AFRL-HE-AZ-TR-2006-0025

Air Force Research Laboratory

AOC EMBEDDED PERFORMANCE MEASUREMENT AND ASSESSMENT

William Stock
Brian T. Schreiber
Todd Denning
Don Cain

L-3 Communications
Link Simulation and Training
6030 South Kent Street
Mesa AZ 85212

JUNE 2006

Final Report for October 2001 to May 2006

Human Effectiveness Directorate
Warfighter Readiness Research Division
6030 South Kent Street
Mesa AZ 85212-6061

NOTICES

TECHNICAL REVIEW AND APPROVAL

AFRL-HE-AZ-TR-2006-0025

This technical report has been reviewed and is approved for publication.

// Signed //

CHAD TOSSELL, 1Lt, USAF
Contract Monitor

// Signed //

HERBERT H. BELL
Technical Advisor

// Signed //

DANIEL R WALKER, Colonel, USAF
Chief, Warfighter Readiness Research Division
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
June 2006	Final	October 2001 to May 2006

4. TITLE AND SUBTITLE

AOC Embedded Performance Measurement and Assessment

5a. CONTRACT NUMBER
F41627-97-D-5000 T.O. #33

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
62205F 63231F

6. AUTHOR(S)

William Stock
Brian T. Schreiber
Todd Denning
Don Cain

5d. PROJECT NUMBER
1123 4924

5e. TASK NUMBER
AS AS

5f. WORK UNIT NUMBER
04 01

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

L-3 Communications
Link Simulation & Training
6030 South Kent Street
Mesa AZ 85212

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Training Research Division
6030 South Kent Street
Mesa AZ 85212-6061

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL; AFRL/HEA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-HE-AZ-TR-2006-0025

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report contains a review of relevant team assessment literature for potential methods/tools/measures not already known to researchers at the Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division (AFRL/HEA); an in-depth feasibility analysis of embedded measurement for the Attack Coordinator position in the Aerospace Operations Center (AOC); and an overview of a modeling approach (e.g., MicroSaint) to measuring performance of the AOC. Each of these efforts is discussed in turn.

15. SUBJECT TERMS
Aerospace Operations Center; AOC; Command and control; Computer-based training tools; Embedded measurements; Mission Essential Competencies; Performance assessment; Performance measurement; Training methods

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			1Lt Chad Tossell
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED	28	**19b. TELEPHONE NUMBER** *(include area code)*

This page intentionally left blank.

AOC EMBEDDED PERFORMANCE MEASUREMENT AND ASSESSMENT

This report addresses:

1. a review of relevant team assessment literature for potential methods/tools/measures not already known to researchers at the Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division (AFRL/HEA),
2. an in-depth feasibility analysis of embedded measurement for the Attack Coordinator(AC) position in the Aerospace Operations Center (AOC), and
3. an overview of a modeling approach (e.g., MicroSaint) to measuring performance of the AOC. Each of these efforts is discussed in turn.

LITERATURE REVIEW

One objective was to conduct a constrained literature review as a basis for an assessment of methods and measures used to investigate team performance, with particular emphasis on application to the AOC. The objectives of this effort were twofold, and included: (a) forming a cross-classification of methods of investigation (methodology) with types of measures used in the literature, and (b) completing a step-back assessment of the likelihood that each method-measure combination can be effectively applied in the AOC.

Method.

A literature review was conducted to explore team-building, mediated team collaboration, and workspace flow and assessment methodologies. To accommodate time constraints, a consensus was reached to use the following search criteria:

1. Earliest date of published work: 1995.

2. Definition of search parameters:

   a. Specifically include authors Nancy Cooke (CERI), Joan Rentsch (University of Tennessee), and John Mathieu (University of Connecticut),

   b. Specifically include the following keyword search terms;

      (1) Computer-mediated collaboration
      (2) Network-mediated collaboration
      (3) Workspace flow and assessment
      (4) Team building

3. Exclude business literature

Prior to starting the literature search, the customer reviewed the search term list, adding and/or deleting select search terms, and approved the final search term list. A preliminary search for the three specific authors was conducted on Academic Search Premier and Lexis Nexis Academic. Only one article--by Nancy Cooke--was located.

The principal literature search was conducted using behavioral science search resources at Arizona State University (PsychInfo–of which PsycLit is an integral part). A search for post-1995 journal articles by Nancy Cooke located eight articles. However, when the keywords "computer mediated collaboration," "network mediated collaboration," "workspace flow and assessment," or "team building" were employed, the number of articles dropped to zero. A search for Joan Rentsch produced no articles. A search for John Mathieu located 18 articles. However, when Mathieu was combined with any of the given keywords, the number of articles dropped to zero. The search was modified to include just the keywords in articles after 1995. In this search, six articles were found using the keyword "computer mediated collaboration" and none for either "network mediated collaboration," or "workspace flow and assessment." Using the keyword "team building," 103 articles were located. The abstracts of these 103 articles were reviewed. Most dealt with education and/or sports teams and were judged to have a focus that was not applicable to the AOC, and were rejected. Nine articles by authors other than Mathieu, Rentsch, and/or Cooke were selected for further review. With respect to Mathieu, Rentsch, and/or Cooke, the parameters of the original search were relaxed. An attempt was made to locate all of the references identified in the name only search, whether or not the article appeared in a business or business-related journal, but still requiring the focus of the article to be somehow related to the intent of the keyword domains. If the source was a book chapter, we were constrained by time to limit our collection to those books available in the ASU library. Eleven articles by Mathieu, three articles by Rentsch, and seven articles by Cooke were retrieved. Thus, a total of 30 articles (including nine by other authors) formed the initial literature for review.

Each of the 30 articles was read and a summary judgment made regarding its applicability. In general an article was rejected from further consideration if it addressed issues judged irrelevant to measurement in the AOC. Three of nine articles by other authors were rejected for this reason, as well as five, two, and two of the articles by Mathieu, Rentsch, and Cooke, respectively. Thus, the findings and conclusions related below are based upon a corpus of 18 articles (see both accepted and rejected source lists).

<u>Accepted Sources</u>

1. Brawley, L. R. & Paskevich, D. M. (1997). Conducting team building research in the context of sport and exercise. Journal of Applied Sport Psychology, 9, 11-40.

2. Cooke, N. J., Kiekel, P. A. & Helm, E. E. (2001). Measuring team knowledge during skill acquisition of a complex task. International Journal of Cognitive Ergonomics, 5, 297-315.

3. Cooke, N. J., Kiekel, P. A., Salas, E., Stout, R., Bowers, C., & Cannon-Bowers, J. (2003). Measuring team knowledge: A window to the cognitive underpinnings of team performance. Group Dynamics, Theory, Research, and Practice, 7, 179-199.

4. Cooke, N. J., Rivera, K., Shope, S. M. & Caukwell, S. (1999) A synthetic task environment for team cognition research. Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting, 303-307.

5. Cooke, N. J., Salas, E., Cannon-Bowers, J. A. & Stout, R. J. (2000). Measuring team knowledge. Human Factors, 42, 151-173.

6. Cooke, N.J., Salas, E, Kiekel, P.A. & Bell B. (2004). Advances in measuring team cognition, in Fiore, S.M. & Salas, E. (Eds) *Team cognition: Understanding the factors that drive process and performance* (pp 83-107) Washington, DC: American Psychological Association. xi, 268pp.

7. Fischer, F. & Mandl, H. (2005). Knowledge convergence in computer-supported collaborative learning: The role of external representation tools. The Journal of the Learning Sciences, 14, 405-441.

8. Hart, R. K. & McLeod, P. L. (2003). Rethinking team building in geographically dispersed teams: One message at a time. Organizational Dynamics, 31, 352-361.

9. Hathorn, L. G. & Ingram, A. L. (2002). Cooperation and collaboration using computer-mediated communication. Journal of Educational Computing Research, 26, 325-347.

10. Huang, W. W., Wei, K., Watson, R. T. & Tan, B. C.Y. (2002). Supporting virtual team-building with a GSS: An empirical investigation. Decision Support Systems, 34, 359-367.

11. Marks, M. A., Mathieu, J. E. & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. Academy of Management Review, 26, 356-376.

12. Marks, M. A., Zaccaro, S. J. & Mathieu, J. E. (2000). Performance implications of leader briefings and team interaction training for team adaptation to novel environments. Journal of Applied Psychology, 85, 971-986.

13. Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Cannon-Bowers, J. & Salas, E. (2005). Scaling the Quality of Teammates' Mental Models: Equifinality and Normative Comparisons. Journal of Organizational Behavior, 26, 37-56.

14. Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E. & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and effectiveness. Journal of Applied Psychology, 85, 273-283.

15. Rentsch, J. R. & Klimoski, R. J. (2001). Why do ‘great minds’ think alike?: Antecedents of team member schema agreement. Journal of Organizational Behavior, 22, 107-120.

16. Salas, E., Rozell, D., Mullen, B. & Driskell, J. E. (1999). The effect of team building on performance: An integration. Small Group Research, 30, 309-329.

17. Smith-Jentsch, K. A., Mathieu, J. E. & Kraiger, K. (2005). Investigating Linear and Interactive Effects of Shared Mental Models on Safety and Efficiency in a Field Setting. Journal of Applied Psychology, 90, 523-535.

18. Tesluk, P. E., & Mathieu, J. E. (1999). Overcoming roadblocks to effectiveness: Incorporating management of performance barriers into models of work group effectiveness. Journal of Applied Psychology, 84, 200-217.

Rejected Sources

1. Conte, J. M., Mathieu, J. E. & Landy, F. L. (1998). The nomological and predictive validity of time urgency. Journal of Organizational Behavior, 19, 1-13.

2. Cooke, N.J. & Shope, S.M. (2002). The CERTT-UAV Task: A synthetic task environment to facilitate team research. *Proceedings of the Advanced Simulation Technologies Conference: Military, Government, and Aerospace Simulation Symposium* (pp. 25-30). San Diego, CA: The Society for Modeling and Simulation International.

3. Cooke, N. J., Neville, K. J. & Rowe, A. L. (1996). Procedural network representations of sequential data. Human-Computer Interaction, 11, 29-68.

4. Gilson, L.L, Mathieu, J. E., Shalley, C. E., & Ruddy, T. M. (2005). Creativity and standardization: Complementary or conflicting drivers of team effectiveness? Academy of Management Journal, 48, 521-531.

5. Heffner, T. S., & Rentsch, J. R. (2001). Organizational commitment and social interaction: A multiple constituencies approach. Journal of Vocational Behavior, 59, 471-490.

6. Rentsch, J. L. & Rentsch J. R. (2002) J-u-s-t-i-f-y to explain the reasons why: A conditional reasoning approach to understanding personality and behavior. Chapter 9, pp.223-250. in B. Schneider & B. Smith,(Eds.) *Personality and Organizations*, Mahwah, NJ: Lawrence Erlbaum.

7. Jessup, C. M. (2002). Applying psychological type and “gifts differing” to organizational change. Journal of Organizational Change Management, 15, 502-511.

8. Mellor, S., Mathieu, J. E., Barnes-Farrell, J. L., & Rogelberg, S. G. (2001). Employees' nonwork obligations and organizational commitment: A new way to look at the relationships. Human Resource Management, 40, 75-88.

9. Mohammed, S., Mathieu, J. E. & Bartlett, A. L. (2002). Technical-administrative task performance, leadership task performance, and contextual performance: Considering the influence of team- and task-related composition variables. Journal of Organizational Behavior, 23, 795-814.

10. Rawlings, Diane (2000). Collaborative leadership teams *oxymoron or new paradigm?* Consulting Psychology Journal: Practice and Research, 52, 36-48.

11. Tesluk, P. E., Vance, R. J. & Mathieu, J. E. (1999).Examining employee involvement in the context of participative work environments. Group and Organization Management, 24, 271-299.

12. Walther, J. B. (1997). Group and interpersonal effects in international computer-mediated collaboration. Human Communication Research, 23, 342-369.

Each source was read with an emphasis on the methodology and measurement attributes of the principal intervening and outcome measures. In no case was an outcome measure ignored if it was judged to have some relevance to application in the AOC. Six sources were retained that did not report results from original research. A review article by Brawley and Paskevich (1997) provided a useful summary of measures and methods in sport and exercise studies of team building. A Proceedings report by Cooke, Rivera, Shope, and & Caukwell (1999), provided a detailed description of the capabilities and flexibility of the synthetic task environment. An article (Cooke, Salas, Cannon-Bowers & Stout, 2000) and a chapter (Cooke, Salas, Kiekel & Bell, 2004) reviewed and outlined measurement issues with respect to the topics of team knowledge and cognition. An article by Marks, Mathieu and Zaccaro (2001) outlined a comprehensive, temporally-focused, functional model for team research. Finally, a meta-analysis by Salas, Rozell, Mullen and Driskell (1999) integrated research on the effect of team building on performance. Each of these six sources informed the findings and conclusions of the present effort.

Results.

The review of the 18 retained sources is summarized in Table 1. Due to page-size limitations, abbreviations are used to report information about the measures in a compressed format. These abbreviations appear in a key at the foot of the table. The information in the table is briefly summarized in the following subsections. Table 1, Summary of Relevant Information, follows:

Target Author	Source Number	Type	Observational Unit	Setting	Causality	Experimental	Type	Intervening Variable	Type	Outcome	Type
C	2	33 AF ROTC students in 11 teams (N = 3) participated in Uninhabited Aerial Vehicle (UAV) mission completed in a moderate fidelity synthetic task environment.	Team	L	DL			TP TK TK TK TmWk Tskwk Tskwk(Consensus)	SM EM,S EM,A1 PP PP PP,A1,S PP	Team Performance	OM
C	3	108 students in 36 teams (N = 3) participated in helicopter rescue missions in a synthetic task environment. Pathfinder	Team	L	DL	Type cross-training	XT	Tskwk Knowledge Tskwk Knowledge TmWk Knowledge TmWk Knowledge	CM,S CM,A1 PP,A1 PP,S	Mission Completion	OM
C	4	General description of synthetic task environment.									
C	5	Review on measuring team knowledge. Judged relevant to performance measurement in the AOC.									
C	6	Chapter on measurement of team cognition. Judged relevant to performance measurement in the AOC.									
M	11	Review article with a useful explication of a model to guide classifying, identifying, and selecting measures. Judged very relevant to the AOC.									
M	12	237 students participated in 79 three-person teams. Manipulations of team-interaction training & leader briefing were conducted prior to participation in simulated tank battles.	Team	L	DL	Interaction Briefing Novelty	XT XI XM	Interaction Interaction	CM,C SM,S	Pillboxes destroyed and/or captured	OM
M	13	140 students participated in 70 two-person teams in a low-fidelity PC-based F-16 flight simulator and completed two training sessions ending with practice missions, and then six 10-minute study missions.	Team	L	DL			TP Task MM Team MM Quality:Task Model Quality:Team Model	SM CM CM CM CM	Survival Waypoint checks Threats killed	OM OM OM
M	14	112 students comprised 52 two-person teams who flew F-16 missions in low fidelity simulators.	Team	L	DL			TP MM	SM CM	Survival Waypoint checks Threats killed	OM OM OM

Target Author	Source Number	Type	Observational Unit	Setting	Causality	Experimental	Type	Intervening Variable	Type	Outcome	Type
M	17	Persons from 47 shift-tower "teams' (N = 2+) of Air Traffic Controllers filled in PP measures of task- & teamwork, used to predict subsequent safety & efficiency measures for tower.	Tower	F	DL			TskWk TmWk	PP, C,A2 PP, C,A2	Safety Efficiency	AD AD
M	18	473 members of 88 road maintenance crews (N=5-12), plus 88 foreman and 21 managers completed a variety of PP measures.	Crew	F	CN			TmWk Self-management Leadership Performance Barriers Crew Action Crew Cohesion Satisfaction	PP PP SM PP SM PP PP	Crew Performance	SM
O	1	Review article on team-building research in sport and exercise. Judged to have moderate methodological relevance to measurement in the AOC.									
O	7	64 students participated in 2-person dyads solving complex educational problems.	Dyad	L	CN	Graphic Tool Condition	XM XS			Knowledge Discourse Patterns	PP DA
O	8	E-communications among all possible pairs of workers (N unspecified) in seven teams from three organizations were examined. Interesting finding on message frequency/length.	Team	F	CN					E-mail	DA
O	9	12 students formed four groups (N=3) to solve problems in computer-mediated environment.	Team	L	CN	Collaboration	XI			Interdependence Synthesis Independence	DA DA DA
O	10	48 teams (N = 5) solved open-end business problem in face-to-face or virtual settings with or without goal setting instructions a part of their instructed task.	Team	L	DL	Environment Goal setting	XS XM			Cohesion Commitment Collaboration Decision Quality N Decisions Created	PP PP PP PP OM
O	16	A meta-analysis. Findings judged relevant to performance measurement in the AOC.									
R	15	315 persons in 41 teams (N = 2-27)	Team	F	CN					Teamwork schema Team Effectiveness Team Experience Demographic measures	CM PP PP PP

Key

Targeted Authors (Column 1)

C = Cooke
M = Mathieu
R = Rentsch

Observational Setting (Column 5)

F = Study conducted primarily in a field or a natural setting.
L = Study conducted primarily in a laboratory setting – includes experiments, quasi- or non-experiments.

Temporal Arrangement of Conditions (Column 6)

CN = Concurrent measurement of intervening and outcome measures.
DL = At least some delay between measure of intervening and outcome measures.

Experimental Manipulations (Column 8)

XI = Experimental manipulation of instructions.
XM = Experimental manipulation of materials.
XS = Experimental manipulation of setting.
XT = Experimental manipulations in training.

Intervening Variables (Column 9)

MM = Mental Model
TK = Team Knowledge
TmWk =Team Work
TP = Team Process
Tskwk = Task Work

Attributes of Measures (Columns 10 and 12)

AD = Archive data: Public records, government statistics, work performance records, etc.
CM = Concept maps: Scaling measures including Pathfinder and multidimensional scaling.
DA = Discourse analysis: Content analysis of written or recorded discourse, including computer-mediated discourse.
EM = Embedded measures: Queries as to expectations or situation awareness.
OM = Objective measures: Time, outcome, error measurements, etc., initiated by actual behavior in real or simulated environment.
PP = Paper & pencil measures such as surveys, questionnaires, and self-reports, as well as on-line versions of same.
SM = Subject matter expert rating of target behavior, including ratings, rankings, or other judgments.

A1 = Estimate of team accuracy.
A2 = Estimate of team agreement.
C = Estimate of team consistency.
S = Estimate of team similarity.

Source, Observational Unit, Setting, Causality, and Experimental Manipulations. Cooke, Mathieu, Rentsch, and Other authors contributed two, five, one, and four empirical studies, respectively. Cooke and Mathieu each contributed one experimental study, while Other authors contributed three experimental studies. Dyads (3 studies), triads (4 studies), and teams of larger or varied size (5 studies) were the primary observational units, with laboratory studies (8 studies) outnumbering field studies (4 studies) by a two to one margin. Both empirical studies by Cooke, four of the five empirical studies by Mathieu, and one of the four empirical studies by Other authors involved at least some delay between the experimental manipulations and/or collection of intervening variable information prior to collection of outcome information. Thus, there is a reasonable effort to introduce either experimental or temporal control over the measures collected in this body of research.

Experimental manipulations included cross-training (1 study) team interaction (2 studies), settings (2 studies), tools (1 study), goal setting (1 study), and mission novelty (1 study). It was the judgment of the reviewer that these manipulations have limited relevance to performance measurement in the AOC, and are not discussed further.

Measurement of Intervening Variables. Intervening variables measured may be classified into three domains: team processes, team work, and task work. There is some variability in terminology, used to represent these domains, and alternate classifications are applied to the overall domain of intervening variables, such as team knowledge (further subdivided into team process/role knowledge and situation awareness knowledge), team and/or task mental models, and team management measures (self-management, leadership, cohesion, action, and satisfaction).

Four primary methods were used to measure intervening variables: judgments or ratings by subject matter experts (SMEs), concept mapping (of which Pathfinder is considered a related method), paper-and-pencil measures, and embedded measures. In addition, for each of these methods, additional analytic processes were used in some of the studies to derive indices that distinguished agreement, accuracy, consistency, and/or similarity. It should be noted that all but observational judgments conducted by SMEs involved data collection prior to, between, or interrupting the study missions or trials–regardless of how the work segments of the team were defined–and could not be characterized as embedded measures collected in a manner transparent to the participants. That is, these measures could not be characterized as being obtained while team work proceeded uninterrupted.

Measurement of Outcome Variables. Outcome variables included mission performance (5 studies), archival safety and efficiency indices (1 study), knowledge (1 study), discourse attributes (3 studies), and team attributes (2 studies). One investigation (Rentsch & Klimoski, 2001) actually collected knowledge, team attribute, and demographic variables in a single-shot case study involving paper-and-pencil measures, including a concept mapping task.

Six different methods of measurement were used to obtain the outcome measures among the sources; objective measurement (6 studies), archival data (1 study), judgments by subject matter experts (1 study), discourse analysis of voice communications, e-mail, or other computer transcripted messages (3 studies), paper-and-pencil instruments (3 studies), and concept mapping (1 study).

All of the studies employing objective measurement were also studies conducted in a synthetic task environment (2 studies) or in low-to-moderate-fidelity simulation environments (3 studies), and all of the individual measures (mission performance, mission completion, pillboxes destroyed, captured and/or rebuilt, survival, waypoint checks, and threats killed) in these environments were judged to be summaries at a high level of performance. That is, they are summary indices, not indices of ongoing processes. We note this characteristic in counterpoint to the breadth and depth of information collected transparently by the Performance Effectiveness Tracking System (PETS) in the F-16 test bed, Mesa, AZ.

Summary of Findings in Table 1. Among the set intervening and outcome measures, there is a substantial number of knowledge measures collected via paper-and-pencil and/or concept mapping techniques. A major thrust of these studies appears to be what team members know about their own tasks, fellow team members (knowledge of other team roles), the status of the current problem (situation awareness), etc., and how such knowledge measures mediate the overall performance of the team, as measured by global outcomes. A number of authors argue that substantial gains in understanding and control of performance may gained by the use of "mental models." Nevertheless, the focus of most of the sources reviewed was on team performance at such an abstract level that it is hard to see a direct application of these knowledge constructs to performance measurement in the AOC (that is not already more precisely articulated in the performance indicators and measurements developed by Aptima for three positions in the AOC test bed, Mesa, AZ).

A number of studies used discourse analysis to attempt to demonstrate "convergence" of team member knowledge, increases in collaboration, or to identify patterns of communication among team members. As implemented in these studies, discourse analysis was time-consuming and labor intensive. This is a drawback in the context of regular, on-going, time critical (i.e., for delivering performance debriefings immediately after training or live experiences). However, one study did employ a computer-based communication tool which suggests a method of analysis that might be employed in extended research investigations exploring the nature of communications in any AOC. The communication tool threaded all communications throughout the studies duration: The examination of threaded communications might be a productive domain of <u>research</u>, as opposed to a realistic performance evaluation tool for AOC test and training beds.

Relevant Elements of Non Empirical Sources. With respect to measurement in sport and exercise, Brawley and Paskevich (1997) noted "…investigators would need to consider the relevance and applicability of the techniques before borrowing them." This same advice is well-taken with respect to the set of measures reviewed for consideration in performance measurement in the AOC test bed. Cooke, Rivera, Shope, and Caukwell (1999) in an extended description of the Uninhabited Aerial Vehicle (UAV) synthetic task environment, stated, "Similarly, the STE should be flexible in supporting a variety of cognitive and performance measures. Specific measurement capabilities should include embedded performance measures, logs of, and post processing routines for, computer and communication events, and rapid data access for immediate analysis." While this prescription describes precisely the nature of embedded measurement in the F-16 test bed, the promise of such embedded measurement in the STE was not on display in any of Cooke's work that was located in the present search. Nevertheless, the prescription by Cooke et al. precisely captures the types and availability requirements for performance measures in the AOC test bed.

In a review and a chapter Cooke and her colleagues (Cooke, Salas, Cannon-Bowers, & Stout, 2000; Cooke, Salas, Kiekel, & Bell, 2004) provide comprehensive descriptions of measurement of team knowledge and cognition. Among others, they make two interesting observations relevant here. First, they write, "The on-line methods of observation and process tracing appear more promising for rapidly changing team knowledge and, specifically, for team situation models." Certainly, the AOC can be characterized as a setting in which relevant knowledge changes rapidly, and in which peak situation awareness by all members of the AOC will lead to more optimal performance. Second, Cooke et al. write, "The measurement of situation models and knowledge in more dynamic domains has received relatively little attention. The application of a broader spectrum of knowledge elicitation measures to the problem of measuring team knowledge should open the door to the measurement of varieties of team knowledge that have been previously unexplored." Cooke et al. (2004) explicitly call for an expansion of automated and embedded measures (p. 90-91).

The theoretical paper by Marks, Mathieu, and Zaccaro (2001) provides an excellent functional framework for conducting investigations of team processes, including the performance of individuals in specific roles. A thorough presentation of this framework is beyond the scope of the present effort. However, some of the highlights of this paper include the following. They distinguish team "processes" from emergent states, where the latter are qualities that represent attitudes, values, and cognitions, and are in a dynamic state as a function of context, inputs, processes, and outcomes. These may be contrasted with team interactions. That is, emergent states are not process because they do not describe role-based interactions among team members. Next, Marks et al. impose an input-process-output framework on the decomposition of team behavior, and go on to discuss the rhythm of team task accomplishment. Both of these conceptions descriptively fit well to the activities of a real or test bed AOC. Finally, they identify and classify ten team processes into three superordinate categories. Mission analysis, goal setting and strategy formulation are labeled transition processes. Monitoring goal and systems progress, team and backup monitoring, and coordination are identified as action processes. Conflict management, confidence building, and affect management are grouped together as interpersonal processes. This framework is judged to be quite relevant as a guide to measurement of performance in the AOC. The Marks et al. framework is used as a basis for the ideas articulated in Section 3..

In their meta-analysis of team building research, Salas, Rozell, Mullen, and Driskell (1999) found that the effect of team building effects on performance decreased as team size increased. This finding was true for both objective and subjective measures of performance. While team building interventions generally involve teams in sport, education, and business, the notion that team size should be minimized for maximal effect is consistent with Air Force objectives for manning the AOC.

Conclusions.

The measures described in sources located in our search do not involve approaches that are not already incorporated in the total package of measures elicited, either overtly or unobtrusively, in the F-16 test bed. Even the embedded, automatic measures suggested by Cooke et al. (1999) for implementation in the STE are consistent with the types of measures now collected by PETS in the F-16 test bed. Thus, for nearly all measures, making content changes (relevant to the AOC) in measures used in the F-16 test bed would result in a set of measures that map well to the measures being used in academic team research in field or laboratory settings, in STE, simulator, or real work situations.

An examination of the usefulness of threaded messages is one interesting concept that does not seem to have a counterpart in the suite of measures used in the F-16, nor, to the reviewers' knowledge, being contemplated for the suite of measures for the AOC test bed.

The literature search did uncover an important paper by Marks et al. (2001) that provides a general framework within which to conceptualize tasks, processes, and their measurement. Given current AF views and articulations regarding the AOC, this framework might provide a useful overview for guiding investigations and evaluations of performance in AOC.

FEASIBILITY OF EMBEDDED MEASUREMENT FOR THE ATTACK COORDINATOR POSITION

Introduction.

As part of the feasibility analysis, we assessed the potential for embedded measurement of the Attack Coordinator position. By embedded measurement, we mean methods and tools that perform computer-based assessment (CBA). These embedded techniques could take different forms or approaches, but once development is completed the embedded techniques result in largely unassisted (i.e., automatic) CBA producing useful measurements for both AOC operator feedback and for AOC training research (i.e., useful in aggregate forms, longitudinal use, and for development). As the effort desired was a feasibility analysis, AFRL SMEs were called upon instead of operational SMEs. Therefore, we consider the feasibility analysis reported herein as exploratory.

Method.

Participants. Both of the two participants serving as the SMEs have had operational AOC experience. One was also a current and qualified air operations integration trainer for dynamic targeting operations in the 505th Command and Control Wing.

Materials. Aptima previously conducted workshops to identify Performance Indicators (PIs) for various positions within the AOC, including the Attack Coordinator. These PIs were identified by operational SMEs as contributing to successful mission performance. As measurement of each PI is therefore desirable, the current effort focused the feasibility analysis on those 64 PIs identified by Aptima for the Attack Coordinator position.

Procedure. Each of the 64 relevant Attack Coordinator PIs was reviewed for embedded measurement feasibility. As the current effort was a feasibility analysis (i.e., exploratory), the procedure for each subsequent step was decided upon and undertaken upon completion of the prior step.

1. A researcher interviewed the SMEs to summarize the situation/context in which each of the 64 PIs could potentially be observed. The goal of this step was to determine relevant conditions for assessing that PI (i.e., context, observable behaviors, goal states, tools/people involved, etc.). Commonalities amongst the PIs were noted, and easily the most apparent of which was that, for the Attack Coordinator position, measurement on observable behaviors was not straightforward. Proper assessment for many of the PIs would require integrating information from more than one resource. The pertinent notes from this step are embedded as notes in the associated Excel document.
2. In Step #2, the resources needed for complete and thorough assessment were identified for each PI. By complete and thorough assessment, we mean that in order to fully and correctly assess a given PI, any measurement technique (embedded or otherwise) would need to take into account information from all the identified resource sources (e.g., current information on a display can only be interpreted correctly when combined with Intel information from the brief and Voice communication that took place 30 seconds prior). If information from even just one source is not taken into account, a measurement method could incorrectly assess an operator on that PI. The SMEs and researcher all concluded that the following five resources were most applicable: **V** = Voice communication; **C** = Chat (text); **A** = ADOCS; **T** = TBMCS; **O** = Other (e.g., briefing materials or working documents). This code was used in identifying the needed resources for each of the 64 Attack Coordinator PIs.

Results.

For the 64 Performance Indicators, we generated a few descriptive statistics. Table 2 shows the number of PIs that require one, two, three, four, or five resources for thorough and accurate assessment of the Attack Coordinator position.

Table 2. Number of PIs Requiring One, Two, Three, Four, or Five Resources.

Number of Resources	# PIs of 64 (%)	Cumulative # (%)
5	4 (6.3%)	5 resources: 4 PIs (6.3%)
4	7 (10.9%)	4 resources or more: 11 PIs (23.4%)
3	28 (43.8%)	3 resources or more: 35 PIs (60.9%)
2	22 (34.4%)	2 resources or more: 56 PIs (95.3%)
1	3 (4.7%)	1 resource or more: 64 PIs (100%)

Table 3 shows the number of PIs that contain each resource. That is, how many times a given resource (alone or in conjunction with other resources) is associated with a Performance Indicator.

Table 3. Number of PIs Containing Each Resource

Type of Resource for Thorough and Accurate Assessment	Number of PIs for which that resource was identified (alone or in combination) as necessary for thorough and accurate assessment.
Chat	50 of 64 (78.1%)
Voice	45 of 64 (70.3%)
ADOCS	40 of 64 (62.5%)
TBMCS	21 of 64 (32.8%)
Other	23 of 64 (35.9%)

Discussion and Recommendations.

During the process of determining opportunities for embedded measurement, the SMEs identified the AC position as the most difficult position within the AOC for such an effort. The results obtained from the resource analysis clearly back the SMEs' opinion that the AC position presents significant embedded measurement challenges. Embedded measurement is most feasible when assessment can be accomplished by using a single resource, but the results of the resource analysis in Step #2 reveal few opportunities for this. A full 95.3% of the PIs require two or more resources for assessment. Though embedded measurement can be accomplished by integrating information from multiple resource sources, the complexity and associated time, effort, and financial commitment is substantially higher to do so.

Just 3 of the 64 AC PIs require a single resource for thorough and complete assessment. Interestingly, in Table 3 we see that more PIs involve Chat than any other single resource (78.1%), but Chat was a single resource in none of the PIs. For the three single source resource PIs, each are discussed in turn:

T2-2, "Recognize current status of ATO execution—refresh rate": Changes are not frequently made in TBMCS. When they are, the SME assessment considering the way changes are made is that they are very rarely wrong. Good target for embedded assessment here is assessing possible errors by continuously monitoring refresh rates of ESTAT. Long refresh rates, especially during critical times (e.g., at the initiation of a DT event, when the ATO data would have to be at the most current to provide effective optimization of the solution) would be considered poor performance. Shorter refresh rates would be inferred as good performance.

T2-27.1, "Review risk guidance for target and target environment": Mostly cognitive. If AC feels comfortable with knowledge of guidance, no visible action would be taken. Only if AC feels unsure of specific guidance for the attack he/she is planning, he/she may check the guidance to improve planning confidence and effect. Either way (mentally or visually on the data wall HUD) the AC will check guidance outside measurable domains, so no real opportunity for assessing degrees of performance exists. No identifiable method for embedded measurement.

E-8.1, "Select appropriate COP display": A possible measurement strategy could pull the linked track numbers of tasked assets (to access the accompanying information of position) and the position of the target to ascertain if the display field of view includes the area of interest for mission monitoring. However, new mission tasks could necessarily divert the AC to another spot while he/she has delegated tracking of the completed planning task to a superior (Dynamic Targeting Cell [DTC] Chief) for mission monitoring, leading to potential measurement errors.

Because of the large percentage of PIs requiring multiple and diverse resources for thorough assessment, the SMEs and researcher concluded that a fair portion of the performance for the Attack Coordinator position could be ascertained by observing errors in performance. It was concluded by both SMEs and the researcher that much of the Attack Coordinator's job is cognitive and good performance is often inferred. This is done not so much by what was observable, but rather what was not observed (i.e., a *lack* of observable errors). Hence, with remaining contract time and funding, Todd Denning and Brian Schreiber are systematically identifying those observable behaviors that would be considered as errors and/or would likely highly correlate (negatively) with AOC performance. This parallels some of the efforts undertaken in the air combat performance measurement work (e.g., counting communication step-overs as errors, which have been demonstrated to be significantly different on Monday vs. Friday benchmarks). These error measurements for the AC position could then be linked to the Mission Essential Competency (MEC) skills and PIs.

Upon critical review of the potential error measures and pending a government's positive decision to pursue error-based embedded measures, we would then recommend that where viable error measurements were identified, a software engineer should investigate the specific software inputs/outputs of those error measurements to determine whether or not the appropriate software code is available (i.e., not proprietary) in order to program those error measurements.

By examining errors, we would remove some of the integration challenges associated with PIs having multiple resources identified. Pending positive results from the software engineer investigation, an embedded error analysis could be a promising approach. Many of the PIs have clear indications of likely errors that would negatively correlate with mission performance. Some early identified examples include: Failure to access and refresh the ATO during an evaluated event, failure to acknowledge tasking of a DT event from the DTC chief, and failure to present an attack solution. Notice that, in contrast to the thorough and accurate assessment approach described earlier requiring multiple resources, tracking the errors associated with a given PI often requires just one resource. This error measurement approach would greatly simplify the embedded measurement and reduce development costs. But, this approach also has drawbacks with a system such as the AOC.

Though the error measurement approach appears to be a viable embedded CBA approach, there exist some serious potential drawbacks. First, any embedded technique (error analysis or other method) relying on host AOC software is directly tied to that software. That is, any change in the host software (i.e., ADOCS, TBMCS) can directly impact the embedded measurements. As an example, if a given embedded software package currently utilizes information from a particular window in ADOCS, that window may be modified (or may not even exist) at the next ADOCS software update. Given that changes to the host AOC software will almost certainly occur (e.g., ADOCS to JDOCS, TBMCS to TBONE, and future updates) and the fact that a full 71.9% (46 of 64) of the AC PIs rely on either ADOCS and/or TBMCS, consequences of deriving embedded measures from the host software should not be underestimated. This inevitable evolving software configuration has several potential negative impacts on long-term standardized embedded measurement, the most serious threat to competency-based research being #3 below:

1. If a new version of the host software is released, one or more measures could easily need to be modified. In the best case scenario, there would only be a short-term delay while measurement modifications/updates are made to reflect the host software changes. In the worst case scenario, these modifications could cause additional technical issues, creating potential additional contractual and/or financial obligations. Either way, time gaps in capturing the measures would result.

2. If a new version of the host software substantially changes the function (or its execution) underlying the measurement for a given PI, some prior embedded measurements may disappear altogether or would have to substantially be recoded. Resulting issues here are the same as those for #1 above, but to a greater extent.

3. Ramifications from #2 (and to a lesser extent, #1) are serious. If the measures change, then longer-term standardized competency-based assessment is severely compromised. Individuals over time would not be assessed with the same metrics for those PIs/skills. Longitudinal research of any form becomes problematic and <u>competency-based assessment across individuals over time would not be standardized</u>. Furthermore, changing of metrics also complicates opportunities for moving towards an adaptive training system in a continuous learning environment.

To somewhat mitigate this challenge, embedded measurements tied to the host AOC software should be metrics a level or two up in abstraction. As a comparison, effort in the air combat work specifically avoided becoming too platform or software-specific. In air combat, the metrics for skills relating to weapons employment, for example, are the same across missile types and platforms—clear avenue of fire, range preservation (A/F pole), launch parameters, etc., and not, for example, focusing in on exact WEZ display symbols within the F-15 avionic suite. These conceptual metrics at that abstraction level will result in the same metrics and calculations over time. We recommend a similar approach to be undertaken for most measures within the AOC. Though a slightly higher level of abstraction does somewhat mitigate software changes, it does so only to a minor extent. To elaborate by again comparing to the air combat work, most of the air combat work is derived from DIS/HLA data—standards that change very slowly, and most existing data (e.g., TSPI) will likely never go away in those standards; changes there are likely to be expansions, which only help long-term standardized measurement. Embedded metrics for the AOC do not have a similar luxury, as updates to the host software are likely to occur more frequently and therefore impact any embedded software system and its associated competency-based measures.

How PETS might relate with AOC measurement. As it is based on an architecture of Distributed Interactive Simulation/High Level Architecture (DIS/HLA) network traffic, the WRAPMTS ATD performance measurement work up is largely inappropriate for the AOC, with two very notable exceptions.

1. Assessment of "Deconfliction." The new architecture within PETS allows for tracking of all entities—red, blue, planes, munitions, tanks, ships, etc. This tracking counts all entities, keeps track of times, and categorizes according to force and type. Prior work on air combat clear avenue of fire measures have resulted in developed algorithms for assessing potential conflicts in time and proximity (i.e., for air combat it was potential fratricides). By leveraging these efforts, we could easily calculate, for example, a fly-out of a Tomahawk missile and its proximity to other airborne assets. *All* potential pairwise space and weapons fly-out conflicts could be automatically calculated. One could also determine how many and what types of munitions were dropped in the same space at the same time. To assess the deconfliction, WRAPMTS would <u>not</u> even have to be embedded in the AOC (indeed, that would only complicate it), but rather be anywhere on the DMO network exercise that an AOC (e.g., DMOC) is influencing. That is, deconfliction violations would be captured off the DMO network directly. (Note: There may be other opportunities for the other positions, but only the AC position has been investigated thus far and deconfliction surfaced as relevant).

2. As part of integrating/leveraging to the WRAPMTS ATD, it would be highly desirable for all measurements of all individuals to be collected by a single system. The architecture of PETS has been built in theory to do this for any size engagement with any number of players. Even if the measures are not directly taken by PETS (e.g., SPOTLITE or John Hopkins software), the outputs from those other assessment devices could be sent over to PETS via its own local network (an architecture to do this is provided in the Appendix). By doing so, a central performance measurement relational database repository could theoretically exist. This central database for all players would better allow reaching the vision of attaining adaptive, continuous learning for all the players involved, as it would then contain all the performance data needed for learning management of the various participants. That is, a single system would be tracking the performance of all individuals involved in the exercises, and that single system would then be best positioned to calculate training gaps in proficiency. And, only with a centralized single data collection source could training gap trade-offs between participants be spotted. If a central performance database is not created, the automated learning management will become a piecemeal effort according to each stovepipe assessment solution for

given sites and/or weapons systems. A stovepiped approach would likely result in a higher probability of less efficiency in training for all players simultaneously involved (i.e., some players more likely to serve as partial or full training aids). Of course, to accomplish a centralized dataset for centralized learning management, a common link for PETS back to the DMO network would also be required, else a local network to PETS is meaningless. However, as desirable as this end state may be, it is a long term vision. For learning management to occur on a larger scale such as that described, robust data collection and tracking functions of all those players/sites needs to be resolved first—a very challenging task on multiple levels (e.g., security, policy, privacy, etc.).

Due to the complexity of assessing many of the PIs for the Attack Coordinator, it is our recommendation that the majority of PIs be assessed using a subjective tool such as SPOTLITE. The subjective methods will result in the lower initial and ongoing development costs and have the additional benefit of being at a higher level of abstraction, resulting in stable, consistent competency-based metrics over time that can be used again and again without fear of changes due to software upgrades. To aid the burden on an SME evaluator, each PI and skill should be systematically evaluated for "easy target" assessment opportunities using an alternative method. For example with the AC position, the John Hopkins software may easily alleviate the need for an SME to evaluate timing-related measures and ADOCS status changes, we could (assuming software access) easily develop an embedded measure for T2-2, "Recognize current status of ATO execution—refresh rate", and the PETS software may be able to automatically assess potential deconfliction issues that arose. The error analysis, when complete, may show easy targets for some other select PIs/skills. It is also our recommendation that given the highly compensatory team nature of actions and decisions within the AOC that additional research effort be specifically taken to examine the AOC team as a whole system.

A RESEARCH PROGRAM FOR INDIVIDUAL AND TEAM PERFORMANCE MEASUREMENT AND MODELING

Background

As a weapon system, the AOC translates JFC/JFACC guidance into a continuing sequence of discrete, multi-component, Air Tasking Orders (ATO). Within an AOC, authority is vertical, and at each level of authority, function and personnel are separated horizontally. Inputs, outputs and communications flow: (a) up and down levels the levels of authority, and (b) across levels of function and responsibility. A network of software/hardware tools creates a Common Desktop Environment accessible by all personnel. The profile of a specific AOC depends upon the theater of battle for which it is organized. Thus, different divisions, as well as cells within divisions have greater or lesser roles in a specific operational context as a function of the conflict and resources devoted to its resolution. This flexibility creates substantial challenges for: (a) selection and training personnel for the to-be-filled roles, (b) setting performance standards for each role, (c) measuring, evaluating, and documenting performance at an individual, cell, division, and global level, and (d) developing a model to optimize the form and processes of an AOC.

In designating the AOC a weapon system, one should not overlook the fact that the AOC is a team of individuals primarily engaged in decision-making activities, said activities resulting in a continuing sequence of discrete ATOs. The “parts” of an AOC are people whose primary AFSC are for other weapons systems. At the intersection of a level of authority and an area of function is a person fulfilling specific responsibilities. Let us label any such intersection a node – each node populated by one or more persons supported by appropriate software and hardware tools – all of whom have the same function. Optimally, each node operates at a high level of competence by making timely, error-free, decisions that move specific JFC/JFACC guidance (AOD-ATO-TAA) to its end state.

Figure 1 notionally depicts activity occurring at each node of an AOC in an input-process-output (IPO) framework commonly used by academic and industrial researchers who investigate the behaviors of work and other teams (see, for example, the article by Marks, Mathieu, and Zacarro, 2001). Examination of Figure 1 suggests that decisions are shaped by environmental factors like prior constraints and current circumstances, by the available support systems, and by the decision rules and processes in effect. Assuming each node may be meaningfully differentiated in IPO terms, then any particular AOC may be notionally represented by stacking and concatenating requisite nodes. [Hence, more realistic versions of Figure 1 provide a template for quickly building realistic staffing depictions for a given AOC.

A Theoretical Framework

Marks, Mathieu, and Zacarro provide a flexible framework to assess current training research efforts in the AOC. For example, the effort to identify MECs for nodes in the Dynamic Targeting Cell (DTC) produced a comprehensive specification of knowledge and skills (for the DTC Chief, Target Duty Officer, and Attack Coordinator). According to Marks et al., these MEC efforts as an excellent first steps in the specification of node taskwork, hence *MECs are the point of departure for performance assessment of individuals at their taskwork.* However, other aspects of the Marks et al. framework suggest opportunities to expand performance assessment to the level of cells, divisions, and AOC team.

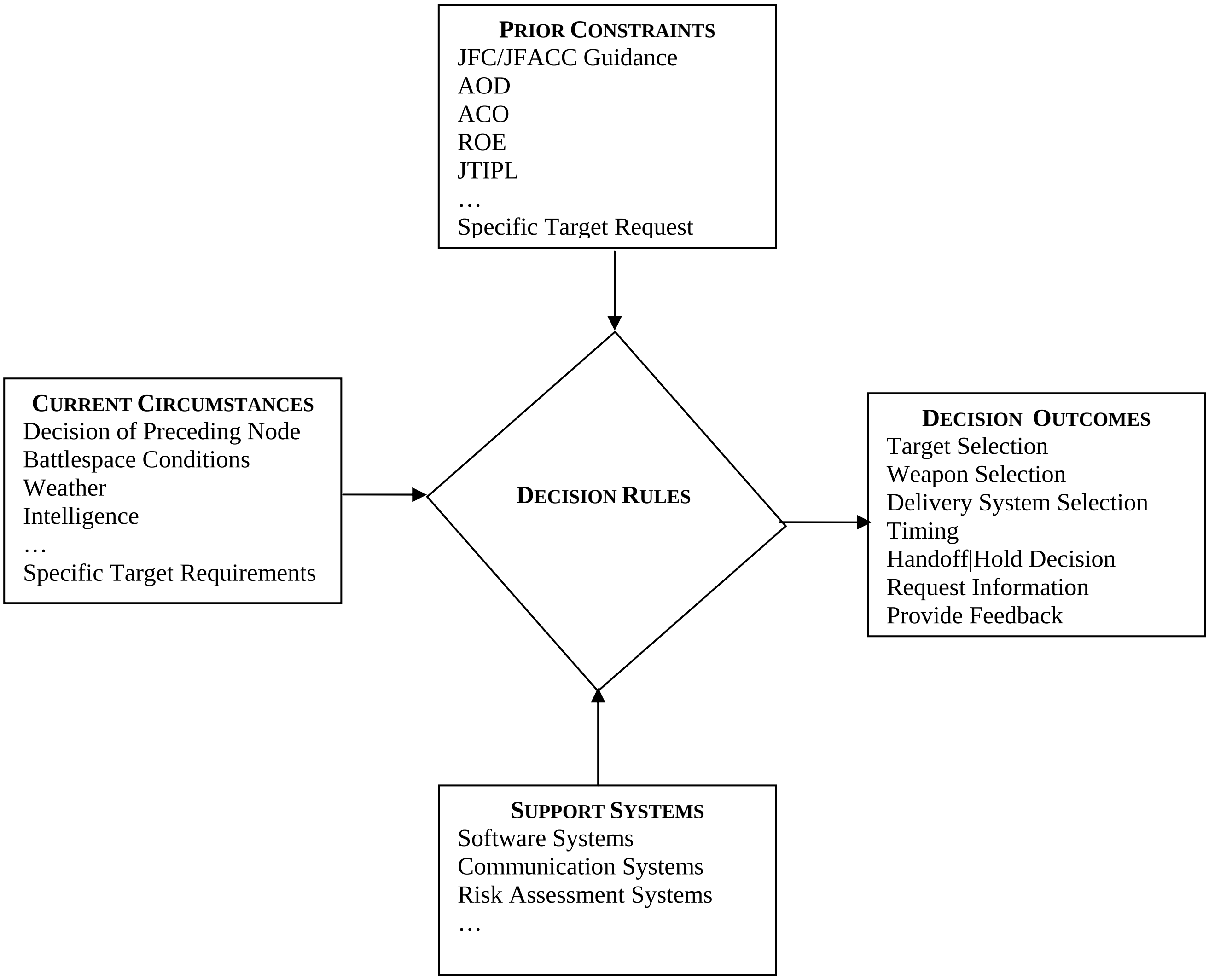

Figure 1. A Notational Depiction of the Node Environment in an AOC

The following elements of the Marks et al. framework suggest a number of lines of research and training investigations of AOC teams and cells:

1. Team activities are composed of episodes having identifiable beginnings and endings. This notion is based on the view that teams "perform in temporal cycles of goal-directed activity". Although Marks et al. make no reference to the AOC, it is clear that "episode" provides an excellent conceptual overlay for activities of individuals in an AOC, independent of the node they occupy. Further, episode is a construct with considerable flexibility, thereby offering a way to characterize activities at a variety of levels of specificity. Thus, the development of a specific ATO in the AOC, the development of the JIPTL, and the disposition of a specific time-sensitive-target in the DTC each can

both be construed as episodes – at different levels of granularity. The notion of an episode maps directly into the need to quickly and automatically construct training scenarios.

2. Similar team performances reoccur across time. That is different episodes must be worked from beginning to end, and teams may work multiple episodes that are running simultaneously, and at different time synchronizations. Again, although Marks et al. make no specific references to the AOC, it is clear that their description provides an excellent overlay for the activities of any AOC team. Marks et al. push this recurrent activity analysis another level by suggesting that each task episode is comprised of a sequence of transition-action phases, and that each transition or action phase is comprised of an IPO sequence. These descriptions also closely match activities in the AOC at various levels, e.g., between divisions, among cells of a division, or between individuals within a cell. Further, the degree of specificity in the Marks et al. framework provides a theoretical framework with sufficient specificity to model activities of a putative AOC team in such temporally-oriented COT simulation software as Micro Saint Sharp.

3. Ten team processes, listed in three super-ordinate categories, are identified as essential to team performance. Mission analysis, goal setting and strategy formulation are transition processes. Monitoring goal progress, monitoring systems progress, team and backup monitoring, and coordination are identified as action processes. Conflict management, confidence building, and affect management are grouped together as interpersonal processes.

We believe identification and classification of "episodes", appropriate delineation of the particulars of team processes, combined effective assessment of individual task work (as defined by MEC specification) will provide the basis for understanding and building AOC teams that perform at optimal levels. As an initial first step, a line of potential research assessment is offered. The research is described in terms applicable to the DTC test bed, Mesa, Arizona.

Specific Research Steps in the DTC Test Bed

The following steps are arranged in their approximate temporal order. It is assumed that the MEC specification process for the nodes of the DTC is complete and essentially implemented with respect to the measurement of individuals.

1. By interview and survey, build a relatively comprehensive list of the "episodes" that comprise the workload of a DTC. In doing so, rank order the episodes from most important and/or most frequently occurring to least important and/or least frequently occurring (i.e., identify the most important episodes).

2. For a small set of the most important episodes, complete a decomposition of the episode into a sequence of transition-action phases. This analysis is completed separately for each DTC node, and involves the following sub steps:

a. Identify important IPO attributes/activities within each transition and action phases, and
b. Identify the team processes that are most likely to be used to successfully complete each IPO action, and
c. Identify what types of behavior typically represent the application of a team process.

3. Construct a set of scenarios matching the episodes chosen at Step 2. Run a small number of teams through these scenarios to provide a data base from which to obtain empirical estimates of time to complete (and other indicators of work flow).

4. [This step may occur in tandem with step 3] Translate the episode-IPO-process specification and decomposition into simulation modeling terms provided by a COT simulation program (e.g., Micro Saint Sharp), thereby creating a model that can be elaborated and refined over time – as the episode-IPO-process space (increasing the set of possible events and actions) and the node space (adding positions) increase. Given fine-grain episode-IPO-process specifications for each node of an AOC team, and reasonable empirical estimates of typical and optimal times to complete, any coherent subset of nodes can subsequently be modeled via the process simulation software system prior to the spin-up of a specific AOC team, i.e., as nodes, and their respective behaviors are modeled, one gains the capability to provide simulated data with which to compare actual outcomes observed in training and exercise environments. As models for individual nodes (e.g., the Combat Operations Chief or the Attack Coordinator), and concatenated nodes (e.g., the Dynamic Targeting Cell) become more robust – with input of performance data derived from the observation of experts operating in these nodes/concatenations – an evolving and improving model of ideal AOC performance emerges. The elegance of this line of research is that it may be: demonstrated at multiple levels (node, cell, multi-cell [e.g., division], AOC team), while simultaneously suggesting performance standards at these levels. Creation of a general model that is improving and evolving provides a formal basis by which to judge training and performance efforts.

5. As an additional beneficial line of research, this effort allows for testing new, alternative training configurations (or technologies) into the AOC. With a large, robust modeled data base in place serving as a cohort dataset, effectiveness evaluations of the alternative methods become possible. New and effective alternative training techniques, methods, or tools shown to be valuable compared to the cohort models could spawn entire new lines of applied research to benefit the operational warfighter. Additionally, the models enable investigation of potential faults. That is, faults could be interjected into the models at specific points to determine their impact on the entire AOC system as a whole (negative impacts would be measured by time delays, changes in transitions, and/or bottlenecks). For those fault areas identified as having the greatest negative impact on overall AOC performance, those areas should receive additional targeted training through MEC-designed scenarios and syllabi. Additionally, those fault areas could provide a launching point for research and development on future 6.3 training tools to mitigate the impact of those critical system fault areas.

This page intentionally left blank.

APPENDIX

AN ARCHITECTURE FOR ASSESSMENT DEVICES TO BE SENT OVER PETS VIA ITS OWN LOCAL NETWORK

For the majority of the objective performance measurement work, we propose adding a new software capability. To date, the Performance Effectiveness Tracking System (PETS) software has focused on utilizing the actions taken by individuals and shared in Distributed Interactive Simulation/High Level Architecture (DIS/HLA) network data, augmented them with algorithms, rule sets, and look-up tables to assess performance within the air combat domain. However, individuals' actions within the Aerospace Operations Center (AOC) are not shared across common DIS/HLA networks, so this DIS/HLA approach cannot be utilized as the primary underlying methodology; the need for AOC measurement dictates a new approach.

New approaches such as an embedded error analysis or outputs from John Hopkins software could be sent over to PETS for centralized data collection. This would be accomplished by using an alternate port as part of the larger DIS/HLA network (black line in Figure 1). This enables summary data of each operator from the AOC measurement network back to the PETS software, thereby allowing integration and time-stamping of AOC operator data with data from other network participants.

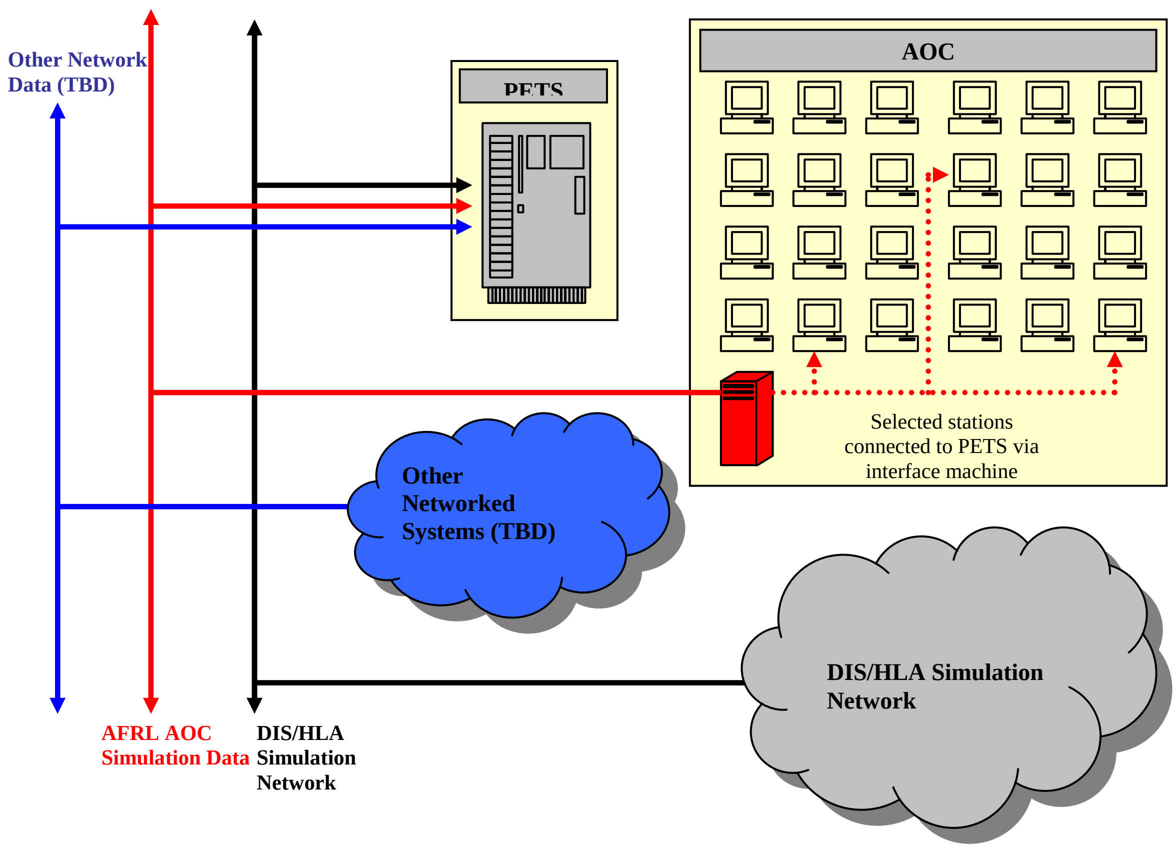

Figure 1. Interoperability Diagram.

This approach allows us to capture the assessments of an AOC operator AND allows us to maintain a centralized data assessment and collection system. Furthermore, to complete this centralized assessment system, the SPOTLITE tool, if resident on a SECRET level tablet, could download its data into this new AOC network component of PETS for integration into the common assessment database. For the new PETS component capability only, a new AOC PETS interface machine would need to be procured.

Referring to Figures 1 and 2, we could create an AOC network (red line & red text) for measurement purposes. On this new AOC measurement network, new code (either locally developed or leveraged from a tool such as John Hopkins) specifically written to assess skills for a particular AOC individual would be developed and installed on that individual operator's machine.

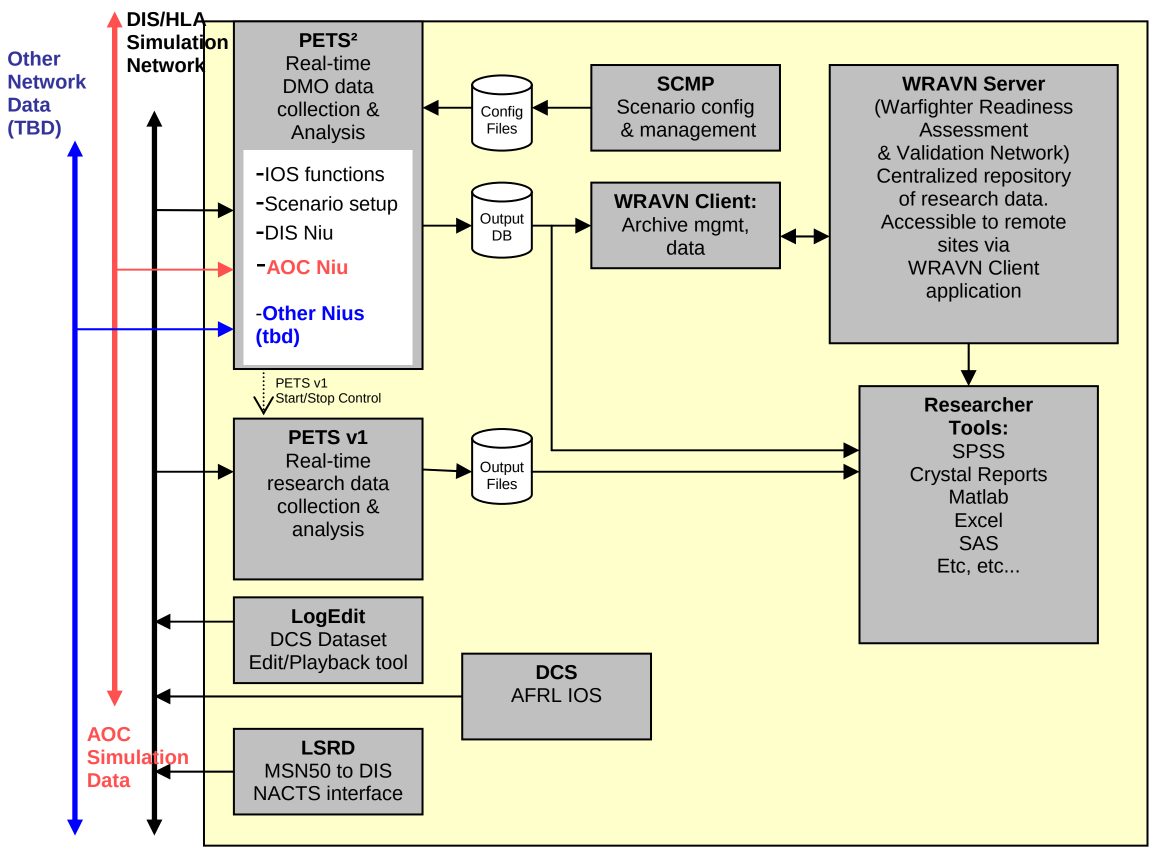

Figure 2. PETS Logical Component Diagram.