

MISSION COMPLEXITY SCORING IN DISTRIBUTED MISSION TRAINING

Todd Denning
L3 Communications, Link Simulation and Training Division
Mesa, Arizona

Winston Bennett, Jr.
Air Force Research Laboratory, Warfighter training Research Division
Mesa, Arizona

Peter Crane
Air Force Research Laboratory, Warfighter training Research Division
Mesa, Arizona

ABSTRACT

At last year's conference (Crane, Robbins, Bennett & Bell, 2001) we presented a first look at an empirically-based methodology for quantifying the complexity of Distributed Mission Training (DMT) scenarios for simulation-based, team training. The outcome of that effort was the development of a scenario complexity index derived from subject matter expert judgments which has been used to assess the effects of training in a building-block syllabus. Over the past few months, training specialists and subject matter experts at Air Force Research Laboratory's Mesa Research site have been characterizing DMT scenarios in terms of specific learning objectives linked to mission-essential competencies and to the underlying knowledge, skills, and experiences that are required for successful combat performance. As part of scenario characterization, we have identified the mission characteristics and levels of those characteristics that are important for complexity indexing. Techniques for determining overall scenario complexity and for relating scenario characteristics to mission essential competencies have been developed. This paper presents results from a validation study comparing the new, analytically-based complexity methodology with the empirically-based approach presented last year. In addition, applications of the new assessment to both F-15 and F-16 weapon system capability and scenario characteristics to learning objectives while controlling overall complexity will be discussed together with plans for developing DMT instructor support systems.

AUTHOR BIOGRAPHIES

Todd Denning commanded the USAF Weapons School F-16 Division 1999-2001 and is now an F-16 training and employment subject matter expert for the Air Force Research Laboratory. He is an 18-year veteran weapons and tactics instructor and training director in USAF and FMS F-16 programs and has produced over 36 basic and advanced F-16 syllabi including the current training and employment standardization used at the USAF Weapons School F-16 Division. He has over 3100 hours in the F-16 with 1800 hours of instructor time. Mr. Denning is an FAA-certified flight instructor and holds degrees in Electrical Engineering and Military Art and Science.

Winston Bennett, Jr. is a senior research Psychologist and team leader for training systems technology and performance assessment at the Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Training Research Division in Mesa, AZ. He received his PhD in Industrial/Organizational Psychology from Texas A&M University in 1995. Dr. Bennett has been an active member of I/ITSEC since 1994 and has served on two conference committees.

Peter Crane is a Research Psychologist at the Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Training Research Division in Mesa, AZ. His major research interest is enhancing effectiveness in Distributed Mission Training systems. Dr. Crane earned a PhD in Experimental Psychology from Miami University in Ohio.

MISSION COMPLEXITY SCORING IN DISTRIBUTED MISSION TRAINING

Todd Denning

**L3 Communications, Link Simulation and Training Division
Mesa, Arizona**

Winston Bennett, Jr.

**Air Force Research Laboratory, Warfighter training Research Division
Mesa, Arizona**

Peter Crane

**Air Force Research Laboratory, Warfighter training Research Division
Mesa, Arizona**

INTRODUCTION

Fighter training program development, as currently conducted, is more art than science. In general, fighter training syllabi produced in the field present theoretical credibility; however, scrutiny of practical application frequently reveals detriments in learning pace as a result of poorly staged training stressors within mission events. Existing training theories and papers highlighted in previous complexity research provide varied approaches to achieving competence in combat operations either using, *or specifically avoiding*, learning objectives (Crane, Robbins, Bennett, & Bell, 2001). The debate for and against objective-based training is interesting, but even in highly structured efforts such as Instructional System Development (ISD), theory falls victim to an expedient and less effective practical application in many military syllabi. In a similar outcome, Simulator Networking (SIMNET) studies showed promise in proficiency gains through voluminous unstructured experiences (Alluisi, 1991), yet Crane et al.'s (2001) analysis of SIMNET still recognizes a need for objective-based, building-block training suggested by ISD for primary and resource-limited training evolutions. Whether structured or unstructured, current training formats fall short of the mark in efficiently attaining true mission competence.

A survey of the practicalities of fighter training points toward a prevailing detractor. The root of the problem in program development *is the typical gulf of separation between the knowledge fields of expert combat operators and instructional system designers*. Researchers at Air Force Research Laboratory (AFRL) have introduced methodology to identify competencies, knowledge, skills, and experiences required to achieve combat success (Bennett, Schreiber & Andrews, in press; Colegrove & Alliger, 2002). Through this effort, both researchers and combat operators charged with training development duties will gain better insight into

learning processes and how they link to competencies required in the field.

Budgetary stress demands improved training efficiency. Selection of an optimized training path for warfighters has economic and security advantages that, in the current climate, demand more than a dusting off and review of an existing syllabus from another mission or aircraft. Despite USAF's embrace of ISD, the core training program for multirole fighters has survived relatively intact from the F-100 era to the present through a review and renew process carried out by teams of combatants-turned-trainers. Combatants are rarely schooled in ISD nor are they able to catalog the complexities resident in many of their chosen training scenarios. In the current budgetary climate, there is no evidence to suggest fiscal limitations will tolerate SIMNET's unstructured approach to proficiency development, particularly for formal training of combatant candidates. A need will continue to exist for structured exposure to warfighting concepts that begins with low complexity and paces the program accurately to higher complexities for maximum training gains in short-term formal courses such as the F-16 Basic Qualification Course and the F-16 Weapons Instructor Course. At the heart of an effective training pace is the accurate selection of scenarios and vignettes that are predictable and measurable in mental and physical task inventories.

Current efforts at the Air Force Research Laboratory, Warfighter Training Division are transforming the art of scenario selection into objective principles through definition of mission essential competencies (MEC) (Colegrove & Alliger, 2002). Continuing research in defining MEC knowledge, skill, and experience (KSE) inventories enables warfighting trainers to dissect scenarios into decision-task sequences (Bennett & Crane, 2002). Each sequence can then be viewed in the

context of a scenario to define required elements and elements subject to modification or elimination in a calculated risk of priorities during combat operations. The decision-task sequence a warfighter chooses to employ in a scenario is observable and predictable when assessing weapons employment in terms of objective standards and MECs. Each task sequence is a finite set of decisions and actions that occur in sequential and parallel execution during employment. These in turn can be mathematically modeled into a relative complexity value using an anchor point of the most elementary actions required in a tactical situation. Given this emerging resolution of complexity, the warfighter and training program developer will be able to accurately scale the complexity in training scenario vignettes used in a structured program.

Research Benchmark

Previous AFRL research into complexity modeling (Crane et al., 2001) established a first look at quantifying complexity by using a variety of operators and levels of experience to stratify a select number of scenarios utilized in DMT. An empirical method was chosen over analytical because the analytical approach was insufficiently definable among subject matter experts. Empirical rank orders of each group demonstrated substantial agreement among evaluations with correlation values from +0.96 to +0.97 between less experienced pilots, instructor pilots, and Air Weapons Controllers. In assigning a normalized complexity value to the rank order and applying it to the existing, and very low resolution grading standard, trainee progress in a building-block syllabus of instruction was quantified. The empirical method demonstrated that a complexity factor would aid training progress visibility, but not without some questions as to its origination, method, and scale anchors. Normalized rank ordering of selected scenarios limits the value definition to the selected set. New scenarios would have to be shuffled into the deck and the deck reassessed by more input from Subject Matter Experts (SMEs) with the new assessment being subject to personal perspective of new SMEs.

Individual pilot perspectives on training and combat experience has, to this point, made training and execution methodology consensus among operators difficult to achieve. This, in turn, makes definition of complexity difficult for researchers to grasp when subjected to operator disagreement or lack of instructional and employment familiarity. In spite of Crane et al.'s concern regarding previous applications of analytical methods for complexity scoring, a more rigorous analytical approach may generate reliable and valid complexity scores while avoiding the difficulties

of empirical scaling. The empirical method is cumbersome to use, sensitive to the pool of scenarios and SMEs selected, and can produce different magnitudes and scalars.

In the last year, key breakthroughs in the areas of MEC definition (Colegrove & Alliger, 2002) together with employment and training standardization within the F-15 and F-16 communities have opened routes for more detailed complexity analysis. Crane et al.'s (2001) critical assessment of analytical scaling was carefully considered during the development and evaluation of this approach. The following sections describe the assumptions and logic for the method of complexity analysis and validation discussed in this presentation.

METHOD

An analytically-based complexity estimator was derived after study of the elements of a tactical problem using the MEC construct to examine actions, decisions, and communications involved in tactical situations. The resultant model was built with the intent to accurately capture performance standards in USAF F-16 training programs. The analytical model aligns to the common training methodology used in fighter programs. Before detailing the model, it is essential to understand the assumptions resident in realistic training.

The tactical objective of an *actual combat mission* resides within the commander's intent as stated by an air tasking order. The objective may not call for complete elimination of adversary forces (although that is commonly fighter pilot's coveted end state). Typically, the objective of real-world missions involving air combat will be to establish air superiority, block incursions into an operating area, and deny the use of selected airspace for some period of time. The degree of aggressiveness of adversaries, combined with logistical constraints of weapons and fuel, and psychological pressures on both sides will temper the degree of confrontation. In actual combat operations, there may be less killing of targets and more chasing away and monitoring of the area. While this is often acceptable for the real combat objective, training to these outcomes is less desirable than pushing to the limits of execution capability.

In realistic training, the objective of a scenario is typically raised to the most stringent execution standard of eliminating all incursions into an area of responsibility. Additionally, adversary conditions are set to predetermined levels of offensive pressure, defensive reactivity, and shootable merit (as measured by rules of engagement) to challenge and improve the

trainee s abilities to operate effectively. This training methodology is hallowed practice of air combat training in all US fighter forces (USAF, USN, and USMC) and is hence the foundational assumption of establishing a tactical objective in complexity definition — the team s job in training is to detect and kill all adversaries coming into the area of responsibility, and they *will* come by scenario definition. Distributed mission training scenarios at AFRL are built upon this customary practice. The assessment of complexity assumes that each adversary group must be assimilated into the tactical plan and dealt with by the fighter team at a time deemed appropriate for the tactics employed on both sides.

Complexity Value Derivation

Physical and mental tasking studies within AFRL and other sources were reviewed in concert with compilations of instructor commentary and training reports from the US Air Force Weapons School (USAFWS) F-16 division over a three-year period. This time span involved six Weapons Instructor Course (WIC) classes, 68 students, and 35 instructor pilots. An additional two-year period was assessed in basic F-16 qualification courses from instructional and training development documentation. Analysis consisted of examining commentary and compiling consensus on attributes of acceptable versus non-acceptable performance based on current F-16 USAFWS tactical employment standards. These attributes were then compared to the KSEs within each MEC to construct basic decision-task sequences or KSE execution paths required to accomplish a tactical objective for each of a finite set of adversary positions, characteristics, and maneuvers. The number of execution paths chosen follows the *element fire and maneuver doctrine* outlined in the F-16 employment standards. Within each execution path, the decisions and tasks were analyzed for sequential and/or parallel characteristics along with an expert analysis of automaticity versus the need for conscious problem solving priority. Models for elemental attributes within a scenario were then constructed using a baseline value for the simplest or most automated decisions and tasks using a mathematical construct to combine the elements into a baseline score for the simplest execution paths. Several differential approaches were considered; however, the situations in air combat do not lend themselves easily to differential analysis with the limited data sets available. Instead, weighted values, combined in products, summations, and limited exponential modeling proved to be a simple and reliable method to express the scale factors sought.

Baseline complexity for the simplest intercept type was molded from a consensus of F-16 WIC instructor considerations defining the most rudimentary tasks and pacing found in an air intercept situation. The baseline attack may take the form of either a simple straight line closure to a weapons envelope, or include a maneuver called the *baseline intercept* to a stern conversion. Air intercepts in distributed mission training and actual combat can rapidly take either shape due to weapons limitations or scenario rules of engagement. In forming the baseline consideration, the lowest task count, load, and pace was considered in the context of the skill sets of both types of intercept to establish the lowest level of complexity to achieve a kill objective. The main criteria in evaluating both intercept types was to define a point from which the tactical experts train forces to employ weapons with the highest probability of weapon effect and the most automaticity in execution. For beyond-visual-range engagements, this position is a collision course to the target followed by a range preserving maneuver. For the visual identification intercepts, expert opinions varied from ending in a position abeam the target to ending in a position aft of the target. However, the baseline intercept remained consistent in skill set and pace for both end states prior to weapons employment in quantitative terms.

A foundation score of 1.0 for an intercept of least complex execution was selected to facilitate calculation of more difficult tasks. The baseline intercept and collision closure were assessed to have comparable task quality and quantity since both required detecting a target, enabling a radar track file through switch activations, and continuously analyzing stable mathematical relationships displayed on the air-to-air radar s B-scan display to adjust for and maintain the desired intercept profile. In the baseline score, assumptions are made for an easily detected and discerned target, no counter-maneuver or awareness of the fighter by the target, no weapon employment restrictions, and no detractors to situation awareness development. The foundation score represents the simplest task set an individual fighter pilot will encounter in a tactical scenario while developing minimum required situation awareness and physical positioning to achieve a kill. Both versions achieve high automaticity levels in early tactical training.

Scenario Complexity Estimate

Models of higher order complexity values for adversary attributes beyond the baseline were initially estimated through analysis of select execution paths used to establish the USAFWS F-16 Division employment standards. These execution paths encompass the realm of options an *element* (two aircraft operating as a team)

would choose from when confronted with varying levels of adversary capability and maneuver. An initial value set was determined by analyzing observed performance in video-taped tactical displays and Nellis Air Combat Training System (NACTS) data collection. Determination included extensive analysis within the expert instructor's debrief of performance evaluation along with assessment of decision and task execution quality as recounted in the mission summary and grading documentation. Weighting factors considered additional mental and physical tasking of individuals within the *element* firepower team, essential communications requirements between team members, and limitations imposed on available resources posed by additional stressors in or near the fight space. Each execution path was then calculated by the model and submitted for independent review by SMEs at the F-16 WIC. As a result, initial estimates of certain weighting factors were reduced in value, some stressors fractured for better resolution, and several were increased in weight value. Attributes assigned to single groups of adversary aircraft include separate weighting factors for number of aircraft, detection range, sort ability at range, weapons carried, awareness of fighter attack, counter-maneuver, countermeasures, and shootable merit based on rules of engagement. These factors were derived from the F-16 tactical standards cubic construct of awareness, reactivity, and identification as measures of stress imposed by the adversary force on a fighter team's tactical problem. Since the single-group construct followed standardized employment methodology, consensus for single group complexity valuation proved consistently high in subject matter expert review.

Estimation of complexity in the empirical method was highly sensitive to the total number of groups in a scenario (Crane et al., 2001). Crane et al. suggest that group number may or may not have a standardized complexity stress; however, the empirical data analysis suggests the experts have significant agreement in estimating group-to-group complexity by numbers alone. The empirical rank order of scenarios was almost completely stratified by total group number with most disagreement in stratification falling primarily in same-number scenarios. In tying groups together in the analytical method, consideration is given to the distance between groups using the same decision process as the engaging pilot. Range sensitivity is expressed by the combatant in terms of delineation points known as *bounding range* and *factor range*. *Bounding range* defines the decision to include or exclude a group from the tactical situation. *Factor range* defines a break point in attention paid to a group when dealing with a primary target or threat. Groups inside *factor range* are given higher complexity weighting as a function of

their proximity to an *element's* primary tactical problem. A *factor group* is part of the tactical problem even if engaged by another force since a *factor group* represents a pending threat that *must* be considered. The analytical segment built for group-to-group interaction provides a straightforward accounting of group positions and ranges for consistent valuation. The construct chosen retains reasonable sensitivity to total numbers of groups, but simultaneously allows for accounting and rational limitation of complexity inflation for groups that are not part of the immediate tactical problem.

Large scenarios (more than four groups) remain problematic in assigning complexity to distant groups being engaged by distant friendly forces and sequential attacks by waves of adversary forces. Initial input from subject matter experts suggested a summation strategy was appropriate. Statistical analysis of this method proved it less appealing than a weighted strategy. The current methodology for scenario analysis is an exponential weighting approach based on MEC KSEs and standardized employment. Standards and MECs were used to resolve an inventory of all tasks, decisions, and inter-flight communications for commonly executed adversary presentations. A procedure for generating complexity estimates for multiple groups was selected that retained the value of attributes of individual groups and assigned a limited exponential curve to increasing numbers of groups. This procedure provided optimum retention of correlation with expert opinion while avoiding statistically troubling characteristics.

ANALYSIS

Trials conducted on the complexity estimate model began with a comparative analysis with the Crane et al. (2001) empirical study, then proceeded to a validation test with new AFRL benchmark training scenarios, and continues with analysis of the benchmark scenarios in DMT along with an analysis of live-fly and simulation scenarios resident in the highly-structured training program of the USAFWS F-16 Division air-to-air training program. DMT syllabus building efforts are also using the complexity estimate as additional guidelines for scenario sequencing.

Empirical Comparison

Complexity estimates were conducted on the same scenario set as presented in the empirical study. The estimates were produced in a blind study without prior knowledge of the empirical stratification. Analytical results were then compared to the empirical study results. A strong positive correlation (+0.96) was achieved with the complexity estimates derived from

subject matter expert rankings in the empirical study. All scenarios except three remained within one standard deviation of the previous study. Complexity index

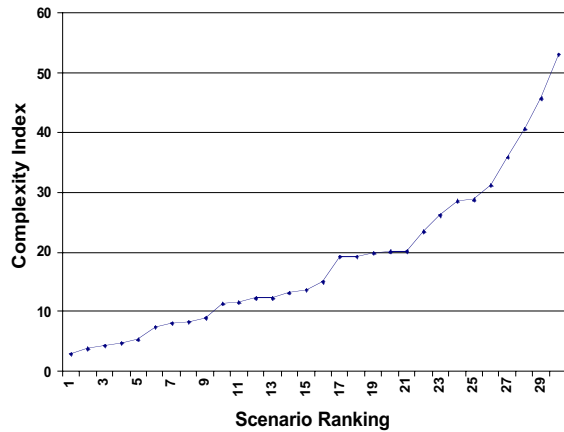


Figure 1. M1 Analytical Complexity Value by Scenario

values were allowed to expand as the model dictated and achieved values from 1.00 to 54.00. Modeling (M1) also initially indicated a capability to visualize training plateaus suggested in Crane et al. s empirical research. The values are shown in Figure 1.

Crane et al. s empirical procedure stratified scenarios in a subjective valuation of the entire scenario as viewed by each study participant. The initial analytical estimates were also computed in a holistic sense using the entirety of scenario attributes in a combined score. As the scenario expanded in size and numbers of stressors, the complexity value inflated exponentially. While this may seem intuitive to the analyst, subject matter experts disagreed with the valuation scheme. The subject matter experts indicated that the entire scenario plays a lesser role in tactical problem complexity than Method 1 analysis delivered.

Additional scrutiny of fighter air-to-air employment methodology suggested that the model needed to be able to break complexity elements into aggregates as the fighter pilot would detect and deal with them. The complexity model has sufficient flexibility to segregate and value engagement vignettes as well as total complexity estimates. A second methodology was examined and recalculated using the subject matter expert inputs. Method 2 (M2) examines the scenario from the weighted approach. Comparisons of empirical and analytical stratification for Method 2 are shown in Figure 2 and 3. Figure 2 shows a summation approach to total scenario complexity while Figure 3 shows the weighted approach.

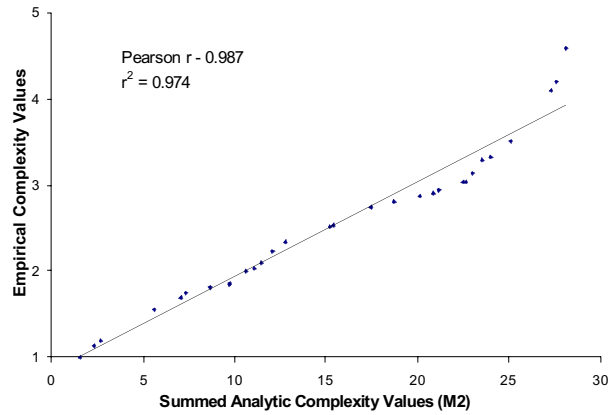


Figure 2. Comparison of M2 Summation complexity estimates with empirically derived estimates

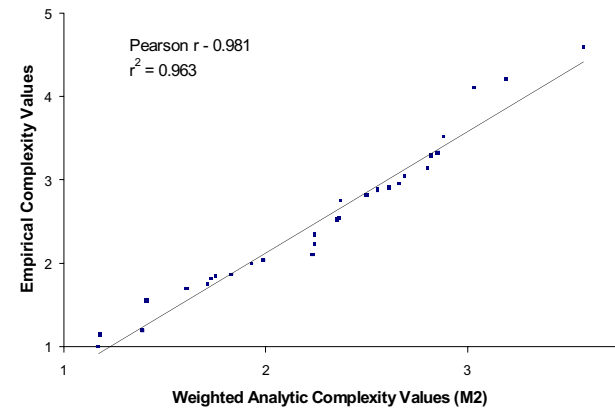


Figure 3. Comparison of M2 Weighted complexity estimates with empirically derived estimates

The results using summation in Method 2 suggest additional investigation is required into the non-linear relationship seen in the comparative graph. The M2 weighted method shows a more random distribution around the regression line with little change in correlation with empirical complexity scores. These analyses suggest the M2 Weighted procedure is a more promising avenue toward complexity factor valuation.

Benchmark Reliability

A second blind study was initiated by DMT researchers to evaluate the ability of the model to reliably compare training scenarios of equal complexity. The researchers provided ten benchmark training scenarios developed by subject matter experts and the AFRL DMT staff to evaluate pilots under the current DMT research program. These scenarios had been scrutinized and flown in DMT to validate the outcomes and pilot

perspectives in the scenarios. The ten benchmarks were actually five parallel or mirror image scenarios that were properly identified by the model. Thumbnail sketches of selected scenarios are shown in Figure 4.

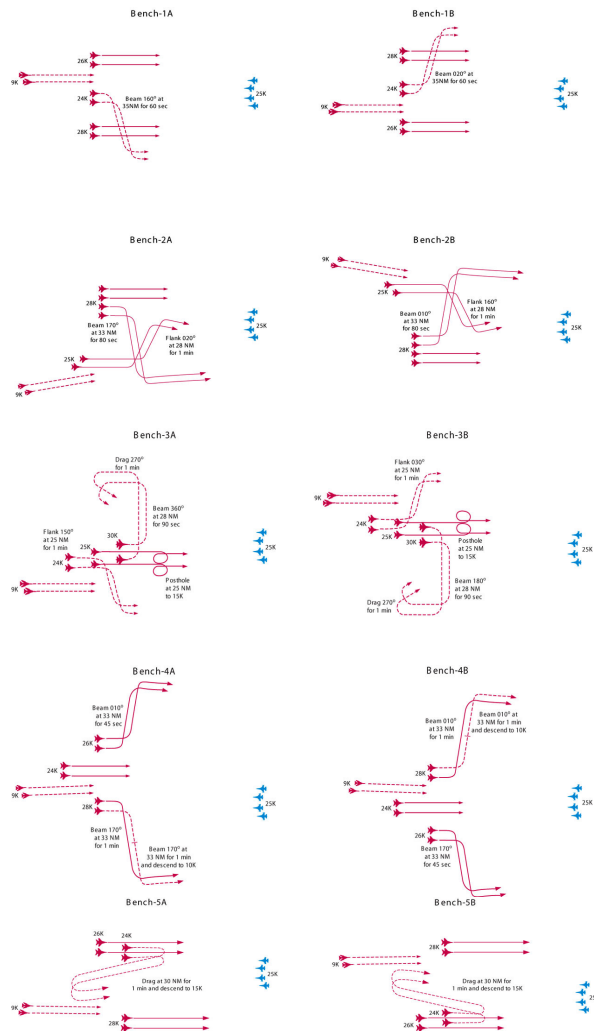


Figure 4. AFRL DMT Benchmark Scenarios 1-5

Calculation of the benchmarks resulted in summed values ranging from 9.04 to 9.40 and weighted values from 2.77 to 2.88. The values for composite scenario complexity are shown in Table 1.

In addition to demonstrating reliability in aggregate calculation, this study also tested the ability to extract complexity scores for individual engagements by element firepower teams. Using the M2 estimate to view and calculate tactical problems as the fighter team sees them, and applying the same calculated risk of disregarding certain *non-factor* geographic attributes as would be done in the immediate situation by the fighter team; the M2 estimate scores reflect the same

Scenario #	Sum Value	Wt Value
1A and 1B	9.04	2.77
2A and 2B	9.15	2.80
3A and 3B	9.40	2.88
4A and 4B	9.40	2.88
5A and 5B	9.06	2.77

Table 1. Summed and Weighted M2 Scenario Complexity values for ten Benchmark DMT scenarios.

complexities of smaller scenarios that were considered accurate in previous subject matter expert reviews. It is essential to understand that the surveillance and targeting risks taken by a fighter team in a geographic sense are attributes found within an attack wave, but either handled by another element team, or outside of factor range to the element team being evaluated.

Another segregation attribute in fighter employment was modeled in the M2 benchmark test based on *non-factor* attributes separated by time. Fighter teams will take on a workload in each engagement that balances situation awareness, optimized firepower, and defensive status. As the team reaches culmination in an engagement, the leader will direct a short retirement known as *cold ops* to separate the fighters from threats and allow time to rebuild range and/or situation awareness for reattack. Similarly, certain tactical situations influence fighters to plan sequential attacks from a maneuver sequence called *the grind*. The model was used in both geographic and time divisions to assess lower levels of complexity internal to the complete scenario.

Separated scores demonstrate not only *element* complexities, but also suggest the model can be used to describe complexity per attack wave as determined by timing and maneuver of each side's tactical plan. In the case of the benchmark scenarios, *element teams* would be compelled to simultaneously attack in azimuth, followed by a short retirement and reattack of remaining adversary forces. The tactical method chosen to break down the internal complexities is a direct reflection of current F-16 WIC school solutions and employment guidelines. The results of the benchmark analysis computed by the complexity estimate model M2 are shown in Table 2 on the following page.

Weapons Instructor Course Analysis

Scenarios provided from the empirical study were of insufficient detail to accurately test the full attributes of any of the analytic complexity estimator versions. Moreover, the narrow scope of the attributes applied in

the empirical method substantially restricted the range of application for, and the long term utility of, the study

Scenario #	N Fighters	S Fighters	Reattack
1A	2.24	2.70	1.76
2A	2.45	2.70	1.76
3A	2.70	2.57	1.76
4A	2.57	2.71	1.76
5A	2.71	2.24	1.76

Table 2. M2 Benchmark Complexity by Fire Team

data. As a result, critical details for estimation were subject to changes in the perspective of the original group of empirical study participants and were considered a possible cause of some of the larger sigma values in Crane et al s (2001) study. An additional concern is that the DMT threat model used to create training scenarios is limited in comparison to live-fly targets. Live-fly threat pilots are briefed to execute precise tactical maneuvers that the threat model cannot duplicate at the present time. Focused training at the F-16 WIC relies heavily on establishing tactical vignettes with precise attributes and considerable threat validity and realism. Any complexity scoring approach must be able to fully account for a full suite of threat and adversary tactics if it is to be useful for valid virtual and live tactical training.

The F-16 WIC formalized a standard approach to adversary presentations beginning in 1988. Since that time, the practice has grown to include all aspects of air-to-air training within the syllabus. Phase managers provide all participants with exact presentation attributes requested for the WIC s structured ascent to graduate-level proficiency. However, the selection of scenarios and attributes in WIC training still relies exclusively on historical precedents of scenarios chosen in a subjective manner and never analyzed in any other method than a brute force of trial and error.

The next evolution of testing for the complexity estimator was to subject its analytical technique to a sample of air-to-air instruction scenarios existing in the F-16 WIC. The sources for these scenarios are the *bandit playbooks* from *Tactical Intercept* training, academically presented scenarios used to teach air combat concepts, as well as graded observations and playbacks of Nellis Air Combat Training System recorded missions from the more loosely structured *Air Combat Tactics*, *Weapons*, and *Mission Employment* phases of instruction.

Twenty scenarios were selected from WIC sources including one well-recognized baseline problem from initial qualification training. The twenty scenarios were

analyzed under the M2 technique using lessons from the empirical comparison and DMT benchmark studies. The goal of this evaluation phase was to compare live-fly complexity values to the consensus opinion of F-16 WIC instructor pilots and training phase managers based on where these scenarios appear in the progressive training of the WIC. It should be noted that the sample size of 20 scenarios is considered too small to draw firm conclusions at this time and additional research using a larger sample will be undertaken to investigate the preliminary outcome of this analysis.

Results of the WIC sample analysis determined M2 weighted complexity values ranging from 1.01 to 2.91. Using the M2 summation technique, the scores ranged from 1.01 to 9.51. Comparison of weight to summation was analyzed to identify how numbers of groups affected the overall rank ordering from technique to technique. The results are shown in Figure 5.

The next phase of analysis was to examine the complexity scores in the weighted method in comparison to the sequence they may appear in the WIC syllabus of instruction. F-16 WIC air-to-air training is broken into segments of multiple rides that examine and train weapons officer candidates in specific employment areas. Training phases are engineered to have increasing levels of stress that may be raised from scenario to scenario or mission to mission to enhance or test the learning pace. The *Tactical Intercept* phase (TI 1-3) is a three-mission block of instruction designed (due to logistics constraints) to ramp students quickly up to radar and weapons proficiency. The first half of the *Air Combat Tactics* phase (ACT 1-3) is the first opportunity to put an entire mission plan to practical application and is hence flown against adversaries with moderately complex attributes. The second half of the *Air Combat Tactics* phase (ACT 4-6) introduces students to the most robust adversary weapons and tactics attributes. The *Weapons* phase copies the air combat problems of the *Air Combat Tactics* phase while introducing the multirole aspects of F-16 employment. The final phase of the course, *Mission Employment* (ME 1-3), provides the largest and most complex problems for students through addition of the requirement to coordinate many friendly platforms other than F-16s into the scenario. Since the complexity estimation technique is yet untested in multi-mission environments, only the air-to-air problems faced by F-16s were extracted from the final two phases for analysis.

Complexity scores for the 20 scenarios were assembled and linked to the missions they relate to either through a

bandit playbook or actual observation. Results of a weighted comparison using M2 are shown in Table 3. The F-16 WIC analysis suggests that the complexity estimation technique follows subject matter expertise

used by the USAFWS to structure the training program for WIC students. The selected scenarios demonstrate attributes that generally follow the easy-to-harder goal of each training segment with some new outlooks on

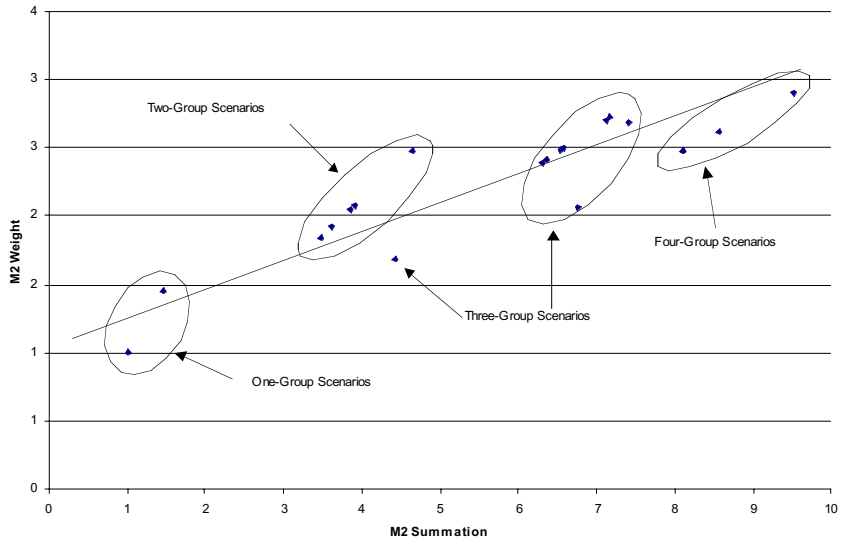


Figure 5. F-16 WIC Scenarios — Complexity Technique Comparison of Weighted Groups versus Summed Groups

Scenario	M2 Weight	M2 Sum	Mission Training Block Linkage			
6	1.01	1.01	IQT			
1	1.46	1.46		TI-1		
5	1.69	4.42		TI-2/3		
24	1.85	3.47		TI-2/3		
2	1.93	3.61		TI-2/3		
18	2.06	3.85			ACT 1-3	
13	2.07	6.75				ACT 4-6
20	2.09	3.91			ACT 1-3	
10	2.40	6.30				ACT 4-6
16	2.42	6.35			ACT 1-3	
12	2.48	8.10				ACT 4-6
14	2.48	8.10				ME 1-3
19	2.48	4.64				ACT 4-6
11	2.49	6.53				ME 1-3
8	2.51	6.57				ME 1-3
7	2.62	8.56				ME 1-3
9	2.69	7.40				ACT 4-6
17	2.71	7.11				ACT 4-6
3	2.73	7.16				ME 1-3
15	2.91	9.51				ME 1-3

Table 3. F-16 WIC Scenario Mission Linkage by Training Block.

training pace. A time-honored tradition of allowing the adversary force substantial latitude in scenario selection and execution within a three-mission training segment inserts a considerable variance in scenario pacing

within the F-16 s ACT, WPN, and ME phases. The limited sample size for this investigation also serves as a warning that more detailed investigation is required to produce a high confidence level in the estimation

technique. At the same time, the analysis also suggests there is reason to reconsider the trial-and-error method as well as adversary latitude of scenario selection and opt for combined techniques using both subject matter expertise and analytical complexity estimation in the selection of scenarios and vignettes for tactical intercept and air combat training.

The next phase of research will be to collect a larger group of WIC scenarios utilizing all bandit playbook options from the *Tactical Intercept* phase and as many observed adversary presentations as possible from the *Air Combat Tactics, Weapons, and Mission Employment* phases. This study is in progress and will be fully explored as a goal to establish the full level of complexity definition and analysis in F-16 air-to-air employment training.

Applicability to Other Aircraft and Missions

Another question being considered in complexity estimation analysis is the applicability of the method to other aircraft executing air-to-air combat and also to F-16 missions that include objectives in addition to or exclusive of air-to-air. Preliminary review by F-15 tactical experts suggests there is differing opinion on complexity valuation in the model. Work in this area requires a complete understanding of the perspective used to create the estimation values from each aircraft's pool of expertise as well as applied analysis similar to the F-16 efforts so far. Researchers at AFRL are preparing for studies in F-15 applicability and valuation as well as F-16 multirole missions that include surface threats, air-to-surface elements, and environmental stressors such as electronic warfare and systems attrition through failure or combat loss.

The physical considerations that drive a high expectation of broader applicability is the strong similarity of current fighter systems and the common set of weapons shared among them. There is also high expectation from an institutional standpoint. The tactical employment standardization of F-16 forces followed from studies and alignment of tactics recommended by the F-15 WIC and tactics manual and the F-18 TOPGUN instructor staff (and corresponding TOPGUN tactics manual). Therefore, along with common systems, all three aircraft enjoy a relatively common employment doctrine and methodology in air combat. The implications of this common doctrine and methodology are considerable generalizability and leveraging of the methods and tools developed in this complexity effort for other domains of similar complexity.

Broadened application to other roles and missions requires investigation of MECs and KSEs for those areas and the establishment of doctrine-based decision-task sequences required to successfully achieve the objectives of each area. Definition of decision task sequence paths is the key to establishing consistently expected and measurable levels of complexity. The focus of complexity research will remain on the F-16 until achieving credible confidence levels for the basic model; however, probes into F-15 employment and F-16 multirole missions are opening the gate to larger fields of application.

CONCLUSIONS

The drive for increased efficiency in tactical training demands that training developers and instructors on the front line understand the role of complexity in training pace. Crane et al's (2001) study highlighted the inability of subject matter experts to agree on valuation and combination of complexity into a single score for a scenario. This observation was made before MEC/KSE linkage was fully developed in the light of decision-task sequences as laid out by the expertise in the USAF F-16 tactical employment standards. Given the advances in understanding execution elements resultant from MEC analysis, subject matter experts are now closing on an ability to not only stratify scenarios by complexity index, but also to provide a qualitative valuation without need for repeated empirical trials.

Along with the ability to remove floating values of complexity resident in empirical methodology, the M2 method also shows promise for internal analysis of selected engagement vignettes to adjust training paces for individual participants within a larger scenario. In other words, we can adjust the complexity and pacing of scenario events and ensure that the training delivered in DMT environments is adaptive to *each* learner's needs and to the learning needs of the smaller team. There are two key goals of complexity analysis. First is to enable training program developers to accurately adjust learning pace in resource-constrained ISD environments. Complexity analysis linked to demonstrated performance will provide substantive data to make the case for enlarging or reducing training resources to meet specified objectives required of the gaining combatant commands. The second goal is to enable the front-line instructor charged with executing DMT training to visualize and optimize progress through learning cycles and plateaus by making real-time changes to training scenarios. That is, it will soon be possible for researchers and instructor pilots to accurately diagnose knowledge, skill and experience gaps and to identify specific scenarios in both virtual and live fly training to target and eliminate the gaps.

By recognizing and systematically emphasizing proper execution in weak performance areas, trainees are likely to achieve a higher probability of success earlier in training and hence reduce total resources required for effective performance in combat operations. In the long term it may well be possible to specify combat readiness and mission qualification at a quantifiable

proficiency-based level of analysis in specific skills as opposed to equating readiness to the less accurate and often demonstrably false measure used in current military planning —*experience* measured as total number of hours accumulated in an aircraft and *proficiency* measured by 30/60/90-day look back hour and event counts.

REFERENCES

Alluisi, E. (1991). The development of technology for collective training: SIMNET, A case history. *Human Factors*, 33, 343-362.

Bennett, W. Jr., Schreiber, B., & Andrews, D.H. (in press). *Developing competency-based methods for near-real-time assessment in a distributed training environment*. Computers in Human Behavior.

Crane, P. and Bennett, W. Jr., (2002). *The deliberate application of principles of learning and training strategies within DMT*. Paper presented at the

NATO SAS-038 working group meeting, Brussels, Belgium.

Colegrove, C. M., & Alliger, G. M. (2002, April). *Mission essential competencies: Defining combat readiness in a novel way*. Paper presented at the SAS-038 NATO working group meeting, Brussels Belgium.

Crane, P., Robbins, R., Bennett, Jr., W. and Bell, H. H. (2001). Mission Complexity Scoring for Distributed Mission Training. In, *Proceedings of 2001 Industry/Interservice Training Systems Conference*, Orlando, FL: National Security Industrial Association