MEASURING THE CONTRIBUTION OF DISTRIBUTED SIMULATION TO UNIT TRAINING

By

Dr. Michael D. Proctor
University of Central Florida

and

Michael J. Lipinski
Major, US Army

Abstract

Simulation is being expanded well beyond procedure training for individuals and crews to address even greater aspects of training of organizations through the use of distributed simulations. Measuring the contribution of these distributed simulations to organizational training has been difficult. Task performance and teamwork are two general measures that provide insight into the contribution that training in distributed simulation makes toward improving organizational performance. This paper reports findings of a field study of US Army units currently training within the Close Combat Tactical Training distributed simulation system. The research examines unit training within this virtual synthetic environment in terms of improvements in teamwork of the formal unit leaders and, secondly, performance of the unit on common tasks. As a portion of this research, a Navy measure for teamwork was refined for application to this study. Data collected during the field study provides insight into the contribution that distributed simulation may have toward these objectives. The results indicate that statistically significant improvements did occur in specific teamwork behaviors and teamwork dimensions by the unit command and control team. Additionally, statistically significant improvements in unit task performance occurred on specified tasks in the virtual synthetic environment. The implications of the study results and methodology of assessment are discussed in terms of evaluation of distributed virtual simulations for unit command and control teamwork training as well as unit task performance training.

A number of distributed simulation programs are underway in all the services that address collective training needs. Austere budgets for training as well as safety and environmental concerns help motivate interests for these programs.

One common objective of these distributed simulation systems is maintaining or raising unit performance. Given the implied trade-off between field training and simulation training, pressure will be on the training, acquisition and testing communities to insure the acquired distributed simulation systems achieve unit training performance objectives. Demonstrating success requires the use of measures relevant to training performance objectives.

Unit performance could be compartmentalized and then measured within system components. But arrogating compartmentalized measures often is not representative of the unit performance. Directly measuring unit performance requires measures relevant to unit tasks and missions.

Attempting to quantify the contribution of distributed simulation system is not a trivial problem since the system is actually a collection of systems and spans the diversity of activities contained in a military unit. Additionally from a training perspective, the unit would benefit if the measures used to quantify performance could provide insight into what processes need improvement.

As an initial effort, this research examines two measures that attempt to quantify the contribution that a distributed simulation system has on unit training.

Work on the Army's Close Combat Tactical Trainer provides the basis of the research. Potential technical performance measures are identified that relate to distributed simulation systems. Finally, analysis is presented of the identified technical performance measures during an actual distributed simulation system exercise.

Distributed Simulation Systems and the Close Combat Tactical Trainer

Quality teamwork demonstrates its tremendous value in athletic competition everyday. In the past the U.S. Armed Forces have had significant opportunity to develop expertise in unit teamwork and mission task performance during field training and field operations using actual equipment and formations. Opportunity for that environment has been reduced significantly due to a number of factors.

All the services and many joint organizations are acquiring distributed simulation systems to help fill the gap created by reduced field training. As an example, the "Distributed Mission Trainer" (DMT) is a "priority" for the Air Combat Command (Hawley, 1998). When fielded in 1999, DMT will add to simulator systems at Eglin, Langley, Shaw and Tinker AFB in an integrated and distributed manner. The integration will provide a complete spectrum of aircraft and facilities needed for unit training and mission rehearsal.

At the same time, the Defense Modeling and Simulation Office is developing a High Level Architecture that will enable multiple simulation federations to exist within and between all the services, joint commands and others. Further, many future distributed simulation systems are planned (Hammond and Edwards, 1998).

Currently being fielded, the Army's Close Combat Tactical Trainer (CCTT) is a more fully developed example of a distributed simulation system acquisition used for training. Like field training exercises, the CCTT "will train Armor, Cavalry, and Mechanized Infantry platoons through Battalion/Task Force on their doctrinal Mission Training Plan collective tasks" (Hammond and Edwards, 1998). Mission training plans identify general and specific tasks with conditions and standards for measuring unit performance against these missions.

Units tend to build ever higher levels of tasks and mission competence through exposure to ever greater challenges in training (CCTT, 1998). In the dynamics of human and unit growth, the learning environment evolves. From learning basic unit tasks, moving on to learn advanced unit tasks, reinforcement of previously learned tasks, and, finally, integration of various combinations of tasks (typically a mission or set of missions), individuals learn through some combination of instruction, discussion and exercises.

Just like in a field or live simulation exercise, the distributed simulation exercise integrates tasks in the form of unit mission scenarios. The training goal is to learn and perfect unit integrated processes like unit tactics, techniques and procedures that are transferable to many different missions. In training the focus is typically not exclusively on the resulting outcome for a particular mission, but rather on the processes that makes it possible for mission success.

Also like field training exercises, senior evaluators and/or unit leaders discuss unit task and mission accomplishments and failures with unit members after the simulation training. This forum is referred to as an "after-action-review." After-action-reviews also provide instruction on process improvements that are aimed at improving overall unit performance.

A simple analogy for this "after-action-review" session might be the discussion that a high school basketball team coach has with his team immediately after a team scrimmage. The emphasis in practice is on processes like individual dribbling and shooting and team plays like setting up a clear three point shot or a fast break. The coach does not focus so much on the score (outcome measure) run up against the scrimmage squad, but rather uses those failures and successes as points to correct specific task errors or reinforcement successes. In combination with personal and other team tasks and plays, their ability to perform these tasks effects their ability to put points on the scoreboard in the real game.

Likewise to avoid poor unit performance in combat, military units focus the after-action-review on training unit tasks. Similar to the basketball scrimmage example, the emphasis is not on the outcome of the scrimmage, but on the processes that can achieve missions during the real conflict.

Unlike field training exercises, "the CCTT-system … consists of networked vehicle simulator manned-modules … in combination with Semi-Automated Forces, Combat Support workstations, computer networks and protocols, and After-Action Review systems" (Hammond and Edwards, 1998). Actual military systems like tanks are not used in the CCTT distributed simulation system. The CCTT may be considered a synthetic battlefield that enables virtual and other synthetic players to interact in mock battles.

Unlike most field training exercises, with the exception of some live simulation sites, sophisticated "after-action-review systems" permit replay of portions of the unit actions that occurred during the exercise. These after-action review systems enhance unit discussion and further enable unit performance improvements.

The CCTT review system allows all events such as utterances between team members, radio communications, vehicle positions, unit actions, gun sight views, etc. to be recreated to exact detail. This allows an observer, evaluator, or trainee to review the simulation and capture the details of each event and task performance. The playback of these tapes provides insight and allows for discussion between the members of the training unit and observers, as to why or how an event unfolded along with its impact (Figure 1).

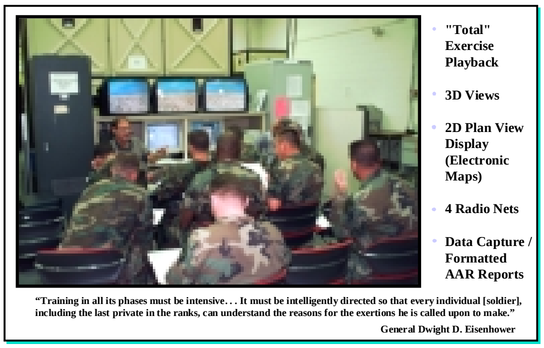

Figure 1. Features of the CCTT AAR Station. Adapted for US Army STRICOM, PM CATT (1998).

As these distributed simulation systems emerge and move toward fielding, the need for metrics to help communicate meaning between trainers and trainees and, secondly, to perform appropriate training evaluation of these unique systems becomes apparent.

Teamwork & Task Performance

Recent research suggests virtual simulations may be effective for training and assessing team performance (Brannick, Prince, Prince, and Salas, 1995). Assessment can be approached in many ways.

The Systems Acquisition Manager's Guide for the Use of Models and Simulations published by the Defense Systems Management College and written by Piplani, Mercer and Roop (1994) identifies numerous technical performance measures for use by acquisition managers of combat systems. In contrast, technical performance measures for unit collective training systems such as the distributed simulation systems are scant if at all present in the acquisition literature. Instead technical performance measures for evaluation of individual training systems have traditionally been submerged within the related combat system acquisition. Typically training systems were justified as trainers for specific aircraft, weapon system, etc. Consequently technical performance measures for individual and crew-training systems have been system specific and system performance oriented.

For unit performance measurement, the most applicable combat system technical measure identified in the Piplani, et al. is the outcome measure Loss Exchange Ratio. Loss Exchange Ratio can be used to judge individual or crew performance improvements. Loss Exchange Ratio (LER) is an outcome measure that compares enemy losses to friendly losses. Using an air warfare analogy, one loss exchange ratio might compare the number of enemy aircraft shot down for every friendly aircraft shot down. The more enemy aircraft shot down for every friendly aircraft shot down, the better your performance. A difficulty in this approach is that it is limited in scope to the individual system and, in a peacetime environment without actual adversaries, LER becomes suspect.

Johnston et al. (1997), Smith-Jentsch et al. (in press a) and Brannick et al. (1995) indicate that “free play” training exercises may produce inconsistent measures such as with LER when measuring unit performance change from training period to training period. Whereas the alternative to "free play" - - structured, calibrated and validated exercises - - were expensive to build, determine the difficulty of, and maintain.

As possible supplemental measures, Glickman et al. (1987), McIntyre and Salas (1995) and others surfaced the influence of teamwork - - a collection of critical behaviors and interpersonal skills - - on unit or collective task performance. These two technical performance measures - - teamwork and unit task performance - - are not widely discussed in the system acquisition literature. As measures they represent analysis of the process as opposed to the outcome.

Johnston et al. (1997) refined these teamwork dimensions and in a second review, Smith-Jentsch et al., (in press -a & b) refined the four teamwork dimensions - - discussed below - - into more reliable and independent dimensions containing sets of specific interpersonal behaviors. Qualitative assessment for each dimension and behavior can be done using behaviorally anchored rating scales. (Johnston et al., 1997).

Use of both these process measures to supplement outcome measures implies a more complete assessment of unit performance than LER alone

(Brannick et al., 1995, Johnston et al, 1997, Smith-Jentsch et al., in press -a, and Stout et al., 1997).

Measuring Teamwork in a Field Study

This research adapted process measurement methodology employed by both Johnston et al., (1997) and Smith-Jentsch et al. (1996a) to examine unit training in distributed simulation. Teamwork was defined as containing four dimensions - - Communication, Information Exchange, Team Initiative/Leadership, and Supporting Behaviors - - with each dimension having associated behaviors.

Dimensions	*Associated Behaviors*
Information Exchange Effective behaviors 1-4	(1) Seeks information from available sources (2) Passes information to the appropriate persons (3) Provides accurate "big picture" situation update
	(4) Accurately informs higher commander
Ineffective behaviors 5-7	(5) Has to be asked for information (6) Provides inaccurate situation update (7) Inaccurately informs higher commander
Communication Effective behaviors 1-3 Ineffective behaviors 4-7	(1) Uses proper phraseology (2) Provides complete reports (3) Adequate Brevity/ Avoids excess chatter (4) Uses improper phraseology (5) Provides incomplete reports (6) Uses excessive chatter (7) Communications are inaudible/garbled
Team Initiative/ Leadership Effective behaviors 1-3 Ineffective behaviors 4-6	(1) Provides effective guidance or suggestions to team members (2) States clear team and individual priorities (3) Appropriately refocuses team in accordance with situation (4) Provides ineffective or unclear guidance or suggestions to team members (5) States ineffective or unclear team and individual priorities (6) Inappropriately refocuses team in accordance with situation
Supporting Behavior Effective behaviors 1-4 Ineffective behaviors 5-6	(1) Corrects team errors (2) Requests backup or assistance when needed (3) Provides backup or assistance when needed (4) Provides constructive feedback (5) Fails to correct team errors (6) Provides or uses nonconstructive feedback

Figure 2. Teamwork Dimensions and Associated Team Behaviors.

Figure 2 lists the four dimensions and the associated twenty-six team behaviors, both effective and ineffective, that were identified as applicable to this research and ground combat. A review of the twenty-six team ineffective and effective associated behaviors enabled grouping of corresponding associated behaviors into fifteen general behavior categories.

Observations on all associated and general behaviors for each teamwork dimension were recorded on a worksheet during the field study (Team Initiative/Leadership Dimension shown as an example at Figure 3). Frequency counts of effective and ineffective associated behaviors for a given teamwork dimension were recorded on the right side of the form. Quality ratings of general behaviors were recorded on the left side of form. Separate sets of Team Observation Worksheet were used for each task event (discussed below). SME rating using a 1 to 5 Likert scale indicated the impact or severity of specific general behaviors.

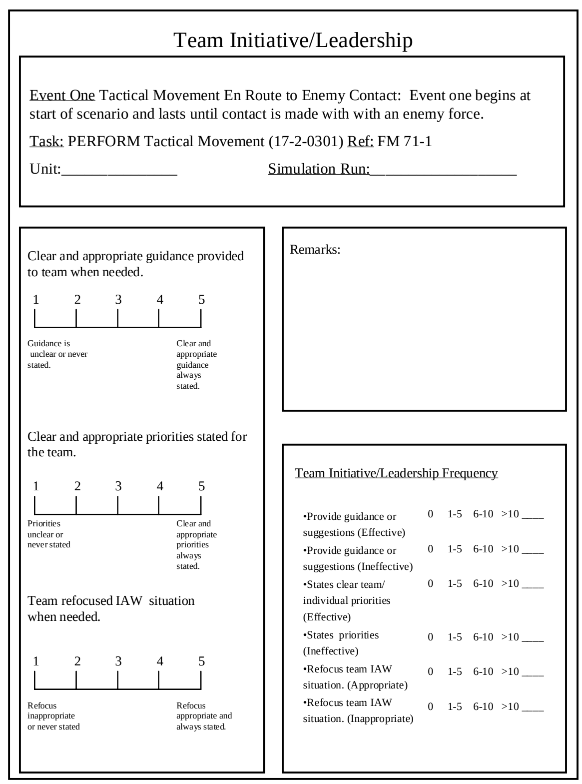

Figure 3. Sample Team Observation Worksheet.

While frequency counts were of interest, quality ratings for the dimensions and general behaviors were used in the final analysis since they can provide additional insight not available by the sole use of frequency counts. While a frequency count documents the occurrence of behavior, it does not consistently reflect the opportunity to correct errors across teams or task events. A quality rating of general behaviors may provide better insight, such as to the accuracy of a situation update or to the appropriateness of stated priorities. The percentage of effective behaviors to ineffective behaviors provide insight into the proportions of respective team behaviors (Johnston et al., 1997).

In order to avoid inconsistency of assessment between multiple observers, one evaluator was trained and validated at 100% proficiency in identification and classification of teamwork dimensions and respective behaviors by using the *"Team Dimensional Training"* computer-based-instructional software program (Smith-Jentsch, in press -b). The same observer assessed team behavior quality ratings for all teams.

Scenario and Type of Unit Training in the Field Study

Stout, Cannon-Bowers, and Salas, (1997) suggest that simulations with scenarios that require both individual cue assessments and relevant and efficient team process skills should be effective for team performance training. Further, Johnson and Noble (1994) indicate that distributed interactive simulation has the potential to effectively train the following primary tasks: Command, Control and Communications; Maneuver and Navigation; Teamwork, and Leadership.

For this research, the research team wished to investigate measures for two of these tasks: teamwork and C^3 task performance. The research team observed these measures during the normal training of two, active-duty US Army battalion task forces within the Close Combat Tactical Trainer (CCTT) facility at Fort Hood, Texas.

For this research, each battalion task force reported to the CCTT facility to conduct training. The battalion task force received familiarization training on the CCTT and then practiced operating and maneuvering manned module vehicles and units within the CCTT.

Training scenarios are available for use within the CCTT. The scenario chosen for the study was a Movement to Contact mission. The scenario had been developed by armor and mechanized infantry subject matter along with instructional system design experts at Fort Knox, KY. The standardized training scenario is available for use by any unit training in CCTT. The scenario provides maximum opportunity to exercise each collective unit task/mission in a realistic "free play" simulation scenario. The intent of the scenario is to provide task cues and conditions that a unit would encounter in an MTC situation. Increasing the performance capabilities and tactical proficiency of the semi-automated-force opposing entities can modify the intensity (difficulty) of the scenario.

In the recorded training exercise, the task force received a Movement to Contact mission and entered its tactical operations planning process. A tactical plan was devised, rehearsed, and then executed in the CCTT.

Next each battalion task force conducted an after-action-review feedback session on unit tactical performance at both the company level and the battalion level. After that, the units repeated the Movement to Contact mission. Upon completion of the second trial, another feedback session was conducted to assess tactical performance.

Participants in the study served in duty positions that included company commander, executive officer, and platoon leaders. For the purpose of this study, this team of leaders was referred to as the tactical command and control team, since these individuals provide the leadership to command and control their units while executing their mission.

For control purposes the same Movement to Contact scenario was used between the first and second simulation run. The selected tasks to be performed were identical between simulation runs. The scenario in each simulation run presented the same mission, enemy force, terrain, time frame, environmental conditions and semi-automated entities coded behavior. The opposing force consisted entirely of semi-automated-forces under the control of an experienced operator.

The semi-automated-force operators used their "free play" prerogative in the second run. Specifically, units typically train against a lesser able opposing force in their initial training. In accordance with the learning objectives of the commanding officer, the semi-automated-force operators typically increase the degree of difficulty by increasing the quality of semi-automated-force tactical operations in subsequent runs (CCTT, 1998).

As indicated above, this common training approach with increasing difficulty in subsequent training exercises was suspected to influence LER relationships. Since the research was aimed at supplementing LER as a technical measure, the research team collected LER data for information purposes.

Measuring Task Performance in the Field Study

Instead of emphasizing outcome measures, an event-based approach was used to evaluate teamwork dimensions and unit processes or sub-task performance of each company tactical command and control team during each movement to contact mission. Each event contains a unique tactical situation that requires team members to coordinate and exchange information in order to assess the situation, make the appropriate decisions, and execute the correct actions. Three specific events were selected which were likely to require the execution of team behaviors. From an analytical perspective, the mission was broken up into three events. The events selected were (1) Perform Tactical Movement (17-2-0301), (2) Perform Actions on Contact (17-2-0304), and (3) Perform an Attack by Fire (71-2-0311) (Department of the Army, ARTEP 71-1-MTP, 1988).

Each event has subtasks or unit processes associated with them. In addition to teamwork measures discussed earlier, success or failure of each process was recorded.

The US Army Mission Training Plan (ARTEP 71-1-MTP) for Tank and Mechanized Infantry Company and Company Team was used as a basis to assess accomplishment of team sub-task or processes. Team sub-task accomplishment ratings were assessed based in accordance with the appropriate Mission Training Plan Task Evaluation Outline. Determining the percentage of critical and total subtasks or unit processes successfully performed within the three events achieved an overall process assessment. The Mission Training Plan provides a systematic methodology for considering units performing collective tasks, but the T, P or U arrogation method is not oriented to focus on process assessment. Each collective Mission Training Plan task is broken down into subtasks. Subtasks are categorized as critical, non-critical, and leader sub-tasks or processes. The percentage of critical sub-tasks (% of Critical Tasks) and total subtasks (% of Subtasks) successfully performed was selected to add a level of discrimination to the Mission Training Plan system.

The rater for team sub-task evaluation had extensive experience (over 20 years) in evaluating military unit task performance at platoon, company, and battalion levels utilizing US Army Mission Training Plan standards. A single rater scored all of the after-action-review tapes and his ratings were used in all data analysis. A subset of the tapes was compared to the results of the unit assessment to determine inter-rater reliability. Agreement on assessment was found in all cases.

Results & Analysis

After training was completed, the CCTT after action review tapes were analyzed to observe, categorize, and record observations of teamwork and sub-task performance for the eight tactical command and control teams. Additionally, team/unit task performance was evaluated for each of the eight company teams. Eight teams executed two MTC missions each, Run #1 and Run #2, for a total of 16 MTC missions. Each after-action-review tape was replayed in real-time to observe each tactical command and control team performance in each scenario. All radio communications and team/unit actions were observed and monitored separately for each tactical command and control team.

Task performance was assessed for each sub-task within each event based on US Army Mission Training Plan for Tank and Mechanized Infantry Company and Company Team mentioned earlier.

Teamwork dimensions and team behaviors were analyzed for indication of improvement.

In addition, the traditional loss exchange ratio measure was evaluated for those missions that had a LER associated with them.

As a means of analyzing Command, Control and Communications (C^3) task performance, a series of matched pairs, one-tailed t tests compared the equality in Critical Task and Subtask Success between simulation Run #1 and Run #2. Matched pairs, one-tailed t tests compared Loss Exchange Ratios differences between runs but due to the nature of the selected tasks, not all tasks involved LER. For all statistical tests a significant difference was declared if the probability of random occurrence was less than or equal to 0.05.

C^3 Task Performance & LER Results & Analysis: C^3 Task Performance indicated statistically significant improvement in Mission Training Plan Critical Sub-Task (ρ = .044) and Total Subtask Success (ρ = .007) as shown in terms of percentage in Figure 4. LER indicates a completely different result.

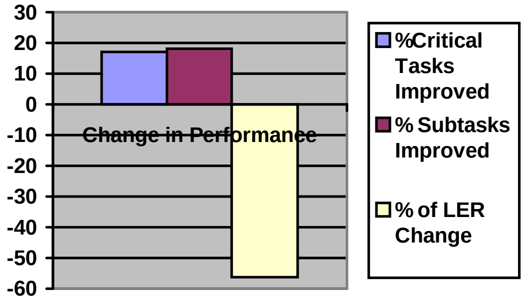

Figure 4: Change in Task & LER Performance

Figure 4 shows a marked decrease in the mean difference for loss exchange ratio (actual loss of 56.25% in LER). That indicates that the unit lost more friendly entities for each enemy destroyed. For LER, from a sample of eight tasks that did involve LER, five teams had an increase in LER, two teams had a decrease in LER, and one team had no change in LER. Statistically LER did not indicate any significant change due to the variability in the standard deviation (2.276) ($\rho = .57$).

Variations in task difficulty in Run #2 due to the free play in the simulation altered some tactical issues (holding of key (advantageous) terrain, enemy initiating contact from a hasty defense/attack-by-fire positions, force ratio of attacking to defending units, etc.). As suspected, these variation in task difficulty between Run #1 and Run #2 may have influenced or confounded LER outcome.

Teamwork Quality Ratings: In order to determine if training in virtual simulation resulted in an improvement in teamwork, teamwork quality ratings were assessed for Run #1 and Run #2. Quality ratings between Run #1 and Run #2 were found to have improved to a statistically significant degree for all teamwork dimensions. Additionally, improvements in quality ratings for the first thirteen of the fifteen general behaviors were found to be significant (Figure 5). From left to right those general behaviors were: (1) Communications Dimension: Improper Phraseology, Brevity/Chatter, Inaudible Communications, Incomplete Reports; (2) Information Exchange Dimension: Seeks Information, Passes Information, Situation Updates, Informs Commander; (3) Team Leadership/initiative Dimension: States Clear Guidance, States Clear Priorities, Refocus Team IAW Situation; (4) Supporting Behavior Dimension: Corrected Errors, Request / Provides Backup or Assistance, Constructive Feedback, Non-constructive Feedback.

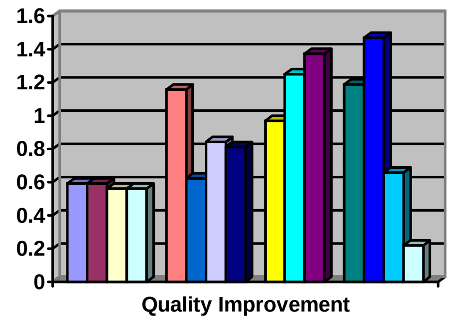

Figure 5: Mean Difference Change for 15 General Behavior Quality Ratings on 1 to 5 Likert scale

Only the last two general behaviors, Constructive Feedback and Non-constructive Feedback did not show a statistically significant improvement.

The percentages of change for the overall quality dimensions are shown in Figure 6.

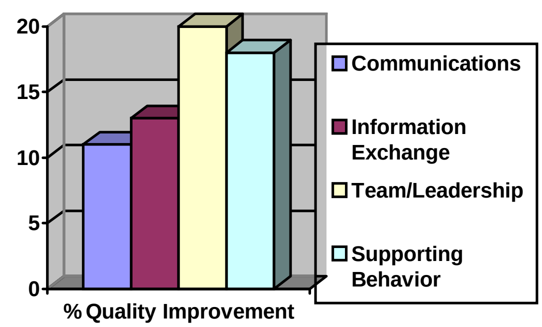

Figure 6: Change in the Quality of the Dimensions for Team Behavior

Conclusions

Whether in athletic competition or in combat, quality teamwork demonstrates its tremendous value. An example process or task worked on by a highly skilled team might be the "no look pass" between Michael Jordan and Scottie Pippen. Their teamwork created many national championships.

In the past the U.S. Armed Forces have had significant opportunity to develop expertise in unit teamwork and mission task performance during field training and operations using actual equipment and formations. That environment has changed significantly due to a number of factors. Despite the change, we don't want to become the Chicago Bulls of 1999.

To compensate our armed forces appear ready to acquire less costly distributed simulation systems in order to help fill the gap created by reduced field training.

These findings indicate that training in distributed simulations systems can result in statistically significant improvements in teamwork and unit task performance. Our findings on the outcome measure LER reflect earlier findings that LER is not a meaningful measure of performance when taken in a training context. Since LER actually decreased from Run #1 to Run #2, an outside observer might think that the simulation actually decreased the performance of the unit and the command and control team.

Specifically, our research indicates distributed simulation systems can help fill at least two gaps - - teamwork and C^3 task performance. Statistically significant improvements in the quality of teamwork were shown while conducting training in a distributed simulation system. Also, C^3 task performance was found to significantly improve between training sessions as evidenced by improved unit processes or Mission Training Plan Critical Subtask and total Subtask Successes.

Our study also indicates traditional outcome measures such as Loss Exchange Ratios do not appear to be appropriate as sole technical measure when evaluating the suitability of simulation systems used for training when those measurements are taken in a training context. No overall statistically significant change in outcome performance between simulation runs as measured by LER was observed though numerically a large negative change occurred. As often the case in training, LER may not be a credible indicator of improved proficiency of the unit as the difficulty of the opposing force fight might increase for training purposes from run one to run two. While duplication of the same scenario and difficulty level is possible in training, this typically only occurs for units that fail that level and need to be retrained.

The findings provide initial indication that both teamwork and task performance are viable and should supplement the training managers' set of technical performance measures for distributed simulation systems. These measures provide the training system manager a more complete assessment of the ability of a prospective distributed simulation system to aid training than loss exchange ratio would by itself. Further, these measures are intuitive and simple, helping to satisfy the challenge of communications.

We identify measures, application approaches and measurement instruments that may prove useful for training system acquisition. As a technical measure it appears appropriate not only for U.S. Army and Navy acquisitions but also for the Air Force in light of their DMT effort. These findings infer that these technical measures and methodologies may be additional tools not only appropriate for the trainer but also appropriate for the astute acquisition manager.

Further research is required to determine if these findings can be confirmed with larger sample sizes, perhaps over time and across other team training distributed simulation systems. Further research may address the application of these findings to other training audiences within distributed simulation such as air wings, ship command and control, higher staffs, and other types of organizations.

Epilogue

Williams and Keaton (1998) evaluate the CCTT 1997-1998 Initial Operational Test and 1997 Limited User Test conducted by, the Test and Experimentation Command. As part of their broad and in depth evaluation, they present findings on two topics of interest to this research. Specifically, Williams and Keaton discuss IOT simulator training in the CCTT and related field exercises at the National Training Center.

As way of background, the testing was designed to support evaluation of unit performance over both simulator training and field exercises. In order to conduct this performance evaluation across both, a panel of military experts selected 27 critical company team and 25 Armor and 27 mechanized infantry platoon level ARTEP MTP subtasks (what we term an event in this paper) that would be conducted in both the simulator training and the field exercise. (Keaton, 1999)

During the actual testing, evaluations were made "of ARTEP MTP subtask performance both during training exercises conducted in the CCTT and during missions at NTC. The evaluators were TRADOC Subject matter experts in the CCTT and both the same Subject matter experts and NTC observer controllers at NTC. The TRADOC Subject matter experts were calibrated on a 10- point scale of selected subtasks performed in offensive, defensive, and movement to contact scenarios. This calibration was done by having the Subject matter experts observe training events together, rate the training events and then discuss and come to agreement about how the unit's performance on each subtask should be rated. In this way, the Subject matter experts were able to come to agreement on the meaning of each point on the 10-point scale." (Keaton, 1999)

The entire CCTT test unit - - referred to as TF4 - - trained in the CCTT. They spent 2 days on platoon-level and 2 days on company team-level exercises. "The Subject matter experts observed the training and rated the platoon-level and company team-level subtask performance. The unit leadership, company commanders and platoon leaders, then spent 3 days conducting battalion level training, but this training was not evaluated in the CCTT." (Keaton, 1999)

As for IOT fixed-site, simulator training in the CCTT, Williams and Keaton report a “range of modest to insignificant gains observed during the CCTT training.” Specifically, “simulator training during the third through seventh weeks of the IOT indicates that few performance gains were achieved by the units undergoing training” based on TRADOC Subject matter experts ratings of the CCTT test unit at the company team level for ten exercises. Williams and Keaton indicate that the lack of gains “are a likely result of the lack of the development of a CCTT training strategy available to the units and the lack of advanced planning by the units.”

Despite this lack of performance gains in the CCTT, Williams and Keaton report, “At the aggregate level across all subtasks, the CCTT units performed significantly better at NTC” than other baseline task forces as shown in Figure 7 below. Aggregating company team performance within each observed task force, "Task Force 4 (TF4), the CCTT test unit, clearly outperformed the three baseline task forces. Examining performance at the individual subtask level, it is clear that the TF4 gave a stronger performance even at the outset, achieving “go” ratings on a preponderance of subtasks.”

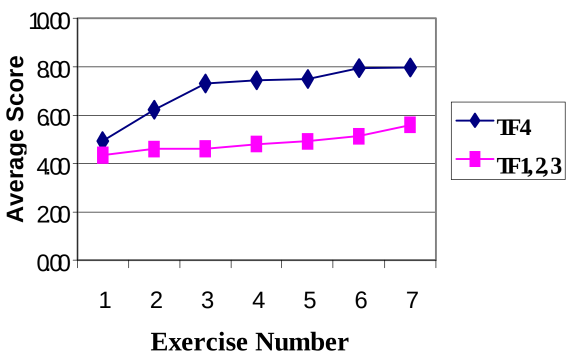

Figure 7. Pure CCTT Co Tms (TF4) vs. All Baseline CO Tms (TF 1,2,3)

The apparent conflict between “insignificant gains” while in the CCTT and “significantly better” NTC performance at the company team level makes one wonder why the disparity exists. In explanation of the above ratings, Williams and Keaton state, “TF4’s average rating at entry (average of ratings during first two missions) appears consistent with its score at exit from CCTT, suggesting consistency between the NTC OC and TRADOC SME ratings.” Recognizing the inconsistency in overall outcome due to the divergence in Task Force performance and the inability of the existing performance-oriented data to explain the difference, Williams and Keaton state, “The initial operation test in-simulator performance data was insufficient to demonstrate a linkage between CCTT training and performance attained in the field.”

The findings of Williams and Keaton indicate a move away from LER and begin to approach unit process measurement, but what they call a sub-task of a mission is what we term an event. As such, the evaluations are still more outcome oriented than our subtasks of events are. We suggest that one consider our newly developed teamwork process measures and revised unit process or sub-task performance measurement methodology. We hope that these methodologies can help explain what is not explained in the traditional outcome-oriented measurement methodologies. Specifically, we have reported above how traditional performance-oriented methods often do not capture training improvements for various reasons. Many of these same reasons were active during the CCTT operational and user tests. Our research suggests that significant teamwork, C^3 and unit sub-task processes might actually have occurred within the Task Force while training in the CCTT but were not captured by the more traditional outcome measures used in the operational and user evaluation. And if gains were made in the CCTT, this suggests that Task Force performance improvements achieved at the NTC were due in part to gains made in teamwork, C^3 and unit sub-task processes achieved as a result of that CCTT training.

Williams and Keaton state that "the TRADOC Subject matter experts noted throughout the initial operational test that units did not have clear training objectives, that subordinate leaders were unaware of what exercise was being trained, and that units commonly started training exercises without prior planning and without conducting mission briefings." The report recommends that TRADOC "institute a training program for unit leaders to show them how to use CCTT…and conduct troop leading procedures and After Action Reviews while using the CCTT." These recommendations are in line with our recommendation to focus more on process.

The above suggestions can not be proved or disproved in this particular case, however, we recommend further research into the implied hypotheses as well as our measurement methodology so that if found valid it could be adapted in the study of future simulation training systems.

Acknowledgments

We would like to thank the regular army participates as well as the STRICOM and OPTEC Commanders and organizations for their tremendous voluntary support for this academic research effort. In particular we would like to thank Dr. Laurie Keaton of OPTEC and Mr. Harry Sotomayor of STRICOM for their strength of character, tremendous professionalism, dedication to perfection, and open-mindedness to new research methods. Finally, we would like to thank University of Central Florida professors Drs. Linda Malone, Kay Stanney and Tim Kotnour for their voluntary contribution of time and professional advice.

References

Brannick, M., Prince, A., Prince, C. and Salas, E. (1995). The measurement of team process. *Human Factors*, 37, 641-651.

CCTT (1998), Close Combat Tactical Trainer, *Interservice/Industry Training, Simulation and Education Conference Demonstration and Presentation.*

Department of the Army (DA), (1993). *Mission Training Plan for the Tank and Mechanized Infantry Company and Company Team. ARTEP 71-1-MTP.* Washington DC: Author.

Glickman, A., Zimmer, S., Montero, R., Guerette, P., Campbell, W., Morgan, B., & Salas, E. (1987). *The evolution of teamwork skills: An empirical assessment with implications for training.* (ONR, Tech. Report 87-016). Orlando, FL: Naval Training Systems Center, Human Factors Division.

Hammond, A.R. and Edwards, M. (1998). *Close Combat Tactical Trainer.* http://www.stricom.army.mil/STRICOM/PM-CATT/cctt.html

Hawley, R.E. (1998). Keynote Address, *Interservice Industry Training, Simulation and Education Conference*.

Johnston, J., Smith-Jentsch, K. & Cannon-Bowers, J. (1997). Performance measurement tools for enhancing team decision-making training. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team Performance assessment and measurement: Theory, Methods, and Applications.* Hillsdale, NJ: LEA.

Keaton, M.L., (1999). *Personal Communication.* United States Army Operational Test And Evaluation Command, Alexandria, Va

Kuhn, Tom. Super-duper Simulators, *Airman*, http://www.af.mil/news/airman/1198/road.htm

McIntyre, R. and Salas, E. (1995). Team performance in complex environments: What we have learned so far. In R. Guzzo & E. Salas (Eds.) Team effectiveness and decision-making in organizations (pp. 9-45). San Francisco, CA: Jossey-Bass.

Piplani, L. K., J. G Mercer, and Roop, R. O. (1994). *Models and Simulations*, Fort Belvoir, Virginia: Defense Systems Management College Press.

Smith-Jentsch, K., Johnston, J., & Payne, S. (in press a). Measuring team-related expertise in complex environments. To appear in J. A. Cannon-Bowers & E. Salas (Eds.), *Decision Making Under Stress: Implications for Individual and Team Training.* Washington, DC: American Psychological Association.

Smith-Jentsch, K., Zeisig, R., Acton, B., & McPherson, J. (in press b). Team dimensional training. To appear in J. A. Cannon-Bowers & E. Salas (Eds.), *Decision Making Under Stress: Implications for Individual and Team Training.* Washington, DC: American Psychological Association.

Stout, R., Cannon-Bowers, J., & Salas, E. (1997). The role of shared mental models in developing team situational awareness: implications for training. *Training Research Journal*, 2, 85-116.

Williams, D. and Keaton, M.L., (1998). *System Evaluation Report (SER) Of The Close Combat Tactical Trainer (CCTT) Initial Operational Test And Evaluation.* United