

# Measuring Team Performance in Simulation-Based Training: Adopting Best Practices for Healthcare

Michael A. Rosen, MA;  
 Eduardo Salas, PhD;  
 Katherine A. Wilson, PhD;  
 Heidi B. King, MS;  
 Mary Salisbury, RN, MSN;  
 Jeffrey S. Augenstein, MD, PhD,  
 FACS;  
 Donald W. Robinson, MD;  
 David J. Birnbach, MD

Team performance measurement is a critical and frequently overlooked component of an effective simulation-based training system designed to build teamwork competencies. Quality team performance measurement is essential for systematically diagnosing team performance and subsequently making decisions concerning feedback and remediation. However, the complexities of team performance pose a challenge to effectively measuring team performance. This article synthesizes the scientific literature on this topic and provides a set of best practices for designing and implementing team performance measurement systems in simulation-based training. (*Sim Healthcare* 3:33–41, 2008)

**Key Words:** Team training, Team performance measurement, Simulation-based training

**T**raining is one of the most flexible and adaptive means for improving performance in organizations.<sup>1</sup> As such, training has been one of the main thrusts in the drive to increase safety in healthcare systems.<sup>2</sup> Specifically, the use of simulation-based training (SBT) to train teamwork skills holds promise as a powerful method to increase safety and effectiveness in healthcare.<sup>3–7</sup> However, to maximize learning in SBT, a systematic and scientific approach must be taken to the development of learning experiences. This is true in general, and especially for training the complexities of teamwork.

Team performance measurement is central to a scientifically rooted approach to SBT for training teamwork skills.<sup>8,9</sup> When executed correctly, team performance measurement provides a systematic means for determining the causes of effective and ineffective performance, providing feedback to correct deficiencies in the individual and team competencies targeted for training, and making decisions about what future training is necessary for the team as a whole or for specific individuals.<sup>10–12</sup> Simulation-based team training without team performance measurement results in unguided practice for teams, which limits the effectiveness of time and resources spent training. Feedback and remediation are nec-

essary for developing the teamwork skills critical to patient safety and team performance measurement is the foundation upon which these rest.

This article provides a set of best practices for developing team performance measures for the purpose of training teamwork skills in healthcare. These best practices are developed from a review of the theoretical, empirical, and practice-based literature concerning team training, SBT, and team performance measurement.<sup>10,13–16</sup> These best practices are not drawn solely from the work done in healthcare; they are drawn from the broader sample of communities and domains applying SBT to train teamwork. This broader knowledge base is presented in the form of best practices for healthcare. Furthermore, we hope this article will facilitate the development of standards for team performance measurement in SBT for healthcare. Before presenting these best practices, we first outline the need for team training and SBT in healthcare, show the importance of team performance measurement to achieving the goals of SBT for teams, and highlight some of the challenges to effective team performance measurement in SBT.

## Why is Teamwork, Team Training and SBT Needed in Healthcare?

Teamwork is central to safety in healthcare as it is often the interactions of healthcare workers that produces effective or ineffective performance.<sup>17,18</sup> Healthcare is an inherently multidisciplinary task where the effectiveness of patient care relies on the interaction of individuals from highly diverse backgrounds in terms of expertise, training, and experience.<sup>19,20</sup> Based on a hospital's historical census and acuity data, a trauma team may consist of several nurses, allied health professionals, a surgeon, an emergency room doctor, and several specialist physicians. All of these individuals have received extensive training in their respective disciplines, but have generally not received formalized training on how to

From the Department of Psychology, and Institute for Simulation and Training, University of Central Florida (M.A.R., E.S., K.A.W.); Miller School of Medicine, University of Miami, William Lehman Injury Research Center (K.A.W., J.S.A, D.J.B.); Office of the Assistant Secretary of Defense (Health Affairs) (H.B.K.); TRICARE Management Activity, Falls Church, VA; The Cedar Institute, Inc., North Kingstown, RI; Army Trauma Training Center, (M.S.); Ryder Trauma Center (D.W.R.); and Department of Anesthesiology, University of Miami, Jackson Memorial Hospital Center for Patient Safety, Miami, Florida (D.J.B.).

This work was supported by the Department of Defense under contract W81XWH-05-1-0372.

Reprints: Eduardo Salas, PhD, Department of Psychology, and Institute for Simulation and Training, University of Central Florida, Florida. (e-mail: esalas@ist.ucf.edu).

The authors have indicated they have no conflicts of interest to disclose.

Copyright © 2008 Society for Simulation in Healthcare  
 DOI: 10.1097/SIH.0b013e3181626276

interact with one another. These heterogeneous teams often function in an environment characterized by high stress, high stakes outcomes, and time pressure.<sup>21</sup> In this context, effective teamwork can serve as a barrier (ideally one of many) to diagnostic and treatment errors.<sup>22,23</sup> Consider the following scenario. An elderly woman, Mrs. L., falls asleep in her chair while smoking. Rescue is called and she is admitted to a Level I Trauma Center with smoke inhalation and burns to her lower body. She is awake, in pain, and upright on the stretcher struggling to breathe in a hot room amid a sea of bustling caregivers. The trauma surgeon calls for morphine sulfate, 8 mg IV, as the anesthesiologist arrives and prepares to intubate. Shortly, Mrs. L. slumps forward and the event recorder calls out “what’s happened to your patient?” The emergency department doctor asks “who gave the morphine?” to which both the anesthesiologist and a nurse respond (unbeknownst to each other, both had administered the medication). Mrs. L. is intubated and proceeds from manual to mechanical ventilation with ongoing sedation. Despite the smooth management of the issue, the same flawed, non-checking interactions allowed this error to occur, continue, and go unchecked for some time. These types of weaknesses in the system can continue to operate unnoticed until the moment when actions cannot be reversed or managed. What if the medication were an aminoglycoside, two times the dose and what if the patient was a child?

An emphasis on teamwork and team training is a hallmark of high reliability organizations such as those in aviation and nuclear power generation industries.<sup>24,25</sup> The successes in increasing safety via team training in these industries has been translated into successful interventions with team training components in medical domains as well<sup>26</sup> beginning with anesthesiology,<sup>27,10</sup> emergency medicine,<sup>5,28</sup> and others such as the military healthcare system at large.<sup>29</sup> Essentially, it has been recognized that individual competence in clinical skills is not enough; team coordination, communication, and cooperation skills are essential to effective and safe performance.<sup>30,23</sup>

The SBT methodology provides unique opportunities to train and assess teamwork skills. As shown in Figure 1, SBT is an approach to training that incorporates five phases: information, demonstration, practice, feedback, and remediation.<sup>32</sup> Teamwork is inherently dynamic as it depends on the interactions of team members. Because SBT provides opportunities for teams to engage in dynamic practice in contextualized task environments that replicate aspects of the “real world” task environment, the social dynamics of teamwork can be effectively trained and evaluated.<sup>33,8</sup> In the case of Mrs.

L., the team behavior of closed-loop communication (ie, ensuring that information conveyed by the sender is understood by the receiver) concerning medication orders would have prevented the error. During SBT, the presence or absence of this team behavior would be observed and corrective feedback given in a postperformance debrief. SBT provides a compromise between the complexity present in the real world (which is simultaneously a necessity to observe teamwork in action and an obstacle to team performance measurement) and the structure and control necessary to provide systematic measurement and training.

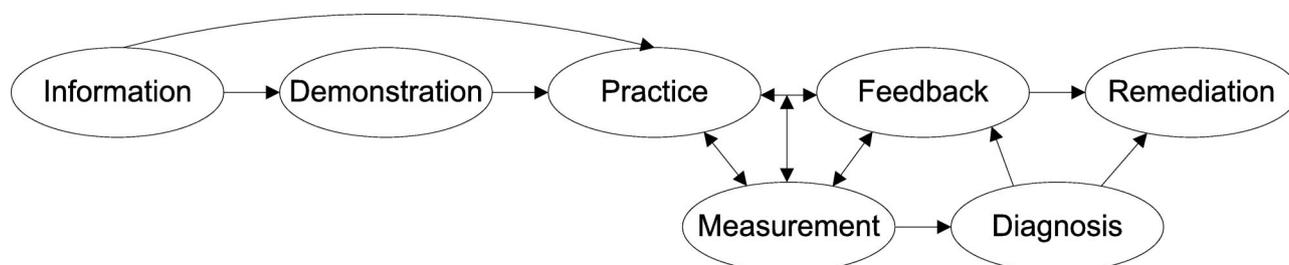
### Why Measure Team Performance in SBT?

Measurement has been identified as the heart of patient safety,<sup>34</sup> meaning that the healthcare system’s ability to improve its processes and create a safer environment, and better patient outcomes must be based upon quality measurement practices. This is particularly true in the context of training where quality measurement practices guide learning and corrective feedback as well as helping to ensure that learners possess the requisite competencies for effective on the job performance. As illustrated in Figure 1, performance measurement plays a central role in SBT.

This article focuses on team performance measurement in SBT as it relates to feedback and remediation as these are two of the most critical uses of team performance measurement in training.<sup>11</sup> To benefit from the practice opportunities embedded in SBT, team members must be guided through their performance processes.<sup>8</sup> That is, they must be provided with feedback that informs them about how they are performing and how to improve their performance.<sup>35</sup> Additionally, practice opportunities in SBT can be viewed as opportunities to detect deficiencies in individual and team competencies. This information can be used to drive the selection of future training to correct these problems.<sup>11</sup>

### What’s so Difficult About Team Performance Measurement in SBT?

Measuring team performance in complex work environments is difficult in general, and medical teams are no exception.<sup>16</sup> It has been shown recently that many of the performance measures that the Medicare system uses as indicators of quality of care delivery are only weakly related to important clinical outcomes such as mortality rates.<sup>36</sup> Does this mean that performance processes do not affect outcomes, that no matter what the hospital staff does, the mortality rates of their patients are only mildly influenced? This is extremely unlikely; the more reasonable interpretation is that the performance measures used by Medicare are not sensitive to



**Figure 1.** The role of measurement and performance diagnosis in SBT (from Salas et al<sup>31</sup>).

the aspects of performance processes that influence patient outcomes.<sup>37</sup> That is, the chosen performance measures do not capture the critical aspects of performance.<sup>38</sup> Clinical outcomes are influenced by many factors, and the performance measures chosen simply do not represent enough of these factors to be useful indicators of quality.<sup>36</sup> This illustrates the danger of using poorly conceived measurement. Deficient measurement begets deficient decisions. Consider the following anecdotal example. The implementation of team training in a large medical center was found to have no impact on operations. This finding was based on measures of the trainee's reaction to the training (the training was evaluated highly), measures of the trainee's knowledge acquisition (pre- and postmeasures indicated that trainees learned the training content), and measures of processes and outcomes in the units (which remained unaffected). However, no observations of teamwork on the units were made because of cost. Post hoc interviews revealed that staff members were not applying the training to the work context. So, it was not that the training was ineffective at building skill and knowledge (as evidenced by knowledge measures), or that teamwork does not improve outcomes (as would be the case if the staff applied their training and outcomes remained unaffected); the real issue (and what was not captured by the original measures) is that the staff did not transfer what they learned to on-the-job performance. Measurement in the original study did not capture all of the critical information necessary to make an informed decision of what had occurred.

Baker and Salas<sup>39,40</sup> detailed several factors that complicate the process of developing effective team performance measures. The complexity of team performance makes meaningful measurement difficult. Team performance involves the dynamic interaction of multiple people, often with heterogeneous knowledge, skills, and attitudes.<sup>41</sup> Consequently, creating a full understanding of team performance requires behavioral, cognitive, and attitudinal measures.<sup>40</sup> The dynamic nature of teamwork means that teams change over time and a single snapshot of performance may not be representative of the team's actual performance.<sup>40</sup> In addition, it is frequently impossible to remove observers or raters from the team performance measurement process and it is difficult to develop and maintain the reliability of observer ratings.<sup>42</sup>

Given these challenges and the importance of team performance to effective and safe healthcare, great care must be taken in developing and implementing team performance measurement systems in SBT. The following section outlines the critical practices for doing just this.

### **Best Practices in Team Performance Measurement in SBT**

Performance measurement can be viewed as an investment in time, effort, and other resources that pays off in terms of decision quality.<sup>43</sup> For the purposes of training, the two most important decisions a measurement system can address are: 1) what meaningful feedback should be given to the team and each individual member, and 2) what further training is required by the team or individual members? This

section presents a set of best practices in the development of team performance measurement systems designed to inform these two decisions. These best practices are summarized in Table 1. The goals of this section are to provide heuristics for increasing the quality of decisions made and the action taken based on the results of team performance measures collected in SBT. This will lead to better training outcomes (eg, increased learning) and ultimately better team performance in the clinical setting.

#### **Best Practice 1: Ground Measures in Theory**

Performance measures should be based on a scientific understanding of team performance. Theory is invaluable to the process of developing quality team performance measures<sup>39</sup>; it is the best guide for answering questions about what to measure. Numerous models and frameworks of team performance and team effectiveness are available.<sup>41</sup> These models outline the inputs, throughputs, and outputs of team performance. Additionally, there is an ever-increasing body of literature dealing specifically with healthcare team effectiveness.<sup>44</sup> These theories and frameworks help to narrow the field of possible measures and focus on the important aspects of performance that determine team effectiveness. For example, a recently proposed theoretical model describes teamwork in terms of five core dimensions (ie, team leadership, mutual performance monitoring, backup behavior, adaptability, and team orientation) and three coordinating mechanisms (ie, mutual trust, shared mental models, and closed-loop communication<sup>45</sup>). Adopting this model of teamwork as the basis for SBT and a team performance measurement system would mean that measures would be developed to capture each of the five core teamwork dimensions and the three coordinating mechanisms. The common trap in team performance measurement is to measure whatever is easy to measure. This is why the most common measures collected during training are the trainee's affective responses to training (ie, did they like the training?). Unfortunately, this information does not provide much insight into the actual performance of the team or how to generate feedback for improvement.<sup>46</sup> Theories of teamwork provide an understanding of what is important to team effectiveness and hence what is important to measure. Unfortunately, there are multiple models of team performance from which to choose. In general, the best model is one that is tailored for a specific domain and has received empirical validation. In the absence of this, general models provide an overview of the critical features of teamwork, but the nature of the task in different domains will dictate the specifics of effective teamwork (eg, differences in leadership structure, interdependencies).

#### **Best Practice 2: Design Measures to Meet Specific Learning Outcomes**

Team performance measurement works best when the goal of measurement is clearly established and considered throughout the development process.<sup>11</sup> For example, a measure developed for the purposes of a high stakes evaluation such as those involved in certifications (eg, medical boards) is not suitable for training purposes. Such an assessment measure is designed to answer the question: does this person meet the minimum standard of performance? Answering this

**Table 1.** Summary of Best Practices in Team Performance Measurement for SBT for Healthcare

## Adopting Best Practices of Team Performance Measurement in SBT

## Ground measures in theory

- Use theory to answer questions about what to measure

- Capture aspects of Input → Process → Output (IPO) models of team performance

## Design measures to meet specific learning outcomes

- Clearly articulate the purpose (ie, specific learning outcomes) of measurement at the beginning of the development process

- Design the measurement system to capture information necessary for making decisions about the learning outcomes (eg, have they been met?)

## Capture competencies

- Explicitly link performance measures to the individual and team competencies targeted for training

## Measure multiple levels of performance

- Performance measures must be sensitive to differences in individual and team performance

## Link measures to scenario events

- Link performance measures to opportunities to perform (ie, events) in the simulation

- Event based measurement affords greater structure and standardization for observers

## Focus on observable behaviors

- Focusing on behaviors decreases the “drift” of observer’s ratings over time (observers are less likely to develop idiosyncratic scoring)

- Measuring observable behavior provides a fine level of granularity which is necessary for providing corrective feedback

## Incorporate multiple measures from different sources

- Different measurement sources provide unique information

- Generate a plan for rapidly integrating multiple sources of measurement

## Capture performance processes in addition to outcomes.

- Team performance measurement should provide information not only about the end result of performance, but about how the team reached that performance outcome

- Team performance measurement should provide information on how to correct team processes

## Create diagnostic power

- Team performance measurement should provide information about the causes of effective and ineffective performance

## Train observers and structure observation protocols

- Establish and implement a program of training and evaluation to ensure observers are accurately and reliably rating performance

- Provide structured protocols to ease the information burden of observation

## Facilitate posttraining debriefs and training remediation

- Team performance measurement should be quickly translated into feedback and decisions about required future training

question at best, identifies if there is a problem, but it does not provide guidance as to what exactly the problem is (eg, what knowledge or skill is deficient) or how to correct the problem. Training is designed to develop a certain set of knowledge, skill, and attitude (KSA) competencies in learners (ie, the learning objectives). The measurement goals subserve the learning objectives; that is, the goal of measurement is to provide information upon which decisions about the learning objectives can be made (eg, do learners possess targeted KSAs; if not, what needs to be done to develop them?). The validity of a measure is defined as “the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests.”<sup>47(p99)</sup> It is the use of the measure that is validated, not the measure in some general sense. Therefore, the intended uses or purpose of the measure should guide the measure’s development from the very beginning.<sup>48</sup> In practice, implementing this best practice requires 1) an explicit statement of what the learning objectives are, and 2) clear linkages between the measures developed and the learning objectives (ie, the KSAs targeted for training).

**Best Practice 3: Capture Competencies**

Team performance measures should be rooted in the individual and team competencies being trained.<sup>13,49</sup> Training is defined as the systematic acquisition of the KSAs necessary for job performance.<sup>1</sup> Therefore, team performance measurement intended to meet this goal should be explicitly tied to the competencies being trained. A general framework of

teamwork competencies—the KSAs necessary for effective teamwork—has been proposed<sup>50,11</sup> and refined over the years.<sup>51</sup> These general listings of teamwork competencies can serve as the starting point for identifying contextualized performance measures that tap into the competencies for a given training program, team, and task environment. It is important to distinguish between individual level taskwork competencies (ie, the KSAs an individual must possess to carry out the parts of their job that do not involve coordination with others such as technical skills) and teamwork competencies (ie, the KSAs necessary for effective coordination and collaboration). These sets of competencies are distinct and should be measured separately.

Because measurement needs to be specific to the competencies being trained and the specific learning objectives, care must be taken in adopting “generic” measurement tools. Remember, measures are validated for specific uses, not in a general sense. When inserting even a “validated” and reliable measure developed for a different purpose without considering whether the measure matches the new context and purpose, the result can easily be a reliable measurement of something completely irrelevant to the decisions being made based on the data. However, measures can be transported to new contexts with careful consideration of its appropriateness (ie, is it capturing competencies targeted for training in the present context?).

#### **Best Practice 4: Measure Multiple Levels of Performance**

Team performance depends on and is influenced by factors not only occurring at the team level (eg, communication) but by levels above (eg, organizational and multiteam system factors such as safety culture) and below (eg, individual level performance such as individual task performance of team members) as well. A better understanding of team performance can be gained by measuring factors at these multiple levels.<sup>52</sup> Fundamentally, team training with SBT requires measures that can distinguish between individual and team performance. For example, consider the case of a team that does not perform adequately based on one specific outcome (eg, patient mortality in a resuscitation scenario). Based on this one measure, it is impossible to determine if the problem was with teamwork skills (eg, the team members did not effectively communicate patient status information) or with individual skills (eg, one individual did not perform his/her task adequately and the simulated patient died as a result). For example, in the case of Mrs. L, the breakdown was at the team level (ie, a lack of closed-loop communication). Team performance is built on top of individual performance, that is, individuals competent at their individual tasks are a necessary, but insufficient, condition for high levels of team performance outcomes. The individuals administering morphine to Mrs. L were each competent at the individual level (ie, each individual correctly administered morphine); however, they were deficient in team level communication behaviors. Capturing performance at multiple levels is necessary for making distinctions between deficiencies in individual and teamwork competencies.<sup>11</sup> Because the ultimate level of team performance outcome is dependent upon both individual and team performances, competencies underlying both of these processes should be targeted for training. Consequently, performance associated with individual performance should be captured to drive individual level feedback (ie, provide guidance for individual skill development) and team level performance should be captured to drive team level feedback (ie, provide guidance for the team as a whole).

#### **Best Practice 5: Link Measures to Scenario Events**

One of the greatest benefits of SBT is that it affords opportunities for practice and observation of performance in situations that faithfully replicate important features of the real world environment.<sup>3,53</sup> That is, SBT brings the complexity of the actual task to a safe and error-tolerant training environment. However, this means that the performance in the simulation must be as complex as it is in the real world. This complexity is a challenge for team performance measurement.<sup>54,55</sup> Event-based training (EBAT) and measurement are generalizable methods for managing this complexity.<sup>56</sup> By developing critical events and inserting them into scenarios, the complexity of performance in SBT is reduced by focusing measurement on predetermined aspects of performance. Critical events are routine or unexpected events that occur during the simulation scenario at a predetermined time. When these events are linked to training objectives, they offer team members the opportunity to perform behaviors associated with the competencies targeted for training.<sup>56</sup> These events enable observers to focus on critical aspects of perfor-

mance that are known to elicit behaviors that are diagnostic of targeted training competencies.<sup>57</sup>

Although EBAT may be the ideal measurement approach in many senses, it is by no means the only method available to capturing team performance. There is no universally accepted best solution, each method has its strengths and weaknesses. For example, although EBAT provides diagnostic power because of explicit linkages between scenario events, team behaviors, and KSAs being trained and increases reliability by focusing on observable behaviors, it is also the most resource intensive to develop. Although a behaviorally anchored rating scale is easier to develop, it provides less information and is more susceptible to rater bias and drift because it requires raters to summate over time and is not explicitly linked to scenario events. In addition, self-report measures are valid means for measuring important moderators of team performance (eg, team cohesion, collective orientation) and some dimensions of teamwork or taskwork knowledge, but they are less suited for capturing the dynamic interactions of team members. In this regard, measurement of team processes during a performance episode is necessary. Table 2 provides a summary of the main measurement methods used for capturing team performance and their associated strengths and weaknesses.

#### **Best Practice 6: Focus on Observable Behaviors**

There are multiple methods of capturing performance. Frequently, global rating scales are used where observers are asked to give an overall assessment of performance or a set of performance dimensions. These types of ratings are known to be plagued with bias and error<sup>59</sup>; high inter-rater reliability can be achieved for global rating scales, but it requires vigilant training. Additionally, global ratings scales do not provide the type of in-depth detail necessary for generating corrective feedback; they are at an inappropriate level of granularity.<sup>31</sup> There are similar limitations for data collected through self-report measures of the trainees. Reducing the amount of subjectivity in observations can provide more reliable and valid measures. This can be accomplished by asking observers to identify discrete behaviors and not make judgments about the quality of performance over time or in some general way.<sup>58,67</sup> By linking the presence or absence of specific teamwork behaviors (eg, the provision of back-up behavior when needed) to teamwork competencies and learning objectives, diagnostic and corrective feedback can be generated rapidly.<sup>68</sup> For example, an observer is rating the quality of team communication on a 7-point scale (1 being poor and 7 being excellent) and rates with a two. When the team receives this rating they know that their team communication was poor, but they do not know how to fix the problem. If, however, the observer was recording the occurrence of closed-loop communication (ie, the receipt of a message being acknowledged by a team member)—an observable behavior—and it was found to be deficient, the team would have information about what needs improvement. In practice, it is necessary to define events (ie, points in time during the scenario) that provide trainees with an opportunity to exhibit performance indicative of the underlying constructs.

**Table 2.** Summary of Main Methods of Team Performance Measurement Used in SBT

Method	Description	Advantages	Disadvantages
Event-based measurement <sup>56–58</sup>	A general method that generates behavioral checklists that are linked to scenario events and KSAs being trained	Maintains explicit connections between measurement opportunities (ie, scenario events), acceptable behaviors, and KSAs being trained  Focuses observers' attention on predefined events  Reduces amount of judgment a rater has to make by focusing on observable behaviors	Development of measures can be time consuming relative to other approaches  Measurement tools must be developed for each scenario
Behaviorally-anchored rating scales (BARS) <sup>60,61</sup>	Provides brief descriptions of behaviors as anchors associated with each particular rating	Amendable to modification Facilitates accurate ratings by providing concrete examples of behaviors	When behavioral anchors contain specific types of behavior, observers tend to focus on these isolated behaviors and miss
Behavioral observation scales (BOS) <sup>62,63</sup>	Generally uses a Likert type scale to rate the frequency of certain team processes	Avoids potential problems with BARS (rating exceptional or isolated performance) by focusing on typical performance	Requires raters to estimate frequencies and consequently ratings may be influenced by recency and primacy effects
Self-report measures <sup>64–66</sup>	Questionnaires administered to each team member individually	Well suited to capture affective factors that influence team performance (eg, collective efficacy, trust, collective orientation, psychological safety)	Does not capture dynamic performance, translating individual scores to team level scores can be problematic

### Best Practice 7: Incorporate Multiple Measures From Different Sources

Triangulation reduces measurement error and provides a more robust picture of team performance. This involves using multiple methods, data sources, and observers for understanding complex phenomenon. If only a single method of measurement is relied upon, that resulting data may be contaminated with variance associated solely with the method and the team's performance. Additionally, team performance is multifaceted, and requires a multifaceted measurement approach<sup>31</sup>; however, a strategy must be in place to make sense of all of this data.<sup>12</sup> To provide timely feedback, there must be a plan for integrating data collected via multiple measures and sources. For example, there must be a procedure in place for integrating ratings from observers and data collected automatically by the simulation to present a clear picture of the team's performance to the trainer who subsequently makes decisions about feedback and remediation. This can be facilitated with the use of computerized rating forms (eg, on PDAs or tablets) whose data can be quickly synthesized with other formats.

### Best Practice 8: Capture Performance Processes in Addition to Outcomes

An outcome measure is a composite measure in the sense that it is the result of many performance processes.<sup>69,70</sup> For example, the state of the simulated patient after the scenario ends is a function of how the team members performed their individual tasks as well as how they communicated and coordinated their actions. A process measure focuses on how the task was accomplished. Performance outcomes can be useful, but they are limited. Any one outcome can be reached by numerous processes. Therefore, it is impossible to tell how a team reached an outcome by considering the outcome alone. This is problematic for training purposes, as the process of performance is what is being trained.<sup>11</sup> Information about team processes (eg, communication, coordination) is

necessary to generate corrective feedback. In practice, this can be accomplished by clearly defining the critical events in the scenario and capturing the behavior responses (eg, communication and coordination) of the team. Feedback on this issue can be facilitated with the use of the review of video and audio recordings at the critical events where good or poor performance occurred.

### Best Practice 9: Create Diagnostic Power

Performance diagnosis is the process of determining the causes of effective and ineffective performance.<sup>11</sup> Information about the causes of performance is critical for training purposes. Once the deficiencies causing poor performance have been identified, they can be corrected. Diagnostic measures can serve as a foundation for providing corrective feedback as well as making decisions about what training is required to correct the deficiencies.<sup>12</sup> The diagnostic capability of a measurement system is developed by creating rich, detailed, and informative performance profiles.<sup>31</sup> For this to be effective, there must exist an explicit linkage between each measure and the underlying KSAs. A team's set of scores on a battery of measures (eg, behavioral responses to critical events, observer ratings of team process) should be interpreted relative to levels of KSAs.

### Best Practice 10: Train Observers and Structure Observation Protocols

Observers are presently a necessity when evaluating the complexities of teamwork.<sup>40</sup> When using observers to rate performance, inter-rater reliability (ie, do different raters produce the same performance measurement scores for a given performance episode?) is always an issue.<sup>42</sup> Over time, the ratings of observers will drift apart. It is imperative to establish a program for building and maintaining observers' rating performance. This involves developing a training and evaluation program and being vigilant in its application.<sup>71–73</sup> For example, each observer should go through a brief training including descriptions of the performance dimensions being rated, examples of a range of

performance (eg, video clips from past scenarios or staged), and an evaluation where they rate sample video clips and are compared with a standard. Observers not meeting the standard should be given feedback on how to adjust their rating. Additionally, structured observation protocols can help with the reliability of observer data as well as consistency of feedback given to trainees. This includes tools that focus the observer's attention on critical aspects of performance. This can be achieved by creating tightly scripted and structured training scenarios and tying measurement opportunities to a priori defined events. This enables observers to focus their attention and increase the reliability of their measurements.

#### **Best Practice 11: Facilitate Posttraining Debriefs and Training Remediation**

As should be clear by this point, team performance measurement for training purposes should provide a basis for generating feedback as well as making decisions about what additional training a team or its individual members needs.<sup>11,12</sup> Therefore, team performance measurement works best when it allows for the timely generation of postperformance review aids to increase the effectiveness of feedback. To this end, team performance measures need to provide accurate and timely insight in the teamwork competencies underlying the processes of team performance. Different types of performance measures have different "informational yields."<sup>59</sup> Meeting the goals of training teams with SBT requires measures with high informational yield—measures that can support decision making about feedback and remediation.

## **DISCUSSION**

The best practices offered in this article are methods for improving the quality of team performance measurement. A central issue of quality in any measure is its reliability. If a measure is not reliable, other properties of the measure cannot be assessed. Because of the heavy reliance on observers in team performance measurement, the issue of inter-rater reliability is paramount. The best practice of "training observers and structuring observation protocols" directly addresses this issue by reducing the amount of variance in team performance measurement scores attributable to the observer. The best practices of "focusing on observable behaviors" limit the requirements for making judgments of quality or summing over time,<sup>60</sup> known sources of error variance. Additionally, by "linking measures to scenario events," observers' attention can be focused on critical aspects of performance, which reduces their overall attentional demands and increases reliability.<sup>60</sup> Other issues of quality of measurement are concerned with the types of decisions that can be made based on the measurements taken. Many of the remaining best practices (ie, ground measures in theory, design measures to meet specific learning outcomes, capture competencies, measure multiple levels of performance, capture performance process in addition to outcomes, create diagnostic power) involve establishing clear connections between the data produced and questions of interest (ie, to what degree do the trainees possess the targeted KSAs?). The remaining best practices of

"facilitating posttraining debriefs and training remediation" involve leveraging the information generated through performance measurement into increased learning outcomes.

These best practices are based upon work in many diverse application domains. Organizations in military and aviation industries tend to employ SBT in a centralized manner, with dedicated simulation centers and the adoption of standardized training practices. Simulation in healthcare differs in this respect with simulation use trending toward distribution throughout the system as opposed to a centralized center. This poses several challenges (eg, access to specialized resources and personnel, the possibility of many people duplicating efforts, etc.); however, it does not change the need to develop quality team performance measurement. In fact, this distribution in healthcare simulation emphasizes the importance of communicating best practices and lessons learned through vehicles such as this article.

## **Concluding Remarks**

The best practices presented in this article offer scientifically grounded guidance for developing team performance measures for the purposes of team training in SBT. Team performance measures are critical to the success of SBT as feedback, remediation, and in the end, learning depends on the effectiveness of the measurement system developed. Systematic application of the SBT methodology—including team performance measurement—to training teamwork skills is a proven method for increasing safety in complex and dynamic environments. The ultimate effectiveness of the adoption of this approach to healthcare can be increased by following the best practices offered in this article.

## **ACKNOWLEDGMENTS**

*We thank David Gaba and three anonymous reviewers for constructive comments on an earlier version of this article.*

## **REFERENCES**

1. Goldstein IL, Ford JK. Training in Organizations. 4th ed. Belmont, CA: Wadsworth; 2002.
2. Kohn LT, Corrigan JM, Donaldson, MS, eds. To Err is Human: Building a Safer Health System. Washington, DC: National Academy Press; 2000.
3. Gaba DM. The future vision of simulation in health care. *Qual Saf Health Care* 2004;13(suppl 1):i2–i10.
4. Shapiro MJ, Morey JC, Small SD, et al. Simulation based teamwork training for emergency department staff: does it improve clinical team performance when added to an existing didactic teamwork curriculum? *Qual Saf Health Care* 2004;13:417–421.
5. Morey JC, Simon R, Jay GD, et al. Error reduction and performance improvement in the emergency department through formal teamwork training: evaluation results of the MedTeams project. *Health Serv Res* 2002;37:1553–1581.
6. Pizzi L, Goldfarb NI, Nash DB. Crew Resource Management and its Applications in Medicine. Rockville, MD: Agency for Healthcare Research and Quality; 2001. Evidence Report/Technology Assessment, Number 43.
7. Jha AK, Duncan BW, Bates DW. Simulator-Based Training and Patient Safety. Rockville, MD: Agency for Healthcare Research and Quality; 2001. No. Evidence Report/Technology Assessment, Number 43.
8. Salas E, Wilson KA, Burke CS, Priest HA. Using simulation-based training to improve patient safety: what does it take? *Jt Comm J Qual Patient Saf* 2005;31:363–371.

9. Headey NA, Undre S, Vincent CA. Developing observational measures of performance in surgical teams. *Qual Saf Health Care* 2004;13(suppl 1):i33–i40.
10. Gaba DM, Howard SK, Fish KJ, Smith BE, Sowb YA. Simulation-based training in anesthesia crisis resource management (ACRM): a decade of experience. *Simul Gaming* 2001;32:175–193.
11. Cannon-Bowers JA, Salas E. A framework for developing team performance measures in training. In: Brannick MT, Salas E, Prince C, eds. *Team Performance and Measurement: Theory, Methods, and Applications*. Mahwah, NJ: Erlbaum; 1997:45–62.
12. Oser RL, Cannon-Bowers JA, Salas E, Dwyer DJ. Enhancing human performance in technology-rich environments. In: Stamford SE, ed. *Guidelines for Scenario-Based Training, Human/Technology Interaction in Complex Systems*, vol 9. CT: JAI Press; 1999:175–202.
13. Salas E, Cannon-Bowers JA. The science of team training: a decade of progress. *Annu Rev Psychol* 2001;52:471–499.
14. Brannick MT, Salas E, Prince C, eds. *Team Performance Assessment and Measurement: Theory, Methods, and Applications*. Mahwah, NJ: Erlbaum; 1997.
15. Salas E, Wilson-Donnelly KA, Sims DE, Burke CS, Priest HA. Teamwork training for patient safety: best practices and guiding principles. In: Carayon P, ed. *Handbook of Human Factors and Ergonomics in Health Care and Patient Safety*. Mahwah, NJ: Erlbaum; 2007:803–822.
16. Zeiss AM. Measuring team performance in health care. In: Heinemann GD, Zeiss AM, eds. *Team Performance in Health Care: Assessment and Development*. New York: Kluwer Academic; 2002: 19–28.
17. Wears RL, Leape LL. Human error in emergency medicine. *Ann Emerg Med* 1999;34:370–372.
18. Risser DT, Rice MM, Salisbury ML, Simon R, Jay GD, Berns SD. The potential for improved teamwork to reduce medical errors in the emergency department. *Ann Emerg Med* 1999;34:373–383.
19. Poulton BC, West MA. Effective multidisciplinary teamwork in primary health care. *J Adv Nurs* 1993;18:918–925.
20. Heinemann GD. Teams in health care settings. In: Heinemann GD, Zeiss AM. *Team Performance in Health Care: Assessment and Development*. New York: Kluwer Academic; 2002:3–17.
21. Orasanu J, Connolly T. The reinvention of decision making. In: Klein G, Orasanu J, Calderwood R, Zsombok CE, eds. *Decision Making in Action: Models and Methods*. Norwood, CT: Ablex; 1993:3–20.
22. Bogner MS. Naturalistic decision making in health care. In: Zsombok CE, Klein G, eds. *Naturalistic Decision Making*. Mahwah, NJ: Erlbaum; 1997:61–69.
23. Salas E, Rosen MA, King HB. Managing teams managing crises: principles for teamwork in the ER and beyond. *Theor Iss Ergonomics Sci* 2007;8:381–394.
24. Wilson KA, Burke CS, Priest HA, Salas E. Promoting health care safety through training high reliability teams. *Qual Saf Health Care* 2005;14: 303–309.
25. LaPorte TR, Consolini PM. Working in practice but not in theory: theoretical challenges of “high reliability organizations.” *J Public Adm Res Theory* 1991;1:19–48.
26. Baker DP, Salas E, King H, Battles J, Barach P. The role of teamwork in professional education of physicians: current status and assessment recommendations. *Jt Comm J Qual Patient Saf* 2005;31:185–202.
27. Holzman RS, Cooper JB, Gaba DM, Philip JH, Small SD, Feinstein D. Anesthesia crisis resource management: real-life simulation training in operating room crises. *J Clin Anesth* 1995;7:675–687.
28. Reznick M, Smith-Coggins R, Howard S, et al. Emergency medicine crisis resource management (EMCRM): pilot study of a simulation-based crisis management course for emergency medicine. *Acad Emerg Med* 2003;10:386–389.
29. Alonso A, Baker D, Holtzman A, et al. Reducing medical error in the military health system: how can team training help? *Hum Resour Manage Rev* 2006;16:396–415.
30. Heinemann GD, Zeiss AM, eds. *Team Performance in Health Care: Assessment and Development*. New York: Kluwer Academic; 2002.
31. Salas E, Rosen MA, Burke CS, Nicholson D, Howse WR. Markers for enhancing team cognition in complex environments: the power of team performance diagnosis. *Aviat, Space, Environ Med* 2007;78(suppl 5):B77–B85.
32. Salas E, Priest HA, Wilson KA, Burke CS. Scenario-based training: improving military mission performance and adaptability. In: Adler AB, Castro CA, Britt TW, eds. *Minds in the Military: The Psychology of Serving in Peace and Conflict*, vol 2: *Operational Stress*. Westport, CT: Praeger Security International; 2006:32–53.
33. Salas E, Cannon-Bowers JA. The anatomy of team training. In: Tobias S, Fletcher JD, eds. *Training & Retraining*. New York: Macmillan Reference USA; 2000:312–335.
34. Joint Commission. Measurement: the heart of patient safety. *Jt Comm Benchmark* 2006;8:4–7.
35. Ende J. Feedback in clinical medical education. *JAMA* 1983;250: 777–781.
36. Werner RM, Bradlow ET. Relationship between Medicare’s hospital compare performance measures and mortality rates. *JAMA* 2006;296: 2694–2702.
37. Horn SD. Performance measures and clinical outcomes. *JAMA* 2006; 296:2731–2732.
38. Ferraco K, Spath PL. Measuring performance in high-risk processes. In: Spath PL, ed. *Error Reduction In Healthcare: A Systems Approach To Improving Patient Safety*. San Francisco: Jossey-Bass; 2000:17–95.
39. Baker DP, Salas E. Principles for measuring teamwork skills. *Hum Factors* 1992;34:469–475.
40. Baker DP, Salas E. Principles for measuring teamwork: a summary and look toward the future. In: Brannick MT, Salas E, Prince C, eds. *Team Performance Assessment and Measurement*. Mahwah, NJ: Erlbaum; 1997:331–355.
41. Salas E, Stagl KC, Burke CS, Goodwin GF. Fostering team effectiveness in organizations: toward an integrative theoretical framework of team performance. In: JW, Spaulding W, Poland J, eds. *Modeling Complex Systems: Motivation, Cognition and Social Processes*, Nebraska Symposium on Motivation, vol 51. Lincoln, NE: University of Nebraska Press. In press.
42. Bakeman R, Gottman JM. *Observing Interaction: An Introduction to Sequential Analysis*, 2nd ed. Cambridge, UK: Cambridge University Press; 1997.
43. Brannick MT, Prince C. Overview of team performance measurement. In: Brannick MT, Salas E, Prince C, eds. *Team Performance Assessment and Measurement: Theory, Methods, and Applications*. Mahwah, NJ: Erlbaum; 1997:3–16.
44. Lemieux-Charles L, McGuire WL. What do we know about health care team effectiveness? A review of the literature. *Med Care Res Rev* 2006; 63:263–300.
45. Salas E, Sims DE, Burke CS. Is there a big five in teamwork? *Small Group Res* 2005;36:555–599.
46. Morgan RB, Casper WJ. Examining the factor structure of participant reactions to training: a multidimensional approach. *Hum Resour Dev Q* 2000;11:301–318.
47. AERA, APA, and NCME. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association; 1999.
48. Baker EL, O’Neil HF Jr. Assessing instructional outcomes. In: Gagné RM, ed. *Instructional Technology: Foundations*. Hillsdale, NJ: Erlbaum; 1987:343–377.
49. Campbell JP, Kuncel NR. Individual and team training. In: Anderson N, Ones DS, Sinangil HK, Viswesvaran C, eds. *Handbook of Industrial, Work, and Organizational Psychology*. London: Sage; 2001: 278–313.
50. Cannon-Bowers JA, Tannenbaum SI, Salas E, Volpe CE. Defining team competencies and establishing team training requirements. In: Guzzo R, Salas E, eds. *Team Effectiveness and Decision Making in Organizations*. San Francisco, CA: Jossey-Bass; 1995:333–380.

51. Salas E, Rosen MA, Burke CS, Goodwin GF. The wisdom of collectives in organizations: An update of the teamwork competencies. In: Salas E, Goodwin GF, Burke CS, eds. *Team Effectiveness In Complex Organizations: Cross-Disciplinary Perspectives And Approaches*. Mahwah, NJ: Erlbaum. In press.
52. Hackman JR. Learning more by crossing levels: evidence from airplanes, hospitals, and orchestras. *J Organ Behav* 2003;24:905–922.
53. Hays RT, Singer MJ. *Simulation Fidelity in Training System Design*. New York: Springer-Verlag; 1989.
54. Arvey RD, Murphy KR. Performance evaluation in work settings. *Annu Rev Psychol* 1998;49:141–168.
55. Austin JT, Crespin TR. Problems of criteria in industrial and organizational psychology: progress, problems, and prospects. In: Bennet W Jr, Lance CE, Woehr DJ, eds. *Performance Measurement: Current Perspectives and Future Challenges*. Mahwah, NJ: Erlbaum; 2006:9–48.
56. Fowlkes JE, Dwyer DJ, Oser RL, Salas E. Event-based approach to training (EBAT). *Int J Aviat Psychol* 1998;8:209–221.
57. Fowlkes JE, Burke CS. Targeted acceptable responses to generated events or tasks (TARGETs). In: Stanton N, Hedge A, Brookhuis K, Salas E, Hendrick H, eds. *Handbook of Human Factors and Ergonomics Methods*. Boca Raton, FL: CRC Press; 2005:53–1–53–6.
58. Fowlkes JE, Lane NE, Salas E, Franz T, Oser R. Improving the measurement of team performance: the TARGETs methodology. *Mil Psychol* 1994;6:47–61.
59. Swing SR. Assessing the ACGME general competencies: general considerations and assessment methods. *Acad Emerg Med* 2002;9:1278–1288.
60. Murphy KR, Pardaffy VA. Bias in behaviorally anchored rating scales: global or scale-specific? *J Appl Psy* 1989;74:343–346.
61. Kendall DL, Salas E. Measuring team performance: review of current methods and consideration of future needs. In: Ness JW, Tepe V, Ritzer D, eds. *The Science and Simulation of Human Performance*. Boston: Elsevier; 2004:307–326.
62. Salas E, Burke CS, Fowlkes JE, Preist HA. On measuring teamwork skills. In: Thomas JC, ed. *Comprehensive Handbook of Psychological Assessment*. Hoboken, NJ: John Wiley & Sons; 2003:427–442.
63. Tziner A, Joanis C, Murphy KR. A comparison of three methods of performance appraisal with regard to goal properties, goal perception, and rater satisfaction. *Group Organ Manage* 2000;25:175–190.
64. Tagger S, Seigts GH. Leader and staff role efficacy as antecedents of collective-efficacy and team performance. *Hum Perf* 2003;16:131–156.
65. Driskell JE, Salas E. Collective behavior and team performance. *Hum Factors* 1992;34:277–288.
66. Malec JF, Torsher LC, Dunn WF, et al. The mayo high performance teamwork scale: reliability and validity for evaluating key crew resource management skills. *Simul Healthcare* 2007;2:4–10.
67. Dwyer DJ, Oser RL, Salas E, Fowlkes JE. Performance measurement in distributed environments: initial results and implications for training. *Mil Psychol* 1999;11:189–215.
68. Fowlkes JE, Burke CS. Event-based approach to training (EBAT). In: Stanton N, Hedge A, Brookhuis K, Salas E, Hendrick H, eds. *Handbook of Human Factors and Ergonomics Methods*. Boca Raton, FL: CRC Press; 2005:47–1–47–5.
69. Yang H, Sackett PR, Arvey RD. Statistical power and cost in training evaluation: some new considerations. *Pers Psychol* 1996;49:651–668.
70. Lane NE. Issues in Performance Measurement for Military Aviation with Applications to Air Combat Maneuvering (No. NTSC TR-86-008). Orlando, FL: Naval Training Systems Center & U.S. Army Research Office; 1986.
71. Holt RW, Hansberger JT, Boehm-Davis DA. Improving rater calibration in aviation: A case study. *Int J Aviat Psychol* 2002;12:305–330.
72. Goldsmith TE, Johnson PJ. Assessing and improving evaluation of aircrew performance. *Int J Aviat Psychol* 2002;12:223–240.
73. Mulqueen C, Baker DP, Dismukes RK. Pilot instructor rater training: the utility of the multifacet item response theory model. *Int J Aviat Psychol* 2002;12:287–303.