Computerized Tests of Team Performance and Crew Coordination Suitable for Military/Aviation Settings

Ben D. Lawson; Thomas W. Britt; Amanda M. Kelley; Jeremy R. Athy; Shauna M. Legan

INTRODUCTION: The coordination of team effort on shared tasks is an area of inquiry. A number of tests of team performance in challenging environments have been developed without comparison or standardization. This article provides a systematic review of the most accessible and usable low-to-medium fidelity computerized tests of team performance and determines which are most applicable to military- and aviation-relevant research, such as studies of group command, control, communication, and crew coordination.

METHODS: A search was conducted to identify computerized measures of team performance. In addition to extensive literature searches (DTIC, Psychinfo, PubMed), the authors reached out to team performance researchers at conferences and through electronic communication.

RESULTS: Identified were 57 potential tests according to 6 specific selection criteria (e.g., the requirement for automated collection of team performance and coordination processes, the use of military-relevant scenarios). The following seven tests (listed alphabetically) were considered most suitable for military needs: Agent Enabled Decision Group Environment (AEDGE), C3Conflict, the C^3 (Command, Control, & Communications) Interactive Task for Identifying Emerging Situations (NeoCITIES), Distributed Dynamic Decision Making (DDD), Duo Wondrous Original Method Basic Awareness/Airmanship Test (DuoWOMBAT), the Leader Development Simulator (LDS), and the Planning Task for Teams (PLATT). Strengths and weaknesses of these tests are described and recommendations offered to help researchers identify the test most suitable for their particular needs.

DISCUSSION: Adoption of a few standard computerized test batteries to study team performance would facilitate the evaluation of interventions intended to enhance group performance in multiple challenging military and aerospace operational environments.

KEYWORDS: team performance, military performance, group performance, shared cognition, performance assessment.

Lawson BD, Britt TW, Kelley AM, Athy JR, Legan SM. *Computerized tests of team performance and crew coordination suitable for military/aviation settings*. Aerosp Med Hum Perform. 2017; 88(8):722–729.

While team performance in the military aviation setting has received attention, performance is critical to the success and safety of military groups in nearly any challenging environment. For example, many of the fundamental cognitive challenges faced by helicopter aircrew performing a sustained operation overlap with those faced by a squad of ground troops occupying a defensive position on a mountain ridge. Efficient shared information processing, communication, and coordination are critical to operations performed by infantry, armor, artillery, aviation, special forces, logistics, medical services, military intelligence, and communications personnel. Faulty team processes are well-documented as a contributing factor to aircraft mishaps[21,35] and command/control-related accidents during aviation and other operations.[2]

As research on performance expanded from a consideration of individual information processing to shared cognition,[29] researchers developed computerized measures to quantify team performance. Better measures of team performance can benefit military and aviation organizations by helping to determine the characteristics of good teamwork, quantifying the effectiveness

From the United States Army Aeromedical Research Laboratory, Fort Rucker, AL.

This manuscript was received for review in November 2016. It was accepted for publication in March 2017.

Address correspondence to: Ben Lawson, Naval Submarine Medical Research Laboratory, Naval Submarine Base New London, Box 900, Groton, CT 06349-5900; benton.d.lawson.civ@mail.mil.

DOI: https://doi.org/10.3357/AMHP.4786.2017

of team training methods,[3] and aiding the development of cost-effective training simulations.[36]

Although many computerized tests have been developed to measure military-relevant team processes and performance, no test dominates the field of inquiry in the way that individual cognitive performance measurement has tended to be dominated by tests derived from the Unified Tri-Service Cognitive Performance Assessment Battery (UTC-PAB).[14,31] In particular, there is limited information concerning which tests are optimal for small teams performing complex missions in extreme environments or demanding situations. The authors are not aware of any comprehensive review of computerized tests of military team performance appearing in an academic journal. The most recent and relevant comparative review in the gray literature was written 10 yr ago by Lamoureux et al.,[22] who wrote a technical report covering 44 potential test platforms. Prior to that, Banner[4] reviewed 7 team performance tests in a brief technical report, while Bowers and Jentsch[6] contributed an indirectly relevant chapter on the potential suitability of 36 commercial computer games for adaptation to team research.

The absence of comparative information on computerized team performance tests contributes to a lack of uniformity in the measurement of team performance and limits comparisons across studies. We conducted a systematic literature review to assess the current state of computerized team performance tests and identify those most suitable for military-related research. We defined the term "test" broadly to refer to computer programs or simulations that automatically collect information on team processes and performance. Our review focused on low-to-medium fidelity systems,[4] which are more widely and easily disseminated than "high-end," custom-built, "one-of-a-kind" simulation facilities, due to factors such as affordability, portability, and configurability.[6,10]

Various approaches have been used to quantify team performance, including surveys, behavioral checklists, and computer tests or simulations.[7,9] This review focused on quantitative computerized tests of team performance based on shared information-processing tasks. This is not a comprehensive assessment of the field of team research, but rather a practical guide to the use of computerized team performance tests. The literature abounds with reviews of the history and state of team performance research, theories concerning team performance, discussions of team coordination and team training approaches, and recommendations concerning what a good measure should be able to do. The present review answers an important but less frequently raised question: "What automated small team performance measures exist that are practically and immediately suitable for military-related research?" To that end, a short list of the most suitable tests for research on military-related teams was identified, in the hope that future team research will become more focused and standardized. The chosen tests assessed important team processes identified by prior researchers (e.g., communication, coordination, planning[26]), as well as indices of team performance.[9] Providing researchers with information on the most appropriate computerized tests of military team performance should facilitate research on important knowledge gaps in teamwork and team dynamics.

METHODS

The present study used a narrative literature review process based upon the guidelines for systematic reviews.[25,28] A four-step process was used to identify those computerized team performance tests relevant to military settings. The steps are described below.

Step 1 involved a comprehensive literature search of articles published in 1990 or later in PsycInfo, PubMed, and the Defense Technical Information Center. The primary search terms were "team" and "performance." Professional librarians assisted, basing their search on an idealized set of criteria exemplified by the following compound Boolean statement:

[Objective / Performance-Based / Quantitative /Automated / Computer / Computerized (NEAR) Measure / Metric / Battery / Test / Assessment / Task / Technology]

(AND)

[(Team / Group / Shared / Squad / Crew / Cockpit) Performance (NEAR) Effectiveness / Cognition / Cognitive / Decision / Attention / Coordination / Resource Management].

Three research psychologists assisted the librarians with the search and then reviewed the literature, excluding items that were unrelated to the focus of the review. The qualitative exclusion criteria rejected items that did not describe the measurement of team cognition or performance, and items that dealt solely with business/management issues, such as team building, personnel selection, or leadership. In addition, items were excluded that dealt exclusively with physical performance (e.g., strength or endurance), unless they also included a cognitive or psychomotor aspect. The review sought to identify computerized performance tests for modern research efforts, so literature prior to 1990 was excluded. Articles before that time would have preceded the mainstream use of personal computers, sophisticated graphical user interfaces for Microsoft© Windows, and hypertext markup language.

Step 2 of the literature review involved obtaining full text articles for a more detailed evaluation. The initial and detailed evaluations gave preference to articles that came closest to meeting the following inclusion criteria:

1. Reports that described systematic, quantitative, computerized performance tasks (as distinguished from surveys, observer-based methods, task analyses, theoretical papers, computer games, or training improvements or guidelines).
2. Reports that described military-relevant tasks (vs. tasks only applicable to nonmilitary situations or general measures of cognitive state). Greatest consideration was given to tasks/measures that appeared to be low/medium-fidelity simulations similar to many tasks that must be performed by military groups (e.g., aircrew, squads).
3. Reports that were relevant to small groups (vs. entire companies or agencies), with a small group being defined as 10 people or fewer.

4. Reports relevant to real-time shared information processing (vs. long-range or strategic planning).
5. Reports that described generalizable tasks or tests not limited to one mission, platform, experiment, project, facility, or course of training/simulation.
6. Reports describing only the more recent versions of a given test (e.g., NeoCITIES instead of the older CITIES). (This last criterion was less important, since most tests have kept their original name as they were modified or have only changed names once.)

Step 3 of the review involved the development of a military technical report summarizing the results of the literature search and the identification of an initial list of final performance tests based on the inclusion criteria identified in Step 2.[24]

Step 4 involved determining whether we had missed any relevant tests or if our information was otherwise incomplete. This was accomplished by presenting the initial findings to fellow researchers at science conferences such as the Technical Cooperation Program's Defense Human Systems Symposium[15] and the Aerospace Medical Association meeting.[23] We also contacted (via email) many team performance researchers. Our purpose was to reach out to team/human performance researchers to ensure we included any relevant findings. Step 4 also included a final check of the literature (as in Step 2), prior to the completion of this manuscript, to identify any further developments (occurring after the Lawson, Kelley, and Athy technical report[24]) concerning computerized tests that met the inclusion criteria in Step 2.

RESULTS

The literature search identified 571 abstracts for further review. From this preliminary list, 73 items merited a detailed review according to the inclusion criteria, and so were obtained as full-text articles and reviewed in detail. These 73 reports contained information regarding 54 potential team performance tests that were summarized in a technical report.[24]

Two main trends were noticed while examining the 73 potentially relevant reports. First, the majority of the initial literature matching the stated search terms (see Methods) consisted of business or management reports, which were opinion pieces of a philosophical, inspirational, or otherwise nonscientific nature (and therefore did not meet the aforementioned inclusion criterion 1 in the Methods). Second, many of the initial matches that appeared to discuss pertinent tests did not actually yield full-text articles describing generalizable tests or test batteries, but rather descriptions of specific research projects, laboratory facilities, or problems surrounding the measurement of team cognition (thus failing inclusion criterion 5 in the Methods).

Based upon the inclusion criteria, the 54 potential tests were narrowed down to 8 tests deemed unanimously (based on the opinion of the first 4 authors) to be most relevant for military research needs. Unfortunately, three of the eight most relevant tests appeared in the literature a dozen or more years ago and did not appear to be readily available for widespread use by researchers. These tests were the Tactical Naval Decision Making System,[12] the Team Interactive Decision Exercise for Teams Incorporating Distributed Expertise,[19] and the Team Performance Assessment Technology.[37] For this reason, we narrowed our list from Steps 1–3 of the search to the five remaining tests.

During Step 4 of the search, we identified four additional tests that were not included in the original search (as well as an updated version of one of the five team performance tests that had been identified). An inspection of the four additional tests revealed that two of the four tests adhered to the inclusion guidelines described in Step 2. Therefore, these two tests were added to the five tests identified in Steps 1–3 of the search, for a total of seven relevant computerized tests. **Table I** shows the seven tests, summarizes important features of each test, lists advantages and limitations of each test, describes the maturity of each test, and provides contact information for procuring the tests for use in research. Mature tests are those that have been in use for a longer period of time and have been used in more research studies and laboratories. While the tests represent contexts for examining the effects of manipulated variables, traditional psychometric indices of reliability and validity were frequently not provided.

The reader should note that throughout this review, the authors observed a paucity of information in many of the publications concerning these tests. When a report described a potentially relevant team performance test (or tests) for inclusion in our review, practical information was often insufficient to make inferences concerning the ease of test administration, time required for testing, ease of access (e.g., is it open-access or available commercially "off-the-shelf?"), maturity of the test (e.g., is it widely used, established, reliable, and valid?), extent of automated and objective scoring, generalizability of the test, or configurability of the test (to different tasks or team sizes). For some of the older reports, it was difficult to determine (by additional web searches or e-mail inquiries) whether the test was compatible with the latest hardware/software, whether there was continued development and use of the test, whether the report described the latest version of the test, or whether the test is available for use or purchase. This lack of practical information has been a problem for human performance measurement in general, which has made applied sources such as the *Human Performance Measures Handbook*[17] particularly valuable.

This section describes the most relevant computerized tests of team performance identified by this review. Below, we briefly present each of the seven tests listed in Table I (in alphabetical order). Six of the tests (i.e., every test except DuoWOMBAT) focused on various aspects of team performance relevant to command and control situations.

Agent Enabled Decision Group Environment (AEDGE)

The AEDGE simulates a weapons director team of an Airborne Warning and Control System (AWACS). This scenario embodies the core characteristics of command/control, including surveillance and communication. Participants must exchange, interpret, and weigh information while coordinating their tactical actions to successfully accomplish the mission.[22]

Table I. Overview of the Final Seven Computerized Team Performance Tests.

TEST	FEATURES	ADVANTAGES	LIMITATIONS	MATURITY	AVAILABILITY
AEDGE[5]	• Military command/control	• Voice recognition/response	• Limited info. on shared knowledge	Well-established	Commercial: 21st Century Systems, Inc.; awilson@21csi.com
	• Individual/team decisions	• Monitors/varies comm. frequency & media	• Mix of computer & observer measures		
	• Manipulates decision style & workload	• AWACS simulation based on SMEs & task analyses	• Specific scenario		
C3-Conflict[18]	• Command, control, & communication	• Automated/multiple communication assessments	• No multiteam capability	Relatively new, but based on an earlier test (C3Fire)	Commercial: rego.granlund@c3fire.org
	• Military peacekeeping	• Realistic military decisions	• No assessment of cognitive ability		
	• Hierarchical teams	• Performance defined by research needs	• Specific scenario		
Neo-CITIES[13,27,38]	• Command, control, & communication	• Can study multiteam groups	• Not military	Actively used as of 2008; should be examined by more labs	Noncommercial; requires collaboration; MMcNeese@ist.psu.edu
	• Civil crisis management	• Multiple measures of communication	• Limited info. on shared knowledge		
	• Distributed decision	• Recommended by Lamoureux et al. (2006)[22]	• Not all measures automated		
	• Manipulates crisis tempo, data rate, & complexity		• Specific scenario		
DDD[16]	• Command, control, & communication	• Automated measures	• Small team size	Well-established; used in multiple labs	Commercial: www.aptima.com
	• Customizable scenarios	• Easy manipulation of task load/demand, authority levels, communication, & information availability	• Limited aspects of cognition		
	• Assesses individual/team		• No multiteam		
	• Manipulates workload, team structure		• Programming to tailor scenarios		
Duo-Wombat[30,32–34]	• Crew coordination & shared awareness	• Tasks map onto basic cognitive abilities	• Only two users	Well-established, reliability/ validity shown	Commercial: AeroInnovation, Inc.; www.aero.ca/e_W_prices_CS.html
	• Attention, psychomotor ability, spatial ability, pattern recognition, memory	• Assesses shared display/ control	• Simplistic interface		
	• Single/dual performance	• Widely used	• May not capture communication		
		• Assesses target tracking plus secondary task	• Some options costly		
			• Specific scenario		
LDS[11]	• Within/between-team work for target engagement	• Multiteam dynamics	• Relatively artificial interface	Relatively new	Noncommerical: Requires IT support; jrh@msu.edu
	• Teams structured into different components	• Precise measures, multiple rounds of performance	• Specific scenario		
	• Teams receive feedback	• Customizable within the scenario of engaging enemy			
		• Online capability			
PLATT[20]	• Between/within team information processing, communication	• Multiteam dynamics	• Scenario must follow a linear sequence	Relatively new, used in several published studies	Noncommercial; requires license; wim.kamphuis@tno.nl
	• Multiple scenarios	• Scenarios customizable	• Further programming to tailor scenarios		
		• Incorporates websites into decisions during scenario			
		• Embedded questionnaires			

AEDGE: Agent Enabled Decision Group Environment; NeoCITIES: the C^3 (Command, Control, & Communications) Interactive Task for Identifying Emerging Situations; DDD: Distributed Dynamic Decision Making; DuoWOMBAT: Duo Wondrous Original Method Basic Awareness/Airmanship Test; LDS: the Leader Development Simulator; PLATT: Planning Task for Teams.

AEDGE is a Java-based technology developed by Barnes, Elliott, Tessier, and Petrov[5] for training and performance research. The task involves human users and computer-generated agents that may adopt any role in a scenario. Any entity (friendly or hostile) not controlled by a human is controlled by the computer. A computer agent makes recommendations for a course of action which the human may or may not choose to view. The system logs and captures the agent's recommendations, allowing comparisons between human and agent with respect to decision making and judgments.

This system is suited for the researcher interested in understanding how one person's decisions are influenced by the decision-making style of his/her "partner" in a team. Many of the options that the experimenter has for manipulation are specific to the type of recommendations made by the agent-controlled entity (e.g., degree of riskiness, degree of certainty).

C3Conflict

C3Conflict takes place within the context of a military peacekeeping operation that includes friendly and hostile forces.[18] The simulation was developed by researchers from the Santa Anna Information Technology Research Institute in Sweden and the U.S. Naval Postgraduate School in Monterey, CA. The authors describe the simulation as a "microworld" wherein 3 to 12 team members coordinate their actions to maximize group performance. The C3Conflict simulation was based on the earlier C3Fire (see www.c3fire.org). The simulation is relatively low in fidelity, but requires realistic decisions to be made. The simulation accommodates a hierarchical organization of the team into a platoon leader, squad leaders, and individual squad members, and allows for one member to be a leader of insurgents working against the friendly forces in the peacekeeping mission. The simulation allows for the analysis of individual and team performance.

Communication among team members is assessed through e-mails and chat text, as well as team member notes written in a shared diary. Common tasks assessed in the simulation include patrolling the simulation grid to identify friendly or enemy forces, finding and destroying insurgents, transporting resources (e.g., fuel, ammunition), and turning angry civilians into nonthreatening bystanders by providing aid. Granlund et al.[18] examined team effectiveness in terms of how well team members worked together on such tasks as land navigation, accuracy of tactical positions, and communication.

C^3 (Command, Control, and Communications) Interactive Task for Identifying Emerging Situations (NeoCITIES)

Of the tests identified in Table I, NeoCITIES is the only test not designed specifically for military-relevant applications, but it was included in our final list because it has some desirable features and it is suitable for paramilitary (e.g., police) situations or for scenarios relevant to national defense (e.g., response to terrorist attack). NeoCITIES was developed to measure team performance, communication, and cognition under pressure. The simulation was first described by Wellens and Ergener.[38] The current version employs a virtual city in a crisis scenario requiring response from emergency services. The shared goal is to respond appropriately to emergency events, prevent devastation, and maintain civil order. Two to three teams of two people each must cooperate to achieve these goals. Each pair is composed of an information manager and a resource manager. Each team monitors changing events and resource allocations. The quantitative outcome variables in the task include communication frequency and type. Data are recorded electronically and can be supplemented by means of a full system structure, including heart monitoring.

The face validity of this simulation appears adequate but further information concerning NeoCITIES test properties is needed. The test is not specific to military applications, but it has relevance to certain military missions. While some of the measures are automated, the nonautomated nature of the audio/video scoring method requires additional equipment and complicates the analysis of communications.

Distributed Dynamic Decision Making (DDD)

The DDD simulation task was developed by Aptima, Inc. to study aspects of team performance and communication in a complex and dynamic scenario. Validation efforts have been underway for decades and military researchers have employed the task for many years, according to Aptima.[1] DDD is marketed as a research tool that can also be used for training. Although originally designed as a simulation of a military command/control environment, the task can be tailored to other contexts. Workload, information availability, and team structure can be manipulated.

More than a dozen publications are listed on the Aptima website, which also supports the DDD. The testing environment is minimally realistic, but the program appears user-friendly and provides multiple resources to demonstrate and guide installation and configuration. The DDD has been adapted to allow participants in different locations to participate in the same real-time mission. A maximum of 50 participants can engage in the mission and "chat" using private or broadcast chat groups, e-mail communication, and voice-communication, albeit voice-communications are not scored automatically.

Duo Wondrous Original Method Basic Awareness/Airmanship Test (DuoWOMBAT)

The DuoWOMBAT is a modified version of the single-user WOMBAT designed to measure aircrew coordination and shared situation awareness. Participants must work cooperatively to accomplish simple shared tasks. The test measures performance with respect to divided attention among multiple information sources, judgment of priorities, ability to estimate probable outcomes, judgment of alternative actions, and prioritization of attention among tasks of varying urgency. The DuoWOMBAT provides an assessment of crew resource management and team coordination for establishing good situation awareness; however, it is limited to two participants and automated quantitative measurement of communication is not included.

The test simulates practical military challenges, such as the need for team effectiveness under conditions of operational stress.[8] The measure is relatively mature and established and has been used in many research laboratories.[34] Participants are seated side-by-side at two WOMBAT consoles separated by a partition, which increases the necessity for coordination. Participants are presented with tasks individually and in dual testing phases.[8] Of the seven tests described in this section, the test most similar to the rudimentary aspects of traditional cockpit flight control tasks engaged in by military aviation crewmembers is the DuoWOMBAT.

Leadership Development Simulator (LDS)

The LDS provides a low-fidelity simulation with a high degree of control over tasks relevant to military and aviation personnel. The simulation records information related to coordinated action, individual team member actions, and performance of up to 14 personnel. Davison et al.[11] used the LDS to examine between and within-team performance, where the overall multiteam objective was to "maximize points scored across a single performance episode"[11 (pg. 814)] by destroying enemy targets and not allowing friendly assets to be destroyed. In their study, the task consisted of multiple rounds where the different teams received an intelligence briefing, then deployed assets to accomplish mission objectives, and finally received feedback regarding their performance. The performance episode lasted approximately 2 h, not counting 1 h of training needed beforehand.

Simulation participants are assigned to one of three teams: 1) a "point component" team that engages enemy positions (with team members in charge of attacking enemies, providing an escort, and providing information about enemy locations); 2) a "support component" team that provides intelligence information about enemies and allies; and 3) a "leadership team" that consists of the mission commander, vice commander, and representatives from the two component teams (a director and assistant director from each team). The three-team system allows researchers to examine how well actions are coordinated within and between teams, in order to maximize the group's performance score.

The LDS appears to be a realistic simulation that is being actively used to address key aspects of within- and between-team processes. The LDS was developed in conjunction with the U.S. Air Force and is relevant to military team research.

Planning Task for Teams (PLATT)

The PLATT provides a flexible platform for examining different team processes and performance outcomes.[20] The authors have examined team processes in scenarios including military evacuation situations, tunnel explosions that may or may not be a result of terrorism, and breakdowns in the water supply. The PLATT provides flexibility for the researcher in developing scenarios for a given study by controlling the roles team members play, their degree of interdependence, and the ways they interact to accomplish the mission. This sets the PLATT apart from other tests that have predefined roles occupied by team members.

Participants responding to a given scenario interact with one another through messaging, e-mail, and visiting shared workspaces. The PLATT does automated coding of behavioral responses, including number of visits to web sites, number of e-mails sent to fellow team members, and shared workspace visits. The program also provides objective assessments of the quality of the team's performance on the scenario.

The PLATT appears to be a flexible and powerful simulation leveraging online resources in the accomplishment of team tasks. The program is customizable to allow for the creation of experiment-specific scenarios, and within- or between-team processes.

In summary, among more than 50 potential tests of team performance, 7 tests were deemed most relevant to small teams performing military- and aviation-related tasks. The seven tests (Table I) each had unique features, strengths, and weaknesses that researchers should consider. The Discussion provides specific recommendations to assist the readers in selecting the most suitable test for their research needs.

DISCUSSION

This review identified computerized team performance tests suitable for military and aviation research on small teams executing a shared task in challenging environments or situations. Practical recommendations for test selection are provided below:

- Researchers examining basic cognitive abilities and display-control problems in two-person cockpit teams should consider employing the Duo-WOMBAT.
- Researchers examining the effects of manipulating communication frequency on individual and team performance using realistic scenarios should consider using the AEDGE, especially if they are interested in operations related to an AWACS.
- For researchers interested in communication, learning, and performance within and between different teams, the LDS or PLATT are good options.
- Researchers interested in communication, learning, and performance within a single team should consider C3Conflict. C3Conflict also could be useful for researchers exploring questions concerning insurgency within a peacekeeping operation, in which case their findings may also contribute to the maturity of the test.
- If researchers are interested in how team members communicate while geographically dispersed, the DDD, LDS, and PLATT are worth considering.
- The PLATT appears to provide researchers with extensive flexibility for defining team member roles, team processes to be examined, and development of team performance measures. Therefore, the PLATT may be a good choice for researchers desiring highly customizable rather than "turn-key" solutions. New research using the test would help contribute to establishing reliability and validity of the platform.
- The NeoCITIES is worth considering by researchers interested primarily in civil emergency response, in which case their findings may also contribute to the maturity of the test.

- In general, the tests judged to be most military- and aviation-relevant, readily available, widely/recently used, and relatively mature in terms of validation are the AEDGE, DuoWOMBAT, and DDD. If the researcher is not sure which of the seven tests is best for his/her anticipated research, it is recommended that he/she evaluate these three tests first.
- It could be helpful for researchers to communicate the usefulness of the team performance tests they employ to the Center for Lessons Learned (CALL). Such feedback could be helpful also to the educational curriculum in the Intermediate-Level Education program at the Command and General Staff Officer's Course (CGSOC) at Fort Leavenworth, KS.

Increased research on the functioning of teams has resulted in a proliferation of tools to assess team processes and dynamics, many created for the sole purpose of a single experiment.[9] Researchers studying military, paramilitary, or aerospace-related team performance should consider the information in this report when seeking to identify tests most appropriate to the specific needs of their planned scientific effort. Important practical questions about team performance will not be answered efficiently by the continual introduction of new team performance test batteries. We recommend refinement, validation, and head-to-head comparison of the existing tests. Until further validation and head-to-head comparisons are done on existing tests, a researcher's choice of which test to employ will be influenced unduly by nonscientific considerations, such as test cost, availability, ease of administration, perceived "realism," novelty, or place of development (e.g., whether the test was "invented here").

Although no single test will be applicable to all situations, there are obvious drawbacks to exploring similar themes in team research separately via dozens of different team performance tests. Single-study, experiment-specific research considerations should be balanced against the multistudy benefits of focusing team research on a few key tests across multiple laboratories. At least within the restricted realm of command/control tasks, it appears possible to narrow the field of tests considerably through future comparisons of this type. A uniform and valid assessment of team coordination and performance will also provide the framework for understanding the team-level effects of challenging, overlapping aerospace challenges, such as hypoxia, high-G, vibration, spatial disorientation, fatigue, and high workload.

ACKNOWLEDGMENTS

The authors thank Ms. Lana Milam for extensive assistance with the preparation of this manuscript.

Authors and affiliations: Ben Lawson, B.Sc., Ph.D., Naval Submarine Medical Research Laboratory, Groton, CT; Thomas W. Britt, M.A., Ph.D., U.S. Army Aeromedical Research Laboratory and Clemson University, Easley, SC; and Amanda M. Kelley, Ph.D., Jeremy R. Athy, M.A., and Shauna M. Legan, M.A., U.S. Army Aeromedical Research Laboratory, Fort Rucker, AL.

REFERENCES

1. Aptima, Inc. Adaptable simulation for sociotechnical systems [Brochure]. DDD® 4.1. Woburn (MA): Aptima, Inc.; 2010.
2. Armstrong JE. Designing, testing, and using command control, communications, computers and intelligence (C4I) systems: what causes the disconnects and what can be done about them? Newport (RI): Naval War College; 1994.
3. Baker DP, Salas E. Principles for measuring teamwork: a summary and look toward the future. In: Brannick MT, Salas E, Prince C, editors. Assessment and measurement of team performance: theory, methods and applications. Mahwah (NJ): Lawrence Erlbaum Associates, Inc.; 1997:331–356.
4. Banner MJ. Assessing team performance and process: existing tools and tasks in "low-fidelity" environments. McLean (VA): Science Applications International Corporation; 2004.
5. Barnes C, Petrov PV, Elliott LR. Agent-based simulation and support of command, control, and communication decision making utilizing a distributed research network. [Internet] 2002. [Accessed March 2011]. Available from: http://dodccrp.org/events/2002/7th_ICCRTS/Tracks/pdf/113.pdf.
6. Bowers CA, Jentsch F. Use of commercial, off-the-shelf, simulations for team research. In: Salas E, editor. Advances in human performance and cognitive engineering research, vol. 1. Kidlington, Oxford (United Kingdom): Elsevier Science Ltd; 2001:293–317.
7. Brannick MT, Salas E, Prince C. Team performance assessment and measurement: theory, methods and applications. Mahwah (NJ): Lawrence Erlbaum Associates, Inc; 1997.
8. Breton R, Tremblay S, Banbury S. Measurement of individual and team situation awareness: a critical evaluation of the available metrics and tools and their applicability to command and control environments. Toronto, Ontario (Canada): Defense Research and Development Canada; 2007. Report No.: 2003-283.
9. Cannon-Bowers JA, Bowers C. Team development and functioning. In: Zedeck S, editor. APA handbook of industrial and organizational psychology. Vol 1. Building and developing the organization. Washington (DC): American Psychological Association; 2011:597–650.
10. Dahlstrom N, Dekker A, Van Winsen R, Nyce J. Fidelity and validity of simulator training. Theoretical Issues in Ergonomic Science. 2009; 10(4):305–314.
11. Davison RB, Hollenbeck JR, Barnes CM, Sleesman DJ, Ilgen DR. Coordinated action in multiteam systems. J Appl Psychol. 2012; 97(4): 808–824.
12. Dwyer DJ, Hall JK, Volpe CE, Cannon-Bowers JA, Salas E. A performance assessment task for examining tactical decision making under stress. Orlando (FL): Naval Training Systems Center; 1992. Special Report No.: 92-002.
13. El-Nasr S, Jones R, McNeese M. A scalable and extensible interactive scenario architecture for Distributed Command and Control simulations. Proceedings of the 2004 Command and Control Research and Technology Symposium; San Diego (CA); 2004. Ft. Belvoir (VA): Defense Technical Information Center; 2004.
14. Englund CE, Ryman DH, Naitoh P, Hodgdon JA. Cognitive performance during successive sustained physical work episodes. Behav Res Methods Instrum Comput. 1985; 17(1):75–85.
15. Estrada A, Lawson B, Kelley A, Athy J. A review of team performance measures and recent experimental challenges. Presented at the Technical Cooperation Program (TTCP) Defence Systems Symposium; 2010 May; Sydney, Australia. Washington (DC): Technical Cooperation Program; 2010.
16. Galster SM, Nelson WT, Bolia RS. The utilization of synthetic task environments in C2 Domains. Dayton (OH): Wright Patterson Air Force Research Laboratory; 2005.
17. Gawron VJ. Human performance measures handbook. Mahwah (NJ): Lawrence Erlbaum Associates; 2000.
18. Granlund R, Smith K, Granlund H. C3Conflict: a simulation environment for studying teamwork in command and control. Ft. Belvoir (VA): Defense Technical Information Center (www.dtic.mil); 2011.
19. Hollenbeck JR, Sego DJ, Ilgen DR, Major DA, Hedlund J, Philips J. Team decision-making accuracy under difficult conditions: construct validation

of potential manipulations and measures using TIDE2 simulation. In: Brannick M, Salas E, Prince C, editors. New directions in team measurement. San Francisco (CA): Jossey Bass; 1997:111–136.
20. Kamphuis W, Essens PJMD, Houttuin K, Gaillard AWK. PLATT: a flexible platform for experimental research on team performance in complex environments. Behav Res Methods. 2010; 42(3):739–753.
21. Katz LC, Kambe G, Kline KF, Grubb GN. Nonverbal communication and aircrew coordination in army aviation: annotated bibliography. Arlington (VA): Army Research Institute for the Behavioral and Social Sciences; 2006. Report No: 1181.
22. Lamoureux TM, Bandali F, Martin LMB, Li Z. Team modeling: review of experimental scenarios and computational models. Ontario (Canada): Defence R&D Canada-Toronto; 2006. Call-up #7911-05.
23. Lawson BD, Kelley AM, Athy JR. Team performance measures for use in military research. [Abstract.] Aviat Space Environ Med. 2011; 82(3): 261–262.
24. Lawson BD, Kelley AM, Athy JR. A review of computerized team performance measures to identify military-relevant, low-to-medium fidelity tests of small group effectiveness during shared information processing. Fort Rucker (AL): U.S. Army Aeromedical Research Laboratory; 2012. Report No: 2012-11.
25. Littell JH, Corcoran J, Pillai VS. Systematic reviews and meta-analysis. New York: Oxford University Press, Inc.; 2008.
26. Marks MA, Mathieu JE, Zaccaro SJ. A temporally based framework and taxonomy of team processes. Acad Manage Rev. 2001; 26(3):356–376.
27. McNeese MD, Bains P, Brewer I, Brown C, Connors ES, et al. The NeoCITIES Simulation: Understanding the design and experimental methodology used to develop a team emergency management simulation. Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting; 2005 Sep 26–30; Orlando (FL). Santa Monica (CA): HFES; 2005:591–594.
28. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. [Improving the quality of reports of meta-analyses of randomized controlled trials: the QUOROM statement.] Lancet. 1999; 354(9193):1896–1900.
29. Ntuen CA. An experiment in collaborative team decision making in complex information processing environments. Greensboro (NC): North Carolina A&T State University; 2006. Project No: 4-41243.
30. Odle-Dusseau HN, Bradley JL, Pilcher JJ. Subjective perceptions of the effects of sustained performance under sleep-deprivation conditions. Chronobiol Int. 2010; 27(2):318–333.
31. Reeves DL, Thorne DR. A synopsis of UTC-PAB (Unified Tri-Service Cognitive Performance Assessment Battery). Bethesda (MD): Naval Medical Research Institute; 1986. Available at http://www.stormingmedia.us/23/2381/A238102.html.
32. Roscoe SN. An aid in the selection process – WOMBAT. Civil Aviation Training. 1993; 4(2):48–51.
33. Roscoe SN. Predicting and enhancing flightdeck performance. In: Telfer R, Moore P, editors. Aviation training: learners, instruction, and organization, PP. Aldershot (United Kingdom): Avebury; 1997:195–208.
34. Roscoe SN, Corl L, LaRoche J. Predicting human performance, 5th ed. Saint-Laurent (Canada): Helio Press; 2001.
35. Salas E, Burke CS, Bowers CA, Wilson KA. Team training in the skies: does crew resource management (CRM) training work? Hum Factors. 2001; 43(4):641–674.
36. Simpson H, Oser RL. Evaluating large-scale training simulations. Mil Psychol. 2003; 15(1):25–40.
37. Swezey RW, Hutcheson TD, Swezey LL. Development of a second-generation computer-based team performance assessment technology. Int J Cogn Ergon. 2000; 4(2):163–170.
38. Wellens AR, Ergener D. The C.I.T.I.E.S. game: A computer-based situation assessment task for studying distributed decision-making. Simul Games. 1988; 19(3):304–327.