Running head: GUIDELINES FOR PM IN SBT

Guidelines for Performance Measurement in Simulation-based Training

Eduardo Salas

&

Michael A. Rosen

Department of Psychology, and
Institute for Simulation and Training
University of Central Florida

Prepared for:
Janet D. Held
Navy Personnel Research, Studies, & Technology (NPRST/PERS-13)
Bureau of Naval Personnel
5720 Integrity Drive
Millington, TN 38055-1000

29-10-2007

Address correspondence to:
Eduardo Salas, Ph. D.
Department of Psychology, and
Institute for Simulation & Training
University of Central Florida
3100 Technology Parkway
Orlando, Fl 32826
(407) 882-1325
(407) 882-1550 fx
(407) 970-1375 cell
esalas@ist.ucf.edu

Introduction

The complexity of jobs is ever increasing in the military and civilian sectors. Employees are required to manage more demanding individual tasks (e.g., handle more sophisticated equipment), coordinate their actions in a rapid and dynamic fashion (e.g., maintain concurrent membership on multiple teams), and adapt to more complex environments than those encountered in the past. Simulation-based Training (SBT) is a valuable tool in developing the knowledge, skills and attitudes (KSAs) necessary for personnel to excel in these types of jobs. Fundamentally, SBT is unique in that it provides opportunities to practice performance in environments that faithfully replicate important features of the 'real world' so that the acquisition and transfer of KSAs is facilitated (Salas, Priest, Wilson, & Burke, 2006). Performance measurement is a critical component of this process, as illustrated in Figure 1. Detailed descriptions of the importance of performance measurement in SBT can be found elsewhere (see Salas & Rosen, 2006; Salas & Rosen, 2007; Cannon-Bowers & Salas, 1997). For the purposes of this paper, we summarize these issues in three overarching principles of performance measurement in SBT.

First, *training goes hand in hand with performance measurement.* Training is defined as the "systematic acquisition of skills, rules, concepts, or attitudes that result in improved performance in another environment" (Goldstein & Ford, 2002, p. 1). Performance measurement is critical to ensuring that training is *systematic.* It is the means by which trainees receive standardized and structured feedback based on their actual performance; it forms the basis of making informed and standardized decisions

about what future training the student needs; and, it is the means by which the effectiveness of a training program can be evaluated.

Second, *performance measurement goes hand in hand with learning objectives.* A central question of any exercise in measurement is deciding and clearly defining what is important to measure. By clearly and thoroughly stating the learning objectives of a training program, performance measurement can be tied to the purpose of training. When this is done, the feedback generated from performance measures can accelerate learning and improvement in the aspects of performance targeted in the learning objectives. Similarly, the trainee's progress can be tracked in relation to these learning objectives.

Third, *the sequence and structure of training can be optimal, but in the absence of performance measurement, learning opportunities can be lost.* Training should be progressive; complex skills should be trained after the simpler tasks or task components have been acquired. SBT should build from providing information, to demonstrations, to practice, to feedback and remediation (as shown in Figure 1). However, even if the training conforms to this ideal structure, it will not provide optimal learning outcomes if trainee performance is not measured through the training program. As training progresses, performance needs to be diagnosed in order to determine if someone is ready to progress to more complex aspects of training. Providing feedback based on performance measures that are diagnostic will also increase learning and move people to more complex levels of training faster.

This paper summarizes and synthesizes best practices from the SBT and performance measurement scientific literatures (see Salas & Rosen, 2006) as well as best practices developed from observations and interviews at several Navy and Air Force

training schools (see Salas & Rosen, 2007). This synthesis is presented in the form of guidelines for developing and implementing performance measurement systems in SBT.

Guidelines

The guidelines presented in this section are empirically, theoretically, and practically based; that is, they are rooted in the science of learning and outline a step by step process for creating useful, reliable, valid, and diagnostic measures of performance in SBT. These guidelines are intended to be for ideal, optimal, or the best possible conditions. Practical considerations such as access to or availability of resources and time constraints may not allow for all of these to be implemented in full. Guidelines are listed in Table 1 along with tips for successfully meeting each guideline.

Guideline #1: Conduct a Cognitive Task Analysis.

For a measurement effort to be successful, the 'what' that is being measured must be clearly and accurately defined. For the purposes of performance measurement in SBT, the 'what' that is being measured comprises the KSAs, or competencies, targeted for acquisition. Therefore, defining the KSAs that underlie effective performance is the first step in the process of developing an effective performance measurement system and ultimately to developing a successful training system. There are several methods available for deriving a list of KSAs (see Goldstein & Ford, 2002); however, methods of Cognitive Task Analysis (CTA) have become a necessity when dealing with the complex knowledge-based performance that is increasingly characteristic of the modern workplace. Derived from the cognitive sciences, CTA is a set of tools, techniques, and methods for eliciting, analyzing, and representing the knowledge and cognitive processes used during task performance (Schraagen, Chipman, & Shalin, 2000; Crandall, Klein, &

Hoffman, 2006). CTA can be use to generate an understanding of how experts perform tasks, and consequently, an understanding of the KSAs that should be targeted for training.

Guideline #2: Develop measurable learning outcomes.

This list of KSAs forms the foundation of the training program and must be translated into a set of learning outcomes that can be assessed and measured. Learning outcomes are the goals of the training program; that is, they are the criteria by which the success of the training program can be assessed. Measurable learning outcomes are essential for two primary reasons. First, a trainee's performance must be assessed throughout training in order to provide systematic feedback and remediation. Decisions about what feedback to give and what type of training is required should be based on an assessment of the trainee relative to the measureable learning outcomes (e.g., has the trainee reached the learning outcome? If not, what is required to progress to that point?). Second, the effectiveness of the training program is evaluated relative to its ability to meet these learning outcomes. Without measuring trainee progress relative to learning outcomes, questions about the quality of and effectiveness of the training program cannot be answered.

Guideline #3: For each learning outcome, derive a set of specific metrics.

Because the purpose of performance measurement in SBT is to support learning and assessment, it is critical that there connections between learning objectives and the performance measures. A frequent mistake in performance measurement is to capture whatever is easy to measure; however, the learning objectives should drive the development of performance measures from the beginning. Performance measures

should be developed or chosen based on the amount of information they provide about the learning objectives. Performance is multidimensional (Campbell, 1999); that is, performance is determined by multiple KSAs. This is especially true in the complex types of environments trained using SBT in the military. Therefore, in order to adequately measure performance, each dimension must be sampled. If a critical dimension of performance is not measured, than systematic decisions about a trainee in regards to that performance dimension can not be made (e.g., what feedback should be given to correct insufficient performance or reinforce correct performance; what further training is needed; does the trainee possess adequate levels of targeted competencies?). So for each of the learning objectives defined, a least one metric specific to that learning objective must be developed. Learning objectives that involve complex performance will necessarily involve a more complex approach to measurement (i.e., more measures will likely be necessary to capture the performance targeted for acquisition).

Guideline #4: Develop behavioral markers of performance for each learning outcome.

One of the greatest advantages of SBT is its support for contextualized practice activities. SBT allows trainees to practice tasks in complex situations representative of the real world task environment. Consequently, the performance that must be measured during practice in SBT is complex as well. Behavioral markers are very useful for capturing complex performance. Essentially, these behavioral markers are descriptions of what good or poor performance look like; that is, they are forms of performance process that indicate the presence or absence of KSAs targeted for acquisition. Behavioral markers can serve to guide the development of the content of measurement

tools such as Behaviorally Anchored Ratings Scales (BARS), Behavioral Observation Scales (BOS), and event-based measurement tools.

Guideline #5: Develop metrics that are diagnostic of performance.

Performance diagnosis is the process of determining the causes of effective and ineffective performance (Salas, Rosen, Burke, Nicholson, & Howse, 2007; Cannon-Bowers & Salas, 1997). The ability to systematically understanding why a trainee's performance was effective or ineffective drives learning. For a summative evaluation, measurement must answer the question: does this trainee meet the set learning outcomes? However, for training to maximize the acquisition of KSAs throughout the training process, measurement must not only answer the above stated question, but it must also reveal the reasons underlying observed performance. Therefore, metrics must be developed relative to standards of performance and afford inferences about why observed performance occurred.

Guideline # 6: Use multiple data sources and types to capture performance.

In capturing multidimensional performance, it is necessary to employ various sources and types of data. Different measurement approaches may be better suited for capturing different aspects of performance (Shadish, Cooke, & Campbell, 2002). For example, consider the case of a training program designed to build leadership skills. Part of being a good team leader is being able to motivate team members. This can be measured by observers capturing the leader's efforts at motivating team members as well as through self-report measures of motivation levels from the team members; however, it would be extremely difficult to capture this information from some type of automated performance measurement tool embedded in the simulation (e.g., one that captures

frequency or efficiency of behaviors). By not relying on one method or one source of data, the complexities of performance can be better represented in the measurements taken. The fit between the measures employed and the nature of the performance of concern is critical as well (e.g., leadership is a social process and will require an observer knowledgeable about leadership skills; a technical skill can often be captured with accuracy and efficiency measures).

Guideline # 7: Capture performance at multiple levels of analysis.

Not only is performance multidimensional, it is a multilevel phenomenon as well (Kozlowski, Gully, Nason, & Smith, 1999). Individuals must frequently interact interdependently within a team. This poses an important issue for performance measurement: Performance measurement must be able to separately capture individual performance (i.e., aspects of an individual's performance that are not dependent upon the actions or inputs of others) as well as team level performance (i.e., aspects of performance; Cannon-Bowers & Salas, 1997). This is important for the purposes of feedback as individuals need feedback on their own performance as well as the team's performance. If a team has a low level of performance outcome, it may be unclear whether this was due to the poor individual performance of one or several team members or whether it was due to poor teamwork (e.g., communication and coordination) of the team as a whole. Performance measures need to be sensitive to these differences.

Guideline # 8: Capture the processes of performance in addition to outcomes.

Performance outcomes (e.g., mission success, efficiency, accuracy, etc.) are no doubt important. However, the processes of performance are critical in SBT; that is, the purpose of training is to develop the ability in the learner to correctly reach a

performance outcome (i.e., the process of performance), not simply to reach the correct outcome (i.e., the outcomes of performance are not trained). So, if the process is trained, and not the outcome, the process needs to be measured to ensure that trainees are acquiring the correct processes of performance and not just reaching an outcome through luck or sub-optimal processes.

Guideline # 9: Create a plan for integrating multiple sources and types of measurement.

The preceding guidelines have advocated a complex approach to measuring performance. This is justified because of the level of complexity of performance trained in SBT. However, there must be a plan in place to make sense of all of this data (Oser, Cannon-Bowers, Salas, & Dwyer, 1999). For feedback to be most effective, it needs to be delivered in a timely manner (i.e., as close to immediately post-practice as possible). Tools can be developed to organize the measurement data into after-action review aids for instructors. This process is facilitated if all measurement is captured in electronic form (i.e., automated performance measurement, observers using PDAs or other type of computer to rate performance).

Guideline # 10: Automate as much of the performance measurement collection and analysis as possible.

The availability of instructors/observers is frequently a rate limiting factor in throughput of training programs. Therefore, developing automated systems for the collection and analysis of performance measurement data is extremely useful. This type of automation can run along a continuum from the point of embedded instructional tutors that diagnose performance and make feedback and remediation decisions autonomously (e.g., Zachary, Cannon-Bowers, Bilazarian, Krecker, Lardieri, & Burns, 1999), to more

simple approaches such as having instructors make ratings on PDAs or computers so that the information can quickly be integrated and displayed. When designing these automated systems, it is important to build in flexibility. The nature of the task being trained will likely change with the introduction of new tools and technology. If the automated performance measurement tool can not be adapted to the changes in the nature of the performance being trained, it will become obsolete.

Guideline # 11: Develop and implement training programs for observers/instructors.

The role of the instructor/observer is critical in SBT; especially for complicated aspects of performance whose measurement and diagnosis can not be automated easily (e.g., leadership or interpersonal skills). Therefore, it is imperative that instructors and observers receive training. This includes, but is not limited to, training on rating performance so that there is high inter-rater reliability. Other important aspects of training include methods for developing new scenarios (so that they are linked to learning objectives and performance measures) and how to manage, interpret, and leverage performance measurement into effective feedback and training decisions. Instructors should also be trained on how to help establish a sense of realism in the simulations (discussed more later).

Guideline # 12: Provide structured tools/protocols for observations.

In addition to training designed to increase the reliability of ratings, a well developed and structured observation protocol can greatly facilitate higher quality performance measurement. These types of tools focus the attention of observers and increase the consistency of what is observed across different instructors/observers. A good measurement protocol associates the structure of the scenario with the structure of

measurement opportunities. A well developed observation tool will guide the instructor through the scenario. It should be ordered in time with either specific events or general phases of performance listed in the sequence they occur in the scenario. Each of these events will be associated with the ratings or observations the instructor should make at that time. This reduces the overall workload of the instructor (i.e., the instructor doesn't have to pay attention to every aspect of performance, just the critical ones that have been identified and targeted for training).

Guideline # 13: Use checklists for observations that link discrete behaviors to scripted events.

As discussed above, an extremely successful approach for developing reliable and valid measures of performance in SBT is to use the structure of the scenario as a guide for measurement opportunities. This works best when critical events are embedded into the scenario. These events are opportunities for the trainee to exhibit performance in a way that provides insight into the underlying KSAs being trained. For example, a critical event in a scenario for pilot training could be an engine failure. The pilot's response to this engine failure can provide insight into the KSAs that pilot has. For that event, a set of acceptable behaviors are defined before hand (i.e., things the pilot can do that are appropriate responses to an engine failure). During the simulation, the instructor knows when the critical event will occur, an engine failure in this example, and what behaviors are acceptable. These should be listed on a checklist, and all the observer has to do is check whether or not the acceptable behaviors occurred.

Guideline # 14: Focus performance measurement on discrete and observable behaviors.

Different types of measurement will require the instructor/observer to use more or less judgment in rating performance. For example, if an instructor is asked to give a rating of overall performance (e.g., to summate over time) there is more opportunity for biases to surface and contaminate measures as well as increases in differences between instructors' rating performance (Bakeman & Gottman, 1997). By focusing on discrete behaviors (i.e., did the trainee do behavior x?), observers only have to make a very simple judgment: did or did not some behavior occur? This helps to eliminate much of the bias associated with subjective ratings. Consequently, the task of training and maintaining high inter-rater reliability is simplified.

Guideline # 15: Create and maintain a systematic and organized representation of performance.

Many of the guidelines presented so far advocate a process of explicitly defining the discrete behaviors that are acceptable ahead of time (e.g., creating checklists of behaviors that are expected responses to events in the scenario). In order for instructors to create new scenarios with new events and accompanying measurement tools consisting of discrete responses to events which are linked to training objectives, there must be an understanding of good and poor performance. This understanding has to be systematic and organized so that performance measures are consistent across the different scenarios developed.

Guideline # 16: Do not over-burden observers; maintain a good ratio of observers to trainees.

Several guidelines offered so far provide the means for increasing the reliability and validity of observations made by instructors (e.g., training, protocols and checklists).

These are valuable tools, but they can only do so much. There is a definite capacity to how much an instructor can be expected to rate. Therefore, there must be a good ratio of observers/instructors to trainees. This ratio will change with the nature of the training and the type of performance being assessed.

Guideline # 17: Create realism and exploit the trainee's sense of presence.

A central strength of the SBT approach is the practice opportunities. This practice is so effective as a learning tool because the situations in which trainees practice are designed to closely resemble the situations of job performance. To fully capitalize on this opportunity, the trainees must feel as sense of presence; they must feel that the training environment is engaging and realistic. This sense of presence comes not only from the physical characteristics of the simulator, but from the attitudes and climate established by instructors. The instructors must convey a sense of realism to the trainees.

Concluding Remarks

SBT is a methodology increasingly called upon to prepare personnel for complex and dynamic tasks. The effectiveness of this methodology and the degree to which it can reliably and systematically improve performance rests in large part upon the performance measurement practices. Performance measurement informs decisions about training and provides the basis for structured feedback. The guidelines offered in this report are scientifically based and practically focused. Following these guidelines will produce better performance measurement in SBT and consequently better training outcomes.

Guideline	Tips	Sources
1. *Conduct a Cognitive Task Analysis.*	• Obtain 'buy in' from experts. • Understand as much of the domain as possible. • Develop a protocol. • Collect data focused on the KSAs underlying performance.	Crandall, Klein, & Hoffman, 2006; Schraagen, Chipman, & Shalin, 2000; Rosen, Salas, Lazarra, & Lyons, in press
2. *Develop measurable learning outcomes.*	• Specify learning outcomes at all appropriate levels (e.g., individual, team) • Set learning outcomes in terms of the KSAs necessary for proficient task performance • Make sure that the learning outcomes are in line with broader organizational goals	Ostroff & Ford, 1989; Goldstein & Ford, 2002; Cannon-Bowers & Salas, 1997
3. *For each learning outcome, derive a set of specific metrics.*	• Assess the psychometric properties (e.g., reliability, validity) of the metrics for each learning outcome • Make sure that each learning outcome is adequately represented by the chosen metrics. • Make sure that measures of different learning outcomes are not contaminated with measurement of other learning outcomes.	Goldstein & Ford, 2002; AERA,APA,NCME, 1999; Vreuls & Obremayer, 1985
4. *Develop behavioral markers of performance for each learning outcome.*	• Understand how the KSAs being acquired are manifested during contextualized performance in a given simulation scenario. • Develop behavioral markers that are responses to critical events during a scenario.	Salas, Rosen, Burke, Nicholson, & Howse, 2007; Fowlkes, Dwyer, Oser, & Salas, 1998
5. *Develop metrics that are diagnostic of performance.*	• Organize metrics around specific standards of behaviors and skills • Collect a robust profile of performance measures • Maintain clear linkages between metrics and the targeted	Salas, Rosen, Burke, Nicholson, & Howse, 2007; Oser, Cannon-Bowers, Salas, &

	KSAs	Dwyer, 1999; Cannon-Bowers & Salas, 1997
6. *Use multiple data sources and types to capture performance.*	• Use trainee observations. • Use automated or semi-automated data collection of 'objective' measures of performance within simulations when available. • Use trainee self-reports.	Salas, Rosen, Burke, Nicholson, & Howse, 2007; Oser, Cannon-Bowers, Salas, & Dwyer, 1999
7. *Capture performance at multiple levels of analysis.*	• When appropriate, measure the performance of: the individual, the team, and multi-team systems.	Cannon-Bowers & Salas, 1997; Ostroff & Ford, 1989
8. *Capture the processes of performance in addition to outcomes.*	• Use automated data collection to capture 'on-line' measures of performance (e.g., discrete communication and behavior). • Identify behavioral markers associated with specific processes (e.g., communication patterns)	Campbell, 1999; Austin & Crespin, 2006; Salas & Cannon-Bowers, 2001
9. *Create a plan for integrating multiple sources and types of measurement.*	• Define the relationships between different performance measures. • Formalize these relationships and automate or semi-automate the combined analysis of multiple measures for use in debriefs.	Hawley, 1984; Oser, Cannon-Bowers, Salas, & Dwyer, 1999
10. *Automate as much of the performance measurement collection and analysis as possible.*	• Use PDAs or tablet PCs that automate as much data entry as possible for the trainers. • Extract meaningful performance data from data collected automatically by simulators. • Have tools for translating electronically entered and stored performance data into debrief aids.	Oser, Cannon-Bowers, Salas, & Dwyer, 1999
11. *Develop and implement training programs for observers/instructors.*	• Provide a tutorial on the performance dimensions being rated/observed. • Provide a reference guide for these dimensions. • Use video/audio recordings of old performance episodes as	Bakeman & Gottman, 1997

	content for guided practice opportunities. • Train observers/raters to a high level of inter-rater reliability on the practice materials before they rate/observe performance during training.	
12. *Provide structured tools/protocols for observations.*	• Guide the observers to critical aspects of performance using performance measurement tools structured around the scenario in order to facilitate higher reliability ratings. • Provide scaffolding (e.g., incorporate descriptions of targeted behaviors into the measurement tool) to increase the reliability of ratings.	Fowlkes, Lane, Salas, Franz, & Oser, 1994
13. *Use checklists for observations that link discrete behaviors to scripted events.*	• Organize performance measurement tools (e.g., checklists) so that they reflect the flow of the scenario. • Create a means for observers to	Fowlkes & Burke, 2005; Fowlkes, Lane, Salas, Franz, & Oser, 1994
14. *Focus performance measurement on discrete and observable behaviors.*	• Create behavioral checklists based upon the developed set of behavioral markers. • Craft the scenario to ensure that trainees have the opportunities to exhibit the behavioral markers	Fowlkes, Dwyer, Oser, & Salas, 1998;
15. *Create and maintain a systematic and organized representation of performance.*	• Develop a document or set of documents that explicitly links KSAs, learning objectives, metrics, and training content (e.g., lecture based material, scenarios, etc).	Towne, 2007; Goldstein & Ford, 2002
16. *Do not over-burden observers; maintain a good ratio of observers to trainees.*	• Maintain records of inter-rater reliability during simulation scenarios to determine how many observers are needed (i.e., find out how low the ratio of observers to trainees can go before seeing a decline in reliability). • During a simulation scenario, reduce the trainer's workload from all other sources (e.g., running the scenario and manipulating events) so their full attention can be devoted to observation.	Bakeman & Gottman, 1997
17. *Create realism and exploit the*	• Represent as much of the detail of the real world	Hays & Singer, 1989;

trainee's sense of presence.	environment as is feasible. • Replicate real world stressors such as time pressure and ambiguity as appropriate. • Develop a sense of presence in the trainee by creating a climate of seriousness through instructor attitudes.	Salas, Priest, Wilson, & Burke, 2006

Table 1. Summary of guidelines.

References

AERA, APA, & NCME. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed.). Cambridge, UK: Cambridge University Press.

Campbell, J. P. (1999). The definition and measurement of performance in the new age. In D. R. Ilgen & E. D. Pulakos (Eds.), *The changing nature of performance: Implications for staffing, motivation, and development* (pp. 399-429). San Francisco, CA: Jossey-Bass.

Cannon-Bowers, J. A., & Salas, E. (1997). A framework for developing team performance measures in training. In M. T. Brannick, E. Salas & C. Prince (Eds.), *Team Performance and Measurement: Theory, Methods, and Applications* (pp. 45-62). Mahwah, NJ: Erlbaum.

Crandall, B., Klein, G., & Hoffman, R. R. (2006). *Working minds: A practitioner's guide to cognitive task analysis*. Cambridge, MA: MIT Press.

Fowlkes, J. E., & Burke, C. S. (2005). Event-based approach to training (EBAT). In N. Stanton, A. Hedge, K. Brookhuis, E. Salas & H. Hendrick (Eds.), *Handbook of human factors and ergonomics methods* (pp. 47-41 - 47-45). Boca Raton, FL: CRC Press.

Fowlkes, J. E., Dwyer, D. J., Oser, R. L., & Salas, E. (1998). Event-based approach to training (EBAT). *The international journal of aviation psychology, 8*(3), 209-221.

Fowlkes, J. E., Lane, N. E., Salas, E., Franz, T., & Oser, R. (1994). Improving the measurement of team performance: The TARGETs methodology. *Military Psychology, 6*, 47-61.

Goldstein, I. L., & Ford, J. K. (2002). *Training in Organizations* (4th ed.). Belmont, CA: Wadsworth.

Hawley, J. K. (1984). *Some considerations in the design and implementation of a training device performance assessment capability.* Paper presented at the 28th Annual Meeting of the Human Factors and Ergonomics Society.

Kozlowski, S. W. J., Gully, S. M., Nason, E. R., & Smith, E. M. (1999). Developing adaptive teams: A theory of compilation and performance across levels and time. In D. R. Ilgen & E. D. Pulakos (Eds.), The Changing Nature of Work and Performance: Implications for Staffing, Personnel Actions, and Development. San Francisco, CA: Jossey-Bass.

Oser, R. L., Cannon-Bowers, J. A., Salas, E., & Dwyer, D. J. (1999). Enhancing human performance in technology-rich environments: Guidelines for scenario-based training. In E. Salas (Ed.), *Human/technology interaction in complex systems* (Vol. 9, pp. 175-202). Stamford, CT: JAI Press.

Ostroff, C., & Ford, J. K. (1989). Assessing training needs: Critical levels of analysis. In I. L. Goldstein (Ed.), *Training and development in organizations*. San Francisco: Jossey-Bass.

Rosen, M. A., Salas, E., Lazzara, E. H., & Lyons, R. (in press). Cognitive Task Analysis: Methods for Capturing and Leveraging Expertise in the Workplace. In W. Bennett, Jr., G. M. Alliger, W. J. Strickland & J. L. Mitchell (Eds.), *Job Analysis: Studying the World of Work in the 21st Century*.

Salas, E., & Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. *Annual Review of Psychology, 52*(471-499).

Salas, E., Priest, H. A., Wilson, K. A., & Burke, C. S. (2006). Scenario-based training: Improving military mission performance and adaptability. In A. B. Adler, C. A. Castro & T. W. Britt (Eds.), *Military life: The psychology of serving in peace and combat* (Vol. 2: Operational Stress, pp. 32-53). Westport, CT: Praeger Security International.

Salas, E., Rosen, M. A., Burke, C. S., Nicholson, D., & Howse, W. R. (2007). Markers for enhancing team cognition in complex environments: The power of team performance diagnosis. *Aviation, Space, and Environmental Medicine Special Supplement on Operational Applications of Cognitive Performance Enhancement Technologies, 78*(5), B77-85.

Salas, E., & Rosen, M.A. (2006). *Performance measurement in simulation-based training: A review and best practices* (No. Navy Personnel Research, Studies, & Technology (NPRST/PERS-13)). Millington, TN: Bureau of Naval Personnel.

Schraagen, J. M., Chipman, S. F., & Shalin, V. L. (Eds.). (2000). *Cognitive task analysis*. Mahwah, NJ: Erlbaum.

Shadish, W. R., Cooke, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Towne, D. M. (2007). *Enhancing human performance via simulation-based training and aiding: A guide to design and development*. Rotterdam, The Netherlands: Sense Publishers.

Vreuls, D., & Obermayer, R. W. (1985). Human-system performance measurement in training simulators. *Human Factors, 27*, 241-250.

Zachary, W., Cannon-Bowers, J. A., Bilazarian, P., Krecker, D., Lardieri, P., & Burns, J. (1999). The advanced embedded training system (AETS): An intelligent embedded tutoring system for tactical team training. *International Journal of Artificial Intelligence in Education, 10*, 257-277.