The Value of Simulation in Army Training

Mark Riecken
Trideum Corporation
Orlando, Florida
mriecken@trideum.com

LTC Johnny Powers, Craig Janisz
PEO STRI
Orlando, Florida
johnny.powers@us.army.mil
craig.janisz@us.army.mil

S.K. Numrich, P.M. Picucci
Institute for Defense Analysis
Alexandria, Virginia
snumrich@ida.org
ppicucci@ida.org

Mike Kierzewski
RDECOM Edgewood Chem-Bio Center
Aberdeen Proving Ground, Maryland
michael.o.kierzewski.civ@mail.mil

ABSTRACT

Intuition tells us that simulation is a valuable tool for Army training; however, proving and quantifying this contribution remains difficult particularly for tasks with subjective performance standards. The Simulation to Mission Command Interoperability (SIMCI) organization, building on past efforts, has undertaken an exploration of this problem focusing on the role of constructive simulation in command and staff training. Establishing the value of simulation is made more difficult by the absence of clear training metrics for brigade staffs and the sensitivity of evaluation data. The approach is designed to incorporate hard and soft data for both cost and performance. A novel approach to cost uses software to capture the "would be" cost of entities in a simulation's scenario had they been live rather than simulated. The performance evaluations use methodologies described by the Government Accountability Office (GAO) and in the Army's Cost Benefit Analysis (CBA) guide. Performance data is captured using specially designed survey instruments. The paper touches on the methodology, data acquisition, data analysis and results. This study is intended to be a first step in developing guidelines and methodologies generalizable to assessing the value of simulation in other application areas.

ABOUT THE AUTHORS

Mark Riecken, PhD Dr. Riecken is a member of the SIMCI Overarching IPT (OIPT) and serves as an architect. He is Chief Engineer for the Trideum Corporation and resides in Orlando, Fla. He has served in numerous leadership roles in modeling and simulation (M&S). He holds a PhD in electrical engineering from the University of New Mexico, MS in physics from Vanderbilt University, and a BS in physics from the University of Evansville (Indiana). Riecken has worked in the information technology and engineering sectors for over 25 years, primarily in the areas of military systems modeling and simulation, test and evaluation, and analysis.

S. K. Numrich, PhD, CMSP Dr. Numrich is a research staff member at the Institute for Defense Analyses where she undertakes studies in areas related to needs of the operational military including training. Prior to coming to IDA, she worked as a research physicist at the Naval Research Laboratory where she served on numerous M&S committees and panels concerned with the value and capability of simulation for acquisition and training. Her last assignment as a Navy civilian was as Technology Director at the Defense Modeling and Simulation Office.

LTC Johnny Powers LTC Powers currently serves as the executive agent for the SIMCI Program in Project Manager Constructive Simulation (PM CONSIM), Program Executive Office, Simulation, Training, and Instrumentation (PEO STRI) and is the APM for Synthetic Environment (SE) Core. LTC Powers served as an officer in Operation Able Sentry (Macedonia) and the IFOR Mission (Bosnia). Powers served in various command positions and was assigned as an Active Component/Reserve Component advisor to the Arkansas Army National Guard. He was deployed for Operation Iraqi Freedom as an advisor. After returning to the United States, he became the simulation officer for 5th U.S. Army/Army North, Fort Sam Houston, Texas, where he was the knowledge management officer for operations in support of the Hurricane Katrina relief effort. For the Army Staff, G-37, he served as the lead for the Mission Command Training Support Program responsible for Management Decision Package management for Army Constructive and Virtual training programs. Before assuming his current role, he was the contracting officer representative for Warfighter Focus Efforts 22 and 42, supporting operations for Operation Enduring Freedom in Afghanistan. LTC Powers attended the University of Arkansas at Little Rock. He received his master's degree in modeling, virtual environments and simulations from the Naval Postgraduate School in Monterey, California.

Craig Janisz Mr. Janisz is the SIMCI Project Director for PEO STRI. Mr. Janisz is a 1976 graduate of Michigan State University with a BSEE and a computer science minor. He spent 19 years as an engineer in industry before beginning his civil service career in 1995. Craig received his MBA from Webster University in 2001. He was board-selected to DoD's Competitive Development Group in 2003 and is a graduate of the Army Management Staff College. He is Level III certified in Program Management and Systems Engineering.

Mike Kierzewski Mr. Mike Kierzewski is the Branch Chief for Modeling Simulation and Analysis at the US Army RDECOM, Edgewood Chemical and Biological Center and is a retired Army Chemical Officer. As branch chief, Mr. Kierzewski leads a diverse team of engineers, physicists, statisticians and mathematicians as they support customers with OR analyses and generalized data analysis. He has been involved with operations research and systems analysis (ORSA) studies of CBRN concepts and systems for the past 25 years, starting with his assignment as the senior operations analyst at the US Army Chemical School in 1988. His primary experience with the use of models and simulations has been for science and technology (S&T) and acquisition decisions regarding CBRN materiel with recent forays into the realm of simulations for training support. A notable recent product is the Individual Protection System Performance Model (IP SPM), a Joint Science and Technology Office (JSTO)-sponsored project. He has been a past Working Group and Composite Group Chair with the Military Operations Research Society Symposium (MORSS). He holds a master's degree in operations research and systems analysis from the US Naval Postgraduate School and a BS in chemical engineering from Virginia Tech.

Pete Picucci, PhD Dr. Picucci is a Research Staff Member at the Institute for Defense Analyses. He holds a national security studies master's degree from California State University, San Bernardino and a political science doctorate from the University of Kansas. His recent research efforts center on formal and informal socio-cultural modeling in support of DoD operations.

The Value of Simulation in Army Training

Mark Riecken
Trideum Corporation
Orlando, Florida
mriecken@trideum.com

LTC Johnny Powers, Craig Janisz
PEO STRI
Orlando, Florida
johnny.powers@us.army.mil
craig.janisz@us.army.mil

S.K. Numrich, P.M. Picucci
Institute for Defense Analysis
Alexandria, Virginia
snumrich@ida.org
ppicucci@ida.org

Mike Kierzewski
RDECOM Edgewood Chem-Bio Center
Aberdeen Proving Ground, Maryland
michael.o.kierzewski.civ@mail.mil

INTRODUCTION

What is the value of simulation? This patently open-ended question can have many answers. The topic has been addressed by multiple researchers and others within the simulation community with limited success. Given an increased reliance on simulation in many areas, it seems prudent to find a way to answer the question. Expressed in more specific terms, how can we measure the benefits gained through the use of simulation and thereby understand whether scarce dollars are returning sufficiently on their investment potential? If we can succeed at answering the question in that form, then surely we can also use that information to improve our investments and the quality of the simulation itself and how it is applied.

While the Army simulation community recognizes seven areas or communities of practice in which simulation plays a vital role, the SIMCI[1] OIPT[2] early on defined the scope of this effort to the Training community of practice. Consistent with the mission of SIMCI, the scope was ultimately defined as training in support of Mission Command, i.e., collective command and staff training. Furthermore, it was determined that the most suitable objective for this effort would be to develop an executable methodology and assess and validate that methodology with a use case.

Significance of This Work

Simulation continues to offer the promise of enormous value-added to military training. It has the potential to increase effectiveness through higher levels of performance mastered with fewer dollars spent. There is general recognition of this, but senior leaders in Congress and in the military, want more evidence before committing additional resources. Unfortunately, evidence is sparse and is clustered towards the training of the more easily measured individual skills; there is almost no substantiated evidence for the value of simulation in command and staff decision-making process in the military. It is a hard problem somewhat ironically because the success of simulation has made it an integral part of much of this training. If a means of reliably and consistently measuring this value can be developed, the impact will be better use of the available dollars as well as better training outcomes.

Case Study Problem Statement

For the purposes of the case study the following question serves as the problem statement: *"What is the Value of Simulation to the Army as it trains its Brigade Combat Team (BCT) commanders and staffs to execute the Warfighting function and the operations process throughout the Army Force Generation (ARFORGEN) cycle?"*[i]

The problem statement for the case study focuses on the Mission Command and Operations Process training for BCT Staffs as they proceed through the ARFORGEN Cycle, post reset through either availability or deployment. The focus is on the BCT as the primary modular force element for Army combat operations. Active and Reserve

[1] SIMCI: Simulation to Mission Command Interoperability
[2] OIPT: Overarching Integrated Product Team

Component Mission Command Staff training were considered. The Mission Command Training Program (MCTP) at four Active component MTCs[3] and two National Guard (NG) MTCs contributed to the case study.

Related Work

A literature review was conducted to locate material in the following categories: substantive articles, reports, studies, and briefings on simulation ROI,[4] simulation or training cost-based analysis (CBA), staff training effectiveness, value of simulation in training, simulation at CTC[5]s and MTCs, and mission command or battle command training or readiness. The literature review did not reveal any previous studies directly addressing either the value or effectiveness of simulation-based training for military staffs in terms of ROI or CBA. Doctrinal publications, guidebooks, and instructions were useful. The CBA process proved most useful. The literature review informed the Value of Simulation (VoS) team of what was available and what was not available.

METHODOLOGY

The overall approach to addressing the problem of characterizing the value of simulation includes both qualitative and quantitative mechanisms. Ultimately the complete methodology should be such that there is a means of summarizing the results through *quantitative* data that can be legitimately interpreted as such (i.e., a quantitative measure of value), but that will also possess *qualitative* aspects that must be understood for a complete and accurate interpretation. In a broad sense, the survey instruments developed in this effort serve as the qualitative measure, although it is important to recognize that the application of these instruments result in quantitative data, i.e., measures of knowledgeable and expert opinions. The ROI model, which is simply a systematic accounting of the costs (investments) and the monetizable benefits, serves as the principal quantitative (economic) component of the approach. The ROI model may have many of these "cost elements" comprising either an entry on the investment or the benefit side of the ledger. One such cost element is provided by examining the "avoided costs" of real vehicle platforms and weapons by using simulation. A tool called the LVC[6] Cost Counter[ii] can be attached to a federation of simulations to collect these data.

Analytical Overview and Conceptual Model

The conceptual model for the study is shown in Figure 1. This envisions a longitudinal approach in which BCTs are followed as cohorts originating from various posts and their supporting MTCs. The intent is to determine if differences in Mission Command training readiness as evaluated at the CTCs can be correlated with differing simulation aided training environment experiences by the BCTs. This can be expressed in the following hypothesis:

"*Commanders and their battle staff who have been exposed to constructive simulation and have used them well during home station training perform better at CTCs.*"

The most pure application of this approach would be to assess the simulation-aided training environment experienced by a BCT then track that unit's performance at a CTC through to deployment or residence in the contingency/ready pool of units. Because of the timelines involved, such a true longitudinal study tracking specific units was infeasible under the present effort. A modified approach was adopted in which like BCTs from common home stations would be assessed at home station then at a CTC or at the

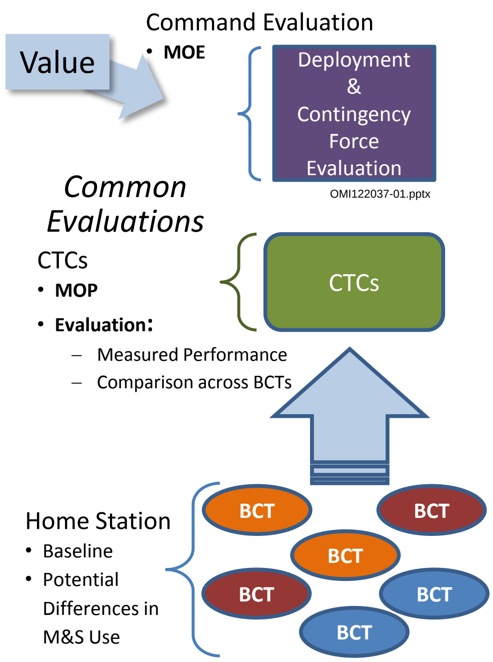

Figure 1. Experiment Conceptual Model

[3] MTC: Mission Training Complex
[4] ROI: Return On Investment
[5] CTC: Combat Training Center
[6] LVC: Live, Virtual and Constructive

contingency force /ready pool end of the ARFORGEN cycle. Therefore the *same* BCT at each portion of the ARFORGEN cycle is not assessed but rather *like* BCTs originating from the same home station. For example, the home station training environment for BCTs originating from Fort Bragg would be assessed and then the performance of other BCTs originating from Fort Bragg would be assessed during their CTC rotations specifically on the collective Mission Command tasks. Note: Due to time and resource constraints, initial coordination with CTCs has been effected but no data collected.[iii] In the absence of external data, self-assessments were used.

Survey Instruments

Evaluation of training has traditionally rested upon the articulation of a task, conditions under which it must be performed, and the standards of performance. Collective training presents a problem in defining standards of performance with appropriate metrics. While the military has learned to develop, train and deploy small teams with great success, determining standards of performance for other types of units has proven far more difficult. Military staffs, essential to the performance of any unit, are one such example where the definition of the conditions and standards of task performance prove challenging.

After-action reviews (AARs) conducted during and after deployments are potential sources of evaluation; however, in most cases, the reviews are an opportunity for the forces to communicate problems, seek remedies or describe changes in tactics, techniques and procedures (TTP). There is generally no repeatable set of questions about training or even topics upon which to focus data collection. In a set of over two hundred after-action reviews recently made available, fewer than 25% even mentioned training. Thus while these documents are valuable sources, their coverage of any single topic is likely to be sporadic and scattered in content.

The most reliable means of acquiring operational evaluations is through use of structured interviews that use a survey as the basis for the discussion. When time and resources do not permit the scheduling of a large number of interviews, on-line surveys, to be answered at the respondents' convenience, can be utilized. Both approaches were used in this study; however, the bulk of command and staff interviews were completed on-line.

The use of surveys can address the issue of not having evaluations from the training sites themselves; however, surveys must be based on the tasks and skill sets that have to be developed by the commander and staff during their pre-deployment training. In the absence of universally accepted tasks, conditions and standards, conversations with former commanders, members of the SIMCI community and trainers focused the research team on the planning process. One set of skills the staff must acquire is their ability to work smoothly together to perform the Army Mission Command and the Operations processes through which the commander and staff determine how to fulfill their operational tasks. This process provided a framework for evaluation and the core of two of the three surveys developed.

The classic and ideal evaluation approach is to measure performance of a skill or skill set before training and then again after completion of training to determine progress or achievement. If computer-based simulation is the chosen delivery mode for training, this approach would translate into comparing training with and without simulation. However the military has employed computer-based simulation for decades and there are few, if any, cases where simulation is not part of the training process. This led to an approach in which comparisons would be made across various home stations, some of which are known for their use of simulation and others not. This method, referred to as the quasi-experimental approach, is used when it is not possible to gather before-and-after or with-and-without evaluations, and rests on comparing similar groups with somewhat different experiences. Designated as Method 1, this approach involves the use to two surveys, Survey 1, dealing with the simulation infrastructure, and Survey 2, involving the self-evaluation of the commander and staff.

A second approach, Method 2, addressing the value of simulation involves a cost-benefits analysis following the general lines set forth in the U.S. Army Cost Benefit Analysis Guide (Office of the Deputy Assistant Secretary of the Army, 2011). A cost-benefits analysis requires a comparative evaluation of several different alternatives against specific organizational values where at least part of the evaluation is based on cost. The alternatives selected for this analysis were three different modes by which training is delivered: instruction, field exercises and simulation. As with Method 1, data was gathered using an on-line survey based on the same skill sets as used for Survey 2, the command and staff survey used in Method 1.

Figure 2 depicts the methods, the surveys and the overlaps in the surveys, in addition to the nominal target audiences for each survey. The term 'command staff' will be used consistently to refer to the commander and his/her staff to accommodate the lack of space for labels in figures, tables and graphs.

Survey 1, the Infrastructure survey, was directed at personnel running active component and National Guard MTCs with the intention of being able to categorize/inventory training center usage of simulation, and relative levels of infrastructure support for simulation usage (including software and frameworks, physical facilities, support personnel, as well as specific hindrances to simulation usage).

Survey 2, the Command and Staff survey, targeted brigade and battalion commanders and their primary staffs who had recently completed a training cycle and were either deploying, deployed, or recently returned from deployment. The intent of the survey was to capture self-assessments of 11 skills needed to support the Army Mission Command and Operations processes, before and after the time period during which they were most apt to use simulation in support of their training as a staff. The survey also mirrored questions in Survey 1 regarding the nature of simulation support, the availability of facilities and the utility of the simulations in supporting the staff's acquisition of specific skills. This user view of the simulation support was intended to provide an additional perspective on the information garnered from the infrastructure survey.

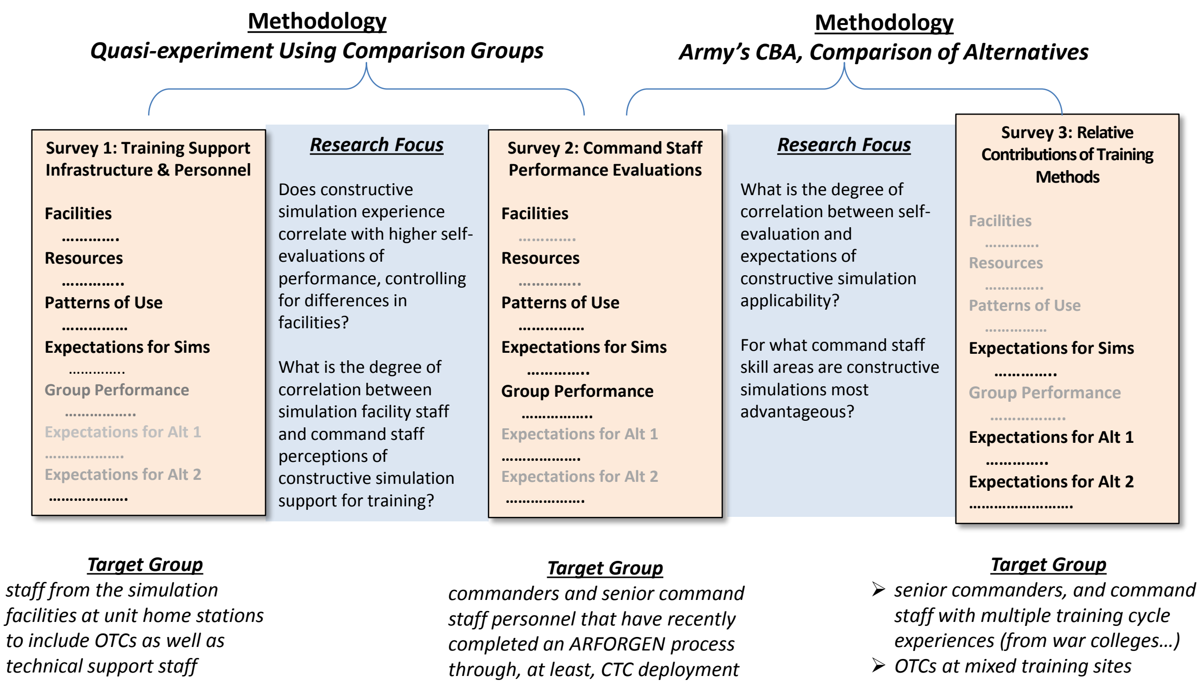

Figure 2. Target Groups, Surveys, and Analysis Methods

Survey 3, designed for Method 2, the cost-benefit analysis, was directed toward senior military officers who had completed multiple training and deployment cycles as commander or member of a staff and could speak to a wealth of accumulated experiences in training for, serving on, or directing command staffs. While Survey 2 focused solely on simulation-based training, Survey 3 was designed to determine whether the senior military officers could differentiate between the relative advantages of instructional, simulation, and live-exercise methods of delivering training.

Each of the three surveys included a demographic collection section that was largely similar to the other two. This duplication was intentional so as to facilitate the linkage of the Infrastructure and Command Staff surveys. The demographic sections were constructed in such a way as to permit identification of the home station training centers for the individual respondents without necessarily identifying specific units and without identifying the respondents themselves.

Cost Accounting and Analysis

A standard economic value calculation is return on investment (ROI). In the context of this effort, ROI is defined as follows:

$$ROI(\%) = \frac{\left(P_{benefit} - P_{Investment}\right)}{P_{Investment}} \times 100$$

Where $P_{Investment}$ is the sum of all costs associated with of a program P and $P_{Benefit}$ is the sum of all monetizable benefits of that program. Obviously the utility of such a calculation depends on how the program P is defined as well as the validity and accuracy of the data elements that comprise the two variables $P_{Investment}$ and $P_{Benefit}$. As part of this effort, a simple ROI spreadsheet was developed that enables an analyst to enter these data elements and automatically calculate the ROI. While in itself this is a trivial application of a spreadsheet, the utility is drawn from the quality of the data element entries. Furthermore it should be stated that since the military is a mission-driven enterprise, ROI alone will never be adequate to characterize value.

General Observations

The experience of seeking pools of respondents confirmed the historical observation that obtaining well-parameterized and documented records of performance is difficult at best. Such records have been traditionally regarded as either promotion-sensitive, or readiness-sensitive, or both, and as such are closely held. One offer of access to marginally useful (for purposes of this study) records looked promising until it became evident that the material needed for comparisons would be redacted because of the above sensitivities. As a consequence, data collection from forces, trained and about to be deployed or having just returned from theater was nearly impossible and the data sets opportunistically acquired were disappointingly small.

Obtaining data for infrastructure studies was comparatively easy. Respondents were not as driven by deployment schedules and spent their time focused on the simulation center, simulation capabilities and supporting the forces during pre-deployment training. When requested to do so, they were willing to be interviewed in focus groups (in person or via telephone) or individually or to respond to on-line surveys.

When considering the results of the data collection, the reader should keep in mind that the purpose of these experiments was to find and test a methodology for determining the value of simulation, and not to actually do the task of measuring the value of simulation. In that context, it was not strictly necessary to have sample populations that were statistically significant and measurably representative of the whole, but merely to have a sufficient number of reasonable respondents, answering validly, to determine the viability of the approach. Internal and external validity tests were conducted and are discussed in the final report.

Note on applicability of the methodology

This methodology was applied to the use case of constructive simulation in support of collective command and staff training. This is arguably one of the more complex and subtle use cases in the Live-Virtual-Constructive-Gaming (LVCG) spectrum. It is therefore expected that this methodology can be applied to other LVCG domains with some adjustments, including the Integrated Training Environment (ITE) as a whole.

RESULTS

The results are examined and documented in order, starting with the first evaluation method which relied on paired responses from Survey 1 and Survey 2. Data collection did not permit the pairing of responses as originally envisioned; therefore, the two surveys are discussed independently. The discussion of Method 2, the cost-benefits analysis, is interwoven with the exposition of results from Survey 3.

Method 1: Survey 1, Simulation Infrastructure

This survey served two primary purposes. The first was the creation of a high level inventory of simulation center capabilities and the second was to use that inventory to distinguish between various simulation centers. The differentiation among sites would then be correlated against the unit assessments from the Command and Staff survey to draw conclusions as to what capabilities (and more importantly what capability gaps) were associated with higher and lower functioning staffs and command and staff training experiences. The initial element of the survey was directed at identifying what simulations and federations were available to the centers and which were most often used in the training of commanders and staffs.

This section was followed up by inquiries into the adequacy of various resources required to support simulation (networking, hardware, software, personnel, AAR facilities etc.) and the resources available to support different levels of multi-echelon training. Additional questions were specific to simulation support personnel. A second portion of the survey was devoted to understanding the training experience as perceived by the simulation center staff. Particular inquiries were made regarding disruptions to training (type, frequency, work-arounds, impact). A number of these questions were mirrored in the Command and Staff survey in order to cross-reference and identify any discrepancies in experience.

Discussion of Results: Infrastructure Survey

Feedback regarding the Infrastructure was generally positive. Basic inventory questions were constructed in a manner that allowed for identification of additional factors not explicitly asked about. There was little or no consistency in such identifications, thus indicating that there are no major gaps in the coverage of relevant simulation center capabilities. The one area of feedback that indicated a survey weakness was that individuals were not provided the means of clarifying why available resources were not used. Respondents requested the ability to differentiate between non-use because of a lack of demand, difficulty of use, or absence of capability. This can be easily incorporated into a revised survey; however, it should be noted that absence of a capability is already captured and the actual inventory effort is agnostic as to the reason why a resident capability is not used.

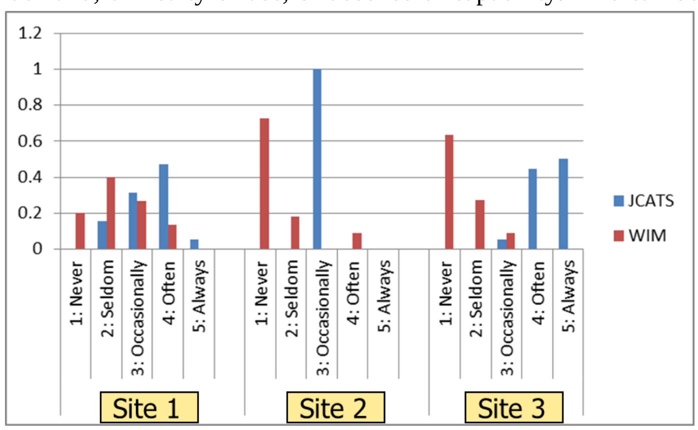

Figure 3. Comparison of relative use of JCATS and WIM at 3 Sites

Crucial to the success of this survey is the ability to differentiate simulation centers from one another. Three areas were tested to demonstrate this capability:

a) *Federation, federate, and simulation presence and usage levels*
b) *Adequacy of various infrastructure resources related to simulation usage for training, and*
c) *Adequacy of support for multi-echelon training*

The survey asked about availability of eight specific simulations (nine counting the ability to specify others of note), as well as how often the component simulations belonging to the Joint Land Component Constructive Training Capability (JLCCTC) Federations, JLCCTC-ERF (Entity Resolution Federation) and JLCCTC-MRF (Multi-Resolution Federation) were used. JLCCTC-ERF has 14 components and the current version of JLCCTC-MRF using WARSIM (JLCCTC-MRF-W) has 9 federates, all of which were named and available as choices for respondents. There are some issues concerning the range of responses within any given simulation center for any given federate. This may be due to the range of respondents or demands of the specific units they support making them more or less familiar with the use of specific federates. Figure 3 shows the frequency of use for two specific federates, the Joint Conflict and Tactical Simulation (JCATS) and WARSIM Intelligence Module (WIM), where the vertical scale is the fraction of responses in that category (total = 1 or 100%). The relative distributions of use across simulation centers as shown in Figure 3 appear to be distinctive. While only JCATS and WIM federates are shown here, results for other simulations display similar differences.

JCATS distributions are quite distinct. JCATS use at Site 1 is clustered around the middle values (2, 3, and 4) with gradually increasing percentages of respondents indicating higher levels of usage. Site 2 is highly consistent with all respondents indicating the middle value of occasional usage. Site 3 is clustered at the high end of usage, almost evenly split between "often" and "always" levels of usage. This data does dovetail nicely with that regarding availability of the simulation or federation: all respondents at both Site 2 and Site 3 indicated availability of JCATS; whereas by contrast, roughly 75% of respondents from Site 1 noted the availability of JCATS. WIM usage is somewhat less distinct with Site 2 and Site 3 appearing very similar to each other while respondents from Site 1 had a more even distribution across usage levels. Results suggest that usage distributions are capable of distinguishing between installations, although with some caution.

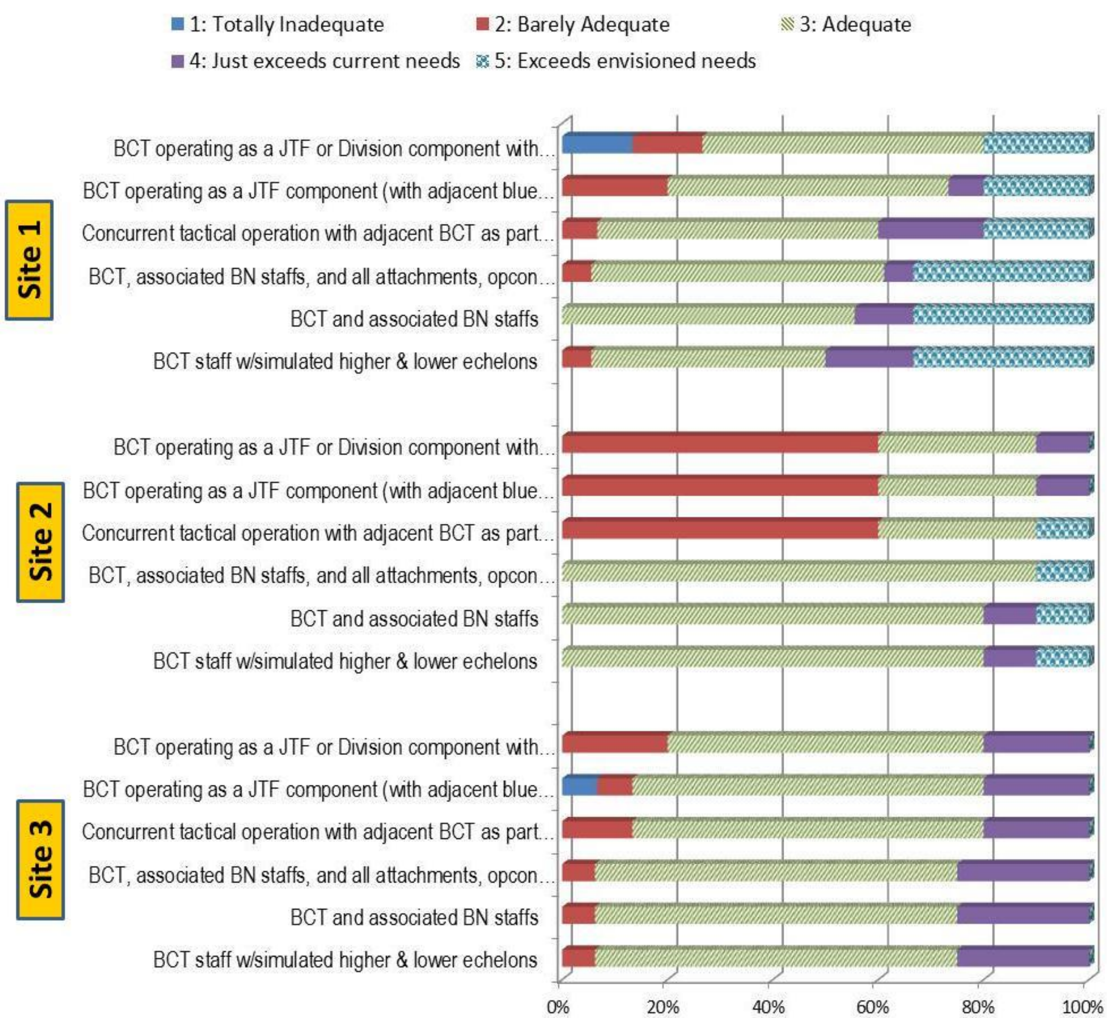

Figure 4 Comparison of the ability of three sites to support multi-echelon training

A further measure of distinction between the centers is their ability to support differing levels of multi-echelon training. Figure 4 shows the responses of the staffs at the same three home stations to this capability. As with the previous indicator, the distribution of responses within each of these three simulation centers does seem to show distinguishing characteristics. Site 2 is distinguished from the other two locations by a very distinctive pattern of support in which all responses to the first three categories are adequate or above; whereas, for the last three categories, 60% of respondents rated the support as barely adequate. The responses from Site 1and Site 3 are similar to one another (as was the case with the resource indicator); however, Site 3 is distinguishable because of the far larger proportion of responses above the "adequate" threshold.[iv]

The differences evidenced in Figures 3 and 4 confirm the initial hypothesis that there are distinguishable differences in simulation use and support at the various home stations – a critical factor for the successful use of Method 1. While the desired paired responses could not be collected, limiting the successful test of Method 1, the next section discussing the command and staff evaluations suggests that those surveys alone may be capable of providing most of the desired comparisons.

Method 1: Survey 2, Command and Staff Evaluations

Survey 2 was composed of three distinct sections. The first asked respondents to evaluate their unit's ability to function across 11 skill sets at both the mid-point and end of their training cycles. The second asked about the relative contributions made by simulation to the training of each of the skill areas, and the third asked about their training experiences.

Skill Assessment. The evaluations in the first section are the heart of this data collection effort. The entering presumption was that impact to command and staff functioning is most likely to occur in the latter stages of training, and that lacking outside objective skill assessments the commander and his/her staff would be the best sources for such evaluations. 11 skill areas were developed as shown in Table 1 in coordination with reviews of the literature on command and staff processes, interviews with subject matter experts, and review by focus groups. They do not represent strictly doctrinal tasks but rather groupings of skills and capabilities, relevant to various aspects of staff functionality. Development focus was on creating a reasonably broad (although certainly not exhaustive) set of skills, applicable across the full range of possible staff functions (although particular attention was paid to skills relevant to the MDMP (Military Decision Making Process) that are reasonably independent of one another.

Table 1. List of skill areas used in survey 2

Question	Skill Areas
16	Ability to develop viable plans in a timely manner
19	Ability to perform good course of action analyses
22	Ability to synchronize forces and resources
25	Ability to manage information to achieve situation awareness
28	Ability to integrate other service and allied assets (forces, ISR, etc.)
31	Ability to adapt plans, actions, processes, rapidly in response to external factors
34	Ability to understand the adversary's capability and intent
37	Ability to understand and anticipate the response of the indigenous population
40	Ability to work under stress
43	Ability to identify, assess, accept, and mitigate risk
46	Ability to conduct execution processes effectively

The methodology is based on the assumption that we should perceive improvements in the skills associated with high-functioning staffs between two points in time bounding the period most likely to emphasize staff training. This period starts from the point at which training shifts focus from basic, individual task skills into the building of a command and staff team that effectively and efficiently combines and coordinates those individual tasks. There is no clear point at which this occurs and differentiation is blurred because of the natural rotation of personnel assignments to the command staff. As a consequence, the "midpoint" of the unit's ARFORGEN rotation was chosen as the initial assessment point. While unit training is never ending, there is a definite point at which focus shifts from preparation to operation. This was captured by placing the second assessment point at unit deployment/unit transition to a CTC prior to actual deployment.

Simulation's Contributions to Training / Level of Simulation Support for Training. The second portion of the survey was directed at determining which skill areas had better or worse simulation support. Inclusion of this section is necessary to demonstrate attribution of improvements to the presence of simulation-based training. Demonstrating skill improvements over time is insufficient; there must also be indication that these improvements are tied to the use of simulation.

The Training Experience. The last portion of the survey was directed at understanding the training experience and resources that were available to the unit for simulation-supported commander and staff unit training. A number of these questions mirrored portions of the Infrastructure survey, omitting those areas that would require detailed training center expertise. While intended primarily as a check on data collected in the Infrastructure survey, additional questions were included in order to provide possible avenues for sub-population determination. Differentiation of answers between Survey 1 and Survey 2 (within the population matching training center to home station) would also provide a measure of the degree of disjuncture between perceived experiences, possibly pointing to areas of needed improvement.

Skill Assessments. While there was indication that some respondents misinterpreted this element (by noting that there was marginal improvement despite having given different ratings) and response rates for it were low, this format was able to provide meaningful differentiation of ratings across the two time points. Evaluations increased by an average value of 0.30 points between the two time periods. The lowest increase was for the "Ability to manage information to achieve situational awareness" (0.23 increase). The greatest increase was for the "Ability to work under stress" (0.36 increase). There was substantial differentiation of responses across both the population and across the skill sets.

In nearly all instances there was evidence of skill improvement over the designated period. Responses for the sub-population groups as well as the total population indicate a larger percentage of respondents indicating "0+" or better than for "0" or lower for all but three (out of 22) instances. This provides at least circumstantial evidence that the survey instrument was sensitive enough to produce viable results and that the choice of evaluation time points was appropriate.

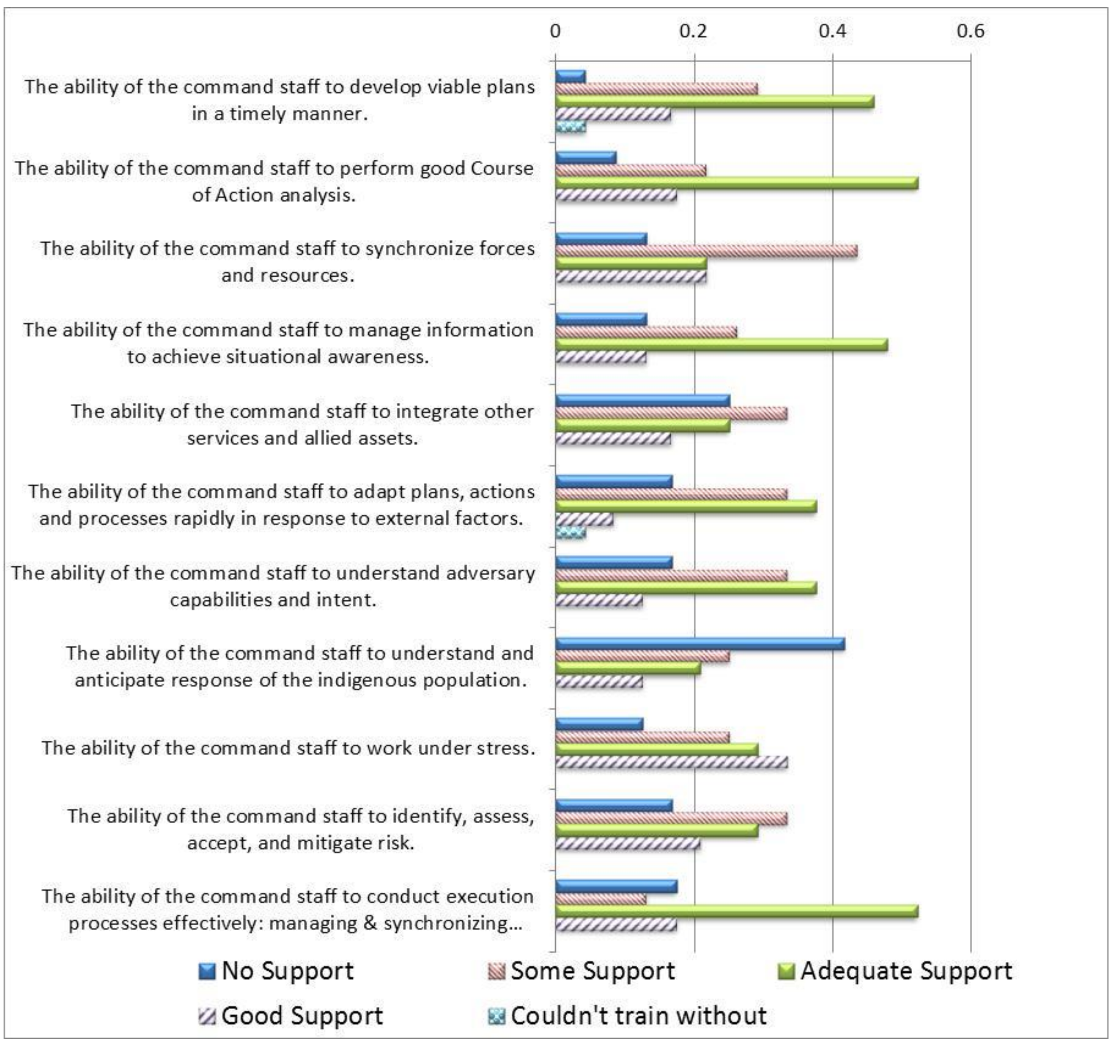

Figure 5. Assessment of the degree to which simulation supports development of specific skills

Simulation's Contributions to Training / Level of Simulation Support for Training. For each of the 11 skills, respondents were asked to rate the extent to which simulation supported the development of that skill.

Figure 5 shows the skills listed along the vertical axis and the responses for each skill displayed as bar graphs. It is easy to see the predominance of the response "adequate support" represented by the green line. The dominance of the response 2 for question 22 is clearly visible. The skill addressed in question 22 is the ability of the commander and staff to synchronize forces and resources.

The fact that these comparisons can be made by using Survey 2 only is an indication that a version of Survey 2, with questions limited to the primary 11 and retaining a few questions regarding perceptions of simulation support, should produce much of the data on performance associated with Method 1.

The Training Experience. Unfortunately, without corroborating data from the Infrastructure survey there is no means of assessing the ability of this section to check that data or to provide an indicator of how different the perceived experiences are between the simulation professionals and those undergoing training. However, the data collected on access to simulation and on the ability of simulation to support gains in proficiency is indicative of the ability to differentiate the population of respondents to Survey 2 meaningfully into those who received significant simulation support and those who did not. Further, these two subsets appear to be correlated with responses regarding degree of simulation support for specific skill sets, as well as with improvements in performing that skill set. Initially this was thought to be impossible; as the presumption was that the use of simulation was so pervasive that we could not elicit a control group from just the respondents to Survey 2. Assuming ubiquity of simulation, this differentiation implies a degree of sophistication on the part of the respondents – they are able to distinguish between general simulation usage and the use of constructive simulations (constructive simulation was specified in the questions). While the research team is inherently suspicious of the ability of the population to distinguish the presence of constructive simulation, the match of data between Q49[7] (see above) and Q60[8] and the levels of perceived simulation support for individual skill areas is highly suggestive that Q49 does bifurcate the population into useful test and control groups.

One would suspect a faulty instrument if those who had access to simulation did not rate average support by simulation higher than the general population. Further, one would suspect a bad instrument if those who indicated a positive contribution to proficiency did not rate average support by simulation higher than those that indicated a lack of contribution. Those values were computed and are shown in Table 2. The questions whose numbers are listed in the first column are those referring to simulation support for the 11 skills (see Table 1). The second column lists the average when all respondents are taken as a single group. Question 49 is the one used to divide the population into the two sets, only one of which had simulation available. Question 60 asked whether or not simulation provided support for skills development. Under the column labeled "Q49: Yes," are listed, on the left, the responses of all who indicated that simulation was available, and in the center the average response for all who answered yes to Q60 as well. Note that the responses in the darkly shaded column are uniformly higher than the average of the entire set of respondents (full population). This should be expected. For the respondents who did not have simulation available (Q49: No, last three columns), the average response is uniformly lower than the average of the full population; however, for those who answered 'no' to Q60, the average is markedly lower than that of the full population. For those who indicated that simulation supported skills

Table 2 Assessment of the degree to which simulation supports development of specific skills

	Full Population	Q49: Yes			Q49: No		
Question[11]	Full	Q60: Y or N (Grp A) Q49Y	Q60: Y	Q60: N	Q60: Yor N Q49N	Q60: Y (Grp B)	Q60: N (Grp C)
16	2.87	3.22	3.25	3	2.50	2.83	2.17
19	2.78	3.22	3.12	4	2.27	2.60	2.00
22	2.52	2.78	2.62	4	2.5	3.00	2.00
25	2.61	2.89	2.87	3	2.6	2.80	2.40
28	2.33	2.67	2.62	3	2.09	2.67	1.40
31	2.50	2.67	2.62	3	2.36	2.67	2.00
34	2.46	2.78	2.87	2	2.36	3.00	1.60
37	2.04	2.67	·2.75	2	1.81	2.17	1.40
40	2.83	3.33	3.25	4	2.73	3.00	2.40
43	2.54	3.33	3.37	3	2.18	2.67	1.60
46	2.70	3.33	3.37	3	2.54	3.00	2.00
	Middle	Highest			Lowest		

[7] Q49: "Were constructive simulation capabilities made available for your training as a staff?"
[8] Q60: "Did the available constructive simulations provide the capability to support the commander and staff in developing proficiency in their tasks?"

acquisition (presumably from other experiences), their responses were generally lower than those for whom simulation was available during their current training experience. Thus, the ability of the survey to detect two distinct groups whose answers corresponded to the logical responses expected of that division is empirically supported.

Infrastructure and Command Staff Survey Intersection

As indicated previously, the voluntary nature of participation did not result in sufficient overlapping populations between the two surveys to appropriately test whether the differences between simulation center facilities could be correlated with differences in unit self-assessments. However, both populations were queried as to their perceptions of the adequacy of simulation support for various types of missions.

Substantial agreement does seem to exist with respect to the paucity of support for Cyberspace Operations; otherwise, the two communities are very far apart. None of the seven mission types listed (except Cyberspace Operations) were within 15% of each other in terms of responses indicating adequate or better support. Particularly telling are the areas of Entry Operations where nearly 80% of Infrastructure responses rated simulation support as adequate or better while less than 40% of Command and Staff responses equivalently estimated said support. Support for Counter-Insurgency was similar. Almost 90% of Infrastructure responses rated support as adequate or better – while only 45% of Command and Staff responses rated it that high. Bearing in mind that this data does not match simulation center to the military unit, the contrasts are nevertheless quite distinct and suggestive of non-trivial differences in perception between the population that operates the simulation centers and the units that use those centers for home-station training.

Method 2: Cost-Benefit Analysis Approach

The cost-benefit analysis was designed for military officers who had multiple experiences with pre-deployment training and subsequent deployment. It contained no self-evaluation of performance, but focused instead on the relative contributions of several different methods of delivering training to the same 11 skill sets used in Survey 2. Since the respondents to this survey were senior personnel who likely have knowledge of the costs associated with the various methods of training, they are positioned to make a qualitative cost/benefit comparison.

Overview of Survey 3

The questions in the survey broke down the eleven skill sets into sub-elements and first asked the respondent to evaluate the importance of each of the subsidiary elements in performing that skill. Following that question for each of the subsidiary elements, the respondent was asked to rate the contribution of each of the different training methods and then assess their cost effectiveness in imparting that skill.

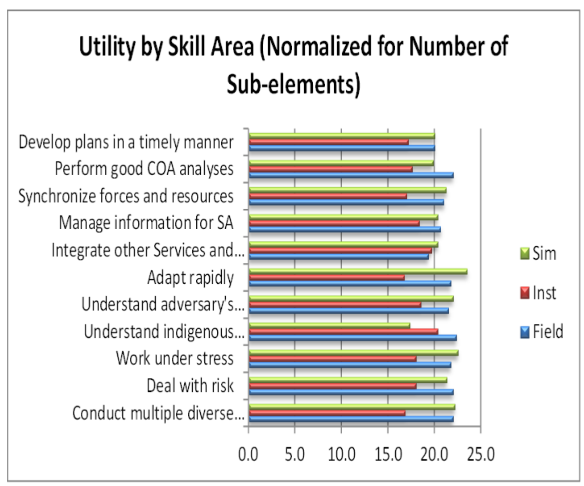

Figure 6 Comparison of the utility of three training methods for eleven skills

The first skill was the ability of the commander and staff to develop viable plans in a timely manner. For each of the sub-elements, the respondent was asked to rate the relative utility of the three types of training (instructional, field exercises, simulation) and their cost-effectiveness. This process was repeated for all 11 skills. There was very little differentiation among the subsidiary elements as to their importance in building the parent skill; therefore, it was reasonable to sum the unweighted scores to get a summary score for the skill. With so little differentiation, it seemed appropriate to eliminate the subsidiary elements in any future survey and evaluate at the parent skill level. This was the same assessment made for Survey 2.

Review of Data from Survey 3

While the sample size was small (6 respondents, only three of which completed the survey), the results pointed to the viability of this method to capture at least the relative value of simulation for commander and staff training and at the same time acquire an estimate of its comparative cost-effectiveness. The results for the utility and cost-effectiveness for each of the 11 skills are shown in Figures 6 and 7. The skill areas are identical to those presented for Survey 2; however, to aid legibility, the labels on the vertical axes of the following graphs display a condensed wording of those skill areas.

In Figure 6, note that simulation and field exercises are first and second for all skills except that of understanding the indigenous population. Warfighting simulations currently in use do not represent well the human element, particularly the non-combatants. In most cases this skill is effectively trained by bringing individuals with the same ethnic background as the indigenous population and having them roleplay in exercises and in classroom seminars. The response is logical and represents a known reality.

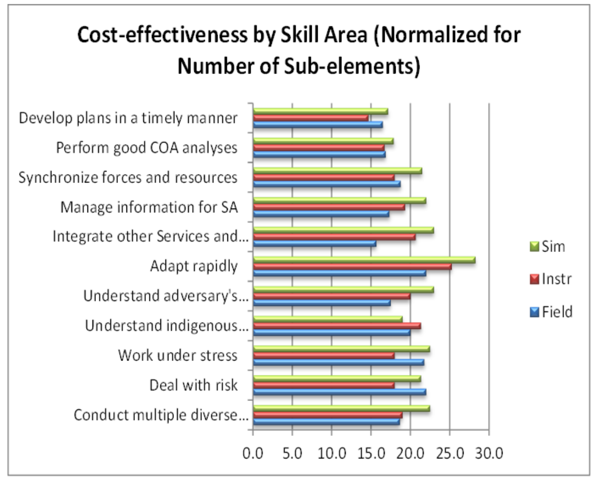

Figure 7 Comparison of cost-effectiveness of training methods for eleven skills

Figure 7 shows the cost-effectiveness for field exercises, instructional training and simulation for the same skill area shown in Figure 6. There is more differentiation in the responses for cost-effectiveness, reflecting the fact that while exercises are a favored mode of learning for the military, experienced officers recognize that there are significant costs involved in staging field exercises, and simulation is often a valued option. As with the assessment of utility, the skill of understanding the indigenous population presents a different picture from the other skills and instructional training is rated as most cost-effective.

A cost-benefit analysis provides an approach to determining a relative, but not absolute, value of simulation. There are questions about how much differentiation will be possible as most training is a combination of all three methods; however, those who took the survey seemed able to perceive a difference. The method also encompases a natural approach to including cost in the assessment, even though the cost is perceived rather than actual. This method will not provide a set of curves or expressions that address the question of how much training value for what increment in cost.

Observations on the Cost Accounting Component

The principal cost data element that was examined in this effort is what can be thought of as cost avoided using simulation to represent vehicles, munitions, and soldiers in a constructive simulation based training exercise. The LVC CostCounter software was employed for this purpose. The plan was to stress the LVC CostCounter software during an Integration Event for a fairly large simulation federation. JLCCTC-ERF was chosen because of its large number of simulations (up to 20) and simulated entities (over 50,000). The data collected from the bulk of a 2 week Integration Event was used only as a proof of principle. A specific Force-on-Force scenario was run to collect and analyze the data. In preparation for this, each entity type and munition was mapped to a specific Cost Model within CostCounter using DIS[9] Enumerations. For the sample event, there were 105 unique Types (87 Entities and 18 Munitions), with multiple variations within certain Types. Once the mappings were complete, a Cost Model was established for each Type that was mapped.

For the compilation to be useful and believable the costs associated with entities in the simulation should be realistic. The cost data chosen for this effort were the values in the TRMIS[10] database maintained by HQDA[11] and

[9] DIS: Distributed Interactive Simulation
[10] TRMIS: Training Resource Model Information System
[11] HQDA: HeadQuarters, Department of the Army

FORSCOM.[12] The costs contained in TRMIS are three-year running averages updated annually for ~473 major end items based on actual costs incurred as the end items have been used in training.
For comparative purposes, using the LVC Cost Counter, the following costs were estimated for a BLUFOR BCT[13] arrayed against an OPFOR BDE[14] for an event of 90 minutes duration. The total costs accrued in the cost counter tool for this simulated combat was over $2.9 million dollars. This figure is comparable to the reported costs associated with actual live training events at a CTC.

The value of simulation methodology described in this paper shows promise as a means of describing both the qualitative and the quantitative value of constructive simulation in the training of BCTs to conduct MDMP-related training. The survey instrument, in particular, has the demonstrable ability to measure some aspects of value in this context. The cost accounting component of value can be addressed through an ROI calculation, particularly if the calculation is populated with authoritative data. A novel approach to collecting cost avoidance data is the use of the LVC Cost Counter. It is also possible to combine the major components of value into a single measure although it must be remembered that any summary quantitative measure must be accompanied by qualitative information as well in order to fully appreciate the value measured.

FUTURE WORK

The survey instruments show promise in obtaining data as well as allowing for meaningful analysis. This was the intent of the present work, and shows it is possible to isolate the effects of a particular kind of simulation, the real utility may be in a broader application. In addition, the somewhat more straightforward (but problematic due to cost data sensitivities) component of the methodology, cost accounting, can be significantly enhanced by identifying and populating the constituent data elements.

Given that, here are some directions for future work:

- Mature the ROI model; employ the use of the cost counter tool at multiple federation locations
- Further investigate the potential to access unit performance data (e.g., AARs, USRs[15]) that would serve to augment the survey data
- Continue to collect data on a volunteer basis
- "Commoditize" the current methodology and make available for broader use across the BCTs as a standard procedure
- Expand the investigation to LVC-G in the ITE context

A more important consideration is whether using this technique, or any other technique, to collect data on a one time basis is sufficient to answer what is really at the heart of the study question: how do we measure value so that we can improve the use of simulation, finding and maintaining its proper share of funding, for its intended purposes and do so in the most cost-efficient manner possible? It seems clear that the best way to answer this question is to persistently collect data using a proven methodology. Given this current work in characterizing the value of simulation, it seems beneficial to consider a transition to ultimately "commoditize" this approach across the Army training community. Using techniques identified in this paper, data may be collected with little effort or interference and done so anonymously. These data, collected over time, may prove invaluable in understanding how simulation dollars can best be spent and improving the quality of training for the Warfighter.

ACKNOWLEDGEMENTS

The VoS team would like to acknowledge the support of a panel of senior advisors from both the M&S and Mission Command communities. Membership included Dr. Amy Henninger, Mr. Milton Eng, Mr. James Shiflett, Dr. Mike Macedonia, Ms. Cynthia Harrison, Dr. Randall Shumaker, Dr. Brian Goldiez, Dr. John Sokolowski, Mr. Bill Waite, Mr. Steve Horner, and Mr. Marshall Reed. Mr. Frank Wysocki, Mr. Jim Ruth, Mr. Dominic Pham, and Mr. Jared Lee also made significant contributions to this effort. Finally, the VoS team would like to acknowledge the many units and organizations within the U.S. Army that supported this effort.

[12] FORSCOM: US Army Forces Command
[13] BLUFOR BCT: "Blue Forces" Brigade Combat Team
[14] OPFOR BDE: Opposing Force Brigade
[15] USR: Unit Status Report

REFERENCES

United States Government Accountability Office, *Designing Evaluations*, GAO-12-208G, January 2012.

Office of the Deputy Assistant Secretary of the Army (Cost and Economics), U.S. Army Cost Benefit Analysis Guide, 2nd Edition, April 8, 2011.

Oswalt, Ivar; Feinberg, Jerry; Cooley, Tim; Gordon, Steve; Waite, William; Waite, Elliot; Lightner, Gary; Severinghaus, Richard, Calculating Return on Investment for U.S. Department of Defense Modeling and Simulation, M&S Journal, Fall 2012.

SIMCI Value of Simulation, to be published

---

[i] To reiterate, this task's objective was to create, assess, and validate a methodology for addressing this problem statement; it was not intended to, nor does it, provide an authoritative answer.

[ii] LVC Cost Counter is a product of Calytrix, Inc that collects the true costs of training activities if real vehicles, fuel, munitions, etc were used instead of simulated entities. LVC Cost Counter monitors the simulation network in real time, and using user-defined costing models, updates running totals in various cost categories and displays the results in a browser. For the study, Cost Counter collected costs of most vehicles by miles driven, along with costs of munitions fired (using actual costs of "dummy" munitions where available).

[iii] In lieu of CTC generated evaluations, collection of which proved untenable, unit surveys asked respondents to evaluate their skills prior to the onset of major simulation usage (mid-point of the ARFORGEN cycle) and just prior to deployment (corresponding with CTC attendance).

[iv] In theory differentiation between training centers could be quantified (difference of means statistical tests, cluster analyses, etc.) given broad enough collection. However it isn't clear that quantification provides much additional value. At face value the MTCs provide very different training environments, specific to the units they process; this project demonstrated that the surveys provide enough descriptive detail to distinguish between them whether or not a difference of means tests shows significance at some arbitrary level.