Evaluating the Impact of Classroom Education on the Management of Septic Shock Using Human Patient Simulation

Geoffrey K. Lighthall, MD, PhD; Dona Bahmani, MD; David Gaba, MD

**Background:** Classroom lectures are the mainstay of imparting knowledge in a structured manner and have the additional goals of stimulating critical thinking, lifelong learning, and improvements in patient care. The impact of lectures on patient care is difficult to examine in critical care because of the heterogeneity in patient conditions and personnel as well as confounders such as time pressure, interruptions, fatigue, and nonstandardized observation methods.

**Methods:** The critical care environment was recreated in a simulation laboratory using a high-fidelity mannequin simulator, where a mannequin simulator with a standardized script for septic shock was presented to trainees. The reproducibility of this patient and associated conditions allowed the evaluation of “clinical performance” in the management of septic shock. In a previous study, we developed and validated tools for the quantitative analysis of house staff managing septic shock simulations. In the present analysis, we examined whether measures of clinical performance were improved in those cases where a lecture on the management of shock preceded a simulated exercise on the management of septic shock. The administration of the septic shock simulations allowed for performance measurements to be calculated for both medical interns and for subsequent management by a larger resident-led team.

**Results:** The analysis revealed that receiving a lecture on shock before managing a simulated patient with septic shock did not produce scores higher than for those who did not receive the previous lecture. This result was similar for both interns managing the patient and for subsequent management by a resident-led team.

**Conclusions:** We failed to find an immediate impact on clinical performance in simulations of septic shock after a lecture on the management of this syndrome. Lectures are likely not a reliable sole method for improving clinical performance in the management of complex disease processes.

**Key Words:** Lecture, Teaching methods, Resuscitation, Clinical competence, Simulation, Medical education

Septic shock is the leading cause of mortality in the intensive care unit (ICU) with more than half a million new cases per year in the United States; with an aging population, this figure will only increase. Clinical guidelines for the management of severe sepsis are structured around interventions that have reduced its morbidity. Despite attempting to standardize the treatment of severe sepsis and shock through the dissemination of such treatment guidelines by the Society of Critical Care Medicine and others, little attention has been directed to the decision making and performance of individuals or teams in the act of managing sepsis. Likewise, the degree to which different educational strategies can facilitate uptake and deployment of best practices has not been adequately examined.

Experiential learning, considered broadly, includes activities that actively engage learners in aspects of or related to the concepts and material to be learned or their application to real-world situations. In addition to technology-based simulation, experiential learning includes role playing, problem-based learning, storytelling, encounters with standardized patients, and psychomotor task training. Generally, even when experiential learning is planned as a critical element of teaching clinical care, it is combined with traditional modalities of learning—typically a lecture (in person or by video) targeting the levels in Miller’s pyramid of “knows” or “knows how.” The effectiveness of a lecture in embedding knowledge can be measured with a written test (knows) but that cannot determine the degree to which such knowledge becomes “deployable” (knows how).

Because we have created high-fidelity simulations of critical illness, including septic shock, to train house staff in the management of patient care emergencies and later developed a scoring tool to assess the quality of teamwork and medical management in these scenarios, we know that statistically distinguishable categories of performance can be identified. Although many investigators had previously attempted to use simulation to impart knowledge to learners, we turned this paradigm around to use simulation to study whether knowledge had become deployable and...
useful. In this study, we make use of the scoring rubric to evaluate the impact of classroom instruction—a lecture—on performance when managing simulations of septic shock.

The intensive care clinical rotation includes a recurring lecture series, 1 day of which is replaced by a 4-hour simulation session. Because the scheduling of specific lectures varies according to the availability of faculty, the lecture on sepsis preceded the simulation of septic shock approximately half of the time. Therefore, we first performed a retrospective analysis of whether receiving the classroom lecture on septic shock before the simulation of septic shock was associated with higher performance on this scenario. We then conducted a prospective investigation of whether a redesigned lecture—one that was geared more toward improving clinical performance than to discussion of physiologic concepts—would lead to better performance on subsequent simulations of septic shock.

METHODS

The experiment presented here consisted of 2 phases. In phase I, data presented in our initial report on assessing the management of septic shock simulations were analyzed in terms of whether the simulation runs that followed a lecture were associated with higher management scores. Phase II analyzed performance in simulations after a redesign of the lecture to a more deliberate focus on a widely accepted algorithm of managing septic shock. Cards with the printed algorithm were distributed to those attending the lecture, and a quiz was administered before and after the lecture to assess comprehension of information presented in the lecture.

Setting and Participants

The human subjects committee at Stanford University School of Medicine approved the protocol; informed consent was waived for the use of videotapes made originally for debriefing and teaching purposes. Study subjects were internal medicine interns and postgraduate year 2 to 4 residents in surgery, anesthesiology, and internal medicine who were rotating for 1 month in the medical/surgical ICU. The critical care unit is a 15-bed semiclosed unit at the Veterans Affairs Palo Alto Health Care System. Structured activities consist of 2 hours of morning multidisciplinary bedside rounds, afternoon sign-out rounds, and a lecture administered each weekday by the critical care faculty. Lectures are mandatory for interns and residents, and topics are those particular to intensive care, including management of respiratory failure, shock, sepsis, and postoperative patients. The lecture series was constant in content from month to month, and the same faculty member gave the lecture on septic shock during the duration of this experiment. In addition, a 4-hour course on critical care crisis management is conducted in a dedicated simulation center each month where standardized scenarios of septic shock and respiratory failure are managed by teams of house staff and allied personnel. Although timing of lectures sometimes varied according to faculty availability, the simulation course took place on the second Wednesday of each month. We attempted to schedule the shock lecture earlier in the month to be before the simulation session, but this could be accomplished only approximately half of the time.

Educational Intervention

In phase I, simulator performance was compared between house staff who received a lecture before the simulation with those who had the simulation first. Phase II involved an educational intervention for which a 2-part lecture on the pathophysiology of shock was changed to a single-hour clinically focused discussion of septic shock that was altered in the following ways.

1. Minimized discussion of the finer physiologic and biochemical details of organ dysfunction.
2. Emphasized “early goal-directed therapy” for severe sepsis and septic shock, as advocated by treatment guidelines; pocket cards with an algorithm for early goal-directed therapy were distributed.
3. Presented an interactive problem-based learning discussion regarding severe sepsis and septic shock, where the patient’s vital signs are projected on a screen and changed according to interventions requested by trainees. This “hypothetical patient” was nearly identical to the scenario encountered in the simulation, but this was never stated.

In phase II, a quiz was administered to trainees both before and immediately after the lecture and was used to evaluate: (1) the acquisition of knowledge resulting from the lecture and (2) a general approximation of the knowledge that each team had before the simulation. Participants placed a self-generated identification code on both quizzes, so they could be matched for before-and-after comparisons while assuring anonymity. No attempt was made to evaluate the impact of the time interval between the lecture and simulation.

A scoring system that quantitatively rates the management of a simulated patient with septic shock was developed and validated previously. The scenario called for a period in which the ICU interns were the sole physicians managing the patient (first 10 minutes) and a later period where residents directed patient management. This allowed for the formulation of management scores for both interns and residents. In addition to the main hypothesis that a revised lecture would improve performance in sepsis management, we also hypothesized that the interns—being relatively new to clinical medicine and having the most to learn—might be influenced to a greater extent by an improved educational strategy.

Simulation Exercise

All participants were naive to the scenario content and were asked to preserve the novelty for others and to sign confidentiality agreements. The sepsis scenario follows a standard script of signs, symptoms, and vital signs at baseline and for patient deterioration or improvement according to clinician interventions. The scenario begins with 2 interns along with a critical care nurse who are called in as first responders. Senior residents, fellows, subspecialty physicians, respiratory therapists, and pharmacists could be summoned if desired; senior physician help was invariably requested but, by design, was kept out of the simulation room until the 10-minute point. Thus, for each scenario, 2 groups of physicians were evaluated: (1) the interns present during the first 10 minutes and (2) residents who took over the management of the scenario at the 10-minute time point.
Data Sources and Analysis

Video recordings of scenarios were analyzed by 2 trained raters, and averages of the 2 scores were used for the comparisons described. Nontechnical ratings were based on skills such as leadership, communication, contingency planning, and resource use. These skills may dictate success in dynamic environments and have been described by a number of studies in the literature. Previous analysis demonstrated a high degree of interrater reliability for both technical and nontechnical rating systems (κ = 0.96 and κ = 0.88, respectively). Comparisons between “lecture first” and “simulation first” for technical and nontechnical scores for the septic shock scenarios were made by analysis of variance (ANOVA). For all tests, P < 0.05 was taken as statistically significant.

Quizzes covering the key findings and therapy of septic shock were administered before and after the redesigned lecture in phase II; comparisons of scores used Student t test for paired samples. The association between quiz scores and simulation scores was assessed by Spearman test for correlation.

Postcourse surveys covering perceived educational value of simulation, realism, likes and dislikes, as well as suggestions for improvement were administered to participants, as is customary for all courses in our simulation center.

Temporal trends were examined to assess whether simulation scores were influenced by timing within the academic year or simply by time, as may be expected if ambient or nontimely scores were influenced by timing within the academic year for either interns and residents.

RESULTS

In the present study, we analyzed intern and group scores in relation to whether a lecture on shock was presented before or after the simulation session. A total of 56 scenarios were analyzed in this experiment; 23 were from phase I. Phase II produced 33 scoreable scenarios during a 41-month period; scenarios from 8 courses scattered throughout the study period were not scoreable when technical problems, illnesses, and other unforeseen events prevented the scenario from being conducted according to experimental standards.

Averages of all scores from phases I and II were compared and showed a statistically significant higher score in phase II (interns 6.8 vs. 9.6 and residents 8.5 vs. 11.2; P < 0.0001 for both). Because of this difference, “lecture first” and “simulation first” scores were compared separately within each phase of the experiment. Data from the “technical” (medical management) scoring system is presented for both interns and full care teams led by subsequently arriving residents. Phase I data presented in Table 1 indicate that neither technical scores nor nontechnical scores were significantly different when the simulation session was before the lecture versus after the lecture.

When comparing scores of “lecture first” versus “simulation first” groups in phase II, we again found no statistically significant benefit of the lecture but surprisingly a higher average score for groups who completed the simulation without the previous lecture (Table 2). The modest 17% improvement in score was statistically significant for only the “resident” group (not for interns). Nontechnical scores were not significantly different in phase II (Table 2, lower row of data).

Prelecture quiz scores correlated with level of training as shown in Figure 1, where significant differences were noted between medical students and all levels of house staff. Most items on the quiz were directly relevant to the elements on our technical rating scale for simulations.

Significant improvement in postlecture quiz scores was seen for subjects at all levels of training. No significant differences existed between training-level groups when postcourse test scores were compared (dark bars in Fig. 1). All groups had average scores of approximately 7.5 of 10; thus, students showed the greatest overall improvement in scores, followed by interns and residents.

To complete the inquiry into the relationship between knowledge and clinical performance, we compared the simulation scores to the covariate of quiz scores at the time of the simulation course (this score provides a rough estimate of the subject’s knowledge about sepsis). Thus, for the group participating in the simulation before the lecture, the average of their “prelecture” quiz scores was used to estimate subject knowledge, whereas the average of “postlecture” quiz scores was used to estimate knowledge of groups receiving the lectures before simulator participation. With data paired in this manner, we found no correlation between test knowledge and simulation performance. However, caution should be exercised in interpreting this finding. This analysis looked at only groups of individuals, not the actual pairing between individuals managing the sepsis scenario and their scores.

Our commitment to administering quizzes with anonymity prevented us from making the more accurate pairing between quiz and video performance.

The influence of time on the scenario scores was also assessed. Figure 2 presents the simulation scores when sorted into successive quarters of the academic year and when the scenario scores are plotted sequentially during the course of the experiment. There seems to be no significant improvement during the course of the academic year for either interns or residents. When all scores throughout the experimental period are viewed in sequence, the scores generated by interns steadily decline during the 41-month period, whereas the average resident scores were unchanged.

On surveys administered at the conclusion of the simulation workshop, a mean (SD) of 90% (0.1%) recorded either “agree” or “strongly agree” to the question of whether they “felt the simulation environment and scenarios prompted realistic responses.” Responses regarding the realism of the...
scenarios did not differ between the groups with the highest (top 5) and lowest (bottom 5 of 23) levels of performance for both interns or for resident teams. Likewise, responses regarding realism did not differ between “simulation first” or “lecture first” groups.

**DISCUSSION**

Simulation techniques have created exciting opportunities for the enhancement of medical education; however, their use as a tool to evaluate aspects of traditional medical education is relatively novel. This is an important goal because the ultimate purpose of providing medical knowledge to trainee clinicians is in fact to allow them to conduct patient care effectively.

**Veracity of Findings**

From the survey data after simulations and voluntary comments concerning the quality of lectures, it seems that the data are a fairly credible representation of participants’ behavior. We therefore believe that the results require consideration of a number of factors related to lectures and classroom instruction including quality, context, and cognitive processing. The results also prompt consideration of our own goals for such instruction and how realistic it is to expect behavioral change from these lectures.

Could the results be explained by a lecture poor in quality or content? All items present on the technical scoring checklist were discussed in the corresponding lecture and queried on the quiz. Residents have repeatedly identified the lectures as being a high quality, relevant, and valued component of the ICU rotation. However, it is interesting that no group was able to achieve an average of 10/10 on the postlecture quiz.

Could temporal trends influence results? Our expectation that collective knowledge increases over time or increases by the number of months into internship year seems to be unfounded (Fig. 2). The latter is somewhat surprising because we assumed that there was some exposure to the management of sepsis would be influenced by in-terventions. Although we hold the belief that simulation approximates real-life behavior, it is still true that physicians, patients, and the health care system in general probably behave differently when real lives are at stake. Although we make great efforts to recreate the “motivational structure” of the real world in simulations, it is impossible to really know whether this has been achieved.

**Implications of Findings**

In phase II, all trainees at multiple levels demonstrated an increase in theoretical knowledge on the quiz, so the lack of transfer to the simulated patient was completely unexpected. We assumed that the lecture would have a positive impact on the technical scores of the interns. This assumption was based on the idea that interns would have had the least experience managing critical illness, and therefore, their management of sepsis would be influenced by instruction to a degree greater than that of more senior residents. Supporting this notion is the fact that quiz scores showed the greatest improvement in lower levels of training (Fig. 1). Our finding is not a unique observation. For example, Rodgers et al \(^\text{16}\) rated performance of nursing students during simulations of cardiac arrests and found no association between simulator performance and score on patients. Participants can address questions about skin color and warmth, capillary refill, pupils, neck veins, and mucus membranes to the simulation operators and receive the information via an overhead speaker. Nonetheless, we grant that some may use these and other clues to form a subconscious picture of a patient and his clinical course that may not be possible with a mannequin.

Our scoring rubric presents a list of ideal actions and behaviors, but even executing them perfectly cannot guarantee a good outcome in all cases. Similarly, the sequence of actions (such as provision of fluids before vasopressors) could be important in real life, whereas our scoring system did not distinguish between order of interventions. Although we hold the belief that simulation approximates real-life behavior, it is still true that physicians, patients, and the health care system in general probably behave differently when real lives are at stake. Although we make great efforts to recreate the “motivational structure” of the real world in simulations, it is impossible to really know whether this has been achieved.

| Table 2. Impact of a Classroom Lecture on the Management of Simulations of Septic Shock |
|-----------------------------------|-----------------------------------|-----------------------------------|
|                                  | Lecture First | Simulation First | Significance |
| Interns’ technical score, mean (SD) | 9.2 (1.6)     | 10.3 (1.6)       | n.s.          |
| Residents’ technical score, mean (SD) | 10.5 (1.6)   | 12.3 (1.2)       | \(P < 0.05\)  |
| Residents’ nontechnical score, mean (SD) | 23.58        | 26.95            | n.s.          |

Phase II data; n.s., not significant.

**FIGURE 1.** Average prelecture and postlecture quiz scores are shown for each level of training. Light bars indicate prelecture scores, and dark bars are the postlecture tests. All groups showed improvement in score as a result of the lecture (\(\ast P < 0.05\) by Student test for paired samples). Intergroup comparisons showed a trend toward higher prelecture scores according to the level of training, with all postgraduate scores significantly different from the medical students’ scores (\(\ast\ast P < 0.05\) by ANOVA with Bonferroni correction for multiple samples). No statistical differences existed amongst the posttest scores.
previously administered tests. The present study adds to the latter work by showing a similar disconnection between acquired and deployable knowledge in multiple levels of postgraduate trainees. The degree of standardization of the scenarios for all participants and consistency of content among the lecture, quiz, simulation and rating tool are unique to this study and improve the credibility of the experimental findings.

We have no clear explanation as to why apparent uptake of clinically relevant material did not manifest itself by higher simulation performance. A number of explanations exist and may be reasonable targets for further investigation. Perhaps, the knowledge gained needs time to be processed, consolidated, and compared with existing mental models of disease management before becoming deployable. It is possible that the conditions presented to the trainees during the simulation were sufficiently stressful that a slower, more deliberate method of processing the new information was required but not easily accessible. As postcourse surveys and discussions do reveal that the trainees consider the simulations intense and challenging, this explanation seems plausible.

A contrasting explanation is that the day or two separating the lecture and the simulation was too great of an interval and that more immediate practice and hands-on experience with newly learned material is a requisite for reliable incorporation into practice. For example, “just in time” and “booster” training provided immediately before simulations improves the efficacy of chest compressions in simulated pediatric arrests. Unfortunately, it is hard to predict what type of patient problem will present at a given moment, so pairing training to immediate clinical need is impractical. Nonetheless, the sequence of instruction followed by guided practice may provide superior training over separation of these components and deserves further evaluation.

We started this experiment hoping to show that the content of lectures could be altered to create improvements in clinical care and ended up questioning the value of lectures in general. Lectures, both in the traditional in-person form and in the video-based “flipped classroom” paradigm, have been thought to have significant value in conveying “knows” and “knows how” information at a rather low cost. Hundreds of learners can be in the audience in person; an unlimited number might view the lecture on video. Yet, if such activities do not result in deployable knowledge and skill, this leverage is useless.

If lectures do not work, then what does? Trainees may need to be “shown the way” to a greater degree than we are

FIGURE 2. A, Technical scores were sorted for interns (dark bars) and residents (light bars) according to successive quarters of the academic year. B, Intern and resident scores were arranged in sequential months of the 41-week study period to assess for nonspecific increases in knowledge or performance over time. For interns, the trend toward lower scores with time was statistically significant (Spearman r = −0.56, P < 0.001).
REFERENCES


