VALIDATING SCENARIO-BASED TRAINING SEQUENCING: THE SCENARIO COMPLEXITY TOOL

Robb Dunne, PhD
Innovative Reasoning LLC
Orlando, FL
Robert.Dunn@innovativereasoning.com

Stephen A. Sivo, PhD
University of Central Florida
Orlando, FL
Stephen.Sivo@ucf.edu

Nathan Jones
MCSC PM TRASYS
Orlando, FL
Nathan.Jones1@usmc.mil

ABSTRACT

Effective and efficient Scenario-Based Training (SBT) is sequenced using well-grounded instructional strategies and learning theory. The primary instructional strategy employed by the Military requires that SBT is sequenced in a "crawl-walk-run" trajectory. For software to sequence scenarios effectively and efficiently in this manner, SBT needs objective, computational values of a scenario's complexity, but designers, software engineers and trainers operate without the necessary tools to objectively calculate Scenario Complexity (SC). This results in subjectively sequenced SBT that may be ineffective, inefficient, or designed without attention to sound instructional practices.

To address this issue, research in education, task complexity, task framework and cognitive resource principles was integrated and an innovative SC tool (patent pending) comprised of an algorithm and supporting process, was developed to objectively and computationally define SC. This paper presents findings from the use of the SC tool to validate a training matrix embedded in the United States Marine Corps' M1A1 Advanced Gunnery Training System.

To establish that the SC tool is accurate and effective, it was first necessary to determine how consistent the Subject Matter Expert (SME) evaluations of the scenario's characteristics were. Then, using the results of their input to the SC algorithm, determine how well the SME sequencing matched that of the training matrix. The objective was to use the SC tool to verify and validate the "crawl-walk-run" sequencing of the training matrix and identify any areas in need of adjustment.

After employing the SC tool, quantitative analyses showed that the SMEs were very consistent in their formulations. Importantly, the SC tool revealed that the training matrix deviates alarmingly from "crawl-walk-run" sequencing. This paper also presents the study's methodology and algorithm, lessons learned and the future impact that this innovative SC tool may have upon design, development and evaluation of SBT and automated, adaptive training.

ABOUT THE AUTHORS

Robb Dunne, PhD holds a PhD in Instructional Technology from University of Central Florida (UCF) and is the Instructional Systems Specialist for the Combat Vehicle Training Systems Program under USMC Systems Command (MCSC) Program Manager Training Systems (PM TRASYS). Dr. Dunne spent 2 years at the Institute for Simulation and Training in the Applied Cognition and Training in Immersive Virtual Environment lab where he was lead investigator and author of two published papers focused on scenario complexity.

Stephen Sivo, PhD is a Professor of Methodology, Measurement and Analysis at UCF. He serves as the Coordinator of Methodology, Measurement and Analysis program overseeing the PhD program track, the Advanced Quantitative Methods certificate for PhDs, and the scheduling of all Masters and Doctoral level Research Methods courses in the College of Education at UCF.

Nathan Jones is the Functional Lead for Instructional Systems at Marine Corps Systems Command PM TRASYS. Mr. Jones has over 15 years of experience in human performance assessments and acquisition program support. He is responsible for overseeing PM TRASYS front end analyses, verification, validation and accreditations, training effectiveness evaluations and other training domain expertise for PM TRASYS.

VALIDATING SCENARIO-BASED TRAINING SEQUENCING: THE SCENARIO COMPLEXITY TOOL

Robb Dunne, PhD	Stephen A. Sivo, PhD	Nathan Jones
Innovative Reasoning LLC	University of Central Florida	MCSC PM TRASYS
Orlando, FL	Orlando, FL	Orlando, FL
Robert.Dunn@innovativereasoning.com	Stephen.Sivo@ucf.edu	Nathan.Jones1@usmc.mil

DEMAND SIGNAL

Scenario Based Training (SBT) simulators present their own problems and limitations. Many simulators require experienced and skilled operators to realize full training efficiency and effectiveness (Good, 2003; Bremner, Aduddell, Bennett, & Van Geest, 2006; Parker & Myrick, 2009). Currently, the Instructor/Operator (IO) of a United States Marine Corps (USMC) SBT system is often heavily tasked and cannot dedicate the cognitive resources necessary to accurately sequence scenarios. Typically, when the IO can dedicate their efforts and originate the scenario sequencing, bottlenecks are created in the training process (Zook et al., 2012). In the USMC, experienced, trained instructors are in constant turnover due to rotation, operations tempo and manpower requirements. Consequently, expensive equipment with notable training affordances is often under-utilized (PM TRASYS, 2014).

Currently, automated SBT matrices, intended to mitigate this condition by embedding “crawl, walk, run” training sequences in the simulators software, provide sequencing that is based on Instructional System Designer (ISD), software engineer, or IO subjective perceptions of difficulty. This subjectivity may lead to training that is not sequenced accurately, effectively or efficiently. To accurately sequence training exercises, automate and adapt to the performance of the trainee, software must computationally define Scenario Complexity (SC). That is, an objective, computationally actionable value must be attributed to each scenario so that advancement or remediation along the trajectory can be grounded in well-researched learning principles and sound instructional strategy.

OVERVIEW

This paper presents results of an effort to employ an innovative evaluation tool for validating SBT sequencing. This validation tool uses a process and an associated algorithm to objectively define a scenario’s level of complexity. Results of this effort will assist SBT designers and engineers in their development, delivery and evaluation of automated SBT exercise matrices. This process and algorithm is hereafter referred to as the SC tool (patent pending).

The SC tool’s algorithm was previously tested and satisfied conditions of a proof of concept (Dunne, Schatz et al., 2010a, 2010b). Based on the outcome of that research adjustments were made to the original algorithm. These adjustments enabled easier computation and better accounted for cognitive factors and the influence of the scenario’s task framework (Campbell, 1988) on the total SC value. Although tailorable to other SBT environments (games, computer based instruction) this SC tool was customized for Military application.

In the following sections, brief definitions of the central concepts and the algorithm’s components are given followed by an elaboration on the instructional foundation upon which effective SBT sequencing is grounded. The study’s methodology is then laid out with results, discussion and implications presented afterwards. Lessons learned and proposed ways forward conclude the paper.

Scenario-Based Training

SBT emphasizes learning by doing (Reigeluth, 1999) and employs real-world problems that focus on performance outcomes in the context of the real work environment (Kindley, 2002). “Train as you will fight,” is one of the fundamental principles upon which USMC training is based (USMC, 1996) and SBT is an ideal vehicle for training within the USMC as SBT includes real-world problems like the pressures and “fog of war”.

Adaptive Automation

Adaptive automation refers to technology that can modify and adjust delivery of content and operations dynamically. In the case of SBT, sequencing is not codified during adaptive systems processes; human integration and machine agents of the system are dynamic during the operation of the system (Parasuraman, 2003).

Adaptive computer-based tutoring has been pursued for almost 50 years (Hartley, 1973; Kinshuk, Patel, & Scott, 2001). Adaptive technology represents the next step in the evolution of automation (Scerbo, 1996) and can sustain an interactive learning and training environment with an automated, adaptive tutoring system at the core (Kenny, 2006). Levels of automation can differ in type and human-system integration, from organizing information sources to suggesting decision options or even carrying out the necessary actions (Parasuraman & Sheridan, 2000). Although automation can objectively and efficiently carry out what is typically a subjective and time-consuming task it does not necessarily eliminate the human role (Parasuraman & Riley, 1997).

Scenario Complexity

SC is defined as the objective quality of a scenario which interacts with individual characteristics (such as trainees' expertise) to yield an individual's perception of the scenario's difficulty (Lum et al., 2008). A literature review was conducted to seek out research defining complexity, task complexity, SBT difficulty and other related terms; the combination of three characteristics offered promise. SC is calculated based upon three scenario characteristics that are *extrinsic from* rather than *intrinsic to* trainees: Task Complexity (TC), Task Framework (TF) and Cognitive Context Moderators (CCM). In turn, these characteristics are comprised of base variables that establish the foundation of the SC tool's algorithm expressed as:

$$SC = (TC \times CCM) + TF \qquad (1)$$

Instructional Foundation

A good instructional progression requires each succeeding task is more complex than those preceding (Merrill, 2007). This concept, the "crawl-walk-run" instructional strategy of increasing difficulty to optimize learning acquisition, is supported by two learning theories: Vygotsky's Zone of Proximal Development (ZPD) (1978) and Krashen's, *i*+1 (1982).

ZPD represents a phase in development that Vygotsky describes as "the distance between the actual development level as determined by independent problem solving and the level of potential development as determined through problem solving under guidance or in collaboration with more capable peers." (Vygotsky, 1978 p. 86).

Krashen describes a learner's zone of current understanding or comprehension (*i*) and the step above but not beyond understanding or comprehension (+1). Krashen states, "the input they [caretakers] provide for children includes *i* + 1, but also includes many structures that have already been acquired, plus some that have not (i + *2, i* + 3, etc.) and that the child may not be ready for" (Krashen, 1982, p. 23).

Algorithm Characteristics

The SC tool's algorithm is comprised of three characteristics (TC, TF and CCM) and associated base variables. It is from the values tied to these variables and established by input from Subject Matter Experts (SME) that the outcome of the SC algorithm is determined.

Task Complexity

The literature review found several accepted definitions of TC (Kohn & Schooler, 1978; Wood, 1986; Frese, 1989; see Campbell, 1988 for a review). However, Dunne, Schatz, Fiore, Martin et al. (2010) and Dunne, Schatz, Fiore, Nicholson et al. (2010) used Wood's 1986 definition of TC. Wood asserts TC is operationalized and manipulated by increasing or decreasing the number of sub-characteristics present in the task itself and that TC is the sum of two sub-characteristics: 1) component complexity and 2) coordinative complexity. Thus TC accounts for the number of tasks, learning objectives and required acts that may be necessary for successful task completion (Wood, 1986). Tasks may stand alone or have accompanying subtasks. Component Complexity (TC_1)**,** accounts for each task and subtask as well as those that may require monitoring of a number of information cues. An information cue is a discrete source of information that must be monitored and/or processed from the environment (Dunne et al., 2010). The second TC sub-characteristic, Coordinative Complexity (TC_2) accounts for the number of subtasks which are interdependent upon the successful completion of another subtask. To address computational constraints, TC_2 is calculated based on the number of interdependent subtasks and is either an exponential or multiplicative function depending on the amount of information cues. This characteristic is expressed as:

$$TC = TC_1 \text{ x } TC_2 \qquad (2)$$

or, broken down to base variable:

$$TC = (TC_r + TC_{ic}) \text{ x } (TC_s \text{ x } TC_{is}) \qquad (3)$$

or

$$TC = (TC_r + TC_c)^{(TCs \text{ x } TCis)} \qquad (4)$$

Where TC_1 = the sum of the number of required acts (*r*) in each task and number information cues (*ic*); and TC_2 = the number of subtasks (*s*) multiplied by the number of interdependent subtasks (*is*) when (*is*) $\leq$ 10 or TC_2 = the number of (*s*) raised by the number of (*is*) when (*is*) > 10.

Task Framework

Second is the TF characteristic and expresses whether a task is well- or ill-defined by determining the number of task paths and task outcomes. TF accounts for the relation between task paths and the outcome associated with each, and addresses which outcomes are possible in a given task (Campbell, 1991). Campbell (1988) proposes four defining characteristics of task framework:

1. The presence of multiple potential ways (i.e., paths) to arrive at a desired end-state.
2. The presence of multiple desired outcomes (i.e., end-states) to be attained.
3. The presence of conflicting interdependence among paths to multiple outcomes.
4. The presence of uncertain or probabilistic links among paths and outcomes.

A task path is the number of possible ways to arrive at the desired objective. Having numerous ways to reach the desired objective decreases the complexity of a task unless, as is often the case, the presence of multiple correct paths is illusionary or an efficiency criterion is embedded in the task. Task outcome is defined as a criterion that must be satisfied in order to reach the desired objective; again, to use the targeting example, the objective may require that the Gunner (GNR) identify target, choose ammunition and fire within a certain time frame *and* successfully overcome the target. Conflicting outcomes are those where successful achievement of one criterion can conflict with achieving another. For example: to arrive quickly at a Battle Position (BP) without intruding on infantry Area of Operations may require taking a long, less efficient route to the BP which is necessary to avoid interruption of friendly activity. Uncertain linkages accounts for increase in cognitive processing as a result of ambiguity, which increases the task's complexity.

The algorithm for this characteristic is:

$$TF = p^u + o \qquad (5)$$

or

$$TF = p^* u + o \qquad (6)$$

Where TF = the number of task paths (p) raised by the degree of uncertainty in the paths (u) plus the number of task criteria (o) while u $\leq$ 10; the number of task paths (p) multiplied by the degree of uncertainty (u) in the paths while u > 10 plus the number of task criteria (o).

Cognitive Context Moderators

The third characteristic, CCM, are external stimuli that affect the operator by increasing load and reducing cognitive resources for the task, thus causing less complex decisions to appear more complex (Dunne et al., 2010) and requiring careful sequencing to avoid overloading and reducing training quality. Dunne et al., (2010) suggest it is this characteristic where the interplay of decision preparedness, performance and complexity is most acute. Tasks with high CCM likely affect trainee ability to build and maintain situation awareness if the complexity exceeds available working memory processing and storage resources (Jodlowski, 2008).

The cognitive task analysis body of literature suggests that cognitive task analyses yield information about the unobservable thought processes that underlie observable task performance (e.g., Schraagen, Chipman, & Shute, 2000). Further, researchers already attempt to articulate the component parts of cognitive tasks. For instance, Campbell (1991) offers articulated details for how to conceptualize decision, judgment, problem, and fuzzy tasks. Researchers such as Skehan and Foster (2001) as well as Robinson (2003) propose the manipulation of a series of cognitive task design factors to achieve different levels of task complexity, this supports the inclusion of the CCM variable in the SC formula. These factors inform two cognitive models of task complexity: Skehan and Foster's Limited Attentional Capacity Model (2001), and Robinson's Cognition Hypothesis (2003).

Dunne, Schatz, Fiore, Martin, et al. (2010) noted that care must be taken to discern the difference between external stimuli and subtasks. Driving an M1A1 is moderated by unreliable terrain, but communicating with the TC through the Combat Vehicle Communications device is a subtask and not a cognitive context moderator unless, as previously mentioned, the communication is unnecessary to the performance of the task or superfluous.

It is necessary to add 1 to the range of values since the exponential component of an equation cannot equal 0. For the purpose of this study values where derived from three conditions: no distraction, very high distraction, and somewhere in between. Thus values from zero to three were assigned from least to most distracting. The equation for this characteristic is:

$$CCM = \{0,\ 1,\ \text{and}\ 2\} + 1 \qquad (7)$$

METHODOLOGY

The purpose of this research was to validate the sequencing of an embedded SBT exercise matrix by employing a tool Educators or Trainers, ISD and software Engineers could use to objectively and computationally define SC. Part of this process included a period of instruction given to the SMEs in the identification of the base variables that comprise the SC algorithm. Using this knowledge, SMEs then determined the values of those SC base variables. These base variables, once calculated, determined the total SC values which were then ranked and compared to the SBT exercise matrix embedded in the M1A1 Advanced Gunnery Training System (AGTS) simulator. Figure 1 is a representation of the SBT exercise matrix.

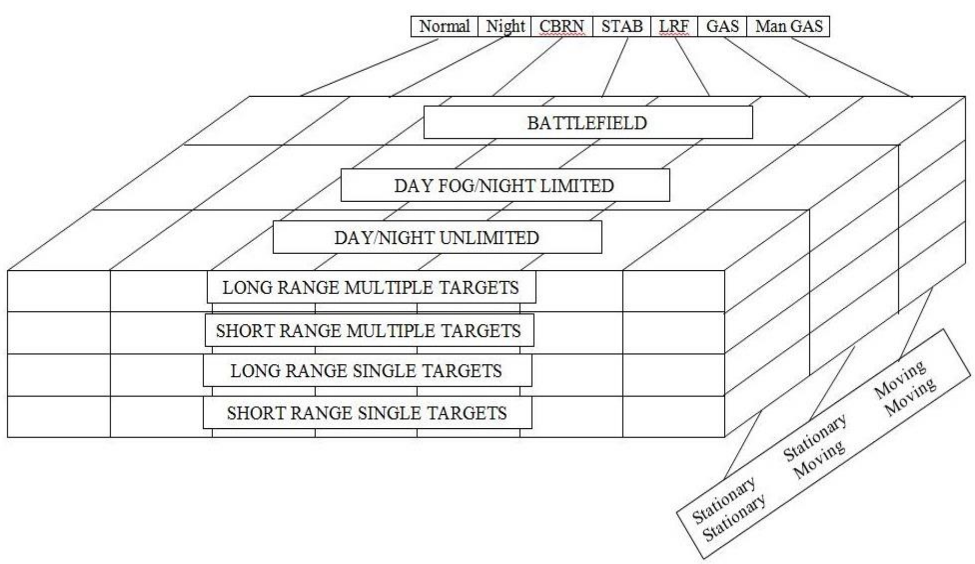

Figure 1. Training Matrix

The SBT exercise matrix sequence is intended to guide a trainee from scenarios with a stationary Own Vehicle (OV) against a stationary, short range, single target under normal and unlimited day conditions (crawl) to a moving OV against moving, long range, multiple targets under battlefield conditions where manual, battlesight gunnery techniques are necessary (run).

The data collection and research procedure for this study followed three distinct steps: demographics, period of instruction and evaluation phases.

Table 1. Research Design

Demographics	Period of Instruction	Evaluation Phases
Protocol was read and participants were given paper, pen or pencil, and Demographics Survey.	Participants were given instruction on definitions and identification of SC characteristics and instructional "cheat-sheet" -an outline of the SC tool for easy identification of base variables for reference during evaluation phases.	Each participants was given three phases with 10 scenarios in each and calculated a total of 30 SC characteristic values.

Limitations and Constraints

At the time of this research there were only 202 USMC M1A1 active duty and reserve unit tank crews. From this pool only a few are chosen to be I/Os and fewer are chosen to be Senior I/Os (SI/Os). Although the M1A1 AGTS is also employed by the US Army and the Saudi Arabian National Guard there may be fewer than 100 certified M1A1 AGTS I/O and SI/O worldwide. Whereas a larger research group would have benefitted the determination of generalizability, it was not possible to gather a larger contingent. Although this may be a limitation and constraint to human subject research, the employed statistical regime focused on the algorithm's items and not the participants and thus mitigated this concern.

Participants

This study was conducted with participants (n = 5) from the USMC 2nd Tank Battalion stationed at Camp Lejeune, NC. These participants were I/O and SI/O that supervise and deliver M1A1 AGTS training, combat veterans and well-grounded in the training requirements of the M1A1 crew. Each participant provided a total of 260 raw data points resulting in a robust total of 1,300 raw data points.

Research Questions and Statistical Analysis

The following research questions were asked:

Q_1: How consistent were the SME ratings of the items?
Q_2: How well do the SME rankings correlated to the SBT exercise matrix?

To investigate Q_1, coefficients derived from implementation of Generalizability (G) Theory (Brennen, 1992) were attained from each of the SMEs' input from the three evaluation phases. The sequences of the SME scenarios were then compared to each other to determine how consistent they were among themselves.

G Theory is a statistical framework for conceptualizing, investigating, and designing reliable observations. It is used to determine the reliability (reproducibility) of measurements under specific conditions (Cronbach, Nageswari, & Gleser, 1963). By using G Theory, individual formula characteristics, or base variables, can be examined to determine their level of error or importance. That is, G Theory reveals if some factors in the SC algorithm are extraneous or more important than others by illustrating levels of reliability. If low levels are indicated then those factors may need to be eliminated or adjusted or their influences on the algorithm's outcome reduced. If high levels are indicated then those factors may be considered validated and well-constructed.

An often overlooked consideration, in regards to degree of reliability, is the level of acceptability when interpreting results. Nunnally (1978) recommends that instruments used in basic research have ≥ .70 reliability but increasing reliabilities beyond .80 is unnecessary. However when the research is high consequence, as in the case of research involving warfighter training, reliability should be at least .90, preferably .95 or better (Nunnally, 1978). Therefore, only reliability levels ≥ .90 were considered acceptable for this research.

To investigate Q_2 both Spearman's rho and Kendall's tau *b* correlations were conducted on the total SC values derived from each of the SMEs scenario inputs. Each of the SME's SC values were then sequenced from least complex (crawl) to most complex (run) and compared to the embedded SBT exercise matrix. This determined the correlations between the SMEs SC rankings and the training matrix's scenario difficulty rankings. The closer the correlation between the SME's rankings and the training matrix's the more aligned the matrix is to the doctrinal crawl, walk, run instructional strategy and supporting instructional theories.

RESULTS

Results were analyzed from the total SC down to the base variable factors (sub-tasks, task paths, etc.) in order to identify and isolate the influence of each of the contributing individual components. This section delivers the wave-top results with discussion and implications of these results presented in later sections.

Q1: How Consistent Were the SME's Ratings of the Items?

Scenario Characteristics

G coefficient analyses were conducted for each of the SC characteristics. The data for the analyses was derived from the base variables' values provided by the SMEs and calculated by the algorithm. The results were mixed.

Results of the G coefficients from the first of the scenario characteristics, TF, were very good (.923) in the first phase but dropped to poor (.716) and (.625) in the second and third phase respectively.

Analysis of the G coefficients for the second characteristic, TC, showed a poor level (.625) in the first phase, very good (.902) in the second, but then dropped to very poor (.005) in the third phase.

Analysis of the CCM G coefficients from the first phase was stopped due to statistical parameters; the range of possible responses was too restricted to render determination of consistency. That is, where the range of responses is restricted and the responses almost identical, no reliable or valid determination of consistency is available. Enough variance was present in the second phase to indicate poor (.490) and excellent (1.00) in the third phase.

Table 2. Summary of Generalizability Coefficients by Evaluation Phase

Evaluation Phase	1 G coefficient	2 G coefficient	3 G coefficient
Total Scenario Complexity	.625	.872	.005
Scenario Characteristics			
Cognitive Context Moderators	*	.490	1.000
Task Framework	.923	.716	.625
Tack Complexity	.625	.902	.005
Scenario Sub-Characteristics			
Component Complexity	.989	.990	.996
Coordinative Complexity	.625	.907	.623
Base Variables			
Acts	.981	.987	.998
Cues	.966	.977	.990
SubTasks	.986	.939	.998
InterDependentTasks	.982	.944	.998
Task Paths	.948	.917	1.000
Task Criteria	1.000	.928	1.000
Unknown/Conflicting Paths	1.000	.920	.998
Distractors	*	.363	1.000

* Responses were too restricted in range to allow determination of variance

Scenario Sub-Characteristics

The G coefficients for the TC sub-characteristics (TC_1 and TC_2), derived from their associated base variables, were revealed to be, in the case of Component Complexity, very good in all three phases (.989), (.990) and (.996) respectively. Coordinative Complexity G coefficients showed as poor in first phase (.625) very good in the second (.907) and poor again (.623) in the third and last phase.

Base Variables

Each of the base variable G coefficients were calculated and, in order of phase, were shown to be very good for all three phases of the required acts base variable (.981, .987 and .998), very good for all three phases of the information cues base variable (.966, .977 and .990) and very good for all three phases of the sub-tasks base variable (.986, .939, .998). Each of the phases of interdependent tasks were also very good (.982, .944, .998) and task paths continued the trend with very good G coefficients for all phases (.948, .917, and 1.00). Task criteria yielded excellent and very good results (1.00, .928 and 1.00) as did unknown/conflicting paths (1.00, .920, and .998). The CCM G coefficients varied from the first phase, where execution of the statistical analysis was stopped by software parameters, to a second phase poor (.363) and final phase excellent (1.00).

Table 3. SME to SME Correlations

Phase 1										
SME	**1**		**2**		**3**		**4**		**5**	
Stat	**tau b**	**rho**	**tau b**	**rho**	**tau b**	**rho**	**tau b**	**rho**	**tau b**	**rho**
1	1.000	1.000	.314	.352	.103	.104	.922	.952	.946	.967
2	.314	.352	1.000	1.000	.759	.783	.338	.404	.396	.417
3	.103	.104	.759	.783	1.000	1.000	.095	.087	.196	.201
4	.922	.952	.338	.404	.095	.087	1.000	1.000	.923	.952
5	.946	.967	.396	.417	.196	.201	.923	.952	1.000	1.000
Phase 2										
SME	**1**		**2**		**3**		**4**		**5**	
Stat	**tau b**	**rho**	**tau b**	**rho**	**tau b**	**Stat**	**tau b**	**rho**	**tau b**	**rho**
1	1.000	1.000	.922	.928	.712	.762	.937	.971	.661	.771
2	.922	.928	1.000	1.000	.800	.874	.864	.940	.747	.829
3	.712	.762	.800	.874	1.000	1.000	.667	.765	.555	.655
4	.937	.971	.864	.940	.667	.765	1.000	1.000	.548	.718
5	.661	.771	.747	.829	.555	.655	.548	.718	1.000	1.000
Phase 3										
SME	**1**		**2**		**3**		**4**		**5**	
Stat	**tau b**	**rho**	**tau b**	**rho**	**tau b**	**Stat**	**tau b**	**rho**	**tau b**	**rho**
1	1.000	1.000	.614	.634	.614	.634	.854	.948	.854	.948
2	.614	.634	1.000	1.000	1.000	1.000	.584	.450	.584	.450
3	.614	.634	1.000	1.000	1.000	1.000	.584	.450	.584	.450
4	.854	.948	.584	.450	.584	.450	1.000	1.000	1.000	1.000
5	.854	.948	.584	.450	.584	.450	1.000	1.000	1.000	1.000

In evaluation phase I SME #1 had significant correlations at the .01 level to SMEs #4 and #5 (.922 to .967). This means that over 90% of the time, these SMEs agreed on the sequencing of the scenarios. SMEs #4 and #5 also had significant correlations at the .01 level (.923 to .952). SMEs #1 and #2 had a significant correlation at the .01 level (.922 to .928).

In evaluation phase II SME #1 had a significant correlation at the .01 level to SMEs #2 and #4 (.928 to .971) and a significant correlation at the .05 level to SME 5 (.661). This means that over 90% of the time, SMEs #1, #2 and #4 agreed and over 60% of the time SME #1 and #5 agreed to the sequencing of the scenarios. SME #2 had significant correlations at the .05 level to each of the other SMEs.

In evaluation phase III SME #1 had a significant correlation at the .01 level to SMEs #4 and #5 (.948) and a significant correlation at the .05 level to SME #2 (.634). This means that over 90% of the time, SMEs #1, #4 and #5 agreed and over 60% of the time SME #1 and #2 agreed to the sequencing of the scenarios. SMEs #2 and #3 had 100% agreement (1.00), SMEs #4 and #5 had 100% agreement (1.00).

Q2: How Well do the SMEs Rankings Match to the SBT Exercise Matrix?

Correlation of SME scenario complexity evaluations, were conducted in three stages; each phase's results were then correlated by SME to the AGTS exercise matrix sequencing.

Table 4. SME to Matrix Rankings Correlations

	Matrix Phase 1		Matrix Phase 2		Matrix Phase 3	
SME	**Kendall's tau *b***	**Spearman's rho**	**Kendall's tau *b***	**Spearman's rho**	**Kendall's tau *b***	**Spearman's rho**
1	.000	.045	-.398	-.500	.090	.140
2	-.159	-.232	-.424	-.549	-.135	-.164
3	.000	-.018	-.424	-.525	-.135	-.164
4	-.141	-.154	-.349	-.462	.200	-.285
5	-.097	-.075	-.432	-.524	.200	-.285

In evaluation phase I there was no significant correlation between SME rankings and matrix rankings in either Kendall's tau *b* or Spearman's rho results. This was also the case for the other two phases; the second phase shows the greatest disparity between SME #2 and matrix correlation (-.549).

DISCUSSION

Q1: How Consistent Were the SME's Ratings of the Items?

The results of the G study affirm that the SMEs were consistent in their rating of the items across scenario base variables and, in most cases, increased from phase to phase. However, trends also indicate that this consistency loses potency with successive algorithm functions. It is reasonable to expect that with such a strong, consistent foundation the upward results would be of equal consistency, but this was not the case. It is concluded that further modification of the SC tool and algorithm will yield more robust results.

For example, results of the first phase of the CCM characteristic indicated that the SMEs were unable to distinguish any difference in the level of distraction among any of the 10 scenarios. After examining the outputs it is concluded that the range of possible responses (1-3) is too restricted to allow reliable responses. A greater degree of granularity and a higher level of definition regarding the moderators that influence cognitive processing is necessary to allow a wider range of possible responses. Crew discipline and personnel levels of experience, environmental states and level of battle chatter and obstacles should increase the range of distractor value possibility.

There were two groups that were not mutually consistent. Consistency was greatest among SMEs #1, #2 and #5. SMEs #3 and #4 were more consistent between themselves. Demographics show that the ranks of SMEs #1, #2 and #5 were either Sergeant or Staff Sergeant with the primary crew position of Tank Commander. SMEs #3 and #4 were Lance Corporals with the primary crew position of Gunner. This suggests that the Commander, even when evaluating a gunnery-intensive situation, has a more complete or holistic appreciation of the details involved for successful completion of the task. Furthermore, SMEs #3 and #4 had one year or less AGTS IO experience. SME #4 also responded that he was only responsible for training once a month. It is therefore reasonable to assume that with greater experience, preferably in the position of Tank Commander, an IO or SIO would respond more consistently, providing responses similar to the #1, #2, #5 group findings.

Results also suggest that the TC_2 sub-characteristic outweighs the other factors and plays an undue role in the calculation of the final SC value.

Q2: How Well do the SME's Rankings Match to the SBT Exercise Matrix?

The SBT exercise matrix sequencing deviates significantly from and does not correlate to the SME sequencing derived from the SC tool. Some deviation was expected between the subjectively sequenced matrix, and the objectively sequenced scenarios determined by the SMEs. However, the degree of disagreement is alarming. In such a high consequence environment the "crawl, walk, run" sequencing strategy supported by Vygotsky and Krashen is vital. Results indicate this doctrinal strategy is severely violated.

IMPLICATIONS

Outcomes from this research indicate several areas of improvement in the SC tool; equation weighting needs to be reassessed. In the area of CCM, a more granular and less restricted range of possible values would remove statistical limitations, benefit the consistency and reliability of the equation and give CCM a larger and more accurate moderating role.

Proceduralizing the SC tool within a sequencing engine would allow automation to occur on three levels; the IO could initiate changes between automated and manual modes or functionality (i.e., an adaptable system); the IO and the system together could initiate changes (i.e., an adaptive system) (Scerbo et al., 2003) or only the system could initiates changes (i.e., an automated adaptive system).

The SC tool can be considered a partial training equation. That is, it is designed and intended to account only for the tasks that are conducted and experienced by a single trainee and does not integrate other trainee's tasks into the SC calculation. Extension of this algorithm to a larger, more complete training equation that incorporates a whole team is an avenue of future research. The value of an effective SC tool may be greater for collective and team training sequencing, and in a Live, Virtual, and Constructive (LVC) environments where training is even more complex.

Military and Other Training Environments

Fiscally restrained environments reduce availability of operational funds related to live training in general and gunnery qualification in particular. Cost avoidances that justify simulation will support increase in implementation of SBT systems. Automating SBT sequencing reduces work-load upon the IO and enables them to pay attention to detailed areas of instruction, increase trainee throughput, and more effective training.

Beyond the immediate implications of this research and realizing the stated initiatives of the Military to move toward LVC training environments, efforts in the near future should build on these findings and be directed toward a collective, rather than individual definition of SC. Extending this SC tool to account for the increase in complexity that occurs during multiple scenario integration is a likely downstream avenue of research for proper integration of primary, secondary, and tertiary audiences in LVC training scenarios.

While individual training may progress along a single-string trajectory, it is relatively simple when compared to progression along multiple-string trajectories encountered in collective training and LVC environments. An effective SC tool would provide high degrees of fidelity between simulated warfighter scenarios and the actual complexity of battlefield conditions.

Validation of this formulation's capability to discriminate between scenarios of differing complexity as well as its successful identification of appropriate scenarios for adaptation based on performance levels is another area of future research and experimentation.

LESSONS LEARNED AND CONCLUSION

It is possible that with successful extension of this research, complete automation of scenario sequencing could be achieved enabling effective and efficient training. The purpose of this research was to examine the efficacy of the SC tool to enable Educators or Trainers, ISD and software engineers to objectively and computationally define SC.

This research advances the field of SBT by suggesting a method of designing SBT that adapts and automatically initializes, constructs, and sequences scenarios based on trainee performance and grounded by researched instructional strategies and learning theory. The SC tool enabled an objective ranking of scenario values which was correlated to the matrix. More research is necessary but promising directions have been shown. Whether these future efforts are conducted by this researcher or others who take up the call the outcome will be additional contributions to the ISD, SBT and educational fields.

Also, as it was beyond the scope of the investigation, no attempt at a longitudinal design to determine if the lessons learned by the participants continued over time, or that they employed those skills in the design and sequencing of their own instruction. In addition, turning the experiment around and asking other SMEs to subjectively verify the sequence objectively determined by these SMEs was not conducted but would be a next step for future research.

The highly proceduralized tasks allowed for a high degree of agreement among participants, but this also made it possible that subjectivity played a less influential role here than it may in other circumstances and within other domains.

Due to the schedule and limited availability of these SMEs this research needed to be conducted within a compressed time frame which created a cognitive constraint. Each of the three phases lasted at least 45 minutes. Results from the third and last phase are, in some cases, less consistent than the others even though an upward trend had been previously indicated. Future research of this type should account for the extensive cognitive resource requirements.

The G Coefficient results implied that consideration must be given to the role experience plays when the IO sequences their exercise scenarios. While the given significant difference between SMEs #2 and #3 and SMEs #4 and #5 implied further development of the SC tool is needed to account for SME experience, it also draws attention to any assessment process reliant on SME expertise. Use of G coefficients or similar statistical regimes may need to be considered more often as part of validation of SME data.

ACKNOWLEDGEMENTS

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of PM TRASYS. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon. The authors wish to acknowledge the vital cooperation of the USMC Tank community that provided and granted use of the M1A1 AGTS to serve as the test-bed, and especially USMC Captain (Retired) Samuel Oliver of Innovative Reasoning LLC for his exemplary leadership and inspiring, relentless dedication.

REFERENCES

Bremner, M., Aduddell, K., Bennett, D., & Van Geest, J. (2006). The use of human patient simulators: Best practices with novice nursing students. *Nurse Educator, 31*, 170-174.

Campbell, D.J. (1988). Task complexity: A review and analysis. *Academy of Management Review, 13,* 40-52.

Campbell, D. (1991). Goal levels, complex tasks, and strategy development: A review and analysis. *Human Performance, 4,* 1-31.

Cronbach, L.J., Nageswari, R., & Gleser, G.C. (1963). Theory of generalizability: A liberation of reliability theory. *The British Journal of Statistical Psychology, 16,* 137-163.

Dunn, W.E. & Lantolf, J.P. (1998). Vygotsky's zone of proximal development and Krashen's 1+1: Incommensurable constructs; incommensurable theories. *Language Learning, 48,* 411-442.

Dunne, R., Schatz, S., Fiore, S., Martin, G., & Nicholson, D. (2010). Scenario-based training: Scenario complexity. *Proceedings of the 54th Annual Conference of the Human Factors and Ergonomics Society*. San Francisco, CA.

Dunne, R., Schatz, S., Fiore, S., Nicholson, D., & Fowlkes, J. (2010). Optimizing decision preparedness by adapting scenario complexity and automating scenario generation. *Proceedings of the ModSim International World Expo 2010*, Hampton,VA. http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20110012060_2011012533.pdf.

Evensen, L.S. (2007). 'With a little help from my friends'? Theory of learning in applied linguistics and SLA. *Journal of Applied Linguistics, 4,* 333-353.

Frese, M. (1989). Theoretical models of control and health. In S. L. Sauter, J. J. Hurrell, & C. L. Cooper (Eds.), *Job Control and Worker Health* (pp. 107-128). Chichester, UK: Wiley.

Good, M. (2003). Patient simulation for training basic and advanced clinical skills. *Medical Education, 37,* 14-21.

Hartley, J.R. (1973). The design and evaluation of an adaptive teaching system. *International Journal of Man-Machine Studies, 5,* 421-436.

Jodlowski, M.T. (2008). Extending long term working memory theory to dynamic domains: The nature of retrieval structures in situation awareness. Dissertation. Mississippi State University. Starkville, Mississippi.

Kenny, C. (2006). Automated tutoring for a database skills training environment. Unpublished master's thesis, Dublin City University, Ireland.

Kindley, R. (2002). The power of simulation-based e-Learning (SIMBEL). *The Elearning Developers Journal. The Elearning Guild, 17,* 1-8.

Kinshuk, P.A., & Scott, B. (2001). Intelligent tutoring: From SAKI to Byzantinum. *Kybernetes, 30,* 807-818.

Kohn, M. L., & Schooler, C. (1978). The reciprocal effects of the substantive complexity of work and intellectual flexibility: A longitudinal assessment. *American Journal of Sociology, 84,* 24-52.

Krashen, S.D. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon.

Lantolf, J.P. (2008). SLA, i+1, SCT, the ZPD, and other things: A response to Evensen. *Journal of Applied Linguistics, 5.2,* 215-219.

Lum, H., Fiore, S.M., Rosen, M.R., & Salas, E. (2008). Complexity in collaboration: Developing an understanding of macrocognition in teams through examination of task complexity. *Proceedings of 52nd Annual Meeting of the Human Factors and Ergonomics Society, New York, NY.,* 1425-1429.

Merrill, M. (2007). A task-centered instructional strategy. *Journal of Research on Technology in Education, 40,* 5-22.

Nunnally, J.C. (1978). *Psychometric Theory*. New York: McGraw Hill.

Parasuraman, R. (2003). Neuroergonomics: Research and practice. *Theoretical Issues in Ergonomics Science, 4,* 5-20.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39,* 230-253.

Parasuraman, R., Sheridan, T.B., & Wickens, C.D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 30.

Parker, C. & Myrick, F. (2009). A critical examination of high-fidelity human patient simulation with the context of nursing pedagogy. *Nurse Education Today, 29*, 322-329.

Program Manager Training System (PM TRASYS). (2014). Marine Corps Verification & Validation (V&V) Report for the Combat Vehicle Training System (CVTS) M1A1 Advanced Gunnery Training System (AGTS): Relocatable (RAGTS), Mobile (MAGTS), Deployable (DAGTS) and Tabletop (TAGTS) Configurations. Orlando, FL: Program Manager Training System.

Reigeluth, C.M. (Ed.). (1999). *Instructional design theories and models: A new paradigm of instructional theory (Vol. II).* Mahwah, NJ: Lawrence Erlbaum Associates.

Robinson, P. (2003). The cognition hypothesis, task design, and adult task-based language learning. *Second Language Studies*, *21*, 45-105.

Scerbo, M.W. (1996). Theoretical perspectives on adaptive automation. In R. Parasuraman and M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 37-63). Mahwah, NJ: Lawrence Erlbaum Associates.

Scerbo, M.W., Freeman, F.G., & Mikulka, P.J. (2003). A brain-based system for adaptive automation. *Theoretical Issues in Ergonomics Science, 4,* 200.

Schinke-Llano, L. (1993). On the value of a Vygotskyan framework for SLA theory and research. *Language Learning, 43*, 121-129.

Schraagen, J.M., Chipman, S.F., & Shute, V.J. (2000). State-of-the-art review of cognitive task analysis techniques. In J. M. Schraagen, S. F. Chipman & V. J. Shute (Eds.), *Cognitive Task Analysis* (pp. 467-487). Mahwah, NJ: L. Erlbaum Associates.

Sheridan, T.B. (1987). Supervisory control. In G. Salvendy (Ed.), *Handbook of Human Factors* (pp. 1243-1268). New York, NY: Wiley.

Sheridan, T.B., & Verplank, W.L. (1978). Human and computer control of undersea teleoperators. MIT Man-Machine Systems Laboratory, Cambridge, MA: Tech Rep.

Skehan, P., & Foster, P. (2001). Cognition and tasks. *Cognition and second language instruction*, 183-205.

United States Marine Corps. (1996). Marine Corps Reference Publication (MCRP) 3-0A. Unit Training Management Guide. Washington, DC: Department of the Navy.

Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wickens, C.D., Mavor, A.S., McGee, J. & United States National Research Council. (1997). *Flight to the Future: Human Factors in Air Traffic Control*. Washington, D.C: National Academy Press.

Wood, R. (1986). Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes*, *37*(1), 60-82.

Zook, A., Lee-Urban, S., Riedl, M.O., Holden, H.K., Sottilare, R.A., & Brawner, K.W. (2012). Automated scenario generation: Toward tailored and optimized military training in virtual environments. In *Proceedings of the International Conference on the Foundations of Digital Games* (pp. 164-171). ACM.