness using the PALS-based debriefing table (Table 1). Finally, the use of internally-derived, novel PALS and ESM tools is a limitation. Without prior implementation in a validity argument for SBT in a similar cohort, their use here limits our findings and argument. As noted above, we do intend to use them in the next phase on a larger cohort of novices. Also, interrater reliability was not high for all components of the checklists, which needs to be addressed in future work before application of these tools.

This project represents one of the first efforts in healthcare to explicitly embed differences in novice and expert clinicians in cue recognition and utilization into SBT. Although this approach is initially focused on the recognition of sepsis, it is likely transferable to other critical, time-dependent medical conditions, e.g., myocardial infarction. As currently constructed, this training course provides an opportunity for instructors to collect (scoring inference) and interpret evidence surrounding learner assessment, recognition and management of sepsis in a simulated, or test, setting (generalization inference). To fully accept or reject its argument for use in our population, future study must infer what test performance implies for real life performance (extrapolation inference) at the bedside in critically ill children and then decide whether this course should be made available, made mandatory or not implemented (implications inference) based on multiple factors, including effectiveness, usefulness, cost, and effort. As Kane notes, ‘a decision procedure that does not achieve its goals, or does so at too high a cost, is likely to be abandoned even if it is based on perfectly accurate information.’⁴²

Conclusions

Although incomplete, evidence from initial testing suggests that the simulations of pediatric sepsis were sufficiently valid to justify their use in training novice pediatric physicians in the assessment, recognition and management of pediatric sepsis. In simulated cases, physicians who adhere closely to PALS guidelines are more likely to recognize sepsis, and recognition is associated with increased compliance with early sepsis management principles.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was funded by the Agency for Healthcare Research and Quality, grant #1R18HS020455-01.

References

1. Wheeler DS, Zingarelli B, Wheeler WJ, Wong HR. Novel pharmacologic approaches to the management of sepsis: targeting the host inflammatory response. *Recent Pat Inflamm Allergy Drug Discov.* 2009; 3:96–112. [PubMed: 19519586]
2. Angus DC, Linde-Zwirble WT, Lidicker J, Clermont G, Carcillo J, Pinsky MR. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. *Crit Care Med.* 2001; 29:1303–1310. [PubMed: 11445675]
3. Martin GS, Mannino DM, Eaton S, Moss M. The epidemiology of sepsis in the United States from 1979 through 2000. *N Engl J Med.* 2003; 348:1546–1554. [PubMed: 12700374]
4. Dombrovskiy VY, Martin AA, Sunderram J, Paz HL. Rapid increase in hospitalization and mortality rates for severe sepsis in the United States: A trend analysis from 1993 to 2003. *Crit Care Med.* 2007; 35:1244–1250. [PubMed: 17414736]
5. Weiss SL, Fitzgerald JC, Pappachan J, Wheeler D, Jaramillo-Bustamante JC, Salloo A, Singhi SC, Erickson S, Roy JA, Bush JL, Nadkarni VM, Thomas NJ, for the Sepsis Prevalence, Outcomes, and Therapies (SPROUT) Study Investigators and the Pediatric Acute Lung Injury and Sepsis Investigators (PALISI) Network. *Am J Respir Crit Care Med.* 2015; 191:1147–1157. [PubMed: 25734408]

6. Barrow RE, Jeschke MG, Herndon DN. Early fluid resuscitation improves outcomes in severely burned children. *Resuscitation*. 2000; 45:91–96. [PubMed: 10950316]
7. Booy R, Habibi P, Nadel S, de Munter C, Britto J, Morrison A, Levin M, Meningococcal Research Group. Reduction in case fatality rate from meningococcal disease associated with improved healthcare delivery. *Arch Dis Child*. 2001; 85:386–390. [PubMed: 11668100]
8. Kumar A, Roberts D, Wood KE, Light B, Parrillo JE, Sharma S, Suppes R, Feinstein D, Zanotti S, Taiberg L, Gurka D, Kumar A, Cheang M. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Crit Care Med*. 2006; 34:1589–1596. [PubMed: 16625125]
9. Pollard AJ, Nadel S, Ninis N, Faust SN, Levin M. Emergency management of meningococcal disease: eight years on. *Arch Dis Child*. 2007; 92:283–286. [PubMed: 17376933]
10. Odetola FO, Rosenberg AL, Davis MM, Clark SJ, Dechert RE, Shanley TP. Do outcomes vary according to the source of admission to the pediatric intensive care unit? *Pediatr Crit Care Med*. 2008; 9:20–25. [PubMed: 18477909]
11. Han YY, Carcillo JA, Dragotta MA, Bills DM, Watson RS, Westerman ME, Orr RA. Early reversal of pediatric-neonatal septic shock by community physicians is associated with improved outcome. *Pediatrics*. 2003; 112:793–799. [PubMed: 14523168]
12. de Oliveira CF, de Oliveira DS, Gottschald AF, Moura JD, Costa GA, Ventura AC, Fernandes JC, Vaz FA, Carcillo JA, Rivers EP, Troster EJ. ACCM/PALS haemodynamic support guidelines for paediatric septic shock: An outcomes comparison with and without monitoring of central venous oxygen saturation. *Intensive Care Med*. 2008; 34:1065–1075. [PubMed: 18369591]
13. Carcillo JA, Kuch BA, Han YY, Day S, Greenwald BM, McCloskey KA, Pearson-Shaver AL, Orr RA. Mortality and functional morbidity after use of PALS/APLS by community physicians. *Pediatrics*. 2009; 124:500–508. [PubMed: 19651576]
14. Larsen GY, Mecham N, Greenberg R. An emergency department septic shock protocol and care guideline for children initiated at triage. *Pediatrics*. 2011; 127:e1585–92. [PubMed: 21576304]
15. Lane RD, Funai T, Reeder R, Larsen GY. High reliability pediatric septic shock quality improvement initiative and decreasing mortality. *Pediatrics*. 2016; 138:e20154153. [PubMed: 27604184]
16. Balamuth F, Weiss SL, Fitzgerald JC, Hayes K, Centkowski S, Chilutti M, Grundmeier RW, Lavelle J, Alpern ER. Protocolized treatment is associated with decreased organ dysfunction in pediatric severe sepsis. *Pediatr Crit Care Med*. 2016; 17:817–822. [PubMed: 27455114]
17. Crandall B, Getchell-Reiter K. Critical decision method: a technique for eliciting concrete assessment indicators from the intuition of NICU nurses. *ANS Adv Nurs Sci*. 1993; 16:42–51. [PubMed: 8311424]
18. Reischman RR. Critical care cardiovascular nurse expert and novice diagnostic cue utilization. *J Adv Nurs*. 2002; 39:24–34. [PubMed: 12074749]
19. Hoffman KA, Aitken LM, Duffield C. A comparison of novice and expert nurses' cue collection during clinical decision-making: verbal protocol analysis. *Int J Nurs Stud*. 2009; 46:1335–44. [PubMed: 19555954]
20. Brilli RJ, Gibson R, Luria JW, Wheeler TA, Shaw J, Linam M, Kheir J, McLain P, Lingsch T, Hall-Haering A, McBride M. Implementation of a medical emergency team in a large pediatric teaching hospital prevents respiratory and cardiopulmonary arrests outside the intensive care unit. *Pediatr Crit Care Med*. 2007; 8:236–46. [PubMed: 17417113]
21. Funk D, Sebat F, Kumar A. A systems approach to the early recognition and rapid administration of best practice therapy in sepsis and septic shock. *Curr Opin Crit Care*. 2009; 15:301–7. [PubMed: 19561493]
22. Lewis R, Strachan A, Smith MM. Is high fidelity simulation the most effective method for the development of non-technical skills in nursing? A review of the current evidence. *The Open Nursing Journal*. 2012; 6:82–89. [PubMed: 22893783]
23. Nguyen HB, Daniel-Underwood L, Van Ginkel C, Wong M, Lee D, Lucas AS, Palaganas J, Banta D, Denmark TK, Clem K. An educational course including medical simulation for early goal-directed therapy and the severe sepsis resuscitation bundle: an evaluation for medical student training. *Resuscitation*. 2009; 80:674–9. [PubMed: 19395143]

24. Hansel M, Winkelmann AM, Hardt F, Gijsselaers W, Hacker W, Stiehl M, Koch T, Muller MP. Impact of simulator training and crew resource management training on final-year medical students' performance in sepsis recognition: a randomized trial. *Minerva Anesthesiol.* 2012; 78:901–9. [PubMed: 22504855]
25. Nishisaki A, Hales R, Biagas K, Cheifetz I, Corriveau C, Garber N, Hunt E, Jarrah R, McCloskey J, Morrison W, Nelson K, Niles D, Smith S, Thomas S, Tuttle S, Helfaer M, Nadkarni V. A multi-institutional high-fidelity simulation “boot camp” orientation and training program for first year pediatric critical care fellows. *Pediatric Critical Care Medicine.* 2009; 10:157–162. [PubMed: 19188876]
26. Littlewood KE, Shilling AM, Stemland CJ, Wright EB, Kirk MA. High-fidelity simulation is superior to case-based discussion in teaching the management of shock. *Med Teach.* 2013; 35:e1003–10. [PubMed: 23126242]
27. Evans KH, Daines W, Tsui J, Strehlow M, Maggio P. A novel, mobile, online, simulation game that improves sepsis recognition and management. *Acad Med.* 2015; 90:180–4. [PubMed: 25517703]
28. Dugan MC, McCracken CE, Hebbar KB. Does simulation improve recognition and management of pediatric septic shock, and if one simulation is good, is more simulation better? *Pediatr Crit Care Med.* 2016; 17:605–614. [PubMed: 27144833]
29. Qian J, Wang Y, Zhang Y, Zhu X, Rong Q, Wei H. A Survey of the First-Hour Basic Care Tasks of Severe Sepsis and Septic Shock in Pediatric Patients and an Evaluation of Medical Simulation on Improving the Compliance of the Tasks. *J Emerg Med.* 2016; 50:239–45. [PubMed: 26275743]
30. Patterson MD, Militello LG, Bunger A, Taylor RG, Wheeler DS, Klein G, Geis GL. Leveraging the Critical Decision Method to Develop Simulation-based Training for Early Recognition of Sepsis. *Journal of Cognitive Engineering and Decision Making.* 2016; 10:36–56.
31. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Medical Education.* 2015; 49:560–575. [PubMed: 25989405]
32. Endsley, MR. Direct Measurement of Situation Awareness: Validity and Use of SAGAT. In: Endsley, MR, Garland, DJ., Mahwah, NJ., editors. *Situation Awareness Analysis and Measurement.* Lawrence Erlbaum Associates; 2000. p. 1-21.
33. Bolstad, CA., Endsley, M. R: Single versus dual scale range display investigation (NOR DOC 90-90). Hawthorne, CA: Northrop Corporation; 1990.
34. Endsley M. Measurement of Situation Awareness in Dynamic Systems. *Human Factors.* 1995; 37:65–84.
35. Endsley, MR., Garland, DJ. *Situation Awareness Analysis and Measurement.* New York, NY: CRC Press; 2000.
36. Wright MC, Taekman JM, Endsley MR. Objective measures of situation awareness in a simulated medical environment. *Qual Saf Health Care.* 2004; 13(Suppl 1):i65–71. [PubMed: 15465958]
37. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977; 33:159–174. [PubMed: 843571]
38. Gazarian PK, Henneman EA, Chandler GE. Nurse decision making in the prearrest period. *Clin Nurs Res.* 2010; 19:21–37. [PubMed: 19955229]
39. Fackler JC, Watts C, Grome A, Miller T, Crandall B, Pronovost P. Critical care physician cognitive task analysis: an exploratory study. *Crit Care.* 2009; 13:R33. [PubMed: 19265517]
40. Militello L, Lim L. Patient assessment skills: assessing early cues of necrotizing enterocolitis. *J Perinat Neonatal Nurs.* 1995; 9:42–52. [PubMed: 7636711]
41. Crandall B, Wears RL. Expanding perspectives on misdiagnosis. *Am J Med.* 2008; 121:S30–3.
42. Kane, MT. Validation. In: Brennan, RL., editor. *Educational Measurement.* 4th. Westport, CT: Praeger; 2006. p. 17-64.

Table 1

Debriefing Table Based on PALS Fundamentals

System	Rapid Cardiopulmonary Assessment			
Airway	Patent, self-maintained	Patent, maintained by adjuncts ¹		Obstructed
Breathing	Normal	Distress	Failure ²	Arrest
Circulation	Normal	Distress	Failure ³	Arrest

Notes:

¹ Adjuncts can include head positioning, suctioning, NP airway or oral airway

² Clinically, respiratory failure would be defined early in resuscitation by the need for positive pressure to maintain oxygenation and/or ventilation. This would include CPAP, BiPaP, or BVM ventilation. Respiratory distress can be thought of as a child who is tachypneic and showing signs of distress (retractions, flaring, etc.) but who is maintaining oxygenation on room air or by non-invasive oxygen delivery systems (nasal cannula, simple face mask or non-rebreather face mask).

³ A five-letter word for circulatory failure is SHOCK, which if recognized would require early in resuscitation the use of rapid infusion techniques (push-pull, Belmont rapid infuser, etc.) where fluid is pushed quickly (i.e. over 5 minutes) and an immediate reassessment is done. Circulatory distress can be thought of as the child who is tachycardic and moderately dehydrated, but has still maintained mental status, distal skin perfusion, and blood pressure. These patients can get aliquots of volume given over 20–60 minutes.

Table 2

Assignment of participation to simulation scenarios

Level of expertise	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5 (garden path)	Scenario 6 (SVT)
Novice	3	3	3	3	6	6
Intermediate	3	3	3	3	6	6
Expert	3	3	3	3	6	6
Total	9	9	9	9	18	18

Brief sepsis scenario descriptions:

1. 6-month old male infant who presents to emergency department with respiratory symptoms and fever [compensated]
2. 9-month old infant who presents to emergency department's shock trauma suites with respiratory distress [uncompensated]
3. 6-year old African-American child with developmental delay who is on the inpatient floor 28 hours post right humeral surgery [compensated]
4. 3-year old with acute lymphoblastic leukemia admitted through oncology clinic to floor with neutropenic fever [uncompensated]
5. 2-year old Hispanic male with end stage renal disease transferred from an outside hospital to floor as direct admission with presumed gastroenteritis [garden path scenario]

Table 3

Recognition of sepsis for each participant (at 3rd SAGAT) by experience and scenario

Participant	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	# Times Sepsis Recognized
Expert						
1		No		Yes	No	1
2	Yes		No		No	1
3			No	Yes	No	1
4	No			No	No	0
5		Yes	Yes		Yes	3
6	No	Yes			No	1
Intermediate						
7		Yes	No		No	1
8	No	Yes			No	1
9		No		Yes	No	1
10			Yes	Yes	Yes	3
11	No			Yes	No	1
12	Yes		No		No	1
Novice						
13	Yes		No		No	1
14		No		No	No	0
15			No	No	No	0
16	Yes	No			No	1
17		Yes	No		No	1
18	No			Yes	No	1
Total (%) recognition of sepsis by scenario						
All	4 (44%)	5 (56%)	2 (22%)	6 (67%)	2 (11%)	19 (35%)

Table 4

Comparison of novice, intermediate and expert physicians in recognition of sepsis during simulated cases adjusted for severity

1st SAGAT	Odds Ratio	Lower 95% CL	Upper 95% CL
Faculty vs. interns	1.59	0.27	9.40
Faculty vs. 3 rd -year residents	0.67	0.13	3.58
3 rd -year residents vs. interns	2.38	0.42	13.55
2nd SAGAT			
Faculty vs. interns	3.02	0.48	19.05
Faculty vs. 3 rd -year residents	0.74	0.15	3.71
3 rd -year residents vs. interns	4.08	0.65	25.59
3rd SAGAT			
Faculty vs. interns	2.59	0.49	13.77
Faculty vs. 3 rd -year residents	0.76	0.17	3.47
3 rd -year residents vs. interns	3.41	0.64	18.09

SAGAT = situation awareness global assessment tool

CL = Confidence Level

Table 5

Recognition of sepsis by scenario

#	Scenario description	Severity level	Number run	Number (%) recognized as sepsis
1	6-month old male infant who presents to emergency department with respiratory symptoms and fever	Compensated	9	4 (44)
2	9-month old infant who presents to emergency department's shock trauma suites with respiratory distress	Uncompensated	9	5 (56)
3	6-year old African-American child with developmental delay who is on the inpatient floor 28 hours post right humeral surgery	Compensated	9	2 (22)
4	3-year old with acute lymphoblastic leukemia admitted through oncology clinic to floor with neutropenic fever	Uncompensated	9	6 (67)
5	2-year old Hispanic male with end stage renal disease transferred from an outside hospital to floor as direct admission with presumed gastroenteritis [garden path scenario]	Compensated	18	2 (11)

Table 6

Assessment and management scores by experience, scenario and severity Characteristic

Characteristic	PALS assessment checklist 15 items Mean (SD)	PALS global performance score 1–7 Likert Mean (SD)	ESM checklist 6 items Mean (SD)	ESM global performance score 1–7 Likert Mean (SD)
Overall (n = 54)	12.89 (1.30)	5.23 (1.11)	4.07 (1.16)	4.76 (1.31)
Experience				
Interns (n = 18)	12.50 (1.38)	4.33 (1.18)	3.67 (1.33)	4.17 (1.37)
Residents (n = 18)	13.17 (1.15)	5.47 (0.74)	4.06 (1.06)	4.61 (1.08)
Faculty (n = 18)	13.00 (1.33)	5.89 (0.76)	4.50 (0.99)	5.50 (1.15)
Scenario				
Scenario 1 (n = 9)	13.33 (0.50)	5.78 (0.71)	4.44 (1.13)	5.28 (0.94)
Scenario 2 (n = 9)	12.33 (1.41)	4.89 (1.11)	4.67 (1.22)	5.39 (1.11)
Scenario 3 (n = 9)	12.33 (1.50)	4.50 (1.22)	3.67 (0.87)	3.50 (1.15)
Scenario 4 (n = 9)	12.56 (1.13)	5.33 (1.17)	4.22 (0.83)	5.22 (0.94)
Scenario 5 (n = 9)	13.39 (1.33)	5.44 (1.07)	3.72 (1.32)	4.58 (1.41)
Severity *				
Compensated (n = 36)	13.11 (1.28)	5.29 (1.12)	3.89 (1.19)	4.49 (1.37)
Uncompensated (n = 18)	12.44 (1.25)	5.11 (1.13)	4.44 (1.04)	5.31 (1.00)

PALS – pediatric advanced life support

ESM – early sepsis management

* Compensated Sepsis included scenarios 1, 3 and 5; Uncompensated Septic Shock included scenarios 2 and 4