Establishing Metrics and Creating Standards: Quantifying Efficacy of Battlefield Simulations

Dan M. Davis
for the University of Southern California
5916 Bixby Village Dr, Ste. 88
Long Beach CA 90803
310 909-3487
dmdavis@acm.org

Skander Guizani & Evan Jaksha
Visiting Researchers, ICT, USC
Playa Vista, California
{iguizani, ejaksha}@ict.usc.edu

CAPT Daniel P. Burns, USN, Ret.
Home Port Solutions, LLC
Savannah, Georgia
daniel.p.burns@homeportsolutions.net

Mark C. Davis
Computer Engineer
Mooresville, NC
davismc@ieee.org

Keywords:
Virtual humans, team formation, terminology standardization, proactive NLP

Abstract*:* *This paper asserts that quantification and verification of Battlefield simulations is necessary to assess, verify, and guide the researchers, military commanders, and users in both their development and their implementation. The authors present their observations on previous development activities that were hampered by lack of effective metrics and present their arguments that much of this was driven by a lack of standards. Tracing back using commonly accepted System Engineering practices, they show how lack of such standards makes even to the development of effective metrics problematic. The paper documents the experiences and enumerates the potential pitfalls of these shortcomings. Both the authors' experiences in military service and the technical literature supporting their theses are adduced to support their analysis of the current technical research and development environment. Then the paper evaluates several System Engineering tools to further investigate and establish the ultimate goals of these formalized processes. Using their current project in establishing virtual on-line mentors as an exemplar of the way such tools would be effective, the authors make a case for the needs for metrics standards that both are accepted by consensus and are ultimately directed at providing the warfighter with all of the training possible before putting that warfighters in harm's way and imperiling the missions for which they are putting themselves at risk. Examples of the nature and reaction to simulator training, virtual human interaction, computer agent interfaces and implementation issues are given to further illuminate for the reader the possible extensions of these approaches into the reader's own research as well as calling for a more community-wide recognition of the needs for standards both for implementation and for metrics to assess Battlefield Simulation utility to the warfighter. Future investigations, analysis and action are considered and evaluated*

1. Introduction

The goal of the following pages is to present an analysis of the current state of the art as to evaluating the efficacy of a genre of simulation typically referred to as battlefield or battlespace simulation and to advance the reasons for more quantitatively characterizing their actual impact on their users. The first subject of this investigation will be a short analysis of the discipline itself and the ostensible lack of tools and techniques to provide a quantifiable metrics set in training outcomes. Part of this analysis will be the authors' anecdotal experiences with the incidental training opportunities that they have observed during evolutions not originally designed to achieve training as a primary objective. Then some System Engineering concepts and tools will be applied to further illuminate both the current and potential environments. The need for standards and metrics are then articulated, along with the justification for this line in inquiry. These issues are then related to the current and future activities of the reader, with references to supporting aids in implementing improved evaluation methodologies. The early and continuous participation by the standards community are presented and justified.

1.1. Background of Battlefield Simulation

For millennia, leaders have used various figurative approaches to prepare for the protection of their nations. These have ranged from the ancient game of chess to complex computerize simulations of warfare. When threatened, there is a natural inclination to use what has been successful in the past and there is an opposite tendency to make use of potentially advantageous new developments. Two of the promising technologies available to defense leaders today are high performance parallel computing and advanced data analysis.

Contemporary analysts are faced with increasing pressure to provide better ways to both analyze the present dangers and prepare for the future operations. The largely empty fields of yesteryears' battles are being replaced by the crowded urban warfare environment of today, populated with non-combatants for whom most western civilizations feel an increased sense of moral, ethical, and legal responsibility. Weapons of increased lethality and improved targeting capabilities make it desirable and possible to honor this ethical re-awakening. Planners and trainers demand access to battlespace models of unfettered scale, that are built on increasingly sophisticated environments, with better emulations of foes and more realistic civilian populations. The coordination and synergy of these simulation and analytical capabilities are necessary to deliver insights for the trainee, analyst and evaluator.

Up-to-date terrain databases are now available in multiple resolutions and for nearly every area of the globe. Using workstation and PC technologies hosted on LAN configurations, truly incredible advances have been made in the DoD's ability to provide a realistic and geographically appropriate environment for conducting large operations [1]. Even these capabilities, however, are often limited in two important dimensions: resolution and total area. As the areas of interest broaden for both the analysts and policy makers, the need to have access to representations of any terrain in any season becomes more imperative.

An example of an important feature of current computer practice to which the community has become accustomed is the constraint imposed by the limits of individual processing speed. The desire to represent sophisticated behaviors requires ever-increasing processor power, and this is magnified by the desire to run multiple instances of non-deterministic simulations to evaluate the range of outcomes [2]. The need to represent tens of thousands of entities that are "aware" of each other also impacts performance. In one class of this "awareness," entities are within a range where they can see each other.

A much more extreme case is now of concern: the high altitude intelligence platform with sensors that can "see" virtually every entity in an entire theater of war. Current programming, as exemplified in the SemiAutomated Forces (SAF) programs, handles this location and awareness issue by running an inter-visibility calculation every few milliseconds. Obviously, with a huge number of entities, this represents a huge compute burden. Current practice shows this type of situation can be simulated on a typical single processor of present-day capacities at only a few hundred vehicles. In Millennium Challenge 02, a network of similar PCs on a LAN, experience seems to indicate that the total vehicle count is limited to a few tens of thousands - not enough for military purposes [3].

Moreover, modern battlefields are rarely located on remote plains, and the battles in urban areas are not fought with the destructive abandon of World War II, as it was in Stalingrad, Berlin and Tokyo. Instead, the modern analyst is looking for ways to achieve national goals while operating in populated urban areas with no loss of non-combatant life, minimal destruction of civil infrastructure, and optimal survival of friendly forces. For that reason, the simulations-enabled analyst is faced with the challenge of trying to understand how modern intelligence platforms can view a city full of vehicles and other entities. Clearly, something on the order of a million civilian entities approaches realism; a few thousand does not. Naturally, this leads to significant System Engineering and Standardization issues

Within the authors' experience, analytical approaches have not changed much over the intervening decades, but massive data handling capabilities have increased dramatically. With all of the DoD's increased sophistication in electronically produced simulations, one very common method of strategic deliberation remains the observation, logging and analysis of simulation outcomes by subject matter experts (SMEs). The authors maintain that adopting and implementing analytical techniques used in the behavioral sciences and operations research should enable these experts to be even more valuable. However, these advances will require improved quantification and thoughtful definition of figures of merit.

The earliest computer-generated simulations were often single platform/vehicle simulators, *e.g.* cockpit trainers and tank turret mock-ups and were used primarily for training, but occasionally were used for evaluation of both equipment and personnel readiness.

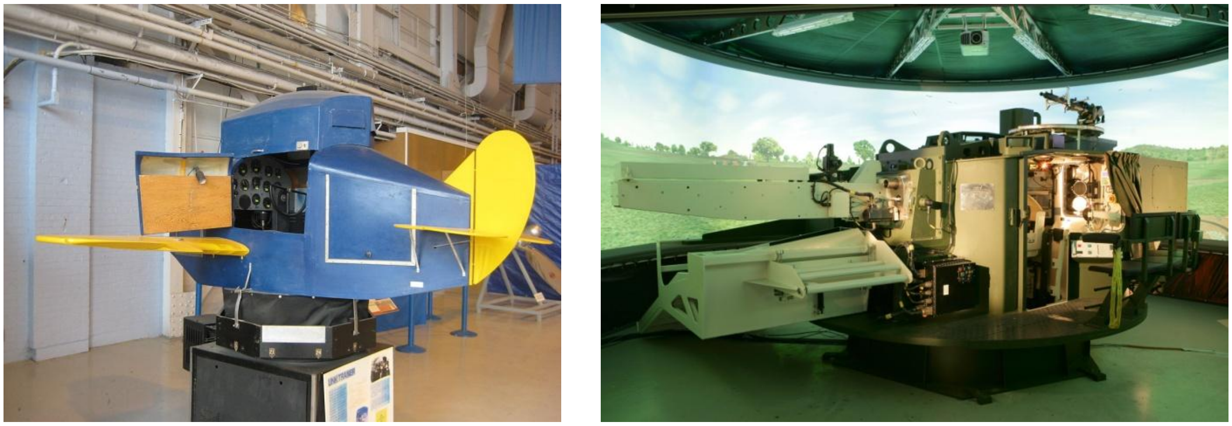
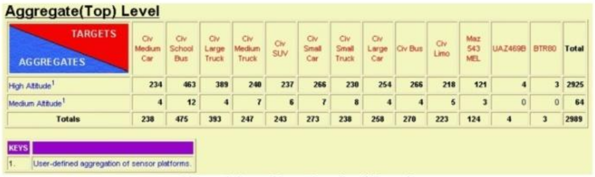


Figure 1 - Link Flight Trainer *circa* 1943 and KMW Tank Turret Trainer *circa* 2005 [4].

Because of the small numbers of trainees in individual systems, analyses of participant performance, training achievements and equipment design were not too difficult. Late in the 20th century, efforts were made to link many of these individual platforms and "vehicles" together to provide interactive and team training. [5]

This led to the recognition of a need to have even more constructive or virtual entities made available via simulation [6], an effort in which several of this paper's authors were intimately involved. Continued pressures for even more simulation objects resulted in the further growth of simulations sizes [7]. These successes of consistently simulating more than 10,000,000 entities generated huge amounts of data [8]. A single exercise could easily store a trillion bytes of data, even after all "redundant and non-essential data" was pruned from the databases. Early attempts at making the distilled simulation insights accessible centered on easily understood presentation of the data.



Aggregate(Top) Level

TARGETS AGGREGATES	Civ Medium Car	Civ School Bus	Civ Large Truck	Civ Medium Truck	Civ SUV	Civ Small Car	Civ Small Truck	Civ Large Car	Civ Bus	Civ Limo	Maz 543 MEL	UAZ469B	BTR80	Total
High Altitude[1]	234	463	389	240	237	266	230	254	266	218	121	4	3	2925
Medium Altitude[1]	4	12	4	7	6	7	8	4	4	5	3	0	0	64
Totals	238	475	393	247	243	273	238	258	270	223	124	4	3	2989

KEYS

1. User-defined aggregation of sensor platforms.

Figure 2 - Sensor Target Scoreboard from JFCOM Experiment [8]

This programming was not cutting edge and it fails to convey, in a graphic and easily internalized way, the salient relationships that are critical. Numerical data in particular, require time to absorb and analyze. This is a time sink that may be available to small-scale simulation analysts and to officers in non-combat environments; however, it presents way too much data for effective analysis of large-scale live or virtual test situations and would of course impose unacceptable burdens on warfighters experiencing the stress of combat. Again, the need for standard terminology, intuitively understood and bolstered by quantified parameters should reduce catastrophic misunderstandings.

1.2. Levels of Simulation

One set of discriminators in the Battlefield Simulation community is the level of human involvement. The levels most commonly accepted are Live, Virtual, and Constructive, listed in order of increasing reliance on and control by computer functions. As with all such trifurcations can be challenged on basis of the blurring at the interfaces between the adjacent descriptions, but in day-to-day efforts this nomenclature is very useful. The following is an example of terminology that has been accepted as standard and is almost always identified by its acronym (LVC). It is defined here straight out of the DoD Modeling and Simulation Glossary [9]:

> "Live - A simulation involving real people operating real systems. Military training events using real equipment are live simulations. They are considered simulations because they are not conducted against a live enemy.
>
> Virtual - A simulation involving real people operating simulated systems. Virtual simulations inject a Human-in-the-Loop into a central role by exercising motor control skills (e.g., flying jet or tank simulator), decision making skills (e.g., committing fire control resources to action), or communication skills (*e.g.,* as members of a C4I team).
>
> Constructive - A simulation involving simulated people operating simulated systems. Real people stimulate (make inputs to) such simulations, but are not involved in determining the outcomes. A constructive simulation is a computer program. For example, a military user may input data instructing a unit to move and to engage an enemy target. The constructive simulation determines the speed of movement, the effect of the engagement with the enemy and any battle damage that may occur. These terms should not be confused with specific constructive models such as Computer Generated Forces (CGF), a generic term used to refer to computer representations of forces in simulations that attempts to model human behavior. CGF is just one example model being used in a constructive environment. There are many types of constructive models that involve simulated people operating simulated systems."[10]

Especially with the advent of new technologies the dividing line between the three categories is becoming increasingly indistinct. The new immersive and mixed reality simulations may be a good example of the need for a concept more like a Venn Diagram representation than and a neatly-ordered list. Yet these are important concepts and drive many new decisions. Live simulations are typically realistic in many ways, but have heavy staffing and physical environment costs. The environment is not easily altered, hence less likely to be able to respond to new situations.

1.3. Uses of Battlefield Simulations

The authors have observed that battlefield simulations have commonly been characterized as being useful in three areas: training, analysis and evaluation. Again, the precision with which these terms may be discriminated may be difficult to ascertain, but the concepts call to mind the need to identify what the simulation is expected to achieve. It may even be argued that a simulation often performs all three functions every time it is run.

Training

Instruction and applied exercises for the attainment and retention of knowledge, skills, and attitudes [11]. In the authors' experience with many service personnel assigned to be the live, real-time human input into simulations being used for analysis and evaluation, they have been privy to many discussions in which the participants recognized the value to their own readiness by such participation. They often encouraged the researchers to advocate for the implementation of a training system relying on such simulations. One General Officer told one of the authors, "This has been the only time in my career that I had the experiences of commanding an entire Regiment in combat, either real or simulated."

Analysis

The process by which information is transformed into intelligence; systemic examinations of information to identify significant facts, make judgments, and draw conclusions. [12] .Many times, the DoD has a need to consider the traditional "What if?" questions. By a careful programming, large scale, independent agent simulations give the analyst a validated "sand box" in which to assess the range of outcomes of various scenarios. The processes involved are characterized by assigning a range of observed probabilities and then allowing a random process to pick from amongst those. These stochastic results can then be expanded by running the program multiple times, and, depending on the extent to which they are constructive, the additional runs are not too costly.

Evaluation

Judging, assigning, or affixing the worth of something [13]. Battlefield simulations are an excellent way to test the impacts of new or even envisioned devices or tactics. Not only are there huge savings, a variety of environments can be invoked easily. These environments fall into the ambit of geography, season of the year, weather conditions, nature of the threat, nature of allied support, and a wide variety of other parameters important to the DoD

2. Standards and Metrics

2.1. Lack of Standards and Metrics

History, both ancient and modern, is replete with examples of failures due to the lack of communication capabilities which arose and endured because the communicators assumed others thought they knew what they meant. This was often exacerbated when more than the originator and the addressee were involved, *cf.* Grant's order to Wallace at Shiloh [14]. On the side of quantification, one would not be ill-advised to cite the account by Professor Tetlock in his book, Super Forecasting: the Art of Prediction [15]. He recounts the story of a CIA analyst asked to form a team to predict whether Stalin would invade Tito's Yugoslavia. After a great deal of study, the team proudly reported to the National Security Committee that there was a "serious probability" of an invasion. Everyone seemed appreciative of that intelligence, but the next day one of the CIA men ran into another member of the NSC and the colleague asked: "What odds did you have in mind when you said 'serious probability' yesterday?". The CIA man said "About 65%." To which the other responded, "Oh, I thought it meant much lower than that!". Troubled, the CIA man, confident in his own assessment, went to his study group members and asked them what they meant when they said those words. The range was from 20% to 80%!! Without a number, the report was basically meaningless.

2.2. Some Suggested Parameters for Assessing Learning Metrics

Engagement and quality of training materials

One of the ways others have evaluated training functions is by recording the frequency of which the content is accessed. It is the authors' position that this useful information (training is certainly of no use unless the subjects use it), but that is not infrequently cited as proof of efficacy, when it is not training the organization wants to foster, it is a change in attitude or an improvement in performance. So, accepting that caveat, this parameter is a quantifiable first look at the impact a program may have. A converse parameter may be the amount of training that is not accessed and nature of the subject content that is skipped. Other parameters in this general area are participation in training sessions, hours engaged, and, finally, the rigor and sophistication of the issues presented. The psychometricians naturally favor survey instruments that generate quantified data, *e.g.* user evaluation instrument responses

Mastery of material and skills

More in line with the established criteria that really respond to organizational needs, identified by using some of the System Engineering tools discussed herein below, is the assessment of competency levels measured by evaluation instrument scores and then comparing progress before and after training. This more effectively gets to the goals sought: indentify the stake-holders, define the goals, and measure real progress toward those goals.[16] Another way of measuring progress via testing is to use tests already validated and accepted, *e.g.* analyzing scores on advancement tests. This technique "off loads" the effort of creating, validating and updating test instruments, without encoring costs on others that they were not already committed to expending to accomplish their mission, in this case determining who is to be advanced.

Improvements in job performance

Albeit the easily utilized and pre-quantified nature of tests and surveys, the central issues is efficacy of the training in improving real life performance, increasing mission accomplishment and preventing personnel losses. One way to do this is to collect peer and command observations of these important characteristics and contributions of the trainees. In addition to the evaluations of the organic resources, the input of other sources, such as cognizant colleagues from other units should be collected. This also raises the obligation of command personnel to pass along constructive comments of other unit personnel to that unit's seniors.

Overall achievement of mission goals

Given the opportunity, time and facilities, the best way to evaluate training is to observed accomplishments longitudinally, *e.g*. recording number of missions accomplished over time. If one of the goals is retention, longitudinally may mean several decades. Shorter periods of time will suffice in other goals, and often even the long-term goals may have interim evaluative events, for instance retention trends may be illuminated by a datum like next enlistment decision. The military is notoriously a dangerous profession, so trends in unit casualties will remain a useful parameter, but care must be taken to not be misled by changes due to outside factors. New combat assignments may be the reason for personnel losses instead of a reflection on the training that preceded the observed delta in casualties. Often, military training is directed to more rapid action, vital in combat. So, one area of quantification that is useful is measuring the response time to alerts. As mentioned, another measure of utility is warfighter retention and re-enlistments, even if that were not the major goal of the training.

2.3. Applicable System Engineering Tools

Stakeholder Diagram – identifying groups with a vested interest in training outcomes:

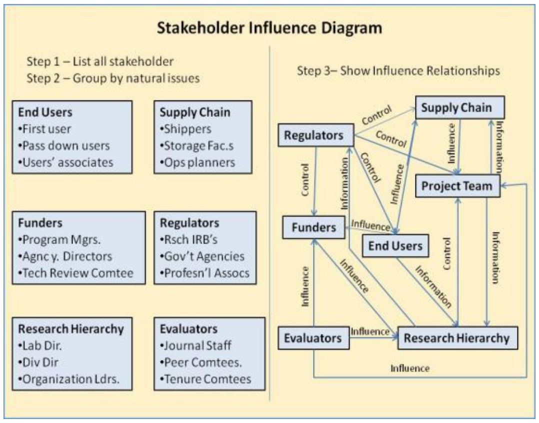


Figure 3 - Stakeholder Diagram

Most of the research efforts in which the authors have served were at least vaguely aware of the people who were interested in its outcome, generally knew the needs of the end users, and were familiar with the command and political entities that would want to be kept abreast of its progress and success. Some of the more astute research managers may even have had an organized list in their minds of who needed to be satisfied, if not actually served, by the project. System Engineering practitioners urge a more formal, but not onerous, process: the creation of an informal, but graphic, representation of such a list. This is often called a "Stakeholders' Diagram." An example of such a diagram is included as Figure 3. Creation of such is reasonably obvious and the authors suggest "brain-storming" early in the project to create such a list and "buy-in" as to its validity and importance. Just the process of creation is often illuminating. This paper supports the thesis that this process could benefit from quantification as to expectations from the identified groups.

Articulating Requirements – converting customer objectives into quantified system requirements:

A vital foundation of any effort is the understanding and characterizing of what is expected and what is required as product. This mandates a rigorous and disciplined process for reviewing, clarifying, and articulating all requirements. This process is critical to precluding "mission creep." Lack of clarify and any conflicting clauses must be rectified expeditiously. Also redundant and nonproductive requirements must be identified and deleted. A careful reading of all controlling documents should be followed by a comprehensive listing of all requirements. These then should be reviewed by all of the cognizant groups within the effort, as a requirement that is well recognized may be unfamiliar to other groups. The reviewers should remember not only the ultimate goal that is envisioned, but the specifications, contractual imperatives, and legal constraints that may drive the system design. Careful attention to these details may have significant impact on project success and in preventing cost over-runs.

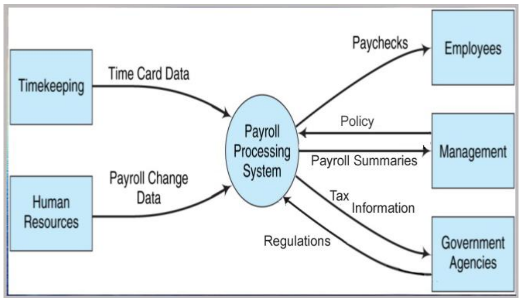


Figure 4 - Context Diagram (Slide Share, 2018)

Context Diagram – visualizing the mutual impacts of external entities and the system:
This tool is designed to highlight how the central concept under consideration fits in with all of the rest of its environment. Its major use it in identifying both the boundaries of the concept and the necessary interfaces that will be required as it will be implemented. It is critical to know what needs to be done, what should not be done and in what ways to associate harmoniously with adjacent entities. An example featuring a payroll function is shown in Figure 4. It shows both the interfacing entities, but identifies the required interfaces that must be defined and implemented.

System Requirements Model – putting requirements into a framework:
Often referred to as Requirements Analysis, this is the process of identifying, characterizing, and articulating all of the requirements of the task ahead. It begins by splitting up Operational Requirements (the end goal and the effort's *raison d'être*) and the Functional Requirements (the individual functions needed to achieve that goal.) Then Functional Requirements are further defined by their constraints, referred to as Non-Functional Requirements (size, energy needs, *etc*.). The actual implementation of this process in any efforts has to be tailored to that effort, but it is essential that the process be rigorous and that it ferrets out any unidentified requirements and that it be revisited periodically to ascertain if any hitherto unidentified requirements have surfaced. Quantification is a must here.

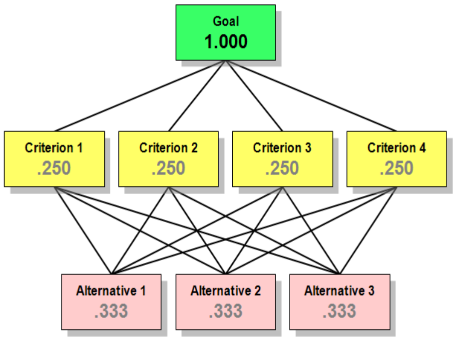
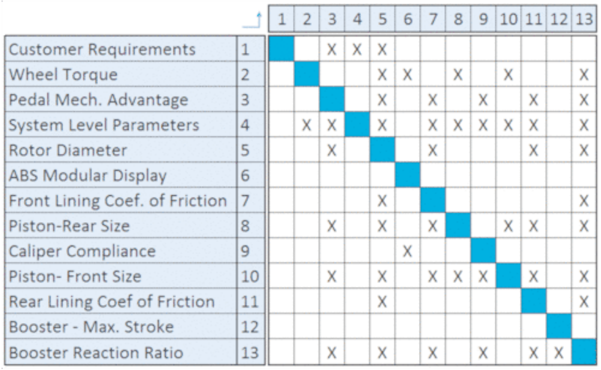


Figure 5 - AHP Chart

The Analytic Hierarchy Process – quantifying the decision-making process:
Another way to approach major decisions is the Analytic Hierarchy Process, commonly referred to as the AHP. It decomposes a single major challenge into many sub-areas for consideration, and then displays them as a hierarchy to better enable the decision makers to understand the values and interrelationships of the various components of the decision. It relies upon assigning numerical values to the various components to enforce a quantitative analysis of issues that are often subjective. It was developed in the latter part of the 20th Century by Professor Thomas Saaty of the University of Pittsburg [17]. As well as the visual aid provided by the process, there have been significant quantitative analyses developed for its implementation. A simple hierarchy is provided in Figure 5, with hypothetical numerical values inserted. This tool is an excellent example of the utility of assigning numerical values to otherwise subjective parameters. Mutual agreement and understanding should not be obviated by fear of imprecision.

Design Structure Matrix – identifying subsystem connections:
A Design Structure Matrix (DSM) is an interesting way to get a grasp of how the various design elements interact. First, the interactions of the various entities are plotted on a flow chart to identify all connectivity. The classic way to view this in a coherent manner is to plot major sub-elements of the overall effort down the vertical axis, and then repeat the same along the horizontal axis. This produces a two dimensional graph, Then, the team goes through the matrix and, if the direction of the interaction is from the entity on the vertical axis to the entity on the horizontal axis, an "X" is

place in the intersecting cell in the bottom/left half, and if the flow is the other way, an "X" is placed in the cell in the upper/right half as is seen in Figure 6 [18]. Even just the identification of these interconnections is helpful on several fronts. It helps define when careful interface specifications are needed, it assists in creating a valid documentation of record for the design, and it is useful for future analyses. With this matrix, any behavior will not be seen *in vacuo*, but will be taken within the interface of all of the relevant subsystems.



		1	2	3	4	5	6	7	8	9	10	11	12	13
Customer Requirements	1	■		X	X	X								
Wheel Torque	2		■			X	X		X		X			X
Pedal Mech. Advantage	3			■		X		X		X		X		X
System Level Parameters	4		X	X	■	X		X	X	X	X	X		X
Rotor Diameter	5			X		■		X				X		X
ABS Modular Display	6						■							
Front Lining Coef. of Friction	7					X		■						X
Piston-Rear Size	8			X		X		X	■		X	X		X
Caliper Compliance	9						X			■				
Piston- Front Size	10			X		X		X	X	X	■	X		X
Rear Lining Coef of Friction	11					X						■		X
Booster - Max. Stroke	12												■	
Booster Reaction Ratio	13			X		X		X		X		X	X	■

Figure 6 - DSM of Automobile Brakes

3. Research

The authors have cited the results of System Engineering research and this section is designed to outline future research goals and suggest some areas in the DoD simulation that would benefit from further work. The use of simulations and emerging technologies like artificial intelligence are recognized, but not explicated due to the paper's space constraints.

3.1. Need a New Look at Metrics

System Engineering has several tools to offer in improving match between training and warfighter performance. Institutions such as the Naval Postgraduate School, the Army War College and the Air Force Institute of Technology are readymade receptacles for this type of Operations Research. What does make a difference? Where can this be assessed? It is the authors; experience and history confirms that peace-time military service often leads to mistaken notions of combat readiness. Successful skill sets in a peace-time armed force is often tragically divergent from that required in a major war. This seems to be true even if there is low-level fighting during the time of peace [19]. It might be useful to promulgate a general request for studies in the area at all of the mentioned institutions. Many of these institutions welcome co-advisors on Master's Theses, which would allow researchers in the area of interest to readers of this paper to "Sheppard" mid-grade officers through studies to help quantify and characterize necessary performance capabilities of warfighters of all ranks and grades.

3.2. Emerging Technologies

There are a plethora of emerging technologies that will help furthering the study hereinabove referenced. Simulations have approached a sophistication level to the point we may soon be able to invoke them as substitutes for human time, which is becoming increasingly dear due overwhelming operations tempos and constraining budgets.

Advances are currently outpacing our ability to use them to the best advantage of the warfighters. Some of these are the ability to make better use of large amounts of data, the new-found capacities in Artificial Intelligence, the advent of Quantum Computing's orders of magnitude ability to do probabilistic computations in areas that do not require a deterministic response, and the prenascent ability for computers to provide a human-like interface that never tires or gets sick.

4. Conclusions

4.1. Standards are Needed

New technologies bring both new capabilities and new ways for people to misunderstand each other. There is a putative need for standard language to be promulgated and adopted. This is even more important in the DoD, as failures here lead to the loss of life and failure of missions rather than the loss of money and the failure of products. The Standards community needs to both respond when needed and to evangelize the need for early consideration of standards of all kinds in the defense community.

4.2. Quantification is Vital

We cannot know where we were, where we are and where we are going by relying on imprecise language. The stakes are too high to expect and hope that everyone knows the minds of everyone else. To optimize something, we need to know its numeric value. Even human emotion can be assessed by physical parameters such as eye dilation, heart rate, *etc*. Again, eternal vigilance is require as it is all too easy to slip back in the habit assuming others know what 'pretty soon", "a little bit", and "not too much" mean.

5. Acknowledgements

Virtually all of the work involving large scale simulations and virtual humans reported in this paper came from a series of research projects of which the Principal Investigators were Dr. Robert F. Lucas, Director, Computational Systems and Technology Division, Director, Computational Sciences Group, Research Associate Professor, Computer Science Department and Dr. Benjamin D. Nye, Director of Learning Sciences at ICT. Without his technical vision and unflagging support, none of this would have taken place. The various projects mentioned were supported by a range of DoD organizations and, while this paper reports insights gained during that work, that was not the focus of the research. Much of this work was supported by grants from the Office of Naval Research's STEM Program (ONR N00014-16-1-2820) and by students supported by the National Science Foundation Research Experience for Undergraduates program (NSF 1560426). Nevertheless, the positions taken in the paper are the authors' own and do not represent in any way the views of the Department of Defense or the US Government.

6. References

[1] Ceranowicz, A., Torpey, M., Hellfinstine, W., Evans, J., & Hines, J. "Reflections on building the joint experimental federation." In *Proceedings of the 2003 Interservice/Industry Training, Simulation, and Education Conference*. December. 2002.

[2] Horne, G. E. "Maneuver Warfare Distillations: Essence not Verisimilitude." In *WSC'99. 1999 Winter Simulation Conference Proceedings*. Vol. 2. pp. 1147-1151. IEEE. December, 1999.

[3] Ceranowicz, A., Dehncke, R., Cerri, T., & Suffolk, V. A. "Moving toward a Distributed Continuous Experimentation Environment." In *Proceedings of the 2003 Interservice/Industry Training, Simulation, and Education Conference*. pp. 1-11. 2003.

[4] Kraus-Maffei Wegmann (KMW), *Turret Trainer*, retrieved from internet on 26 December 2014, from: http://www.kmweg.com/home/training-simulation/gunnery-and-combat-training/turret-trainer/product-information.html. 2014.

[5] Miller, D. C., & Thorpe, J. A. "SIMNET: The Advent of Simulator Networking." *Proceedings of the IEEE*, Vol 83(8). pp 1114-1123. 1995.

[6] Messina, P.C., Brunett, S., Davis, Gottschalk, D. M., Curkendall, D. T., & Seigel, H., "Distributed Interactive Simulation for Synthetic Forces", in the *Proceedings of the 11th International Parallel Processing Symposium*, Geneva, Switzerland. 1997.

[7] Lucas, R.F., & Davis, D.M., "Joint Experimentation on Scalable Parallel Processors". In the *Proceedings of the Interservice/Industry Simulation, Training and Education Conference*, Orlando, Florida. 2003.

[8] K.T. Yao, , Lucas, R. F., Ward, C. E., & Wagenbreth, G., "Data Analysis for Massively Distributed Simulations", In the *Proceedings of the Interservice/Industry Simulation, Training and Education Conference*, Orlando, Florida, 2009.

[9] Department of Defense. “DoD Modeling and Simulation Glossary.” *Modeling and Simulation Coordination Office*. 2011.

[10] Department of Defense. “DoD Modeling and Simulation Glossary.” *Modeling and Simulation Coordination Office*. pp, 119, 159 & 85. 2011.

[11] Department of Defense, Military Handbook. MIL-HDBK-29612-4A, *Glossary for Training*. p 130 Washington DC. 2001.

[12] Office of National Intelligence, ODNI. *U.S. Intelligence –An Overview 2011*. Washington, DC. 2011/

[13] Department of Defense, Military Handbook. MIL-HDBK-29612-4A, *Glossary for Training*. p 43 Washington DC. 2001

[14] Somers, Lucas R., "Major General Lew Wallace at Shiloh" *The Student Researcher: A Phi Alpha Theta Publication*. Paper 6. http://digitalcommons.wku.edu/stu_res_jour/6. p 117. (2014).

[15] Lammers, R. L., Davenport, M., Korley, F., Griswold-Theodorson, S., Fitch, M. T., Narang, A. T., ... & Hamann, C. J. Teaching and Assessing Procedural Skills using Simulation: Metrics and Methodology. *Academic Emergency Medicine*, Vol 15(11). pp 1079-1087. 2008.

[16] Tetlock, P. E., & Gardner, D. *Superforecasting: The Art and Science of Prediction*. Random House. 2016.

[17] Saaty, T. L. (1980). *The Analytic Hierarchy Process: Planning, Priority Setting, Resources Allocation*, 281. New York: McGraw,

[18] DMSO, *The Design Matrix Tool and its Use*. Retrieved from the Internet on 04 December, from URL: https://dsmweb.org/tutorial-overview/

[19] Scales, R. H. *Artillery in Small Wars: the Evolution of British Artillery Doctrine, 1860-1914.* (Doctoral dissertation, Duke University). 1976.

Authors’ Biographies

DAN M. DAVIS is now a consultant for the University of Southern California, focusing on large-scale distributed DoD training, education and avatar mentors. In the decade prior to his retirement, he was the Director of USC’s JESPP project for JFCOM. As the Assistant Director of the Center for Advanced Computing Research at Caltech, he managed Synthetic Forces Express, bringing HPC to DoD simulations. Prior experience includes serving as a Director at the Maui High Performance Computing Center and as a Software Engineer at the Jet Propulsion Laboratory and Martin Marietta. He has served as the Chairman of the Coalition of Academic Supercomputing Centers and has taught at the undergraduate and graduate levels. As early as 1971, Dan was writing programs in FORTRAN on one of Seymour Cray’s CDC 6500’s. He saw duty in Vietnam as a USMC Cryptologist and retired from the reserves as a Commander, Cryptologic Specialty, U.S.N. He received B.A. and J.D. Degrees from the University of Colorado in Boulder.

SKANDER GUIZANI is studying Electrical Engineering and currently focuses his research efforts on hardware, infrastructure, and operations of long-range secure communications systems. Skander is from Tunisia and is a foreign exchange student at the United States Military Academy at West Point, New York. Back in Tunisia, his father is a Colonel and commands a unit in that country. While at the Institute for Creative Technologies, Skander will be working with the MentorPal team to enhance the virtual human conversations for advising high school seniors. He intends to pursue a doctoral degree and is focusing his research on that goal. He anticipates receiving a degree in Electrical Engineering in the spring of 2022 from the United States Military Academy, West Point.

EVAN JAKSHA is studying Computer Science and currently focusing in the use of virtual humans as an effective interface to address a range of critical issues in the military. He comes from a military family with both his father and two siblings currently serving in the US Armed forces. His research interests include cyber warfare and national security. Growing up in San Diego, he is a long-time resident in Southern California. He is scheduled to graduate in 2022 with a degree in Computer Science, with a Minor in Cyber Security from the United States Military Academy in West Point, New York.

DANIEL P. BURNS is a lifelong Systems Engineer, first with the Active Duty Navy, then SAIC, and small business. He served as Naval Chair and Professor of Practice in Systems Engineering at the Naval Postgraduate School (NPS).Captain Burns served as the as the Military Associate Dean and as acting Dean of the Graduate School of Engineering and Applied Sciences at NPS. His research interests center on analyses of both human and resource utilization in defense efforts. He successfully facilitated \ the creation of a new program for Air Force Officers who seek post-graduate degrees. Captain Burns received a BS degree from the U.S. Naval Academy, an MS from the Naval Postgraduate School and an MS from Southern Methodist University. He is currently working with Portland State University on a Ph.D.

MARK C. DAVIS, PH.D. is currently retired after careers in the US Navy and as a computer design engineer for both IBM and Lenovo. Rising to the level of Distinguished Engineer at Lenovo, he was responsible for the design of laptop computer cross-disciplinary technology, including PC architecture, embedded systems, open source and virtualization. Previous work was with IBM in the areas of software development and architecture involving security, storage and virtualization. Dr. Davis has been granted well over fifty patents that were filed during his service at both companies. He is a graduate of the Duke University NROTC program and was commissioned as an Ensign, attended nuclear power school, and served as a Submarine Officer for twelve years, including one duty tour as a classroom instructor. He left the service as a Lieutenant Commander to pursue a PhD. Mark holds a BSEE degree from Duke University and a PhD in Computer Science from the University of North Carolina, where his advisor was Professor Fredrick P. Books.