MISSION COMPLEXITY SCORING FOR DISTRIBUTED MISSION TRAINING

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

Robert Robbins
L3 Communications, Inc., Link Simulation and Training Division,
Mesa, Arizona

Winston Bennett, Jr. & Herbert H. Bell
Air Force Research Laboratory, Warfighter Training Research Division
Mesa, Arizona

ABSTRACT

Distributed Mission Training (DMT), the Air Force's concept of ground-based training using networks of real-time simulators and computer-generated forces, evolved from earlier programs including Distributed Interactive Simulation and Simulator Networking. In these programs, defined training objectives were often deliberately avoided and training benefits were presumed to result from the intrinsic feedback and lessons learned from battle engagements. Previous research has demonstrated that DMT can provide effective training tailored to meet defined learning objectives through careful development and delivery of scenarios that are presented in a building block format over several training sessions. Using this paradigm, however, assessment of trainee performance has been problematic since the scenarios contain many different elements and encompass a wide range of complexity. A process for developing an empirically based degree-of-complexity scale for scenarios is described. The resulting scale allows direct comparison of team performance over several days of training and will support improved linking of scenario content with training objectives.

BIOGRAPHIES

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

Robert Robbins works as an Instructional Systems Designer for Link Simulation and Training at the Air Force Research Laboratory in Mesa, AZ. He is a retired Air Force fighter pilot with 3800 total flight hours, over half as a flight examiner/instructor pilot. He has a Masters degree in Aeronautical Science.

Winston Bennett, Jr. is a Senior Research Psychologist and team leader for the 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.

Herbert H. Bell is a Research Psychologist with the Air Force Research Laboratory's Warfighter Training Research Division. He received his Ph.D. in Experimental Psychology from Vanderbilt University in 1974. Dr Bell's research has included the use of distributed interactive simulation for combat mission training.

MISSION COMPLEXITY SCORING FOR DISTRIBUTED MISSION TRAINING

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

Robert Robbins
L3 Communications, Inc., Link Simulation and Training Division,
Mesa, Arizona

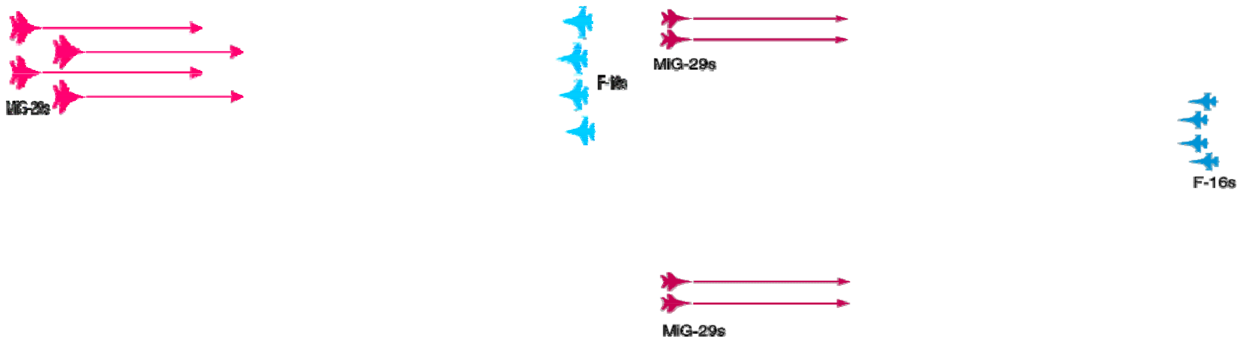
Winston Bennett, Jr & Herbert H. Bell
Air Force Research Laboratory, Warfighter Training Research Division
Mesa, Arizona

INTRODUCTION

The Air Force's Distributed Mission Training (DMT) program is a major advance in ground-based training that will allow pilots and other warfighters to train for complex, multi-player combat operations. Researchers from the Air Force Research Laboratory, Warfighter Training Research Division (AFRL/HEA) are investigating strategies for using DMT to augment advanced flying training in operational units. Effective application of multi-player simulation for enhancing individual and team skills has been demonstrated for F-15 pilots (Berger & Crane, 1993; Houck, Thomas, & Bell, 1991); F-16 pilots (Crane, Schiflett, & Oser, 2000), Tornado pilots and navigators (Huddlestone, Harris, & Tinsworth, 1999); pilots, forward air controllers, and ground forces executing close air support (Bell, et al., 1996); Tornado, Jaguar, and Nimrod crews (McIntyre & Smith, 2000); Air Force Special Operations teams (Nullmeyer & Spiker, in press); and for F-16 pilots upgrading from wingman to flight lead (Crane, Robbins, & Bennett, 2000).

Unlike many other training programs that have been developed using a formal training model such as Instructional Systems Development, DMT emerged from developments in training system technologies notably Simulator Networking (SIMNET) in the late 1980s and Distributed Interactive Simulation (DIS) in the 1990s. In these programs, defined training objectives were often deliberately avoided and presumed training benefits were based on gaining experience, not on meeting specific objectives. Thorpe (1987) described SIMNET as providing, "a rich form of trial and error learning." More formally, Alluisi (1991) stated that in SIMNET exercises, "Training would occur as a function of the intrinsic feedback and lessons learned from the relevant battle engagements." This unstructured approach has been highly successful in demonstrating the value of distributed simulation. For specific training applications, however, incorporating carefully selected elements into design of battle engagements can increase the effectiveness of DMT. Scenario design and the sequence of scenarios experienced, therefore, has become the major avenue for training intervention in DMT.

Crane (1999) described a process for developing DMT scenarios to meet specified learning objectives. Similar processes have been described for scenario development by Army (Campbell, Quinkert, & Burnside, 2000) and Navy (Smith-Jentsch, Zeisig, Acton, & McPhearson, 1998) researchers. More recently, Air Force researchers (Bennett & Crane, 2001 and, Colegrove & Alliger, 2001) have developed a process for specifying the mission essential competencies (MECs) associated with mission ready combat performance. MECs represent the higher-order skills encompassing individual, team, and inter-team competencies required of fully prepared pilots, crews, or flights for successful mission completion under adverse conditions in a non-permissive environment (Colegrove & Alliger, 2001). MECs represent a significant advance in how researchers and operational warfighters think about and define what it means to be combat mission ready. MECs are atypical in that they are not abstract knowledge or general skills, but rather are the specific knowledge, skills and related experience that are required for successful combat mission performance. MECs are also granular enough to permit demonstration in the context of actual missions (or high-fidelity simulated missions) under wartime conditions. MECs are the bridge between mission essential tasks and training task lists and are reflected in each mission area from pre-mission briefings through post-mission debriefs.



a) b)

Figure 1. Examples of non-maneuvering scenarios; a) enemy aircraft in a single group, b) two groups separated in azimuth.

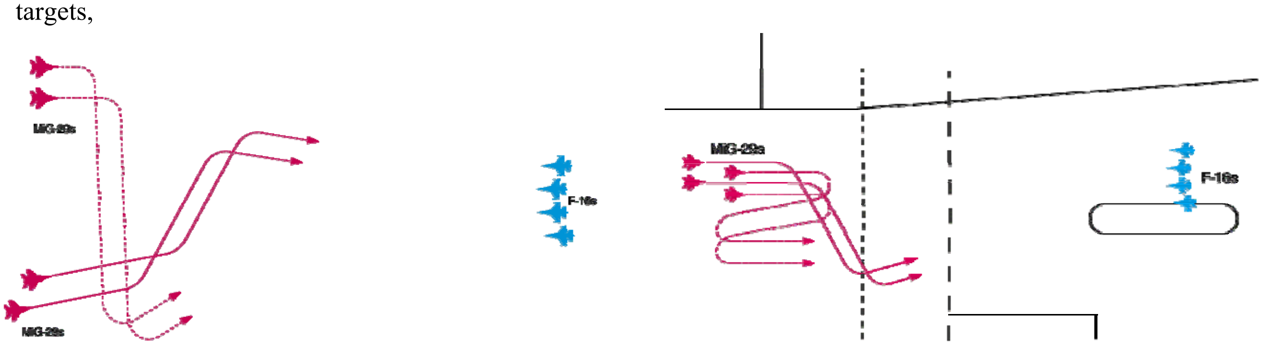
AFRL/HEA researchers are using MECs to drive syllabus development which includes the specification of training and learning objectives and development of valid and reliable measures of performance, both in DMT and in live-fly operations. In this process, subject-matter experts define learning objectives for scenarios and associated trigger events that provide the opportunity for exercising the identified knowledge or skills. In DMT research exercises conducted at the Air Force Research Laboratory in Mesa, Arizona, teams of F-16 pilots and AWACS air weapons controllers develop and brief their mission plans focusing on the specific learning objectives and then fly multiple scenarios seeing different variations of selected trigger events. During replay and debrief, instructors can then key in on the trigger events to assess team performance and provide feedback.

Examples of critical skills are:

- Use the fire control radar to detect multiple airborne targets,
- Individually target and sort in accordance with the briefed plan (i.e., each pilot in the flight identifies and tracks the group/aircraft he has been assigned in the briefing and maintains radar lock even as the target maneuvers), and
- Communicate this radar picture to other pilots and the air weapons controller in accordance with standards.

Examples of scenarios incorporating trigger events for these skills are shown in Figure 1. Airborne targets are presented in different scenarios in a single group, multiple groups in range, azimuth, and altitude, and in combination.

As pilots gain experience with employing their air-to-air radar against non-maneuvering targets, additional MECs can be addressed and by reference new learning objectives can be introduced by developing and delivering more complex and challenging scenarios.



a) b)

Figure 2. Examples of maneuvering scenarios a) Beam and flanking maneuvers, b) Flank and drag maneuvers.

• Use radar to track multiple, maneuvering formations, and
• Determine if target maneuvers require a change in sorting assignment, gameplan or tactics, and
• Adjust air-to-air missile employment as necessary to counter target maneuvers.

Examples of scenarios incorporating target maneuvers as trigger events are presented in Figure 2.

Training research exercises at AFRL typically last for four and one-half days allowing teams to fly nine, one-hour DMT missions. This schedule supports a building-block (crawl – walk – run) approach to training in which learning objectives for missions later during the week are dependent upon mastery of skills exercised earlier. Although evaluations from participating pilots and air weapons controllers have been highly positive (Crane et al., 2000), researchers have experienced two problems using this procedure.

One problem has been in quantitatively assessing changes in team performance over several days of training. At AFRL, proficiency is currently assessed by trained subject-matter experts on a 0 – 4 rating scale which is similar to that used by instructor pilots in Air Combat Command:

0 – Performance indicates lack of ability or knowledge
1 – Performance indicates lack of proficiency; makes errors of commission or omission
2 – Performance is essentially correct; makes and corrects errors
3 – Performance is correct, skillful, and without hesitation
4 – Performance reflects an unusually high degree of ability

As teams move through the syllabus and move from "crawling to walking and on to running", instructors increase the number of learning objectives which also increases the perceived complexity and difficulty of the scenario being used for training. However, ratings of performance remain constant at a 1 or 2 across the syllabus. While everyone acknowledges that the scenarios are increasing in complexity from the beginning of the week to the end, there was no variance detectable in comparisons of performance across the scenarios. For example, mean ratings for overall mission performance and nine skills from teams of pilots and air weapons controllers who participated in training research exercises at AFRL during May through July 2001 are shown on Figure 3. While feedback from participants and their instructors has been highly positive, the data indicate that performance does not improve substantially after pilots and controllers gain initial familiarity with the simulators even though a rating of "2" on a Tuesday scenario, which may be for a fairly straightforward 2 v 2 minimal threat scenario, is qualitatively and quantitatively different from a 4 v 8 high threat scenario that teams might see on Thursday.

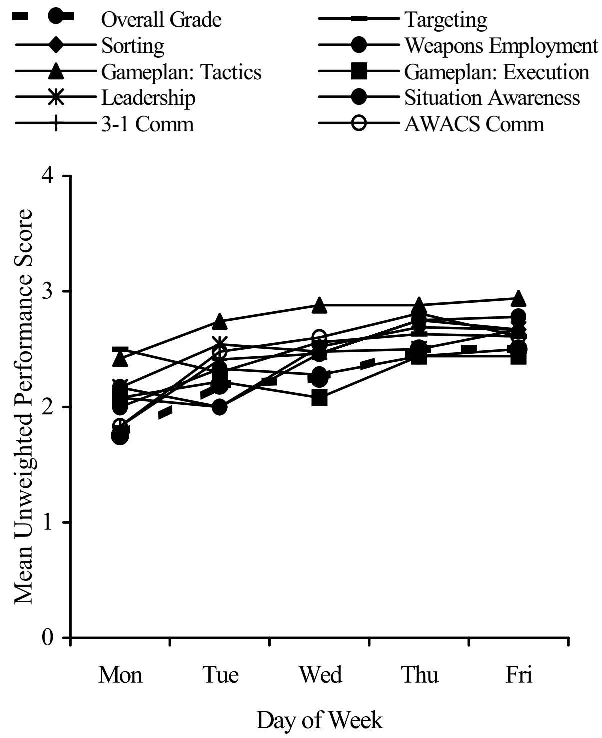
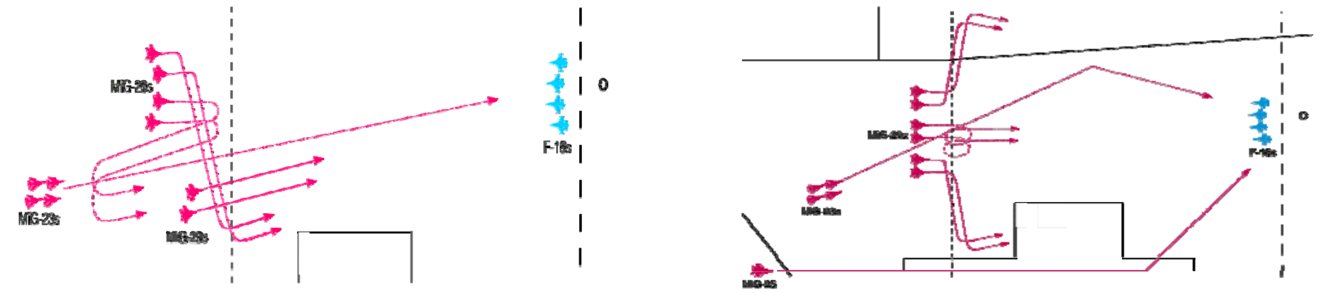


Figure 3. Mean unweighted performance scores for Overall performance and for nine rated skills: Radar sorting, Tactics selection, Leadership, Communication in accordance with standards, Radar targeting, Weapons employment, Executing the gameplan, Situation awareness, and Communication with AWACS.

One way to begin to address this potential problem has been to develop a set of benchmark missions that have been judged by subject-matter experts as similar in overall level of complexity. Examples of benchmark missions are presented in Figures 4a and b. These missions are relatively similar in the number of enemy aircraft, maneuvers, and aggressiveness. In research exercises, teams fly one benchmark mission early in the week and performance is compared to a different benchmark flown later in the week. Performance changes on these two missions allow comparisons from the beginning of training to the end for each of the rated skills and competencies (Crane et al., 2000).



a) b)

Figure 4. Examples of benchmark scenarios.

With scores on benchmark as a starting point, a next and critical step is the establishment of some process whereby the complexity of each scenario can be accounted for in assessments of performance. What is needed is some type of overall degree of complexity scale that could be applied within and across all scenarios in a syllabus. A degree-of-complexity scale would provide an index that could be used as a multiplier to weight performance scores on different scenarios and permit comparisons among scenarios.

A second problem experienced in implementing a building-block training program is assessing and comparing the relative complexity of various scenarios. Variables used in creating scenarios include: mission type (offensive vs. defensive counterair [OCA vs. DCA]) the number of enemy aircraft, the types of aircraft and ordnance, the number of groups, group formations and maneuvers, and the level of enemy aggressiveness and reactivity. Adjusting the mix of these variables to create a scenario that both incorporates the desired trigger events and presents an appropriate overall level of complexity has been problematic. Subject-matter experts disagree about the relative contributions of different scenario variables to overall complexity and the order of presentations required to support building-block training. An empirically based scale of scenario complexity is required to overcome these problems and to support systematic matching of desired learning objectives to specific DMT scenarios.

METHOD

There are two basic approaches to developing scenario complexity scores. In one approach, subject-matter experts review a prototype scenario, list the variables that affect complexity, assign scores to the values of these variables, and combine these scores into a single rating. For example, the basic scenario used at AFRL consists of a flight of four F-16s armed with radar and infrared (IR) guided missiles and supported by an air weapons controller that is tasked to conduct counterair operations. OCA missions might be assigned a score of 1 (less complex) and DCA missions assigned 2 (more complex); a single group of enemy aircraft assigned 1, two groups 2, etc; enemy aircraft armed with IR missiles only assigned 1, short-range radar missiles and IR missiles 2, and longer range radar missiles plus IR missiles 3. While this process has intuitive appeal, implementation has been problematic.

One issue is that the scale values only indicate rank order of complexity rather than the degree of complexity, per se. Observations from previous exercises demonstrate that increasing the number of enemy aircraft groups from two to three adds much more complexity than the increase from one to two groups. A second issue is that some variables change the tactical situation without necessarily changing overall complexity. An example would be four enemy fighters approaching the F-16s in two groups of two, separated in range as opposed to being separated in azimuth. The situations are different but not inherently easier or more complex. On the other hand, changing enemy aircraft armament from shorter range radar missiles to longer range missiles greatly increases complexity without increasing the number of enemy aircraft. Finally, there is no straightforward procedure for combining element scores into an overall score. Attempts to add point values or to aggregate similar scenarios into a limited number of categories have proven to result in inconsistent scales and disagreement among subject-matter experts.

An alternative approach to constructing a complexity scale is to ask subject-matter experts to make direct judgments of complexity or to compare scenarios and to then use mathematical procedures to extract a scenario degree-of-complexity scale from these judgments. Psychological scaling procedures have been developed over many years to describe sensory experience and to assess similarities and differences among stimuli such as consumer products and the contributions of noted scientists. Psychological scaling procedures have been used to assess the perceived resolution of different flight simulator display systems (Crane, Gerlicher, & Bell, 1986) and to determine the requirement for transformation accuracy in simulated ground-mapping radar imagery (Crane & Bell, 1985). To generate a scale of scenario complexity, F-16 pilots and air weapons controllers who had completed nine hours of four-ship OCA and DCA training using the Air Force Research Laboratory's DMT testbed system (see Crane et al., 2000, for a description of this system) were asked to rank 30 training scenarios from easiest to most complex. Mathematical analyses were then conducted to generate a complexity scale that can be used to weight measures of mission performance.

Raters

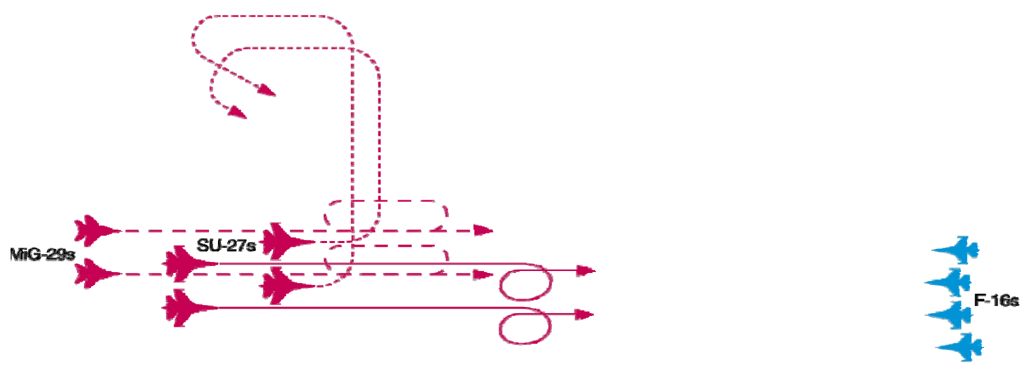
Scenario complexity rankings were obtained from 31 F-16 pilots participating in DMT effectiveness research exercises at Air Force Research Laboratory, Mesa Arizona in June – July 2001. Pilot experience levels ranged from approximately 100 hours in the F-16 for three pilots who had only recently graduated from the formal training unit to an average of 1170 hours for 25 instructor pilots including five graduates from Air Force Fighter Weapons School. Rankings were also obtained from six experienced AWACS controllers who also participated in DMT research exercises at AFRL.

Scenarios

Subject-matter experts at Air Force Research Laboratory, working with instructor pilots from many F-16 squadrons, have developed over 200 OCA and DCA scenarios to meet a wide range of training needs. A subset of 30 scenarios was selected for ranking. The scenarios were selected to incorporate a wide range of complexity levels from a small number of non-maneuvering, non-aggressive enemy aircraft to many aircraft with different types of armament arriving in multiple waves over 20 - 30 minutes. A drawing of each scenario was presented to the pilots together with additional information regarding the type of mission, number and types of enemy aircraft, armament, level of reactivity, awareness of F-16 capabilities, and the rules of engagement. An example scenario is presented in Figure 5. This DCA scenario consists of six maneuvering and reactive aircraft: four Su-27s and two MiG-29s.

Procedure

After completion of four and one-half days of two and four-ship DMT scenarios, each pilot or controller was given a folder containing depictions of 30 scenarios on separate sheets in random order. Each rater was instructed to work individually and to place the scenario sheets in rank order from the easiest to most complex. Level of complexity was defined in terms of meeting mission objectives: a scenario would be ranked as relatively simple if pilots had a high probability of completing their mission objectives using a few, basic tactics; a complex mission would require employment of multiple, advanced tactics to complete the mission. Pilots and controllers had no problems completing the ranking task and asked few questions regarding the definition of complexity.



Types:	Armament	Reactivity	AMRAAM aware	Spike aware	ID/ROE
2 MiG-29	AA-10A/AA-11	High	Yes	Yes	BVR/Hostile
4 Su-27	AA-10A/AA-11	High	Yes	Yes	BVR/Hostile

Note: Three group ladder—15nm range; exploding cantaloupe

Figure 5. Example of scenario rated by study participants. This scenario incorporates six reactive aircraft in three groups performing beam, drag, and posthole maneuvers; all aircraft can be identified as hostile beyond visual range (BVR).

ANALYSES AND RESULTS

Rank orderings were compared from the three groups of raters: highly experienced F-16 pilots, less experienced pilots, and AWACS controllers. Rank order correlations among the groups ranged from $r = +.96$ to $r = +.97$ demonstrating that there was substantial agreement among the groups regarding mission complexity. Subsequent analyses were conducted using all raters combined as a single group.

Using procedures described by Engen (1971), an interval scale of complexity was derived from rankings. For each rater, the scenario judged to be easiest was assigned a value of 1 for that rater and the scenario judged to be most complex was assigned rank 30. The ranks for each scenario were summed and mean ranks computed. The lowest mean rank (easiest scenario) was 1.54 and the highest rank (most complex scenario) was 28.1. The proportion of judgments of greater complexity for each scenario was computed by dividing each mean rank by the total number of scenarios minus one. Thus, the easiest scenario was assigned a proportion of 0.07 while the most complex was assigned 0.98. Based on the assumption that the characteristic of complexity is distributed normally, the proportion for each scenario was converted to a z-score using the unit normal distribution. The easiest scenario has a z-score of -1.51 with $z = +2.08$ for the most complex. Finally, a constant was added to each z-score so that the easiest scenario was set equal to one.

Degree of complexity for each scenario together with the standard errors for ranks is shown on Figure 6. The figure shows that except for the three easiest scenarios and the three most complex scenarios, relative levels of complexity are evenly spaced across scenarios. Inspection of the standard errors shows that the degree of agreement among raters was relatively consistent with the exception that 75% of the raters ranked one scenario (Figure 7) as the most complex resulting in a smaller standard error. Overall, these analyses demonstrated a high level of agreement among raters with respect to relative complexity of DMT scenarios.

Inspection of the scenarios ranked as easy, moderate, and most complex reveals the factors that drive the rankings. The easiest scenarios are characterized by a single wave of aircraft in one or two groups with limited maneuvers. The scenarios depicted in Figure 1 are among the easiest with complexity values of 1.19 (Figure 1a) and 1.14 (1b). The scenarios shown in Figure 2 are weighted as moderately complex with weights of 2.35 (Figure 4a) and 2.52 (4b). These scenarios are characterized by a single wave of four to six aircraft in multiple, maneuvering groups. Scenarios may include fighters armed with longer range missiles and mixed rules of engagement including some aircraft that can be identified as hostile beyond visual range and others that must be visually identified. The benchmark missions shown in Figure 4 were ranked as among the most complex with weights of 2.95 (Figure 4a) and 3.51 (4b). These missions incorporate multiple waves of maneuvering and reactive enemy aircraft with mixed rules of engagement. The scenario rated as most complex, Figure 7, is weighted 4.60.

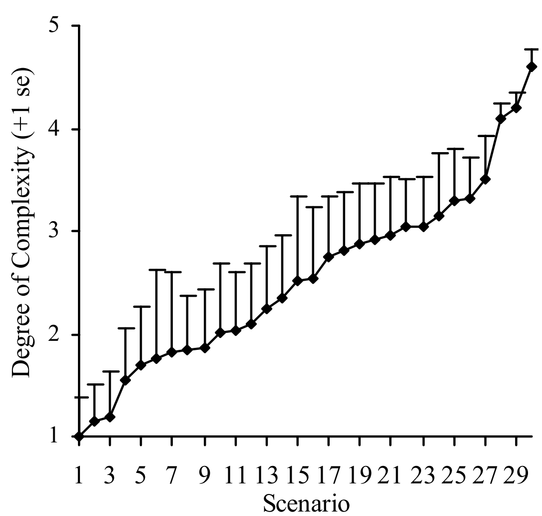


Figure 6. Degree of complexity weights (plus one standard error) for 30 DMT scenarios.

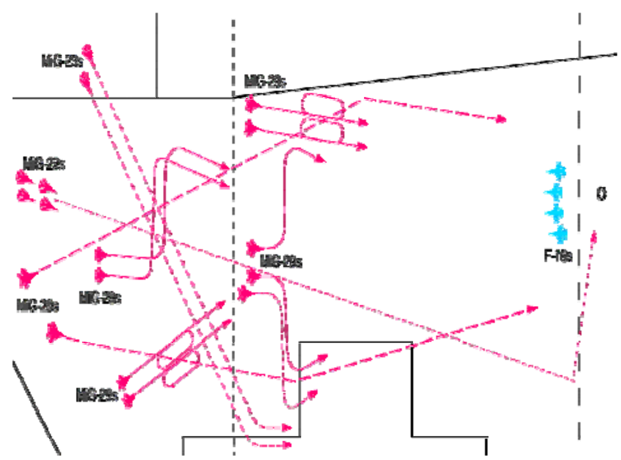


Figure 7. DMT scenario with the highest complexity rating.

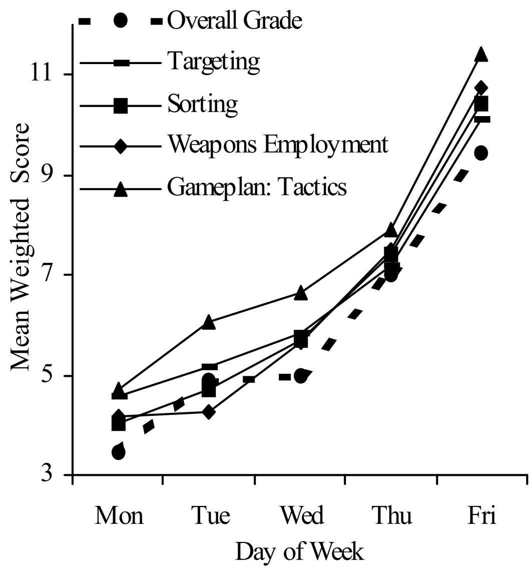


Figure 8. Mean performance scores for Overall performance and four rated tactical skills weighted by degree of complexity scores.

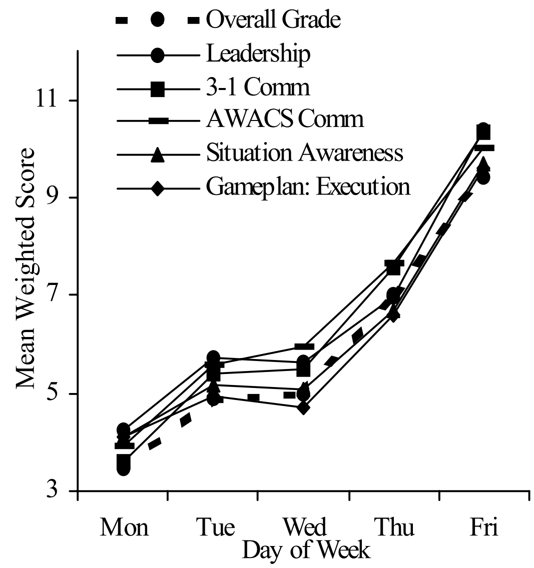


Figure 9. Mean performance scores for Overall performance and five rated team skills weighted by degree of complexity scores.

The degree-of-complexity weights were used to reanalyze the mission performance data shown in Figure 3. Each performance evaluation rating assigned by a subject-matter expert using the 0 to 4 scale was multiplied by the complexity weight for each of the training scenarios. The results of this reanalysis are shown in Figures 8 and 9. The weighted performance scores are in strong agreement with verbal feedback from participants and their instructors. One instructor pilot summarized a DMT training week as, "overcome simisms first, initial proficiency next, and exponential improvement after that." Not all rated skills, however, improved uniformly with experience.

The ratings shown in Figure 8 are for radar skills: targeting, sorting, and weapons employment, and for a mission planning skill, tactics selection when building a gameplan. These tactical skills show steady improvement over the week. Overall performance scores depicted by the heavy dashed line in Figures 8 and 9, show that performance decreases slightly from Tuesday to Wednesday but increases after that. The changes in performance scores shown in Figure 9 illustrate which skills contributed to this temporary decrease in overall performance scores.

In the building-block training format used at AFRL, missions switch from OCA to DCA on Wednesday. DCA missions include: combat air patrol procedures; grinder tactics in which the four-ship flight separates into two elements that alternate attacks against incoming enemy aircraft; pause mechanics when all F-16s turn away from the enemy, increase range, and then turn to re-engage; and de-louse procedures in which one aircraft or element engages an enemy aircraft that is pursuing another F-16. Questionnaire data collected from participants in AFRL DMT research exercises has shown that these tasks and skills are infrequently practiced in the aircraft due to cost, airspace, and resource constraints. As a result, teamwork skills required to perform the DCA mission including flight leadership, communications in accordance with standards and communication with AWACS, situation awareness, and gameplan execution, require an additional day of training before performance scores increase. These data are being used to define and select DMT scenarios that will more effectively complement aircraft training exercises and increase the effectiveness of DMT.

CONCLUSIONS

Assessing performance in training for complex tasks has been an enduring problem. While subject-matter experts are able to identify and describe good vs. poor performance, it has been difficult to quantify these changes in performance. Attempts to define scenario complexity based on analyzing the content of different scenarios have been problematic. Subject-matter experts frequently disagree about what variables drive overall complexity levels and about how to combine ratings on different variables into a single score. An alternative approach based on psychological scaling procedures only requires subject-matter experts to rank scenarios from easiest to most complex. Analyses of ranking data show that disagreements among experts are relatively small using this procedure and that mathematical analyses can be used to derive a useful scale of complexity. Complexity scales based on psychological scaling provide an effective method of quantifying trainee progress in a building block program and help instructors build scenarios to meet specific training objectives.

REFERENCES

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

Bell, H. H., Dwyer, D. J., Love, J. F., Meliza, L. L., Mirabella, A., & Moses, F. L. (1996). *Recommendations for planning and conducting multi service tactical training with distributed interactive simulation technology (A Four-Service Technical Report)*. Alexandria, VA: U. S. Army Research Institute.

Bennett, W., Jr., & Crane, P. (2001, October). *The deliberate application of principles of learning and training strategies within DMT*. Paper presented at the SAS-038 NATO working group meeting, Brussels, Belgium.

Berger, S., & Crane, P. (1993). Multiplayer simulator based training for air combat. In, *Proceedings of 15th Industry/Interservice Training Systems Conference*, Orlando, FL: National Security Industrial Association

Campbell, C. H., Quinkert, K. A. & Burnside, B. L. (2000). *Training for performance: The structured training approach*. (Special Report 45) Alexandria: VA: U. S. Army Research Institute.

Colegrove, C. M., & Alliger, G. M. (2001, October). *Mission essential competencies: Defining combat readiness in a novel way*. Paper presented at the SAS-038 NATO working group meeting, Brussels Belgium.

Crane, P. (1999). Designing training scenarios for distributed mission training. Paper presented at the 10th International Symposium on Aviation Psychology, Columbus, OH.

Crane, P., & Bell, H., (1985). *Flight training simulators: Effects of terrain accuracy on simulated radar image quality*. (AFHRL-TP-85-28, AD A160905). Williams AFB, AZ: Air Force Human Resources Laboratory, Operations Training Division.

Crane, P., Gerlicher, J., & Bell, H. (1986). *Flight simulator: Comparison of resolution thresholds for two light valve video projectors*. (AFHRL-TP-85-43, AD A164577). Williams AFB, AZ: Air Force Human Resources Laboratory, Operations Training Division.

Crane, P., Robbins, R., & Bennett, Jr., W. (2000). Using distributed mission training to augment Flight Lead Upgrade Training. In, *Proceedings of 2000 Industry/Interservice Training Systems Conference*, Orlando, FL: National Security Industrial Association

Crane, P., Schiflett, S., & Oser, R. (2000). *RoadRunner 98: Training effectiveness in a distributed mission training exercise*. (AFRL-HE-AZ-TR-2000-0026). Mesa, AZ: Air Force Research Laboratory, Warfighter Training Research Division.

Engen, T. (1971). Psychophysics II: Scaling methods. In J. W. Kling & L. A. Riggs (Eds.),*Woodworth and Schloberg's Experimental Psychology, (3rd Ed)*. New York: Holt, Rinehart, and Winston, Inc.

Houck, M. R., Thomas, G. S., & Bell, H. H. (1991). *Training evaluation of the F-15 advanced air combat simulation*. (AL-TP-1991-0047, AD A241875) Williams Air Force Base, AZ: Armstrong Laboratory, Aircrew Training Research Division.

Huddlestone, J., Harris, D., & Tinworth, M. (1999). Air combat training—The effectiveness of multi-player simulation. In, *Proceedings of '99 Industry/Interservice Training Systems Conference*, Orlando, FL: National Security Industrial Association.

McIntyre, H. M. & Smith, E. (2000). Training in a synthetic environment for improved operational effectiveness in collective air operations. In, *Proceedings of 2000 Industry/Interservice Training Systems Conference*, Orlando, FL: National Security Industrial Association

Nullmeyer, R. T. & Spiker, V. A. (in press). The importance of crew resource management in MC-130P mission performance: Implications for training effectiveness evaluation. *Military Psychology*.

Smith-Jentsch, K., Ricci, K. E., Campbell, G. E. & Zeisig, R. L. (1998). Shaping mental models of the scenario-based training process: A preliminary validation of shipboard instructor training. In, *Proceedings of '98 Industry/Interservice Training Systems Conference*, Orlando, FL: National Security Industrial Association.

Thorpe, J. A. (1987). The new technology of large scale simulator networking: Implications for mastering the art of warfighting. In, *Proceedings of the 9th Industry/Interservice Training Systems Conference*, Orlando, FL: National Security Industrial Association.