Crone, D., Sanderson, P., Naikar, N. & Parker, S. (2007). Selecting sensitive measures of performance in complex multivariable environments. Proceedings of the 2007 Simulation Technology Conference (SimTecT 2007). Brisbane, Australia, 4-7 June 2007.

Selecting Sensitive Measures of Performance in Complex Multivariable Environments

Mr. David Crone
Air Operations Division, DSTO
david.crone@dsto.defence.gov.au

Prof. Penelope Sanderson
ARC Key Centre for Human Factors, The University of Queensland
psanderson@itee.uq.edu.au

Dr. Neelam Naikar
Air Operations Division, DSTO
neelam.naikar@dsto.defence.gov.au

Dr Simon Parker
Air Operations Division, DSTO
simon.parker@dsto.defence.gov.au

Abstract. The Defence Science and Technology Organisation (DSTO) is required to provide advice to customers for the procurement of complex future military systems. The high fidelity human-in-the-loop simulation (HILS) facility housed in the Air Operations Simulation Centre (AOSC) is increasingly being used to evaluate these novel systems. One of the benefits of the high fidelity HILS facility is the opportunity to explore realistic human and equipment behaviour. However, with the increase in realism there is a commensurate increase in the number of possible equipment and human performance measures that can be selected for use in evaluation. With such a large number of candidate measures a question arises: What measure(s) of performance *should* be used to evaluate the system of interest? Researchers at DSTO have conducted several experiments assessing a new method for selecting measures that will provide a sensitive evaluation of the impact of a new subsystem on human-equipment performance. The new approach emphasises a constraint-based, rather than a task-based, view of human and equipment behaviour. The value of this method is that unknown, future behaviour may be accounted for and explored within the HILS. This paper reports the results from an experiment assessing the sensitivity of measures of performance derived from the new method, and outlines the theoretical position adopted.

1. INTRODUCTION

The problem of evaluating whether a change to a system[1] has been effective may be approached from two distinct work analysis methods. On the one hand, task analysis (TA: [1]) is a widely-used traditional method in Human Factors Engineering (HFE). The task analysis community takes the view that the basic unit of analysis is "the task". On the other hand the newer HFE approach of Cognitive Work Analysis (CWA: [2, 3]) emphasises the constraints of the system. The approaches have been compared on different dimensions (e.g. [4]) but there has been no work that compares them on their usefulness for selecting measures that are sensitive to system change. We are currently addressing this issue in a program of research [5]. A key hypothesis of the research is that CWA-based measures will be more sensitive to a system change than task-based measures when performance is evaluated in a human-in-the-loop simulation (HILS) environment.

In this paper we report the results of applying a CWA-based method for selecting performance measures for an upgrade of a current system. We will show that CWA can be used to select sensitive system performance measures for system evaluation within the HILS environment. After a brief review of the existing Human Factors Engineering evaluation methods, a typical Black Hawk airmobile operation is described. The CWA-based framework that was the precursor for the CWA-based method will then be described[2]. Following this, an outline of the experiments and results will be presented.

2. BACKGROUND

The Defence Science and Technology Organisation (DSTO) provides advice to customers for the procurement of future

[1] In this paper the term "system" is used to encompass human, equipment sub-systems, and the environment. Hence, a change to system could include any modification to hardware, software, the environment and human behaviour.

[2] The actual CWA-based method that was developed to select measures will be reported at a later date.

military systems. The high fidelity HILS facility housed in the Air Operations Simulation Centre (AOSC) is increasingly being used to facilitate the evaluation of these novel systems. One of the benefits of the HILS facility is the opportunity to explore realistic human and equipment behaviour. However, with the increase in realism there is a commensurate increase in the number of possible performance measures that can be used for evaluation. The question arises: What measure(s) of performance *should* be used to evaluate complex military systems?

2.1 Human Factors Engineering methods for selecting human-equipment performance measures

Human Factors Engineering methods for evaluating systems may be broadly divided into task-based and constraint-based work-analysis methods.

Selecting measures for use via task–based methods is influenced by several factors such as practitioner experience, resources available, and criteria including reliability, validity and sensitivity. Irrespective of the stage of evaluation, task-based methods offer a common approach to testing system improvements. This approach involves reproducing the tasks and activities that the system will typically perform, selecting measures that relate to that task, choosing a success/fail criterion for each measure and then testing to see whether or not the system can perform the tasks against the success/fail criterion. For new systems, where there is little knowledge of what tasks and activities may be performed, the analyst develops tasks that represent the best assessment of what the system behaviour may be. These tasks are then used as the basis for testing the system. This is a risky strategy, however, because the analyst's assessment of future tasks may be wrong and the measures chosen (based on the task) may not be the ones that reveal an important system performance benefit or deficit.

Constraint-based methods, as represented by Cognitive Work Analysis (CWA), have been developed to design and evaluate complex systems. CWA is based on an ecological approach to work analysis. The ecological approach stresses "work demands associated with factors external to the worker" as opposed solely to "work demands associated with worker cognitive characteristics" [3].

The aim of CWA is to capture the factors that constrain or shape human behaviour. It contains five analytical phases: Work Domain Analysis (WDA; an analysis of the constraints on human behaviour derived from the physical, functional and purposeful nature of the environment), Control Task Analysis (CTA; an analysis of the constraints that are derived from the actions that have to be performed by the system), Strategies Analysis (SA; an analysis of problem solving strategies of the system), Social Organisation and Cooperation Analysis (SOCA; an analysis of constraints operating on the relationship between actors), and Worker Competencies Analysis (WCA; an analysis of the human cognitive and physical constraints).

CWA may be thought of as a framework in which to identify and group the constraints on human behaviour (as identified in the five analytical phases). It is not a theory of human cognition and is not a behavioural approach (e.g. task analysis). In fact, CWA stands in contrast to task-based approaches. CWA has been used in a number of operational systems contexts over the last 20 years (e.g. [6]). However, it has only been used to determine human performance measures within a microworld environment (e.g. [7]) that is not readily applicable to large scale system evaluation. Of the five phases, WDA and CTA seem most suited to selecting measures for complex systems.

2.2 Comparing HFE work-analysis methods

From a review of the literature it seems that task-based methods are limited in terms of their suitability for analysis of future systems, but they have wide applicability to system evaluation. In contrast, CWA may be potentially of benefit in the analysis of future systems, but is limited in terms of its use for determining human performance measures in operational systems. The challenge is to determine which of the two approaches is more useful in identifying measures of system performance in HIL environments.

The approach we have taken to this problem is to compare an existing task-based HFE method with a CWA-based method developed by the first author. The methods are being compared to each other in terms of the measures that they suggest and in particular whether the measures they suggest are sensitive to system change. This paper only presents the results of applying a CWA-based method to identifying measures for evaluating subsystem upgrades to Black Hawk helicopters. The comparison of the task-based and CWA-based methods will be reported in the future.

3. BLACK HAWK AIR MOBILE OPERATIONS

Australian Black Hawk helicopters and crew typically take part in airmobile operations which involve the transportation of troops and/or equipment into areas where there may be hostile troops, over a variety of terrains and in various weather conditions. An airmobile operation is made up of three distinct phases that may be identified as: mission planning; mission flight; and mission debriefing. Of interest to us the second phase – mission flight.

During mission flight the aircraft is flown along the planned route with the aircrew engaged in many tasks; for example, re-planning as a result of changed orders, engaging hostile troops or dealing with aircraft system failures. During this phase there may be periods of relatively benign flight and periods when the aircraft is engaged by hostile troops (by small arms fire and/or by missiles). Typically, the only indication that an aircraft is being engaged by small arms fire will come from observation by the aircrew. If an aircraft is engaged by a missile then the Electronic Warfare (EW) system will detect and relay that information to the aircrew. Depending on the terrain, weather conditions and other factors (e.g. the aircraft position in the formation) the crew will manoeuvre the aircraft to defeat the threat. Clearly, the actual route taken throughout the mission flight phase is unique if one considers the variation in speed, altitude, heading, pitch, yaw and roll available to the aircraft. In addition, although many aircrew tasks can be predicted, some may be unpredicted – that is, produced as a consequence of unforeseeable events or unique combinations of events.

For example, Figure 1 shows the flight path taken by one crew during a mission (mission scenario 7). The location of the threats is indicated by letters. Significant data points are indicated by numbers. Figure 1 shows that the aircrew progressed along the planned route until data point 4 where they deviated from the planned route in response to a threat[3]. Their new route took them to the left until data point 10 where they aborted the mission because of the likelihood of being shot down by the threats. The decisions to deviate to the left and subsequently to abort the mission *at those specific locations*, *under those specific conditions*, could not have been predicted before the mission was flown. Therefore, any method that aims to identify sensitive measures of performance must take into account a wide range of possible aircrew-system behaviours.

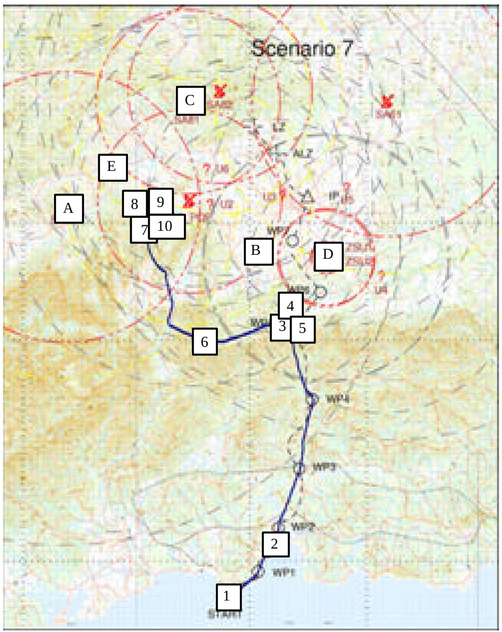

Figure 1: Route flown by aircrew during mission scenario 7

4. A CWA-BASED FRAMEWORK FOR SELECTING SYSTEM PERFORMANCE MEASURES

The CWA-based framework is based on two phases of CWA: WDA and CTA (represented as Temporal Coordination CTA: TC-CTA).

4.1 Work Domain Analysis

WDA presents and indicates the relationship of human values and priorities with software, hardware and environmental components of the work domain and their purpose. The integration of these different components and the ability to reflect on how they function together to achieve system purposes is a unique strength of WDA.

The WDA may be represented in a lattice format (Figure 2) and is made up of a series of nodes that represent different parts of the work domain. Physical parts of the work domain (e.g. the EW sub-system) are represented at the lowest level and functions of the physical parts

[3] It should be noted that the crew had an initial intelligence briefing on the rough location only of the threats - the exact locations were unknown to them.

are represented at the next level. The third level represents functions that characterise the whole domain, e.g. tactical operation. The priorities and values that characterise the domain are shown in the second level. Finally, the purpose of the work domain, in this case “To transport troops and equipment…” is shown at the top level. Measures of performance are derived from finding the most appropriate measure for each node. For example, a measure associated with the node tactical operation could be “The difference between planned time on target (TOT) and actual TOT”. Identifying measures from the WDA ensures coverage of higher-order purposes and low-order physical functions, and ensures that coverage is complete.

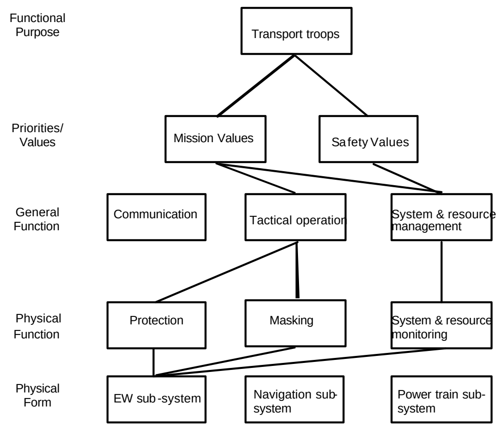

Figure 2: Portion of the Black Hawk Airmobile operation Work Domain Analysis

4.2 Control Task Analysis

CTA presents the limitations on human action as a result of the system control limitations and may be represented as a Temporal Coordination CTA (TC-CTA [8]; see also [9]). Measures associated with TC-CTA indicate how an activity is constrained; for example by time or the action of another activity.

Figure 3 presents an extract of the TC-CTA produced for the Australian Army helicopter. Control tasks can be viewed as a point or interval on a range of time, represented as a double-headed arrow. Each control task has a number of properties associated with it. For example, each has a completion time and an error rate. In addition to these control task-specific properties, there are several properties related to all the control tasks. For example, the properties associated with the control task of “Manage EW system: operate in response to threat” include control task-specific properties (for example, “time task takes” and “start time”) and general properties (for example, task priority and task sequence). Hence, there are many possible interactions within and between control task properties that need to be captured. For example, a change in the “start time” of a control task may cause a change in the sequencing of all the control tasks.

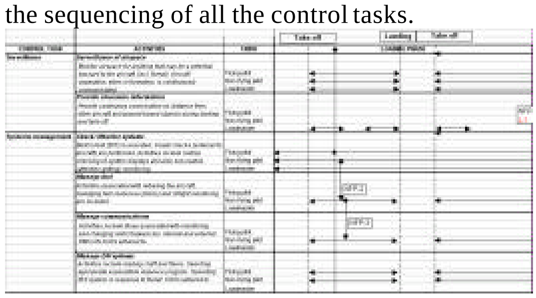

Figure 3: Portion of the Black Hawk Airmobile operation TC-CTA (shown for illustrative purposes only)

4.3 The framework and process for selecting system performance measures

Figure 4 shows the theoretical framework that capture the sources of, and interactions between, system performance measures. On the left side WDA-sourced measures are represented: the human values and priorities with software, hardware and environmental components of the work domain and their purpose are represented. On the right side the TC-CTA-sourced measures are represented.

For example, if an analyst is interested in the relationship between equipment subsystem performance and mission effectiveness, the left side of the framework can be used. The large vertical arrow indicates that a change in the work domain at any level can have an effect on higher or lower levels. However, if the same analyst is interested in the relationship between equipment sub-system performance *and* human behaviour, the right side of the figure can be incorporated. The analyst will move from “sub-system performance measures” to “measures applicable across all activities” and “measures specific to an activity”. The impact of a change can flow within *and* between each half of the framework. Hence, the analyst can cross the framework many times, and at different levels to understand how a change affects the work domain and the activity domain.

Using the framework in Figure 4, a method (not shown) was developed that was used to predict how a change to a part of the system would affect the other parts of the system. The process developed was used to select a number of measures of performance. These measures were then tested empirically for their sensitivity to an EW upgrade in the HILS (see Table 1).

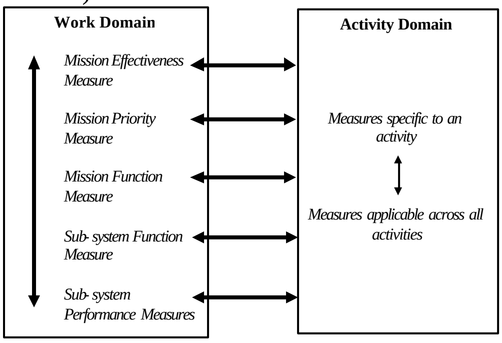

Figure 4: Framework for Human-Equipment performance measures

5. THE EXPERIMENTS

We designed several experiments to compare measures derived with the two different work analysis methods. The first experiment is described here; it was directed at selecting measures for a current sub-system being evaluated for procurement. The sub-system was an EW system mounted as part of Black Hawk helicopter systems fitted for airmobile operations. As part of an EW "upgrade" the range for which a threat could be detected was increased, i.e. the sensitivity level of the EW sub-system was increased. A number of analytical products (e.g. Figure 2 and 3) were developed, based on the EW sub-system, and were used to produce candidate measures that should be sensitive to the upgrade.

The experiments all had several high level requirements:

1. Complex system behaviour should be supported,
2. All airmobile operation stages should be included (mission briefing, flight and debriefing),
3. Aircrews should plan and fly airmobile missions in a tactically realistic environment using the Black Hawk simulator in the AOSC,
4. Aircrew should follow standard operating procedures (SOPs), including flight checks and threat avoidance,
5. Novel human behaviour (but defensible by the aircrew) should be supported,
6. Aircrew should perform their missions under normal operational mission constraints,
7. All human and equipment performance measures, as suggested by the work analysis methods, should be recorded,

To test whether the measures were sensitive to the upgrade, qualified air crew flew a number of missions in the HILS facility in the AOSC either with or without the sub-system upgrade. During the missions, data relating to the measures suggested by the work analysis methods were collected. Following the mission flight, the crews were debriefed and interviewed using a modified critical decision-making methodology (CDM: [10]).

5.1 Experiment results

Table 1 presents some of the results from the experiment. The table lists the measures that were suggested from the WDA and indicates whether those measures were sensitive—specifically, whether the measures produced a significant statistical difference between system upgrade and non-upgrade conditions. The table also maps the measures onto the categories at the left portion of Figure 4. The table shows the results from testing the nine metrics. Two of the measures; "Mission achieved" and "Damage…" were not sensitive to the system upgrade. Under the high sensitivity condition 100% of all missions were achieved, and under the low sensitivity condition 80% of all missions were achieved. Given the total number of mission runs (10) and the training of the aircrew it was unlikely that the difference between experimental conditions on these measures would be large. Of the seven remaining measures, six were sensitive.

6. GENERAL DISCUSSION

This paper has presented some initial results that demonstrate that CWA can be used to select sensitive measures of system performance. The CWA framework is important for several reasons. First, it provides a way to categorise measures into meaningful groups. Second, it provides a mechanism to explore relationships between system components and, third, it is testable.

The results of this first experiment indicate that most WDA-based measures were sensitive to the system upgrade. These limited results suggest that by using the framework and associated method, an analyst can select metrics that are important for test and evaluation purposes. In addition, the framework grounds the selection of measures in a sound theoretical foundation - something that is not achieved in current procedures for selecting measures.

Crone, D., Sanderson, P., Naikar, N. & Parker, S. (2007). Selecting sensitive measures of performance in complex multivariable environments. Proceedings of the 2007 Simulation Technology Conference (SimTecT 2007). Brisbane, Australia, 4-7 June 2007.

Even though this report has been limited to the WDA-based measures, future papers will report on the selection and testing of the activity-based measures and the comparison of the CWA-based metrics to task-based HFE methods. However, even at this stage it seems that applying the CWA framework does helps to answer the question: What measure(s) of performance *should* be used to evaluate complex systems?

7. ACKNOWLEDGEMENTS

The authors acknowledge the following: Australian Army Black Hawk aircrews for participating in the experiments; members of the Air Operations Simulation Centre (AOSC) at DSTO, especially Ms Jodie Doman, for helping to define and implement the experiment requirements; and Richard Fraccaro (formally of StatSoft Pacific Pty Ltd) for advice concerning statistical analysis.

REFERENCES

1. Kirwan, B and Ainsworth, L.K (1992) *A Guide to Task Analysis.* Taylor and Francis. London.
2. Rasmussen, J., Pejtersen, A. M. and Goodstein, L. P. (1994) *Cognitive Systems Engineering*. John Wiley and Sons, Inc. New York
3. Vicente, K. (1999). *Cognitive work analysis: Towards safe, productive, and healthy computer-based work*. LEA. Mahweh NJ.
4. Miller, C. A. and Vicente, K. J (1999) Task "Versus" Work Domain Analysis Techniques: A Comparative Analysis. *Proceedings of the 43rd Annual Meeting Human Factors and Ergonomics Society.* Santa Monica, CA.
5. Crone, D. J., Sanderson, P. M. and Naikar, N. (2003) *Using Cognitive Work Analysis to Develop a Capability for the Evaluation of Future Systems*. Proceedings of the 47th Annual Meeting Human Factors and Ergonomics Society. Denver, CO.
6. Burns, C. M., Bryant, D., & Chalmers, B (2000). A work domain model to support naval command and control. *Proceedings of the 2000 IEEE International Conference on Systems, Man, and Cybernetics*, 2228-2233.
7. Yu, X., Lau, E., Vicente, K. J. & Carter., M. W. (2002). Toward theory-driven, quantitative performance measurement in ergonomics science: the abstraction hierarchy as a framework for data analysis. *Theoretical Issues in Ergonomics Science*, 3 (2),124-142.
8. Sanderson, P., & Naikar, N. (2000). Temporal coordination control task analysis for analysing human system integration. *Proceedings of the Joint Meeting of the Human Factors and Ergonomics Society and the International Ergonomics Association (IEA2000/HFES2000).* Santa Monica, CA: HFES. Vol 1, Pp 206-209.
9. Naikar, N., & Pearce, B. (2003). Analysing activity for future systems. *Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society.* Denver, CO: 13-17 October.
1. Klein, G. A., Calderwood, R. and MacGregor, D. (1989) Critical Decision Method for Eliciting Knowledge. *IEEE Transactions on Systems, Man, and Cybernetics,* Vol 19, 3, May/ June.

	Measures tested (from WDA)	Results: Do measures reflect system performance (different EW sensitivity levels)?
Mission effectiveness measure	Mission achieved (landing at target YES/ **NO)** ?	No
Mission priority measures	Damage – weapon in flight (aircraft damage)	No
	Distance to enemy (distance from current position to enemy position)	Yes
	Detection (chance that you will be detected)	Yes
	Damage – before weapon in flight (enemy radar mode)	Yes
	Time enemy is detecting aircraft (if detected, how long are you detected?)	Yes
Mission function measures	Truthfulness (Accuracy of the system to provide threat location)	No
	Timeliness (Is the information about the threats available at a 'useful' stage of the mission?)	Yes
System function measure	Display of threat information (Number of threats displayed)	Yes

Table 1: Assessing the sensitivity of the WDA-based performance measures