ARL-TR-8620 ● JAN 2019

US Army Research Laboratory

Crew Training Enhancement and Systems Testbed (C-TEST)

by Clayton Burford, Grace Teo, Lauren Reinerman-Jones, Daniel Barber, Mark Riecken, Joseph McDonnell, and Scott Gallant

NOTICES

Disclaimers

ARL-TR-8620 ● JAN 2019

US Army Research Laboratory

Crew Training Enhancement and Systems Testbed (C-TEST)

by Clayton Burford
Human Research and Engineering Directorate, ARL

Grace Teo, Lauren Reinerman-Jones, and Daniel Barber
Institute for Simulation and Training, University of Central Florida, FL

Mark Riecken
Trideum Corporation, Huntsville, AL

Joseph McDonnell
Dynamic Animation Systems, Inc., Fairfax, VA

Scott Gallant
Effective Applications Corporation, Orlando, FL

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
January 2019

2. REPORT TYPE
Technical Report

3. DATES COVERED (From - To)
25 April 2017–25 April 2018

4. TITLE AND SUBTITLE
Crew Training Enhancement and Systems Testbed (C-TEST)

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Clayton Burford, Grace Teo, Lauren Reinerman-Jones, Daniel Barber, Mark Riecken, Joseph McDonnell, and Scott Gallant

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
US Army Research Laboratory
12423 Research Parkway
Orlando, FL 32826

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-8620

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
US Army Research Laboratory
ATTN: RDRL HRB B
Aberdeen Proving Ground, MD 21005-5425

10. SPONSOR/MONITOR'S ACRONYM(S)
ARL/HRED

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

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

13. SUPPLEMENTARY NOTES

14. ABSTRACT

As part of the obligated Phase II work to enhance assessments, we developed a workflow and approach for conducting the network analysis to derive structure from metadata of human–agent teaming (HAT) research, the research domain that was selected in Phase I for our initial work. During performance of Phase II work, we identified an urgency to redirect our efforts from the network approach in favor of a sharper focus on crew and mixed crew (which includes HAT) research. We then defined new aims, objectives, and activities for our effort under the Crew Training Enhancement and Systems Testbed (C-TEST). C-TEST was to be a proving ground for training concepts for crew and mixed crew, and was expected to contribute to the Next Generation Combat Vehicle–System/Software Integration Laboratory efforts. Our ground-laying work on C-TEST included understanding the training trajectory for the crew, situating our efforts within the larger Army science and technology space, specifying the relevant concepts pertaining to human and agent involvement and functions within the crew, and identifying areas of research pertaining to crew training and evaluation. Further work involved descriptions of testbeds that can be modified for use in establishing C-TEST.

15. SUBJECT TERMS
human–agent teaming, network approach, crew and mixed-crew training, training testbed, team-enhancing behaviors

16. SECURITY CLASSIFICATION OF:
a. REPORT Unclassified
b. ABSTRACT Unclassified
c. THIS PAGE Unclassified

17. LIMITATION OF ABSTRACT
SAR

18. NUMBER OF PAGES
57

19a. NAME OF RESPONSIBLE PERSON
Clayton Burford

19b. TELEPHONE NUMBER (Include area code)
407-208-3022

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

1. Introduction

This report describes research conducted in support of concept development for the Crew Training Enhancement and Systems Testbed (C-TEST). This effort is a continuation and extension of groundwork laid in the prior year and executed by a diverse team of researchers in social science, psychology, computer science, and engineering referred to as the C-TEST Team. Results of the team's efforts under Unified Multimodal Measurement for Performance Indication Research, Evaluation and Effectiveness (UMMPIREE) in the first year can be found in Burford et al. (2018). In the first year, the C-TEST Team identified critical challenges with the current state of assessments in human research and anticipated challenges with future assessments across the services, especially with the increased use of technology, autonomous systems, robots, and agents in military operations. The challenges relate to the following:

- Constructs are numerous and diverse. Any intervention, technological or otherwise, typically produces a diverse range of outcomes and implicates multiple constructs. Thus, a full assessment may involve measurement of multiple constructs. Constructs can also be multifaceted and require multiple measures.

- Variability in how assessments are conducted. Different methods for measurement of a construct do not necessarily converge. Variability in assessments can also arise from broad factors such as the design of the assessment plan and processes followed, as well as specific factors such as the conceptual definition and measures used.

- Assessments cover both humans and machines. As more intelligent systems and machines arrive on scene, boundaries of human- and system-centric assessments are being blurred. The scope of assessments is also widening to include Soldiers in different military occupational specialties and ranks performing various activities (e.g., training and participating in operational tasks).

- Multiple contexts and timespans for assessment. Assessments take place at different phases of operations, and different stages of expertise acquisition or system development. They even span different stages of Soldiers' careers.

- Variability in assessment locations. As more and more "opportunistic" assessments are conducted, the likelihood that assessments are conducted outside standard, controlled conditions is increasing. With more military training being decentralized and distributed, assessments could be administered in home stations and various deployed locations, contributing to variability in assessment conditions.

In response to these challenges, our goal was to enhance the quality and impact of assessments, so we developed a three-phased plan. Phase I involved understanding the current state of assessment practices that would inform subsequent direction of the project as well as provide a baseline with which to compare the future state of assessment. The baseline was conducted within a selected research domain that typified the assessment challenges. Phase II was tentatively planned as including the development of a mathematically based network approach, which is a novel, bottom-up approach to extract themes and patterns from the research conducted in the selected research domain. This network approach was to complement theory development in the scientific paradigm. For Phase III, this network approach was to be extended to a separate but related research domain.

For Year 1, the project team focused on Phase I objectives, which translated to the following activities:

- Activity 1: Specify terminology, construct definition, and assessment principles.
- Activity 2: Develop a conceptual framework and architecture for characterizing and guiding assessment.
- Activity 3: Identify a domain that exemplifies assessment challenges and needs.
- Activity 4: Establish a baseline of current assessment needs and practices in the domain that can inform ideas for tool and product development.

These activities culminated in a conceptual framework and architecture for assessments and a baseline of the state of assessments in the human–machine teaming (HMT) research domain, which was identified in Activity 3 as a domain that exemplifies many of the assessment challenges.

In Section 2 of this report, we discuss work accomplished as a continuation of Phase II of UMMPIREE (see Burford et al. 2018). This work comprised a) development of a workflow for network analysis, b) extraction of metadata as part of the baseline establishment and sourcing for a database management system, and c) a pilot study for the network analysis where a preliminary analysis of the metadata was conducted. The remaining sections report on the proposed new direction for the project and our continued work along that new direction.

2. Phase II (Year 2): Linking and the Network Approach

The aim of the network analysis was to identify patterns and structures from the metadata in human–agent teaming (HAT) research that may reveal information that can shape and improve assessments. The network analysis is expected to reveal information that addresses the following questions that pertain to assessments in the HMT domain:

- Which constructs are most studied?
- Are there constructs that tend to be studied together?
- What measures have been used to operationalize the constructs?
- Are there measures more suited for certain research applications (e.g., tasks used in studies to elicit behaviors of interest) than others?
- Are there commonly used clusters of measures?
- Which tasks and environments are more successful for researching the different constructs?
- What constructs/measures tend to be associated with which authors?
- Which authors collaborate within the domain? Outside the domain?
- Which authors tend to open up new sub-areas of research?

Findings related to these questions would reveal constructs that are the most important in HMT research thus far. The measures for these can be reviewed, and recommendations can be made for how these constructs can best be operationalized and measured within the various tasks and environments. New sub-areas of research and suggestions for potential research collaborations may also be gleaned from patterns extracted during the network analysis, providing some direction for HMT research. To guide the work in Phase II, we developed a workflow that articulated the steps to progress the project toward the network analysis (Teo et al. 2018).

2.1 A Workflow for Network-Analysis-Based Structure Discovery in the Assessment Community

The workflow consists of three steps: 1) Data Preparation, 2) Data Analysis, and 3) Structure Discovery from the Network Analysis (Fig. 1). Data Preparation involved extracting metadata on studies that conducted assessments in HMT and formulation of a construct-measure grammar during data extraction. For the Data Analysis step, we organized the data from the previous step and performed preliminary analysis (a pilot study) before the data factoring and network building. We also explored several database solutions (i.e., sourcing of a database management system) that were appropriate for our data type and could enable the type of database manipulation that would facilitate the subsequent step, where methods and tools from network science and mathematics (e.g., set theory and graph theory) will be exploited for Structure Discovery.

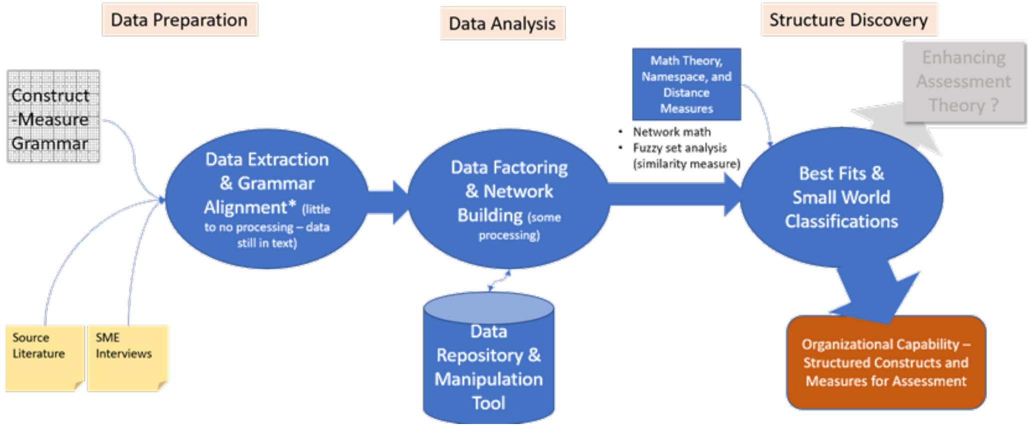

Fig. 1 **Workflow of the network approach**

The data extraction work accomplished in the Data Preparation step. The sourcing of the database management system (i.e., data repository and manipulation) in the Data Analysis step will be described in Section 2.2. Section 2.3 will report on the pilot study conducted to inform the data factoring and network building.

2.2 Data Extraction and Database Management System

2.2.1 Data Extraction

In Phase I, we identified HMT domain as exemplifying many of the assessment challenges. To obtain a baseline of the current state of assessments in the HMT domain, we a) conducted face-to-face semi-structured interviews of several Army and Air Force researchers (subject matter experts [SMEs]) who have published substantial work in HMT and b) reviewed HMT literature. We reported on the general themes and findings of the SME interviews in the Year 1 technical report (Burford et al. 2018).

From our literature review we decided to use of the term "human–agent teaming" instead of "human–machine teaming" since the former described robots, machines, and systems that have some level of agency allowing them to be characterized somewhat more like teammates rather than mere tools (Bruemmer and Walton 2003).

We reviewed 74 empirical studies in which HAT assessments were conducted. These studies included research on HAT-relevant concepts such as cyber security, driving, vigilance, and human–computer interfaces. The authors of the studies reviewed were several US Army Research Laboratory (ARL), US Air Force Research Laboratory (AFRL), and US Naval Research Laboratory (NRL) researchers and their co-authors. The following criteria were used to constrain the literature search:

- Research within the last 5 years (2012–2017)
- Studies by US Defense Department research laboratories (e.g., authors from ARL, AFRL, and NRL)
- Empirical studies that are not meta-analyses since meta-analytic studies incorporate several studies at once, leading to double counting
- Keywords that indicate that the study falls under the HAT domain
- Constructs that indicate that the study falls under or is relevant to the HAT domain

From the studies, we extracted the following metadata fields that will be required for the network analysis planned for Phase II:

- Study title
- Primary author and co-author(s)
- Year
- Keywords
- Research question: this was inferred from the study if not explicitly stated
- Theories/models cited
- Study task
- Testbed/simulation environment/experimental platform used
- Study design
- Study conditions
- Sample size and characteristics
- Study variables/constructs: independent variables (IVs), dependent variables (DVs), and subject variables

- Measures (i.e., operationalization of study constructs)
- Analysis
- Study results

During the collection of the metadata, we encountered several issues that were symptomatic of some of the assessment challenges earlier identified. For instance, there were inconsistencies in some terminology. While some studies used the term "task difficulty", others used "task demand" or "task load" to refer to the same concept. Even as we made every attempt to preserve the original construct labels reported in the studies, it was necessary to impose some consistency and recode some of the terms used to facilitate discovery of trends and patterns in the metadata in Phase II. As a result, we created a "thesaurus of synonyms" to keep track of what had been recoded (see Table 1 for an excerpt):

Table 1 Excerpt of the thesaurus of synonyms for keywords/construct labels

Keyword/construct names	Synonyms
HMT	HATs; human–robot interaction; human–robot teaming; human–robot teams
Unmanned systems	Unmanned vehicles; unmanned aerial vehicles (UAVs); unmanned aerial systems; unmanned ground vehicles (UGVs); remotely piloted aircraft
Task load	Task difficulty; task demand
Cyber security	Cyber defense
Reliability	Automation reliability
Supervisory control	Multi-robot control
Decision making	Team decision making
Workload	Mental workload; real-time workload; operator workload; cognitive load
Stress	Task-induced stress; stress states

A construct-measure "grammar" was constructed to organize the metadata in an Excel worksheet. It described the relationships (i.e., connected constructs to their respective measures) among selected metadata fields that were important for network building. For example, the following is an italicized row in the worksheet:

> *…the (i) Independent variable, (ii) Transparency, which is a (iii) Construct (iv) Operationalized by (v) Level of Information provided by the Agent: 3 levels, (vi) significantly affected the (vii) Scores on the NASA-TLX* [NASA Task Load Index], *which is (viii) an operationalization of (ix) Workload, a (x) Construct that was used as a (xi) Dependent Variable in the* analysis.

The italicized terms (i) through (xi) were the data entered in 11 columns/fields. Data entered this way captured the relationships among the fields in the metadata.

2.2.2 Database Management System

From the data that had been organized according to the construct-measure "grammar", a graph database scheme was derived (Fig. 2).

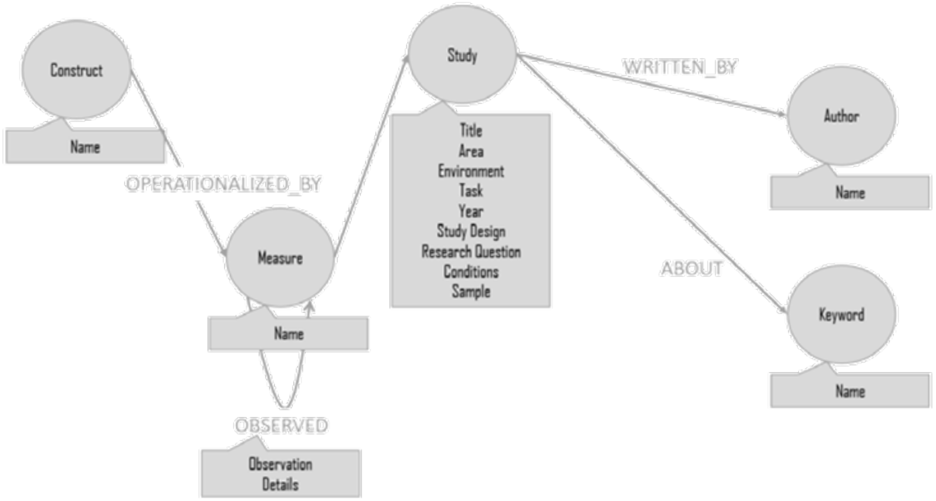

Fig. 2 Graph database schema

With the schema in place, we searched for a database management system that accommodated the type and anticipated volume of data, allowed easy data manipulation, and facilitated the network analysis. The open-source Neo4j graph database management system was identified as a system that potentially met our requirements. It has been used in fraud detection, social networks, knowledge graphing, network and information technology operations (Neo4j, n.d.). The Neo4j system is readily configurable, can host data files on a web server, and has a consistent environment. It has several visualization tools and enables different cypher queries to be formulated to generate customized views of the data (Fig. 3).

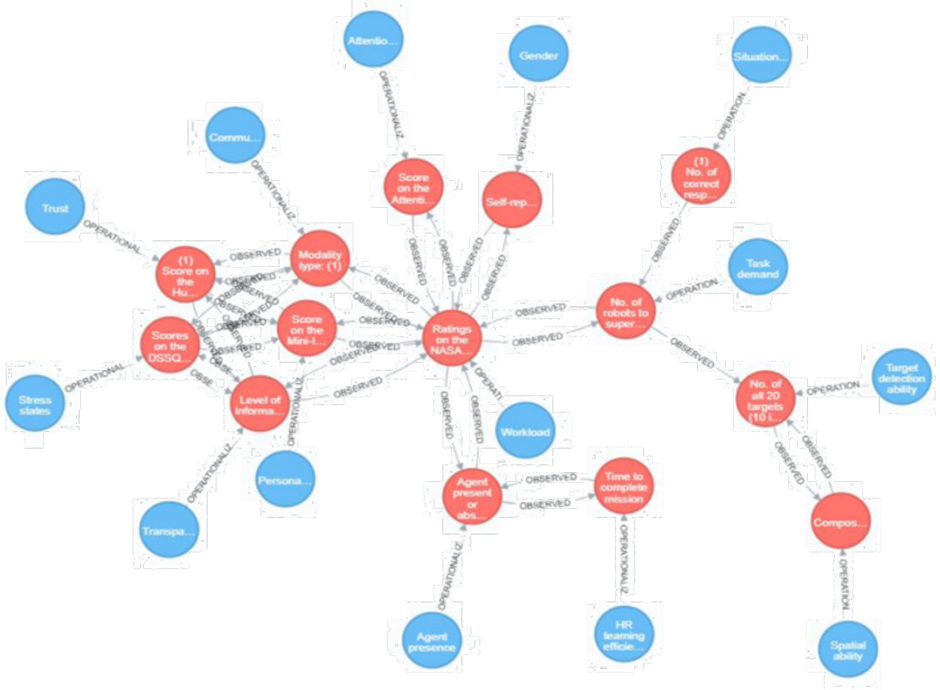

Fig. 3 Neo4j visualization of the constructs and measures of one study

The Neo4j database management system could configure the data appropriately for the network analysis using one of two possible methods: 1) graph theory (West 2005) and network analysis (Borgatti et al. 2009), and 2) set theory (Nelson 1977; Moore and Lodwick 2003). These methods were selected because they are "scale-free" and do not impose assumptions about the data. A comparison of these methods is provided in Table 2).

Table 2 Similarities and differences between the methods for Structure Discovery

Graph theory and network analysis	Set theory
Similarities	
Has "neutral" organizing principles that enables an objective "lens" whereby to view the state of the assessments.	
By not forcing a rigid theoretical structure, the method allows for continual adjustments and "newcomers".	
Method can result in structures that further point to underlying concepts that in turn may lead to increased theoretical systemization.	
Differences	
Identifies authors, concepts, constructs, measures, and the like as nodes. Looks for relationships among and across them. Results must be interpreted. Certain properties of networks are also "scale-free", thereby providing a degree of continuity for this research framework as we incorporate more data.	Describes constructs via set of operationalizations. Considers abstraction and generalization. Explores various set similarity measures. Can be complemented by other related mathematical techniques such as the fuzzy set theory and set similarity.

2.3 Pilot Study: Preliminary Analysis of Metadata

Before submitting the data to the network analysis, we conducted a pilot study involving a preliminary analysis of the data (metadata) to uncover general trends in HAT research from frequency counts of the outcomes (DVs) and factors (IVs) that were most investigated in the studies, as well as the measures and operationalizations of these constructs. We also identified the tasks and testbeds that were used most often in the studies, as the scope of research can be influenced and limited by the ability of the testbeds to simulate agent behaviors and administer different tasks. The following sections summarize the results of the pilot study.

2.3.1 Dependent Variables

There were many different DVs investigated across all the studies, and some referred to the similar outcomes but were labelled differently. For our purposes of identifying trends and patterns, we categorized and grouped certain constructs. For instance, performance outcomes in a threat detection task and target detection task were both categorized as *detection performance*. There were 86 unique DVs/DV categories (outcomes), with the five most commonly studied in HAT research being workload, detection performance, efficiency, stress states, and trust (Fig. 4).

Fig. 4 Top five outcomes examined in HAT research

Each of these outcomes were operationalized by multiple measures. An excerpt of the measures associated with these outcomes are provided in Table 3.

Table 3 Top five most-studied outcomes in HAT research and their associated operationalizations

Dependent Variables	Operationalizations
1. Workload: DV	Overall score, unweighted score, weighted score of the NASA Task Load Index (NASA-TLX) (Hart and Staveland 1988); Multiple Resources Questionnaire (Boles and Adair 2001); cerebral blood flow velocity; regional saturation of oxygen; pupil diameter; eye fixation, electroencephalogram; heart rate variability, approximate entropy; Crew Status Survey; visual dispersion, etc.
2. Detection performance: DV category pertaining to performance in various detection tasks: cyber threat detection, change detection, detection of critical signals, etc.	Proportion of correct detections/responses; correct rejections; false alarms; misses; errors of omission; errors of commission, etc.
3. Efficiency: DV category pertaining to performance with respect to time	Time to mission completion; average response time to correct detections; etc.
4. Stress states: DV	Full version and short version of the Dundee Stress State Questionnaire (Matthews et al. 2002); subscales of task engagement, distress, and worry.
5. Trust: DV	Function Specific Trust in Automated Systems Scale (Parasuraman et al. 2000); Trust in Automation Scale (Jian et al. 2000); Human–Robot Trust Scale (Schaefer 2013).

2.3.2 Independent Variables

As with the DVs, there were multiple IVs investigated in the studies. After applying the same criteria for categorization, there were 89 unique IVs/IV categories (i.e., factors) with the five most commonly studied being technology features, period of watch, task load, task phase, and technology type (Fig. 5).

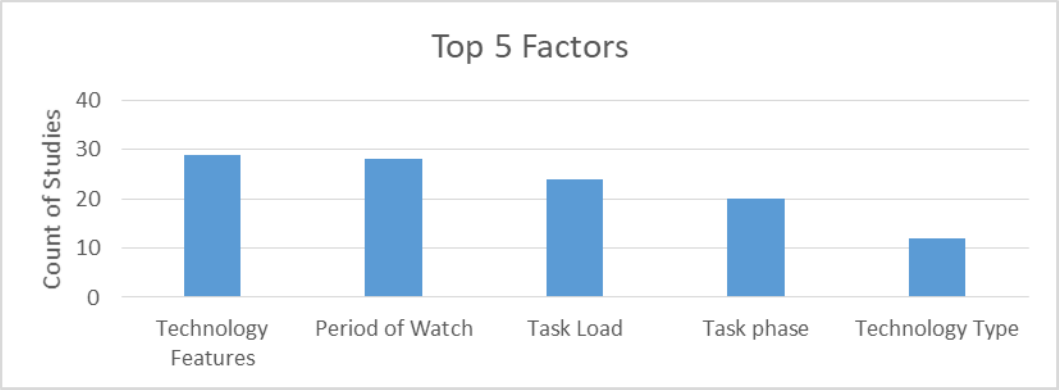

Fig. 5 Top five factors examined in HAT research

These IVs were operationalized in various ways and had different levels. Table 4 show an excerpt.

Table 4 Top five most-studied factors in HAT research and their associated operationalizations

Independent variables	Operationalizations
1. Technology features: IV category pertaining to task and agent parameters: agent reliability, transparency, maneuverability, automation type, automation imperfection type, etc.	Reliability (55%, 60%, 86.7%, 90%, 93%, 100%); transparency (minimal, contextual, constant information transparency levels); automation type (static, adaptable, adaptive, no automation); automation imperfection type (false alarms, misses); etc.
2. Period of watch: IV	2-, 3-, 5-, 10-, 15-min periods on the task
3. Task load: IV category pertaining to task demand: event rate, sensory modality, number of parameters, signal probability, etc.	Event rate (various rates used across studies); sensory modality (auditory, sensory); source complexity (single, dual tasks); number of parameters (single/dual tasks, different number of robots to supervise, different number of digits to hold in memory); signal probability (different rates of signal likelihood); etc.
4. Task phase: IV category pertaining to different phases of the task: task-related change, learning phase, task phase, etc.	Task-related change (pretask, posttask); learning phase (first few trials, last few trials); task phase (braking, replying, recovery phases); etc.
5. Technology type: IV category pertaining to types of agent capabilities: agent presence, display parameters	Agent presence (agent present, agent absent); display parameters (2-D, 3-D, coordinated, uncoordinated displays); etc.

2.3.3 Tasks Used

Analysis of the tasks used in HAT research provided an understanding of the scope of application in HAT. The following are the top 10 most-used tasks in the studies reviewed. Many tasks yielded performance measures that corresponded to the DVs in the category Detection Performance (Fig. 6).

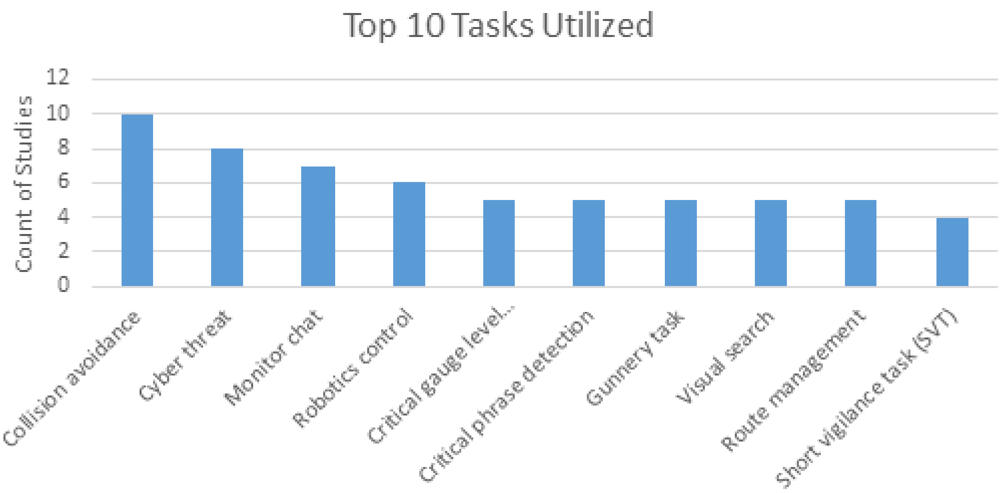

Fig. 6 Top 10 tasks used in HAT research

Collision avoidance was a task in several studies where participants assumed the role of pilots, drivers, and UAV/UGV controllers. Cyber threat detection involved detection of hostile emails or IP addresses. Chat monitoring was often employed as a concurrent or secondary task. The robotics control task required participants to tele-operate a robot, while the critical signal detection task involved detecting an important signal, such as a barrel being longer, a line being longer, an auditory signal, and so on. The route management task entailed the rerouting of agents. Participants who performed the visual search task searched an area visually for specified targets. The gunnery task required participants to search for hostile targets and report their location. In the critical phrase detection task, participants detected a critical phrase from a stream of auditory information. For the critical gauge level detection task, participants detected when a gauge, often symbolizing battery life or fuel level, reached a critical level.

2.3.4 Testbeds Used

The testbed or simulation environment is a large part of the context in which the assessments are conducted since the capabilities and functions of the testbed can constrain the type of tasks and assessments that can be administered. The testbeds most commonly used in the HAT studies examined supported a combination of communications, detection, tele-operational, and navigational tasks (Table 5).

Table 5 Excerpt of the most commonly used testbeds and capabilities in HAT studies

Testbed capabilities	ALOA*	MIX †	MMC‡	SCOUT§	TCU‖	VBS2¶
Supporting communications among team members (e.g., supports chat monitor, coordinate response measure, critical phrase, simple communication)	✓	✓	✓		✓	…
Administerig detection tasks (e.g., gunnery task, threat search, visual search, change detection task, insurgent search, critical gauge, critical phrase)	✓	✓	✓	✓	✓	✓
Displaying map locations	…	✓	…	…	…	…
Supporting teleoperations (e.g., robotics control)	…	✓	…	…	✓	✓
Supporting route management	…	✓	…	✓	…	…

*ALOA = Adaptive Levels of Autonomy, v.3. Research testbed developed by OR Concepts Applied (Johnson et al. 2005)
†MIX: Mixed Initiative Experimental Testbed (Barber et al. 2008)
‡MMC: Multi-modal Communication, a network-centric communication management suite (Finomore et al. 2011)
§SCOUT: Supervisory Control Operations User Testbed (Sibley et al. 2016)
‖TCU: Tactile Control Unit, developed by ARL's Robotic Collaborative Technology Alliance (Chen and Terrence 2009)
¶VBS2: Virtual Battlespace 2 (Morrison 2012)

3. Proposed New Direction for Project: C-TEST

After further discussions and engagements with HAT researchers, we determined that we needed to focus on HAT assessments and research instead of proceeding with the network approach in Phase II and the work for Phase III. The impetus for this redirection also included consideration of the following:

- Army Futures Command's goal to modernize the Army by developing future warfighting concepts to increase lethality and effectiveness. This is to be accomplished under six Cross-Functional Teams (CFTs) (US Army STAND-TO! 2018)
- Chief of Staff of the Army's Modernization Priorities and Associated CFTs (McCarthy 2017)
- Signed EXORD (execute order): "Next-Generation Combat Vehicle (NGCV) must have embedded training and be interoperable with the Synthetic Training Environment (STE)". (BG Lesperance 2018)
- US Army Research Laboratory Essential Research Areas. (US Army Research Laboratory Essential Research Areas 2018)
- The 2016 National Artificial Intelligence Research and Development Strategic Plan

In recognition of the current modernizing of command and push toward initiatives that improve the quality and speed of delivery of new materiel and capabilities to the Soldier (McCarthy 2017), our new work on HAT research is positioned to better serve the transition of HAT research to crew and mixed crew teams that include agents and systems.

3.1 C-TEST Aims and Objectives

Along with the current impetus for developing new configurations of crew and mixed-crew teams (i.e., including agents/robots/machines/systems) to meet the changing dynamics and demands of theater is an equally urgent need for training. Specifically, there is a need to develop training that will enable crew members to work effectively over protracted periods, maximize team-enhancing behaviors (TEBs), and optimize the capabilities of each team member to achieve mission success. For agents, "training" entails machine learning: development of models and algorithms that drive agent behaviors that enhance the team.

Much of the training curriculum and simulations for these HAT/Crew have often been designed with the urgency of war and only meet face validity. This may result in under- or overuse of robotic capability and/or machine intelligence or other improper use of agents by the human crew, not unlike the misuse, abuse, disuse in automation (Parasuraman and Riley 1997), all of which can jeopardize mission success. The training curriculum and simulations are seldom subject to proper upfront assessment to determine the extent to which the training meets learning requirements and transfers to the field.

The work in assessments and HATs accomplished by the project team is relevant and can be applied to the aforementioned problem with crew and mixed crew training. In the new direction, the C-TEST project sought to address the current problem by developing a proving ground for training concepts prior to final system development. The proving ground will provide the capability to assess 1) the training concepts and curriculum, 2) alignment of training objectives to learning outcomes, and 3) ability of the training environment/simulation to deliver the training required. C-TEST will examine a spectrum of training concepts and technologies for HAT/Crew that can enhance TEBs in this context and provide specific recommendations to simulation platform researchers and developers with respect to how these training concepts and technologies can be embedded or attached to the program of record systems (both manned and unmanned systems).

3.2 C-TEST Value Proposition

Before addressing our specific method, further discussion regarding the C-TEST value stream is merited. Figure 7 illustrates a C-TEST "value proposition". This value proposition described an overall flow of value that originates in existing research in HAT and NGCV research and development activities, as well as emerging tactics, techniques, and procedures that will govern military tasks using the HAT-based technologies. The main output of this phase of the value stream were the TEBs and research areas for training that are described in the later sections of this report.

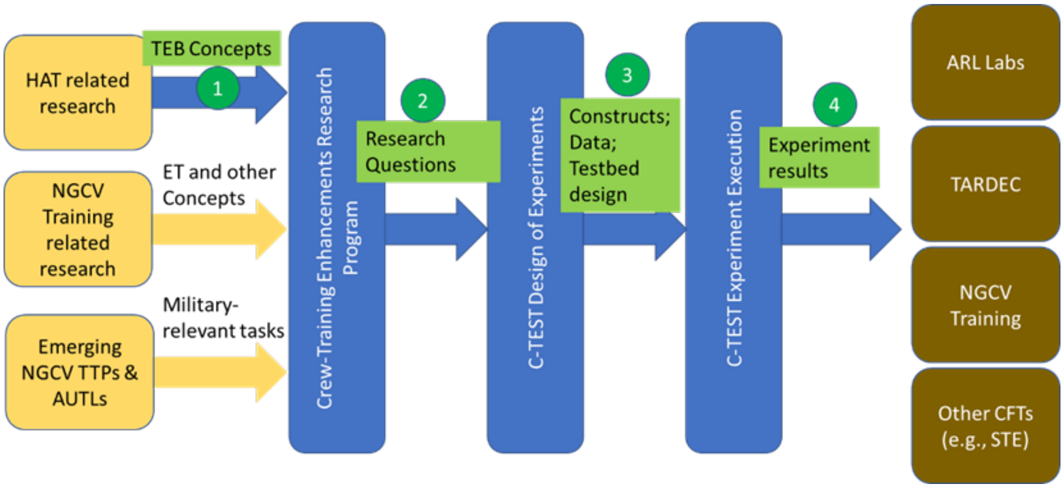

Fig. 7 **C-TEST value proposition**

The numbered circles indicate milestones in the value proposition. Although C-TEST was informed by multiple areas including NGCV training research, it was primarily driven by HAT-related research that produced TEB concepts. The C-TEST program then examined the TEB concepts to generate relevant research questions that can be investigated through design of experiments, which employs constructs, data, and testbed designs to enable experiments to be conducted. These yields empirical results that can be consumed by a variety of stakeholders. One such stakeholder is the NGCV System Integration Laboratory (NGCV-SIL) community, which can benefit from the training concepts evaluated in C-TEST. The NGCV-SIL includes simulating the operational environment of crew in future combat vehicles. As a proving ground within the SIL environment, C-TEST can enhance training-related experimentation and analysis capability in the NGCV-SIL.

3.3 Summary of C-TEST Activities

To serve the goals under this new direction, the C-TEST project team engaged in the activities detailed in Table 6.

Table 6 Rationale for C-TEST activities

Initiative from activity	Why is activity important?	How activity addresses that problem	How activity feeds into tool or knowledge product for C-TEST
Activity 1: Identify other concepts and ideas related to HAT/Crew training			
Progression of training types for HAT/Crew across the Army Force Generation cycle	Current training curriculum for HAT/Crew may not account for the lifespan evolution of the HAT/Crew over time and deployment stage.	Helps with the design of training by adding to the understanding of how training needs of the HAT/Crew change over time.	Contributes to the areas and criteria for evaluation of training concepts under C-TEST. Informs design of training.
Mind-map of science and technology (S&T) concepts in HAT/Crew training	Design of HAT/Crew training must take into consideration important S&T components and other practical constraints.	Maps out the broader context of relevant areas and shows the relationships among these.	Contributes to the areas and criteria for evaluation of training concepts under C-TEST. Informs design of training.
Activity 2: Develop conceptual frameworks for C-TEST			
Taxonomy of agent involvement and framework of teammate functions	Shows the different types and levels of agent involvement in a HAT.	Provides a means to express the relationship between teammates in a HAT in terms of their functions, which are associated with various behaviors.	Contributes to the understanding of the functioning within the HAT.
Taxonomy for TEBs	Training for a HAT/Crew should result in behaviors that enhance the team, which lead to superior team performance.	Articulates the possible effects and areas for TEBs within the HAT/Crew.	Shows the “end-point” behaviors that the HAT/Crew training should relate to, and so doing, informs evaluation and design of training.
Research areas (i.e., knowledge, skills, attitudes) for TEB training	Research and assessments on the training concepts should not be fragmented but be programmatic and based on a framework to encourage development of a knowledge base.	Specifies the training areas of HAT/Crew so research can be planned and prioritized.	Helps to prioritize areas of research for training concepts in C-TEST.

Table 6 Rationale for C-TEST activities (continued)

Initiative from activity	Why is activity important?	How activity addresses that problem	How activity feeds into tool or knowledge product for C-TEST
Activity 3: Support training within NGCV-SIL			
Canvas of testbeds, simulations, and other related applications	To speed up testbed development time and provide the richest possible environment, C-TEST may be able to utilize parts of existing testbeds.	Identifies capabilities, features, and functions of existing testbeds, simulations, and other applications that may be adapted for use in C-TEST.	To support training in the NGCV-SIL, a robust testbed is essential to ensure C-TEST knowledge products are reliable, traceable, and repeatable.

4. Concepts for HAT/Crew Training

Work under C-TEST required understanding of HAT/Crew training in the larger context, which included training across the lifespan of the HAT/Crew and training within the science and technology (S&T) space.

4.1 Stages of HAT/Crew Training over Time

Training needs for the HAT/Crew evolve over the lifespan of the HAT/Crew as depicted in the stages of the Army Force Generation (ARFORGEN) cycle. First, formation of a HAT entails pairing team members that were trained separately previously (for the agent, prior “training” may be certain default or preset algorithms) and bringing them together as a new HAT crew. The initial training as a HAT crew may be general, but this becomes more specific to the mission, geolocation, and type of human–machine partnership as the HAT crew prepares for deployment. To maintain readiness, the HAT crew’s training will be some type of mission rehearsals. After deployment, HAT crew training would be in the form of learning “in the wild” for the machine as the human transfers his/her training to the real world. As shown in the green ovals below the force generation model in Fig. 8, there are multiple possibilities for HAT training along this process (i.e., from individual training at home station through crew training in deployment).

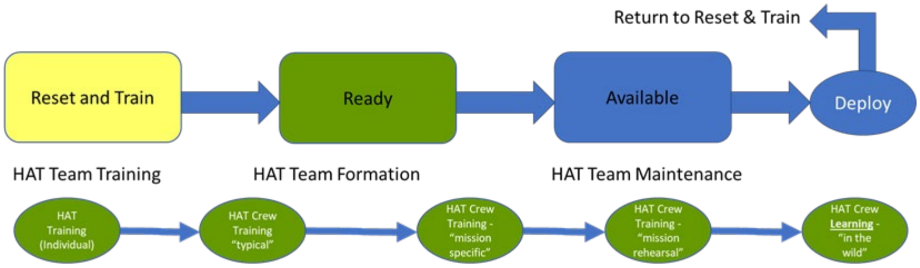

Fig. 8 Types of HAT training across the ARFORGEN cycle

This ubiquity of training needs highlights the importance of conducting research into how to best train HAT and TEBs in the various stages of Warfighter readiness.

4.2 S&T Problem Space of HAT/Crew Training

Figure 9 is an informal “mind map” of the HAT S&T problem space. At the center of the mind map is the idea of a technology prototype that can contribute to the overall HAT S&T research area, but specifically supporting TEBs, as seen in the top right of the mind map.

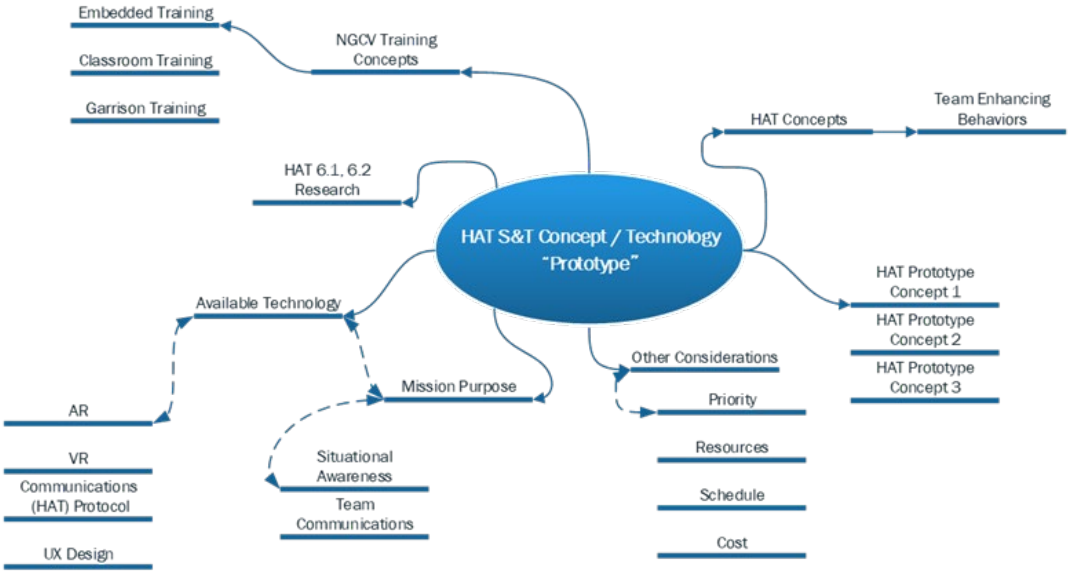

Fig. 9 HAT S&T mind map

The mind map shows the broader context and connectedness of many areas of related work. Starting from the 11 o’clock in the drawing, NGCV training concepts are considered. The possibility of embedded training (ultimately focused on TEBs) is noted. Following the mind map in a counter-clockwise fashion, the next item is fundamental HAT research (6.1 and 6.2). Available technologies such as augmented reality, virtual reality, communications protocols, and user experience design are considerations. Essential to any HAT prototype is the mission purpose. Mission purpose may drive research aspects such as situational awareness or team communications. Practical considerations such as Army priorities, available resources, cost, and schedule are also considered. Various HAT prototypes that may be under development should be considered as candidate “platforms” for TEB research.

5. Frameworks for C-TEST

We derived the following frameworks to inform our work on training concepts for the HAT/Crew in C-TEST. The frameworks will also be useful in guiding future system engineering and development.

5.1 Taxonomy of Agent Involvement and Framework of Teammate Functions

Training concepts for the human teammate in a HAT must be grounded in the understanding of the dynamics between the human and agent teammates while they cooperate on a task. The agent can be considered an enabling technology with which the human collaborates to complete a task. With this in mind, research in automation can be extended to autonomous agents and HATs. Such work includes the extent to which and the ways in which automation assists the human. Drawing from the work on levels of automation (Endsley 1999; Parasuraman et al. 2000), we developed a 10-level taxonomy that described the levels of system or agent involvement in a dynamic control task and the resultant human and agent functions or roles (Table 7).

Table 7 Taxonomy of levels of automation and associated human and agent roles (adapted from Endsley 1999)

Level	Role	Monitoring	Generating	Selecting	Implementing
1.	Manual control	H	H	H	H
2.	Action support	H/A	H	H	H/A
3.	Batch processing	H/A	H	H	A
4.	Shared control	H/A	H/A	H	H/A
5.	Decision support	H/A	H/A	H	A
6.	Blended decision making	H/A	H/A	H/A	A
7.	Rigid system	H/A	A	H	A
8.	Automated decision making	H/A	H/A	A	A
9.	Supervisory control	H/A	A	A	A
10.	Full automation	A	A	A	A

Notes: H = Human; A = Agent

The four functions or roles across the levels (top row of Table 7) loosely correspond to the four-stage model of information processing (Parasuraman et al. 2000), with each level showing the extent of automation involvement in these functions. These four stages of information processing also parallel Boyd’s (1996) Observe-Orient-Decide-Act (OODA) loop, which has been commonly used to describe human decision making (Table 8).

Table 8 Similarities across frameworks describing stages/functions that can be undertaken by human and/or agent

Framework	Stage 1	Stage 2	Stage 3	Stage 4
1. Model of human information processing and application of automation (Parasuraman et al. 2000)	**Human information processing**			
	Sensory processing	Perception/ working memory	Decision making	Response selection
	Type of automation			
	Information acquisition	Information analysis	Decision selection	Action implementation
2. Roles and functions across levels of automation (Endsley 1999)	Monitor	Generate	Decide	Implement
3. OODA loop (Boyd 1996)	Observe	Orient	Select	Act

Within the context of HAT, the OODA loop has been applied describe the stages of the agent instead and the different ways that the human can be kept in the loop. Concepts such as human-*in*-the-loop, human-*on*-the-loop, and human-*inside*-the-loop articulate the level of human involvement in agent processes (Sudit 2015 as cited in James et al. 2017):

- Human-in-the-loop (human between the "Decide" and "Act" stages): while the agent may recommend a variety of decisions, the human makes the final decision to act (James et al. 2017)
- Human-*on*-the-loop (human between the Act and Observe stages): agent operates within specified goals; human can veto algorithmic decisions and plans (James et al. 2017)
- Human-*inside*-the-loop (human involvement between the Observe and Orient stages, and between the Orient and Decide stages): human analyzes and weighs in on agent decisions (James et al. 2017)

These models provide a means to express the relationship between the human and agent in a HAT in terms of their involvement (roles and levels) and functions, which are associated with various behaviors. The next step is to specify how behaviors that enhance the team can come about within the HAT.

5.2 Taxonomy of Team-Enhancing Behaviors

While the preceding section discusses the type and level of human and agent involvement in the team's task, and articulates how the teammates work together fulfilling different roles and functions, the taxonomy of TEBs outlines the different actions/behaviors taken by *each* teammate that enhances the team. In other words, for each of the TEBs to be exhibited, the teammate in question would have been engaged in the stages discussed in Table 8 (i.e., the human or agent teammate would be in the midst of information processing).

For a team to successfully execute a joint activity, all team members must 1) agree to work together, 2) be mutually predictable in their actions, 3) be mutually directable, and 4) maintain common ground (Klein et al. 2004). When applied to HATs, these require the human and agent team member to have a common understanding of the task or mission as well as knowledge of each other's roles, functions, and capabilities (for predictability). The agent should have adequate models of how the human teammate performs the task to be able to support the human with his/her tasks and information about the human's state to possibly render the correct aid to the human at the appropriate time. In turn, the human teammate needs to know the agent's status and understand the agent's functions and ways in which it can malfunction, which will allow him/her to be able to correct minor problems. There also needs to be good communication between the teammates (for "directability") (Klein et al. 2004).

In general, TEBs are behaviors by the human and/or agent team member in a HAT that preserve, magnify, optimize, enhance, and complement human capabilities, resulting in superior task performance. The premise is that although agents may reduce the need for the human to undertake certain tasks/subtasks, they do not supplant the human or render him/her redundant, as there are higher-level, more-meaningful tasks that require human intervention. TEBs are behaviors, not task conditions or factors in the task environment, although these can have a bearing on what TEBs are more effective at any point in time.

Figure 11 depicts the possible behavioral effects/TEBs within a HAT dyad that has been assigned a notional intelligence, surveillance, and reconnaissance task typically used in HAT studies where the agent scouts an area of interest, piping back a video feed of its environment to the human teammate, who searches the environment for predefined targets (Chen and Barnes 2008; Chen and Clark 2008; Baldwin et al. 2010; Sibley et al. 2010; Kidwell et al. 2012; Calhoun et al. 2014; Guznov et al. 2015; etc.).

The TEBs can be either teamwork behaviors or taskwork behaviors (Salas et al. 2008; Crawford and LePine 2013), all of which occur within a task environment. While taskwork TEBs are more task-specific, teamwork TEBs include backup behaviors, performance monitoring, behaviors showing adaptability, team leadership behaviors, behaviors indicating team orientation and trust, and closed-loop communication (Salas et al. 2009).

Specific types of TEBs are described in the following list and in Fig. 10. The vertical arrows refer to taskwork TEBs, while the horizontal arrows refer to teamwork TEBs. Each arrow shows the direction of the initiating effect (pointing *from* the initiating effort). There are two sets of TEBs for each arrow. Within each arrow, the first the set of TEBs depict the behaviors of the teammate initiating the effect, and the second set of TEBs denote the behaviors of other teammate in response to that effect. TEBs exhibited by the human are denoted by odd numbers (i.e., TEB$_{H}$1, TEB$_{H}$3, TEB$_{H}$5, TEB$_{H}$7, and TEB$_{H}$9), and TEBs by the agent are denoted by even numbers (i.e., TEB$_{A}$2, TEB$_{A}$4, TEB$_{A}$6, and TEB$_{A}$8). Subscripts denote the teammate exhibting the behavior ("H" for human teammate and "A" for agent teammate).

- TEB$_{H}$1: Self-management behaviors by the human (e.g., stress coping behaviors and/or emotional/time management behaviors)
- TEB$_{A}$2: Agent behaviors that show its understanding of how the human performs the task (e.g., generation of a model of human performance on the task and/or backup behaviors to support the human's tasking)
- TEB$_{H}$3: Taskwork behaviors by the human performing the task (e.g., detects targets in the environment)
- TEB$_{A}$4: Taskwork behaviors by the agent performing the task (e.g., travels through rough terrain in the area of interest collecting data)
- TEB$_{H}$5: Human behaviors that show his/her understanding of agent processes and functioning, and helps the agent perform the task (e.g., human dislodges agent that was stuck in muddy terrain and/or clears debris occluding its sensors)
- TEB$_{A}$6: Agent behaviors that show its understanding of the human's state and how they are teaming (e.g., agent's aiding behavior based on its model of human task performance and human's workload level)
- TEB$_{H}$7: Human behaviors to communicate to the agent (e.g., explicit and implicit closed-loop communication to agent about his/her state, the task, and how they are teaming)
- TEB$_{A}$8: Agent behaviors to communicate to the human about its state (e.g., agent informing human that it needs a battery recharge)
- TEB$_{H}$9: Human behaviors in response to agent's teaming behaviors (e.g., act on agent's recommendations, silence agent's updates, and adapt to agent's aid)

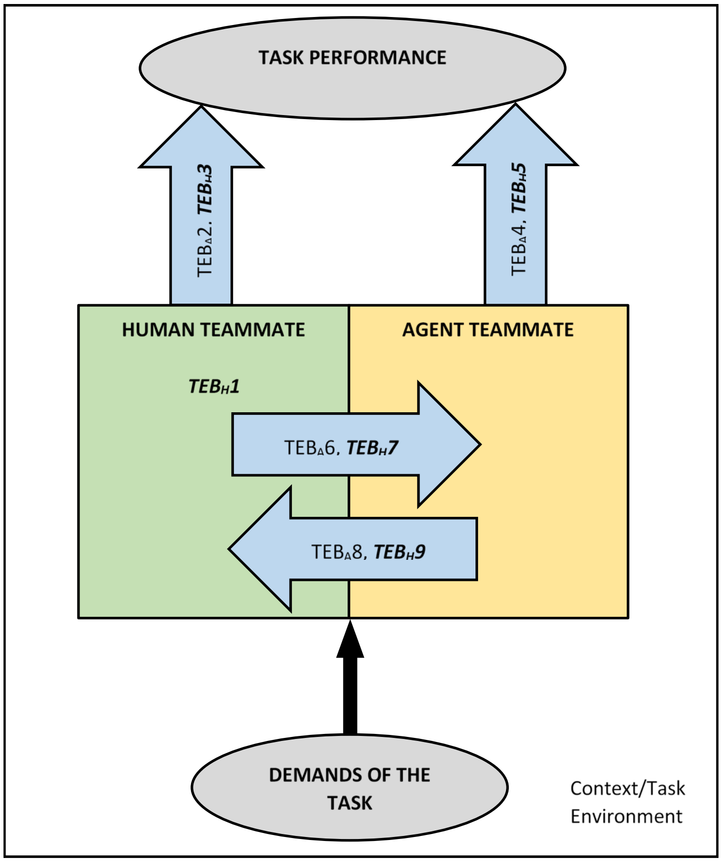

Fig. 10 Taxonomy of TEBs within a HAT (TEBs of the human are in bold italics)

The extent to which a given HAT can exhibit all these TEBs depends on the level and type of teammate involvement and role (see Table 7). For instance, a human teaming with an intelligent scope that highlights potential targets that appear whenever it senses that the human teammate requires detection assistance (i.e., TEB$_A$2) might not be able to detect human state and is unable to exhibit any kind of TEB$_A$6 or TEB$_A$8.

Like most behaviors, these TEBs are outcomes that are the manifestations of knowledge, skills, and attitudes (i.e., competencies) that drive them. Tracing these behavioral outcomes to the competencies that produce them will enable identification of training areas. Mapping TEBs to competencies and training areas can guide the research conducted and evaluation of training concepts in C-TEST.

5.3 Research Areas for Training TEBs in HAT/Crew

Unlike machine learning, training for the human (personnel training) is typically based on competencies, which are the knowledge, skills, and attitudes that drive effective behaviors (Cannon-Bowers et al. 1995; Stout et al. 1997, Hofrichter and

McGovern 2001; Salas et al. 2009). Identifying these desired outcomes will help focus training on the competencies that result in them.

Competencies are the knowledge, skills, and attitudes that enable personnel to accomplish his/her tasks and perform on the job. (Cannon-Bowers et al. 1995; Stout et al. 1997; Cannon-Bowers and Salas 1998). In teams, these knowledge, skills, and attitudes include task-related competencies as well as team-related competencies. Knowledge pertains to the concepts, definitions, models, and facts, and can include metacognitive knowledge (Krathwohl 2002) and implicit knowledge (Boshuizen and Schmidt 1992). These include understanding of team members' roles and responsibilities, cue-strategy associations, and shared mental models (Stout et al. 1997). Skills, which are related to knowledge to an extent, entail the proficiency at which data, ideas, or things are manipulated manually, verbally, or mentally, that allow for the smooth execution of certain well-defined tasks (Gotsch et al. 2012). Examples of skills that facilitate teamwork are adaptability, leadership and team management, interpersonal relations, coordination, communication, and decision making (Cannon-Bowers et al. 1995). Attitudes are the beliefs, inclinations, state of mind, and feelings regarding the task and team. These include willingness to be in the team, team orientation, team efficacy, trust, and cohesion (Cannon-Bowers et al. 1995; Cannon-Bowers and Salas 1998). Some attitudes such as trust may not be explicitly trained, but come about from possessing certain knowledge and skills. For instance, appropriate trust calibration results from knowing how the agent operates (knowledge) and being able to deal with agent failures (skills). Possible TEBs associated with appropriately calibrated trust in the agent include only accepting/rejecting the agent's recommendations after checking on the agent's rationale for recommendations.

Table 9 depicts the framework that traces the TEBs of the human teammate in Fig. 11 to the competencies and areas for which training can be developed:

Table 9 Training areas and associated competencies and outcomes (TEBs)

Focus of training for the human	Training area	Examples of competencies (knowledge, skills, attitudes) associated with TEB	TEB
Self	Train human on self-management Train human on task adaptability, (i.e., how to determine the best way to perform a task)	Self-knowledge, self-awareness. Time, resource management skills. Stress-coping skills. Self-efficacy. Impulse control. Openness to doing tasks differently. Ability to adapt to task situations.	TEB$_H$1
	Train human on the task (e.g., target detection)	Knowledge of target and distractor features. Search strategy and skill.	TEB$_H$3
Agent capabilities	Train human on agent's *tasking* capabilities	Knowledge of agent functioning (i.e., transparency). Skill to work with agent's tasking capabilities (e.g., computer, mechanical skill). Appropriate attitude toward agent (e.g., trust).	TEB$_H$5
	Train human on agent's *teaming* capabilities	Knowledge of agent's communication interfaces and teaming behaviors (e.g., type of adaptive/aiding behaviors). Skill to work with agent's teaming capabilities (e.g., computer, mechanical skill, team leadership skills). Appropriate attitude toward agent (e.g., trust).	TEB$_H$7, TEB$_H$9
Agent failures	Train human on possible agent errors/failures in its *tasking* capabilities	Knowledge of agent functioning (i.e., transparency). Skill to deal with or "fix" agent's tasking capabilities (e.g., computer, mechanical skill). Appropriate attitude towards agent (e.g., trust).	TEB$_H$5
	Train human on possible agent errors/failures in its *teaming* capabilities	Knowledge of agent's communication interfaces and teaming behaviors (e.g., type of adaptive/aiding behaviors). Skill to deal with or "fix"' agent's teaming capabilities (e.g., computer, mechanical skill, team leadership skills). Appropriate attitude toward agent (e.g., trust).	TEB$_H$7, TEB$_H$9

5.3.1 Training Focus No. 1: Self

Before involving the agent teammate, the human requires training in self-management and how to perform his/her task. Although self-management is fundamental and needed for all jobs, in most cases it is assumed that competencies related to self-management (e.g., self-awareness and coping skills) are adequate, so training in them is not needed. However, there may be instances where some training in stress coping skills is beneficial (e.g., tasks that expose the Soldier to potentially traumatizing stimuli). The other area of training enables the Soldier to perform his share of the task.

5.3.2 Training Focus No. 2: Agent Capabilities

Human TEBs that support agent capabilities can be targeted at the agent's taskwork or teamwork. Training the human on the agent's capabilities and functioning in how the agent performs its taskwork allows the human to exhibit TEBs that support the agent's tasking, while training on the agent's capabilities for teamwork, such as the type and ways that the agent may aid the human, will help the human to respond with the appropriate TEBs (e.g., able to deal with introduction or withdrawal of agent aid and/or task reallocations or changes in task pace due to agent intervention).

5.3.3 Training Focus No. 3: Agent Failures

In addition to training the human on the agent's capabilities and how the agent is to operate and function, it is also necessary to train the human about how the agent may fail and the possible errors that it can make in both taskwork and teamwork. This is because agents, like any piece of technology or system, are not perfectly reliable, and there will be occasions when the agent errs, fails, or does not behave as expected. The effects of such agent failure are often catastrophic, due in no small part to the "lumberjack" effect (Onnasch et al. 2014; Baker and Keebler 2017). The "lumber jack" effect is the observation that "more automation yields better human-system performance when all is well but induces increased dependence, which may produce more problematic performance when things fail" (Onnasch et al. 2014, p. 477). This trade-off between the benefits of a reliable agent and the expected costs (i.e., product of true costs and failure probability) when they fail should factor into decisions on what to automate and what capabilities the agent should have (Sheridan and Parasuraman 2000). Having the knowledge that the agent can malfunction and understanding the ways in which the agent can fail is a critical part of preparing the human teammate for unexpected events. Such training contributes to the resilience (Zhang and Lin 2010) of the HAT.

6. Supporting Training within NGCV-SIL

These training areas, while broad, can provide some guidance for developing and evaluating training concepts. The next step to enable research on these concepts requires a suitable testbed with the capability of supporting a variety of tasks undertaken by crew. This includes having real and/or simulated crew interfaces with systems and equipment (e.g., displays, controls, and instrumentation). In the data extraction of metadata from HAT studies performed previously, we identified a number of testbeds and experimental platforms that were used in HAT research (see Section 2.3.4) that supported some of the tasks undertaken by HAT/mixed crew. We expanded this work to develop a canvas of testbeds and platforms (see following section).

From our understanding of HAT research, crew/team roles, tasks, and functions, identification of TEBs as well as training areas for teams/crew, we could select the testbeds and platforms that can support the common crew tasks and the associated behaviors to be trained. By adapting and modifying some of these testbeds and platforms, we would not need to build the C-TEST infrastructure in its entirety, but we could establish the C-TEST proving ground for evaluating training concepts in a timely manner.

We envision that C-TEST's research on crew training concepts will be valuable to the NGCV-SIL community. While the NGCV-SIL focuses on crew performance within future systems, C-TEST would offer insights into the training of current and future crew/mixed-crew within those systems.

As mentioned previously, C-TEST seeks to address the current problem by developing a proving ground for training concepts. Within this proving ground will be a testbed or simulation environment for experimentation and eventual training. To support this, the C-TEST team assembled a list of candidate models and simulations that would be useful for HAT training. In modifying and adapting from existing testbeds, the development time for the C-TEST testbed will be shortened.

The following list was created from our own literature search and from a list of Soldier simulations created by the Natick Soldier Research and Development Center. Some simulations found in the literature search are not listed because they are no longer active programs. (Acronyms are included, as the simulations are usually better known by their acronyms.) Also, our investigations were restricted to the unclassified domain. The following lists the simulations reviewed:

- Mixed Initiative Experimental Testbed (MIX Testbed)
- Visualization Testbed (VTB)

- Robotics Collaborative Technology Alliance Mulitmodal Interface (RCTA MMI)
- One Semi-Automated Forces (OneSAF)
- DUJO Engine
- Wingman Software-in-the-Loop (Wingman SIL)
- SpatialOS
- Testbed for Integrated Ground Control Station (GCS) Experimentation and Rehearsal (TIGER)
- TARDEC Virtual Experimentation Capability (TVEC)
- ARMA: Armed Assault or ARMA 3
- Virtual Battlespace 3 (VBS3)
- RUAG Virtual Cab Simulator – Crew Training Simulator (CTS)
- Universal Mission Simulator
- Common Driver Trainer (CDT)
- VR Forces and related products (e.g., DI GUY)
- Flexible Analysis and Mission Effectiveness System (FLAMES)
- Simigon SIMBox
- Autonomous Navigation Virtual Environment Laboratory (ANVEL)
- Robotic Interactive Visualization Experimentation Technology (RIVET 1.0)
- Gazebo
- Virtual Combat Convoy Simulator
- Operator Driver Simulator
- Infantry Warrior Simulation (IWARS)
- Joint Conflict And Tactical Simulation (JCATS)
- Joint Semi-Automated Forces (JSAF)
- FireSim XXI
- Systemic Theater Operations Research Model (STORM)
- Close Combat Tactical Trainer (CCTT)
- Squad Synthetic Environment (SSE)
- Dismounted Soldier Training System (DSTS)
- Umbra
- Interactive Dante
- Combat Convoy Simulator (CCS)
- Scalable Advanced Graphics Engine (SAGE)
- Soldier Station
- Integrated Training Environment (ITE)
- Command, Control, and Communications Human Performance Model (C3HPM)
- Improved Performance and Research Integration Tool (IMPRINT)
- Human Performance Reliability/Error Analysis (HPRA)
- Integrated Performance Modeling Environment (IPME)
- Total Crew Model (TCM)
- Crew Station Design Tool (CSDT)

- Warfighter Physiologic Status Monitoring (WPSM)
- DIGuy Artificial Intelligence (DIGuy AI)
- Command, Control, and Communications Techniques for Reliable Assessment of Concept Execution (C3TRACE)
- Pythagoras
- Agent-Based Simulation for Network Enabled Capabilities (ABSNEC)
- Recognition Primed Decision-Making-enabled Collaborative Agents Stimulating Teamwork (R-CAST)
- Ascape
- FLAME
- Geographic Information System (GIS) Agent-based Modeling Architecture (GAMA)
- Multi-Agent Simulator Of Neighborhoods (or Networks) (MASON)
- Multi-Agent System Visualization (MASyV)
- Versatile Simulation Environment for the Internet (VSEit)
- Shell for Simulated Agent Systems (SeSAm)
- Simple Platform for Agent-based Representation of Knowledge (SPARK)
- TeamBots
- Man-machine Integration Design and Analysis System (MIDAS)

We evaluated each simulation on multiple properties and criteria. Table 10 is an excerpt of the full set (see Appendix A for complete set of properties).

Table 10 Excerpt of the properties in the review

Area	Category	Field	Description	Values	Interpretation
Technologies	Integration	License/ usage	License for how the product may be used and redistributed. May also specify licensing requirements for derivative works.	Open source, BSD, government-owned, government use rights, or other license name	Name of the license used for the specific product software if available (e.g. BSD, GPL, LGPL), otherwise a brief descriptor of standard use options.
…	…	Protocols	Communication protocols supported for use as part of a distributed simulation environment.	JAUS (version), DIS (version), HLA (version), custom open protocol, custom	List specific protocols (comma-separated) that the system uses and version numbers. Custom Open Protocol means states it uses a proprietary communication method, it is open and documented for use in other systems. Custom implies it is a closed custom protocol not currently available for use.
…	…	Integration with other systems/ protocols	Has the system integrated with systems and protocols?	Protocol, system	List of specific protocol (version number) and application names.
…	…	Modification possible	Is source code, plugin architecture, or other means of editing available to extend the system functionality for other means or integration with other applications?	Y, N, ?	Y for supporting the capability in some way. N for not at all; for not sure, ?.

Notes: BSD = Berkeley Software Distribution; GPL = General Public License; LGPL = Lesser General Public License; JAUS = Joint Architecture for Unmanned Systems; DIS = Distributed Interactive Simulation.

The following are some initial observations from our evaluation of the testbeds and simulations (see Appendix B for the full review):

- Many of the HAT simulations involve a robotic vehicle.
- We did not find a large number of simulations that provided artificial intelligence or intelligent agents in currently fielded systems.
- Integrating numerous models will be difficult outside of a consistent, common representation of the Soldier and the assistive technologies.
- We lack a centralized process for using these simulations and then providing our results back to the simulation developers for inclusion in the next iteration of the software
- There is no common database of simulation use across experiments and the results it has supported. This made us suspect our literature search is incomplete.
- A general C-TEST architecture can help inform common requirements for modeling and simulation in the HAT domain.

7. Conclusion

As the military strives to move forward with developing cutting-edge technologies, assessments and research must keep pace. Following the initial Phase I efforts, our work in Phase II sought to improve assessments by applying a novel network analysis to derive structure and extract patterns of construct operationalizations in research domains. However, current modernizing of command and emphasis on the implementation and delivery of new materiel and Soldier capabilities have necessitated a shift in project direction toward assessment issues related to crew/mixed-crew training and enhancement. The C-TEST project was formulated as a result.

C-TEST recognizes that much of the training curriculum and simulations for HAT/crew have not been appropriately assessed to determine if they meet learning requirements and transfers to the field. This implies that crew/mixed-crew training may not be as effective as claimed, jeopardizing missions and lives. As a direct response to this assessment concern, we seek to build a proving ground for training concepts prior to final system development. This serves as a validity check and a part of risk-reduction effort before training implementation.

This report documents the activities to develop the C-TEST proving ground. Activity 1 involved understanding of the training cycle in ARFORGEN and mapping out of the S&T “HAT training landscape”. Activity 2 entailed 1) developing frameworks to guide research into HAT/crew training concepts and 2) evaluation of HAT/crew training curricula and simulations. Whereas the taxonomy of agent involvement and framework of teammate functions covered the different ways and extent agents can team in a HAT, the taxonomy of TEBs explicated the effects within the HAT/crew and specified the various teamwork and taskwork behaviors by the human and agents that would augment the team. These behaviors were then mapped to areas of training (i.e., competencies) for the human in the HAT/crew. Activity 3 addressed the next step toward establishing the C-TEST proving ground, which was developing the appropriate testbed within the NGCV-SIL environment.

These efforts have culminated in a systematic approach for conducting HAT/crew research and assessments that can ease the process of transition towards development and implementation. Future work will continue along this trajectory to help ensure that deliverables to enhance and deliver capabilities to the Soldier meet expectations in the field.

8. References

Baker AL, Keebler JR. Factors affecting performance of human-automation teams. In: Savage-Knepshield P, Chen J, editors. Advances in human factors in robots and unmanned systems. New York (NY): Springer International Publishing; 2017;499:331–340.

Baldwin C, Coyne JT, Roberts DM, Barrow JH, Cole A, Sibley C, Buzzell G. Prestimulus alpha as a precursor to errors in a UAV target orientation detection task. Washington (DC): Naval Research Laboratory (US); 2010.

Barber D, Davis L, Nicholson D, Finkelstein N, Chen JYC. The mixed initiative experimental (MIX) testbed for human robot interactions with varied levels of automation. Orlando (FL): University of Central Florida; 2008 Dec.

Boles DB, Adair LP. The multiple resources questionnaire (MRQ). Proceedings of the 2001 Human Factors and Ergonomics Society Annual Meeting. Los Angeles (CA): SAGE Publications; c2001 Oct 1.

Borgatti SP, Mehra A, Brass DJ, Labianca G. Network analysis in the social sciences. Science. 2009;323(5916):892–895.

Boshuizen HP, Schmidt HG. On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. Cognitive Science. 1992;16(2):153–184.

Boyd JR. The essence of winning and losing. Unpublished Lecture Notes. 1996;12(23):123–125.

Bruemmer DJ, Walton MC. Collaborative tools for mixed teams of humans and robots. Idaho Falls (ID): Idaho National Engineering and Environmental Laboratory; 2003.

Burford C, Reinerman L, Teo G, Matthews G, McDonnell J, Orvis K, Metevier C. Unified multimodal measurement for performance indication research, evaluation, and effectiveness (UMMPIREE). Phase I Report. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2018 Jan. ARL-TR-8277.

Calhoun GL, Miller CA, Hughes TC, Draper MH. UAS sense and avoid system interface design and evaluation. Proceedings of the 2014 Human Factors and Ergonomics Society Annual Meeting. Los Angeles (CA): SAGE Publications; c2014 Oct 17.

Cannon-Bowers JA, Salas E. Team performance and training in complex environments: recent findings from applied research. Current Directions in Psychological Science. 1998;7(3):83–87.

Cannon-Bowers JA, Tannenbaum SI, Salas E, Volpe CE. Defining team competencies and establishing team training requirements. In: Guzzo R, Salas E, editors. Team effectiveness and decision making in organizations. San Francisco (CA): Jossy-Bass; 1995. p. 333–380.

Chen JY, Barnes MJ. Robotics operator performance in a military multi-tasking environment. Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction; 2008. p. 279–286.

Chen JYC, Terrence PI. Effects of imperfect automation and individual differences on concurrent performance of military and robotics tasks in a simulated multitasking environment. Ergonomics. 2009;52(8):907–920.

Chen JY, Clark BR. UAV-guided navigation for ground robot operations. Proceedings of the 2008 Human Factors and Ergonomics Society Annual Meeting. Los Angeles (CA): SAGE Publications; c2008. p. 1412–1416.

Crawford ER, LePine JA. A configural theory of team processes: accounting for the structure of taskwork and teamwork. Academy of Management Review. 2013;38(1):32–48.

Endsley MR. Level of automation effects on performance, situation awareness and workload in a dynamic control task. Ergonomics. 1999;42(3):462–492.

Finomore Jr V, Popik D, Dallman R, Stewart J, Satterfield K, Castle C. Demonstration of a network-centric communication management suite: multi-modal communication. Proceedings of the 2011 Human Factors and Ergonomics Society Annual Meeting. Los Angeles (CA): SAGE Publications. c2011. p. 1832–1835.

Gotsch AR, Keck CW, Spencer HC. Knowledge, skills, and attitudes (KSAs) for the public health preparedness and response core competency model. Association of Schools of Public Health; 2012 [accessed 2018 Nov 29]. www.midamericaphp.com/wp-content/uploads/2009/12/KSA.pdf.

Guznov S, Nelson A, Lyons J, Dycus D. The effects of automation reliability and multi-tasking on trust and reliance in a simulated unmanned system control task. In: Kurosu M, editor. Human-computer interaction: users and contexts. conference on human-computer interaction. New York (NY): Springer; 2015. p. 616–621.

Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Advances in Psychology. 1988;52:139–183.

Hofrichter DA, McGovern T. People, competencies and performance: clarifying means and ends. Compensation & Benefits Review. 2001;33(4):34–38.

James A, Pierowicz J, Moskal M, Hanratty T, Tuttle D, Sensenig B, Hedges B. Assessing consequential scenarios in a complex operational environment using agent-based simulation. Aberdeen Proving Ground, (MD): Army Research Laboratory (US); 2017 Mar. ARL-TR-7954.

Jian JY, Bisantz AM, Drury CG. Foundations for an empirically determined scale of trust in automated systems. International Journal of Cognitive Ergonomics. 2000;4(1):53–71.

Johnson R, Leen M, Goldberg D, Chiu M. Adaptive levels of autonomy (ALOA) for UAV supervisory control. Wright-Patterson AFB (OH): Air Force Research Laboratory (US); 2005. Report No.: AFRL-HE-WP-TR-2005-0062.

Kidwell B, Calhoun GL, Ruff HA, Parasuraman R. Adaptable and adaptive automation for supervisory control of multiple autonomous vehicles. Proceedings of the 2012 Human Factors and Ergonomics Society Annual Meeting. Los Angeles (CA): SAGE Publications; 2012.

Klein G, Woods DD, Bradshaw JM, Hoffman RR, Feltovich PJ. Ten challenges for making automation a "team player" in joint human–agent activity. IEEE Intelligent Systems, 2004;19(6):91–95.

Krathwohl DR. A revision of Bloom's taxonomy: an overview. Theory into Practice. 2002;41(4):212–218.

Lesperance BG. Memo: letter to industry on desired capabilities for next generation combat vehicle – optionally manned fighting vehicle (NGCV-OMFV). Fort Benning (GA): Next Generation Combat Vehicle Cross Functional Team; 2018 July 5.

Matthews G, Campbell SE, Falconer S, Joyner LA, Huggins J, Gilliland K, Warm JS. Fundamental dimensions of subjective state in performance settings: task engagement, distress, and worry. Emotion. 2002;2(4):315.

McCarthy R. US Army directive 2017-24. Cross-functional team pilot in support of materiel development. Memorandum No.: ARN6101_AD2017-24. https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/ARN6101_AD2017-24_Web_Final.pdf.

Moore R, Lodwick W. Interval analysis and fuzzy set theory. Fuzzy Sets and Systems. 2003;135(1):5–9.

Morrison P. VBS2 release version 2.0: white paper. Nelson Bay (Australia): Bohemia Interactive Australia PTY Ltd; 2012.

Nelson E. Internal set theory: a new approach to nonstandard analysis. Bulletin of the American Mathematical Society. 1977;83(6):1165–1198.

Onnasch L, Wickens CD, Li H, Manzey D. Human performance consequences of stages and levels of automation: an integrated meta-analysis. Human Factors. 2014;56(3):476–488.

Parasuraman R, Riley V. Humans and automation: use, misuse, disuse, abuse. Human Factors. 1997;39(2):230–253.

Parasuraman R, Sheridan TB, Wickens CD. A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans. 2000;30(3):286–297.

Salas E, Cooke NJ, Rosen MA. On teams, teamwork, and team performance: discoveries and developments. Human Factors: The Journal of the Human Factors and Ergonomics Society. 2008;50(3):540–547.

Salas E, Rosen MA, Burke CS, Goodwin GF. The wisdom of collectives in organizations: an update of the teamwork competencies. In: Team effectiveness in complex organizations: cross-disciplinary perspectives and approaches. Abingdon (UK): Rutledge/Taylor & Francis Group; 2009. p. 39–79.

Schaefer K. The perception and measurement of human-robot trust. Orlando (FL): University of Central Florida; 2013.

Sheridan TB, Parasuraman R. Human versus automation in responding to failures: an expected-value analysis. Human Factors. 2000;42(3):403–407.

Sibley C, Cole A, Gibson G, Roberts D, Barrow J, Baldwin C, Coyne J. Adaptive training in an unmanned aerial vehicle: examination of several candidate real-time metrics. Washington (DC): Naval Research Laboratory (US); 2010.

Sibley C, Coyne J, Thomas J. Demonstrating the Supervisory Control Operations User Testbed (SCOUT). Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2016;60(1):1324–1328. https://doi.org/10.1177/1541931213601306.

Stout RJ, Salas E, Fowlkes JE. Enhancing teamwork in complex environments through team training. Group Dynamics: Theory, Research, and Practice. 1997;1(2):169–182.

Sudit, M. (2015). Computer prediction is an unrealistic panacea, cognitive prediction is not scalable, so computer-human prediction is tactically attainable. Presented at the CUBRC workshop, Buffalo, NY.

Teo G, Reinerman-Jones L, Riecken ME, McDonnell J, Gallant S, Hidalgo M, Burford CW. A workflow for network analysis-based structure discovery in the assessment community. In: International Conference on Augmented Cognition. New York (NY): Springer; 2018. p. 341–352.

National Science and Technology Council. The National Artificial Intelligence Research and Development Strategic Plan; 2016 [accessed 2018 Jan]. https: //www.arl.army.mil/www/pages/2401/ERA_Summary_Sheets_Jan2018.pdf.

neo4j. The Neo4j graph platform: The #1 platform for connected data [accessed 2018 July]. https://neo4j.com/.

US Army Research Laboratory essential research areas 2018 [accessed 2018 Nov 29]. https://www.arl.army.mil/www/pages/2401/ERA_Summary_Sheets_Jan2018.pdf.

US Army STAND-TO! Army Futures Command (AFC); 2018 [accessed 2018 Oct 24]. https://www.army.mil/standto/2018-03-28.

West DB. Introduction to graph theory. Vol. 2. Upper Saddle River (NJ): Prentice Hall; 2005.

Zhang WJ, Lin Y. On the principle of design of resilient systems – application to enterprise information systems. Enterprise Information Systems 2010;4(2): 99–110.

Appendix A. Properties of Testbeds and Simulations

This appendix appears in its original form without editorial change.

Area	Category	Field	Description	Possible Values	Interpretation
Technologies	Integration	License/ Usage	License for how the product may be used and redistribution. May also specify licensing requirements for derivative works.	Open Source - BSD, Government Owned, Government Use Rights, or other license name	Name of the license used for the specific product software if available (e.g. BSD, GPL, LGPL), otherwise a brief descriptor of standard use options.
…	…	Protocols	Communication protocols supported for use as part of a distributed simulation environment.	JAUS (VERSION), DIS (VERSION), HLA (VERSION), Custom Open Protocol, Custom	List specific protocols (comma separated) that the system uses and version numbers. Custom Open Protocol means states it uses a proprietary communication method, it is open and documented for use in other systems. Custom implies it is a closed custom protocol not currently available for use.
…	…	Integration with Other Systems/ Protocols	Has the system integrated with systems and protocols.	PROTOCOL, SYSTEM	List of specific protocol (version number) and application names.
…	…	Modification Possible	Is source code, plugin architecture, or other means of editing available to extend the system functionality for other means or integration with other applications?	Y, N, ?	Y for supporting the capability in some way. N for not at all. ? for not sure.
…	…	Live	Has the system been used in or is part of live environment?	Y, N, ?	Y if yes, N for not at all, ? for not sure.
…	…	Virtual	Has the system been used in or is part of a virtual environment?	Y, N, ?	Y if yes, N for not at all, ? for not sure.
…	…	Constructive	Has the system been used in or is part of a constructive environment?	Y, N, ?	Y if yes, N for not at all, ? for not sure.

Area	Category	Field	Description	Possible Values	Interpretation
…	Sensors	Sensor Production	Does the system provide real or simulated sensor data for agents, entities, or other sources. For example, if it simulates a robot, does it provide virtual video feeds, or simulated reconnaissance reports?	Y, N, ?	Comma separated values list of different sensors simulated, N for not at all, ? for not sure.
…	…	Sensor Acquisition	Does the system provide methods for data acquisition from simulated agents, entities, or other sources For example, can it connect to a physiological sensor or video camera and record?	Y, N, ?	Comma separated values list of different sensors simulated, N for not at all, ? for not sure.
…	Mixed Reality	Virtual Reality	Does the system support virtual reality applications?	Y, N, ?	Y if yes, N for not at all, ? for not sure.
…	…	Augmented Reality	Does the system support augmented reality applications?	Y, N, ?	Y if yes, N for not at all, ? for not sure.
…	Modalities	Auditory - Sound Files	Does the system support playback of auditory messages from pre-recorded sound files?	Y, N, ?	Y if yes, N for not at all, ? for not sure.
…	…	Auditory - TTS	Does the system support playback of auditory messages using text-to-speech (TTS)?	Y, N, ?	Y if yes, N for not at all, ? for not sure.
…	…	Auditory - STT	Does the system support speech-to-text (STT) data collection or use within applications?	Y, N, ?	Y if yes, N for not at all, ? for not sure.
…	…	Tactile Cues	Does the system support generation of tactile cues or haptics?	Y, N, ?	Y if yes, N for not at all, ? for not sure.
…	…	Tactile Messages (Belt)	Does the system support generation of tactile messages (e.g. tacton) using a belt or vest?	Y, N, ?	Y if yes, N for not at all, ? for not sure.
…	…	Video	Does the system generate simulated video feeds from unattended sensors, unmanned vehicles, or other sources?	Y, N, ?	Y if yes, N for not at all, ? for not sure.

Area	Category	Field	Description	Possible Values	Interpretation
…	…	Imagery	Does the system generate imagery information to participants? This can be in the form of simple stimulus for a laboratory experiment or as part of reports.	Y, N, ?	Y if yes, N for not at all, ? for not sure.
…	Unmanned Systems	Operator Control Unit	Does the system simulate or provide an operator control unit for unmanned systems. This includes tablet devices for small portable systems up to multi-agent control units.	Y, N, ?	Y if yes, N for not at all, ? for not sure.
…	…	Squad Level	Does the system simulated unmanned systems for squad level interactions?	Y, N, ?	Y if yes, N for not at all, ? for not sure.
…	…	Crew Support Vehicle	Does the system provide or simulate capabilities for crew with unmanned systems?	Y, N, ?	Y if yes, N for not at all, ? for not sure.
…	…	Route Following	Do vehicles and or unmanned systems support route following to a-priori waypoints, goal objects, or other inputs?	Y, N, ?	Y if yes, N for not at all, ? for not sure.
…	…	Target Detection	Do unmanned systems provided perform target detection?	Y, N, ?	Y if yes, N for not at all, ? for not sure.
…	…	Weapon Management	Do unmanned systems provided perform any weapon management or force application?	Y, N, ?	Y if yes, N for not at all, ? for not sure.
…	…	Sensor Management	Do unmanned systems provide multiple sensors and provide management of the data?	Y, N, ?	Y if yes, N for not at all, ? for not sure.
…	…	Formation Following	Do unmanned systems support staying information on routes or in dismounted applications?	Y, N, ?	Y if yes, N for not at all, ? for not sure.
Experimentation & Training	Application	Training	Has the system supported training?	Y, N, ?	Y for representing the capability in some way. N for not at all. ? for not sure.

Area	Category	Field	Description	Possible Values	Interpretation
…	…	Experimentation	Has the system been used in laboratory or field experiments?	Y, N, ?	Y for representing the capability in some way. N for not at all. ? for not sure.
…	Population	Novice	Was the population novice to the domain of study?	Y, N, ?	Y for representing the capability in some way. N for not at all. ? for not sure.
…	…	Warfighter	Have warfighters used the system in training or experiments?	Y, N, ?	Y for representing the capability in some way. N for not at all. ? for not sure.
…	Usage	Number of Studies/ Exercises	How many studies or exercies has the system supported or been used in?	#, 100+, ?, N	# representing a specific number of studies or exercies run, ? for not sure, N for not at all, and 100+ for more than 100.
…	…	Typical Sample Size	Typical sample size for each study or exercise.	#, ?, N	# represents actual mean sample size for exercises run, ? for not sure, N for not at all (no studies run).
…	Logging	Format	Log file formats the system records to.	TXT, CSV, BINARY, N/A	TXT representing human readable text files in either tab, comma, or other delimited format. CSV if known to be comma separated values. BINARY to represent a binary internal representation that is not human readable. N/A for no logging or not applicable.
…	…	Screen Capture / Video	Is screen capture or other video recording supported?	Y, N, ?	Y for representing the capability in some way. N for not at all. ? for not sure.
…	…	Audio (Microphone)	Is audio recording available?	Y, N, ?	Y for representing the capability in some way. N for not at all. ? for not sure.
…	…	GUI Events	Are Graphical User Interface (GUI) events logged. In other words, if the user presses a button, is that event logged and timestamped?	Y, N, ?	Y for representing the capability in some way. N for not at all. ? for not sure.

Area	Category	Field	Description	Possible Values	Interpretation
…	…	Simulation Events	Are simulation driven events logged?	Y, N, ?	Y for representing the capability in some way. N for not at all. ? for not sure.
…	…	Performance	Are performance measures for tasks or other activities within the system calculated and recorded?	Y, N, ?	Y for representing the capability in some way. N for not at all. ? for not sure.
…	…	Analysis Tools	Are tools available to aggegreate data for reporting or statistical analysis?	Y, N, ?	Y for representing the capability in some way. N for not at all. ? for not sure.
…	Scripting & Control	Time Event Triggers	If scripting support for scenarios is available, are events triggered using time?	Y, N, ?	Y for representing the capability in some way. N for not at all. ? for not sure.
…	…	Location Event Triggers	If scripting support for scenarios is available, are events triggered using location data (e.g. user moves character to a location, entity moves to a location).	Y, N, ?	Y for representing the capability in some way. N for not at all. ? for not sure
…	…	Other Event Triggers	Are other event triggers supported or combinations of time and location?	Y, N, ?	Y for representing the capability in some way. N for not at all. ? for not sure
Human Factors	…	Trust	Does the system capture or provide a measure of trust?		Description of measures, and where applicable citation information.
…	…	Stress	Does the system capture or provide a measure of stress?		Description of measures, and where applicable citation information.
…	…	Workload	Does the system capture or provide a measure of workload?		Description of measures, and where applicable citation information. Examples include NASA-TLX, electroencephalogram, and heart rate variability.
…	…	Spatial Awareness	Does the system capture or provide a measure of spatial awareness?		Description of measures, and where applicable citation information.

Area	Category	Field	Description	Possible Values	Interpretation
Mission Forces	Fires	Remote Controlled Fires	Representation of a human remotely controlling a gun of a variety of types. The human is not manually firing the gun via trigger or button, but rather using a device separate from the gun to fire.	Y, N, ?, N/A	Y for representing the capability in some way. N for not at all. ? for not sure
…	…	Manual Fires	Representation of the human firing the gun themselves using the trigger or button.	Y, N, ?, N/A	Y for representing the capability in some way. N for not at all. ? for not sure
…	Defensive Posture	Reaction to Fire	Representation of the forces executing the TTP for reacting to fire.	Y, N, ?, N/A	Y for representing the capability in some way. N for not at all. ? for not sure
…	…	Security	Representation of the TTP for security posture.	Y, N, ?, N/A	Y for representing the capability in some way. N for not at all. ? for not sure
…	Movement	Navigation	Representation of a human or intelligent agent providing navigation (planning of movement)	Y, N, ?, N/A	Y for representing the capability in some way. N for not at all. ? for not sure
…	…	Driving Vehicle	Representation of a human driving a platform that they are physically in/on.	Y, N, ?, N/A	Y for representing the capability in some way. N for not at all. ? for not sure
…	…	Remote Controlled Guidance	Representation of a human remotely providing guidance for a platform such as waypoints, formation, stop/go, etc.	Y, N, ?, N/A	Y for representing the capability in some way. N for not at all. ? for not sure
…	…	Remote Controlled Driving	Representation of a human driving a platform using a remote device and they are not on that platform.	Y, N, ?, N/A	Y for representing the capability in some way. N for not at all. ? for not sure
…	Target Acquisition	Camera Control	Controlling a camera manually or using a remote device to control a camera with the intent of somebody using that visual to detect targets.	Y, N, ?, N/A	Y for representing the capability in some way. N for not at all. ? for not sure

Area	Category	Field	Description	Possible Values	Interpretation
…	…	Detection	Examining the area, a camera, or a sensor in order to detect targets.	Y, N, ?, N/A	Y for representing the capability in some way. N for not at all. ? for not sure
Mixed-Initiative Teaming	…	Intelligent Agents	Use of autonomous entity that receives information via a sensor or data stream and acts according to its goals based on that stimuli.	Y, N, ?, N/A	Y for representing the capability in some way. N for not at all. ? for not sure
…	…	Asset Management	The management of manned and unmanned platforms, sensors, weapons, and defenses.	Y, N, ?, N/A	Y for representing the capability in some way. N for not at all. ? for not sure
…	…	Crew Coordination	The representation of human coordinating amongst themselves and with intelligent agents or robots (mixed teams)	Y, N, ?, N/A	Y for representing the capability in some way. N for not at all. ? for not sure
…	…	Teleoperation	The control of machines using an electronic remote device	Y, N, ?, N/A	Y for representing the capability in some way. N for not at all. ? for not sure
…	…	Collaboration	The representation of humans and intelligent agents/robots working together to achieve goals	Y, N, ?, N/A	Y for representing the capability in some way. N for not at all. ? for not sure
NGCV	…	Autonomous Navigation	Representation of the NGCV concept of platforms executing all of their own navigation via formation, terrain, and sensing of other vehicles and obstacles	Y, N, ?, N/A	Y for representing the capability in some way. N for not at all. ? for not sure
…	…	2 Operators Per UGV	Representation of the NGCV concept of two humans operating each accompanying UGV - one operating the camera/sensors and one remotely driving	Y, N, ?, N/A	Y for representing the capability in some way. N for not at all. ? for not sure
…	…	3 Operators Per UGV	Representation of the NGCV concept of three humans operating each accompanying UGV - one managing operation, one operating the camera/sensors, and one remotely driving	Y, N, ?, N/A	Y for representing the capability in some way. N for not at all. ? for not sure

Area	Category	Field	Description	Possible Values	Interpretation
Future Concepts	Land Carrier	Lizard	Representation of a modular, articulated, multifunctional vehicle that is reconfigurable, modular, and wheeled. Includes many types of modules (medical, missiles, transport, directed energy, communications, etc.)	Y, N, ?, N/A	Y for representing the capability in some way. N for not at all. ? for not sure
…	…	Mothership	Representation of vehicles that include pods, deployable UAS/UGVs. Modular and reconfigurable per use case. Can be combined with Lizard concept	Y, N, ?, N/A	Y for representing the capability in some way. N for not at all. ? for not sure
…	…	Swarming	Representation of manned and unmanned cooperation for behaviors such as reconnaissance, attack, security, or traversing difficult terrain using sensors.	Y, N, ?, N/A	Y for representing the capability in some way. N for not at all. ? for not sure
…	…	Modular UAS Carrier	Representation of a platform that is tracked and joins the force for storing, charging, and deploying UAS	Y, N, ?, N/A	Y for representing the capability in some way. N for not at all. ? for not sure
…	…	Megacity Vertical Fighting Vehicle	Representation of the platform that is oriented and optimized for operations in the vertical plane. Tilt-able carousel launchers, identifying threats, and includes unmanned operation.	Y, N, ?, N/A	Y for representing the capability in some way. N for not at all. ? for not sure

Appendix B. Testbeds and Simulations Reviewed

This appendix is included as an Excel attachment to the published pdf.

List of Symbols, Abbreviations, and Acronyms

ARFORGEN	Army Force Generation
AFRL	US Air Force Research Laboratory
ALOA	Adaptive Levels of Autonomy
ARL	US Army Research Laboratory
BSD	Berkeley Software Distribution
CFT	Cross-Functional Team
CNS	central nervous system
C-TEST	Crew Training Enhancement and Systems Testbed
DIS	Distributed Interactive Simulation
DV	dependent variable
EXORD	execute order
GPL	General Public License
HAT	human–agent team(s) or human–agent teaming (depending on context)
HMT	human–machine teaming
IP	Internet Protocol
IV	independent variable
JAUS	Joint Architecture for Unmanned Systems
LGPL	Lesser General Public License
MIX	Mixed Initiative Experimental Testbed
MMC	Multi-modal Communication
NASA-TLX	NASA Task Load Index
NGCV	Next-Generation Combat Vehicle
NRL	US Naval Research Laboratory
OODA	Observe-Orient-Decide-Act
S&T	science and technology

SCOUT	Supervisory Control Operations User Testbed
SIL	System/Software Integration Laboratory
SME	subject matter expert
STE	Synthetic Training Environment
STT	speech-to-text
TARDEC	Tank Automotive Research, Development and Engineering Center
TCU	Tactile Control Unit
TEB	team-enhancing behavior
TTS	text-to-speech
UMMPIREE	Unified Multimodal Measurement for Performance Indication Research, Evaluation, and Effectiveness
UAS	unmanned aerial system
UAV	unmanned aerial vehicle
UGV	unmanned ground vehicle
VBS2	Virtual Battlespace 2

1 DEFENSE TECHNICAL
(PDF) INFORMATION CTR
DTIC OCA

2 DIR ARL
(PDF) IMAL HRA
RECORDS MGMT
RDRL DCL
TECH LIB

1 GOVT PRINTG OFC
(PDF) A MALHOTRA

1 ARL
(PDF) RDRL HRB B
T DAVIS
BLDG 5400 RM C242
REDSTONE ARSENAL AL
35898-7290

1 ARL
(PDF) RDRL HRB A
J EVERETTE
BLDG E2929
DESERT STORM DR
FORT BRAGG NC
28310-0001

7 ARL
(PDF) SFC PAUL RAY SMITH CTR
RDRL HRA M CLARKE
RDRL HRA I MARTINEZ
RDRL HRR R SOTTILARE
RDRL HRA C A RODRIGUEZ
RDRL HRA B J HART
RDRL HRA A C METEVIER
RDRL HRA D B PETTIT
12423 RESEARCH PARKWAY
ORLANDO FL 32826

1 USA ARMY G1
(PDF) DAPE HSI M SAMS
300 ARMY PENTAGON
RM 2C489
WASHINGTON DC 20310-0300

1 USAF 711 HPW
(PDF) 711 HPW/RH K GEISS
2698 G ST BLDG 190
WRIGHT PATTERSON AFB OH
45433-7604

1 USN ONR
(PDF) ONR CODE 341 J TANGNEY
875 N RANDOLPH STREET
BLDG 87
ARLINGTON VA 22203-1986

1 USA NSRDEC
(PDF) RDNS D D TAMILIO
10 GENERAL GREENE AVE
NATICK MA 01760-2642

1 OSD OUSD ATL
(PDF) HPT&B B PETRO
4800 MARK CENTER DRIVE
SUITE 17E08
ALEXANDRIA VA 22350

ABERDEEN PROVING GROUND

13 ARL
(PDF) RDRL HR
J LANE
J CHEN
P FRANASZCZUK
K MCDOWELL
K OIE
RDRL HRA A
C BURFORD
RDRL HRB
J LOCKETT
RDRL HRB C
J GRYNOVICKI
RDRL HRB D
D HEADLEY
RDRL HRF A
A DECOSTANZA
RDRL HRF B
A EVANS
RDRL HRF C
J GASTON
RDRL HRF D
A MARATHE