# Measuring Vigilance Abilities to Enhance Combat Identification Performance

Clark Shingledecker, PhD, David E. Weldon, PhD, Kyle Behymer, Benjamin Simpkins, & Elizabeth Lerner
JXT Applications, Inc.

Joel Warm, PhD, Gerald Matthews, PhD, Victor Finomore, & Tyler Shaw
Department of Psychology, University of Cincinnati

Jennifer S. Murphy, PhD
US Army Research Institute

**Abstract**

The warfighter's ability to focus and sustain attentional resources in order to detect critical, but often obscure and infrequent events is essential to Combat Identification (CID) performance. This ability, known as "vigilance," was first studied by psychologists during WWII, and subsequent research has identified a number of task variables, notably signal rate and task duration, that strongly affect detection performance. However, the search for the psychological traits that are possessed by individuals who excel at vigilance tasks has yielded equivocal results.

The objective of the ongoing research reported in this paper is to identify a battery of subjective and objective measures of individual differences in personality, performance on brief vigilance tasks, and attitudes toward a task that can predict performance on realistic combat threat detection tasks. The ultimate goal of this work is to identify the characteristics of the vigilant war fighter, develop standardized measures that can be used to optimally select individuals for duties requiring a high level of detection performance, and provide objective targets for training that may increase the vigilance capabilities of all war fighters.

## INTRODUCTION

Vigilance is the ability to maintain one's focus of attention and remain alert for prolonged periods of time. Tasks that require a high degree of vigilance are an integral part of warfare. In addition to conventional surveillance activities, which base detection on monitoring unprocessed sensory input from the environment, the modern warfighter is likely to engage in computer-generated monitoring tasks associated with the control of missiles, unmanned aerial vehicles, or combat robots, and perform detection tasks in efforts to counter terrorist activity. Future battlefields are expected to continue to demand that warfighters move and engage rapidly under conditions where friendly forces are intermingled with those of the enemy, and survival depends upon an ability to detect and locate hostile forces quickly (Warm, 1993). In many cases, these activities include the need to monitor sensor-generated signals regarding enemy positions and troop movements, as well as movements of friendly forces in a widely dispersed battlefield scenario where vigilance is the price of survival. In addition to military situations, vigilance is a crucial aspect of many civilian activities including air-traffic control, airport baggage inspection, industrial quality control, robotic manufacturing, medical monitoring, and public safety assurance.

Early vigilance research was stimulated by the performance of British airborne radar observers while on patrol for enemy submarines. Despite extensive training and motivation, the observers failed with increasing frequency over the course of a watch to detect critical signals. This resulted in submarines going undetected, free to attack allied ships. In efforts to study this problem, Mackworth (1948) developed a simulated radar display called the "Clock Test." Using this task, he charted individuals' performance over time and confirmed that the quality of sustained attention is fragile: it wanes over time. This progressive decline in performance over time, labeled the "decrement function" or the "vigilance decrement," has been confirmed in subsequent investigations. Investigations using various tasks indicate that the decline in performance is complete from 20 to 35 minutes after the initiation of the vigil and at least half of the final loss is completed within the first 15 minutes. Under especially demanding circumstances, the decrement can appear within the first few minutes of the watch. As Dember and Warm (1979) have noted, the most striking aspect of this finding is that it seems to result from the necessity of looking or listening for a relatively infrequent signal over a continuous period of time. Since Mackworth began his earliest research, much has been learned about vigilance, its importance, and the characteristics of people with high vigilance tendencies. Much of this research is summarized in Matthews, Davies, Westerman, and Stammers (2000).

Because individuals vary significantly in their ability to perform vigilance tasks effectively, it is important to assign individuals selectively to tasks with high requirements for sustained attention. While the problem of personnel selection for exceptional vigilance and sustained-attention abilities has been researched for years, only limited success has been achieved in predicting individuals' vigilance performance capabilities based on traditional selection tests. More reliable and valid approaches to assessing vigilance aptitudes and capabilities are needed to optimize assignment of personnel and improve overall operational effectiveness in current and future combat environments.

The overarching purpose of the research program described in this paper is to identify a set of individual difference variables that predict vigilance performance, to assemble selection measures for these variables, and to validate the measures for use in practical, military personnel-selection settings.

## Identifying Individual Differences that Predict Vigilance

Developing a useful solution to the problem of identifying effective individual-difference predictors of vigilance performance requires an examination of past vigilance research findings, their implications for additional research, and the benefits that can be derived from developing a set of predictive vigilance measures that can be easily applied in a routine test environment.

Personality Factors. Davies and Parasuraman (1982) discuss several personality dimensions related to performance efficiency in vigilance tasks. Included in these are introversion-extraversion, field dependence-independence, internal-external locus of control, and the Type A (coronary-prone) behavior pattern. The findings of these studies indicate that, in general, the performance of introverted observers exceeds that of their extraverted cohorts, field-independent individuals perform better on vigilance tasks than field-dependent observers, individuals with an internal locus of control fare better on a vigilance task than those with an external locus of control, and the vigilance performance of Type-A individuals who are characterized by a rushed, competitive, achievement-oriented life style exceeds that of their more relaxed, Type-B counterparts. Additional research has suggested that boredom-prone individuals may be poorer monitors than those who are less boredom prone (Thackray, Bailey, & Touchstone, 1977). In addition, absent-minded individuals, defined by high scores on the Cognitive Failures Questionnaire, have been found to do more poorly in vigilance tasks than non-absent minded observers and to report higher levels of perceived mental workload than the non-absent minded (e.g., Robertson, et al., 1997). Finally, optimists perform more effectively on vigilance tasks than do pessimists (Helton, Dember, Warm, & Matthews, 1999). Such results indicate that personality profiles should be a part of any approach for developing a vigilance test with reliable predictive features.

Performance Sampling as a Predictor. A second promising source of useful predictors of sustained-attention ability is the objective measurement of an individual's performance on vigilance tasks. Traditional laboratory types of vigilance tasks require a lengthy watch period that would make them impractical as selection tests for large groups of examinees. However, recent research has shown that brief, highly-demanding, brief vigilance tasks can be constructed that produce performance changes mirroring the vigilance decrements typically observed during long-term vigils (e.g., Matthews, Davies & Lees, 1990; Temple et al., 2000). These tasks show very rapid perceptual sensitivity decrements over a period of 10 minutes or less, and they demonstrate the key diagnostic indicators of being resource-limited: sensitivity decrement, high subjective workload, and sensitivity to stress and arousal factors. Thus, a high level of performance on a short task may be a good indicator of aptitude for longer vigilance tasks.

Differences in Subjective Responses to Vigilance Task Demands. Finally, recent studies indicate that the perceived workload of vigilance tasks is quite substantial and that workload grows linearly over time (Warm, Dember, & Hancock 1996). Leading researchers (e.g., Johnson and Proctor, 2003) conclude that, rather than being under-stimulating, vigilance tasks place high information-processing demands upon observers. Thus, Resource Theories appear to take precedence over Arousal Theory as a model of the factors that control vigilance performance. However, following Kahneman's original model (Kahneman, 1973), Matthews and Davies (2001) have argued that Arousal and Resource Theories are not necessarily mutually exclusive and that they can be integrated by viewing arousal as the agent responsible for resource production. In support of this view, they have employed a self-report measure of arousal known as energetic arousal (alertness-sluggishness) that correlates with psychophysiological measures of autonomic activity under performance testing conditions. Using this measure, they demonstrated that when confronted with psychophysical challenges designed to increase information-processing load, observers in vigilance tasks who were high on energetic arousal fared better than those who were low on this measure. That is, energetic arousal may provide a marker for resource availability. The finding of an agreement between psychophysiological measures, subjective self-reports, and performance, as predicted by the integrated models, is of significance for selection-test development.

In addition to workload response differences, operators may also differ in the ways that they deploy compensatory effort and coping strategies to adapt to demanding performance environments (Hancock & Warm, 1989). Short tasks are sometimes insensitive to stressor effects, but as time progresses, it becomes increasingly more difficult for the operator to maintain successful coping. Thus, it may be possible to identify useful predictor measures from an operator's reactions to performing a short vigilance task that may offer early warning signs of difficulties in coping. Such difficulties do not necessarily impact performance immediately, but may have adverse effects on a longer, operational task. There are several indices that may be acquired following performance in order to assess whether the operator is being unduly taxed by task demands. Although workload indices are typically used to discriminate *tasks*, they may also be used to assess the level of demands perceived by the *person*. Measures of subjective stress state may also indicate the degree of challenge afforded by task performance. As previously mentioned, subjective energetic arousal relates positively to vigilance. Finally, the operator's style of coping with the pressure of the task may facilitate or prevent effective resource deployment: task-focused coping is preferable to emotion-focused or avoidant coping (Matthews & Campbell, 1998).

A Model for Development of a Vigilance Selection Test Battery

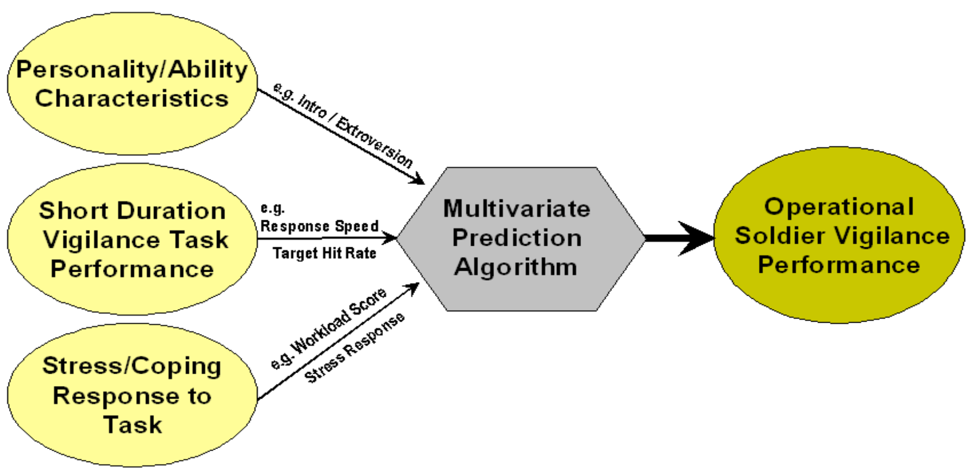



**Figure 1. VIGILANT WARRIOR: An Approach to Developing a Personnel-Selection Tool for Sustained-Attention Ability**

The challenge presented by the current research problem was to apply the concepts of vigilance and its measurement discussed above to develop a reliable and valid vigilance-prediction toolset. To approach this challenge, a multi-dimensional solution to vigilance prediction was conceived that would sample key constructs related to (1) personality and ability, (2) objective task performance, and (3) stress/workload responses to vigilance tasks. A primary goal was to extract the optimal measurement instruments from these complementary approaches and blend them to produce an efficient, low-cost personnel-selection system capable of predicting vigilance performance. A graphic representation of the "Vigilant Warrior" personnel-selection battery concept is shown in **Figure 1**.

## PRELIMINARY RESEARCH

To identify preliminary components for each of the three vigilance-prediction dimensions we evaluated the literature addressing the relationship between various personality and ability variables and vigilance performance. Limitations and strengths of both traditional and more recently explored vigilance predictors were documented, and expert ratings reflecting the degree of research support and projected utility for each dimension were independently assigned by members of the research team. In addition, we examined available brief vigilance tasks that could be included in the battery, as well as subjective rating dimensions and scales that could be used to determine an examinee's perceived workload and coping responses, and attitudes associated with performing the vigilance task. Based on the results of these two analyses, we identified a candidate vigilance-prediction battery composed of personality/ability metrics, brief vigilance task performance metrics, and resource depletion and allocation metrics.

The top five rated personality/ability dimensions that have been shown to have some established relationship to vigilance performance were Extraversion – Introversion, Intelligence Quotient, Boredom Proneness, Cognitive Failures, and Conscientiousness. We added the dimensions of Trait Sleepiness, ADHD, Schizotypy, and Propensity to Daydream to this group because recent studies suggest they hold promise for predicting vigilance.

Two versions of a short vigilance task (SVT) were selected for the battery in order to account for the well known differences in performance and sensitivities to stimulus and environmental variables observed in tasks with (simultaneous) and without (successive) a comparison stimulus being available for classifying an event as a signal or a non-signal. The task is a paired-symbol brief vigilance task. Events are presentations of letter pairs in any combination drawn from the letters D, O, and backward D (e.g., “DO,” “OD,” “ O,” etc.). In the simultaneous trials, the signal is any matching pair (e.g., “DD”). In the successive version, the signal is defined as the occurrence of the pair “OO.”

The Dundee Stress State Questionnaire (DSSQ), the Coping Inventory for Task Situations (CITS), the NASA TLX workload scale, and the Boles Multiple Resource Questionnaire were selected to assess subject attitudes toward, and responses to, performing the brief vigilance task. Dimensions assessed by these instruments are Task Engagement, Distress, Worry, Coping (task focused), Coping (avoidance), Coping (emotion focused), Workload, and Multiple Resource Usage.

We conducted the main preliminary investigation in advance of the full-scale battery validation to validate the short vigilance tests, assess the psychometric properties of the candidate personality, intelligence, and stress/attitude/coping measures, and assess their differential ability to predict vigilance performance on the tasks to be used for the brief vigilance task performance component of the battery. The study was conducted using a sample of 210 participants recruited from psychology classes at the University of Cincinnati.

## Method

Participants completed a series of questionnaires and performance-based assessments according to the following sequence: p*ersonality tests; intelligence tests*; *pre-task stress state*, *12-minute* short vigilance task; and p*ost-task stress state and coping*. During the vigilance task, pairs of characters were presented against a masking background at a high event rate. One hundred five participants performed a simultaneous version of the task requiring a comparative judgment to detect the target. An equal number of participants performed a successive version of the task requiring an absolute judgment to detect the target.

## Results

<u>Validity of the Short Vigilance Task.</u> One objective of this study was to ensure that the brief vigilance task developed for the battery would show classic performance changes over time that are

characteristic of typical longer tasks. **Figure 2** shows the average number of correct detections made by subjects performing the successive and simultaneous versions of the test over the six continuous 2-minute watch periods. As **Figure 2** suggests, the short tasks yielded the expected difference between the task conditions ($F_{(1, 208)} = 19.80$, $p < .001$) and a common decrement in performance over the 12-minute watch ( $F_{(3.91, 813.42)} = 44.74$, $p<.001$.).

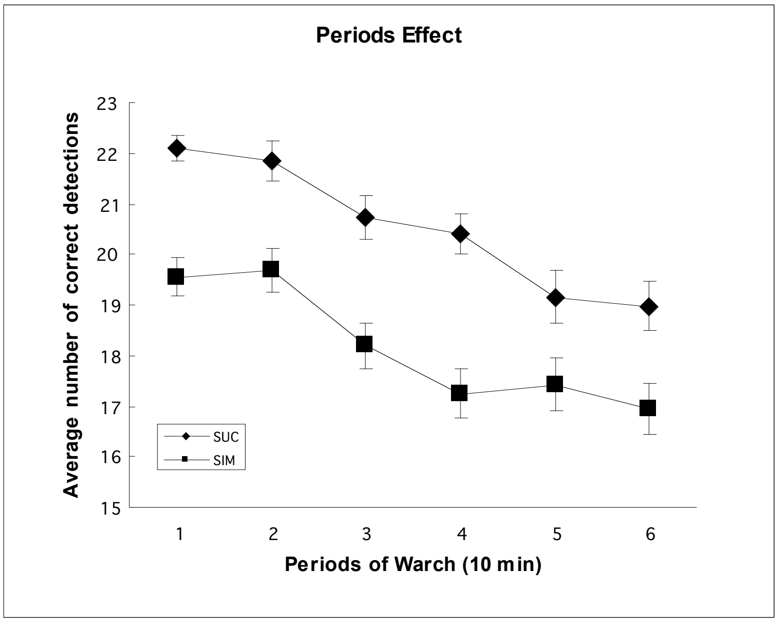



**Figure 2. Mean number of correct detections as a function of periods of watch for both simultaneous and successive conditions (Error bars are standard errors)**

Factor analysis of the Subjective Scales. A factor analysis was conducted to test whether the initial set of questionnaire scales could be reduced to a smaller number of underlying factors. Analysis of the personality scales showed that these individual difference indicators were intercorrelated. A principal factor analysis was run, followed by an oblique (direct oblimin) rotation. On the basis of the scree test and factor interpretability, a four-factor solution was extracted, explaining 63.7% of the variance. Factor loadings are shown in **Table 1**. Factor 1 (labeled Cognitive Disorganization) is defined by various scales linked to disruption of attentional focus, including cognitive failures, mind-wandering and day-dreaming, as well as the O-Life disorganization scale. Factor 2 (Heightened Experience) is defined by O-Life unusual experiences, sensation-seeking, and low internal boredom (i.e., enjoyment of events). It appears to indicate a vivid, excitable mental life. Factor 3 (Sleep Quality) brings together the 3 subscales of the

Pittsburgh Sleep Quality Index used in the study. Surprisingly, the other sleepiness index (Epworth) fails to load on this factor. Factor 4 (Impulsivity) contrasts the two impulsivity scales with low premeditation on the UPPS scale (Urgency, [lack of] Premeditation, [lack of] Perseverance, and Sensation seeking). The factors were intercorrelated, with the highest correlations found between factors 1 and 4 ($r$= .51) and between 1 and 3 ($r$ = .44). Factor 2 was largely uncorrelated with the remaining factors.

| | **Factor** | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| Cognitive Failures | **.858** | -.091 | .123 | -.055 |
| O-LIFE Cog. Disorganization | **.839** | -.053 | .128 | .029 |
| Mind Wandering | **.743** | -.013 | -.046 | .103 |
| Daydreaming | **.552** | .386 | -.070 | -.050 |
| Fatigue | **.543** | -.144 | .266 | .077 |
| ADHD | **.495** | .076 | .152 | .326 |
| Boredom – External | **.463** | .102 | .153 | .096 |
| Epworth Sleepiness Scale | **.440** | -.154 | -.059 | .117 |
| Boredom – Internal | .205 | **-.540** | .075 | .119 |
| O-LIFE Unusual Experiences | **.404** | **.538** | .091 | .012 |
| UPPS Sensation-seeking | -.024 | **.485** | .106 | .121 |
| Pittsburgh Sleep Quality (2) | -.068 | -.001 | **.804** | -.051 |
| Pittsburgh Sleep Quality (1) | -.017 | .044 | **.658** | .031 |
| Pittsburgh Sleep Quality (5) | .116 | -.030 | **.620** | .003 |
| I7 Impulsivity | .112 | .179 | -.029 | **.845** |
| UPPS Premeditation | .071 | .123 | .027 | **-.696** |
| O-LIFE Impulsivity | .157 | .184 | .162 | **.494** |

**Table 1. Factor Analysis of Personality Traits: Factor Pattern Matrix (N=210)**

Prediction of Performance Satisfied that the SVT possessed the fundamental characteristics of a more classical extended-duration task, we used it as a surrogate to conduct a preliminary exploration of the predictive capabilities of the subjective measures. Pearson correlations were computed to examine the extent to which the personality, intelligence, and stress and coping measures could predict performance on the SVT. Personality was represented by regression-model factor scores computed on the basis of the factor analysis. Detection frequencies within each 2-minute period were highly intercorrelated (alpha = .93), so average target detection frequency was used as the performance measure for this analysis. **Table 2** provides a summary of the correlations of the various scales with performance, for simultaneous and successive conditions. There were 105 participants in each condition. It can be seen in this table that the two measures of intelligence positively correlate with performance on the SVT. The advanced vocabulary test is a better predictor of performance on the successive task, while the letter-sets test correlates with both the simultaneous and successive tasks. **Table 2** also indicates that the four personality factors correlate poorly with performance on the SVT.

| [illegible] | [illegible] | [illegible] | [illegible] |
|---|---|---|---|
| [illegible] | [illegible] | [illegible] | [illegible] |
| | [illegible] | [illegible] | [illegible] |
| [illegible] | [illegible] | [illegible] | [illegible] |
| | [illegible] | [illegible] | [illegible] |
| | [illegible] | [illegible] | [illegible] |
| | [illegible] | [illegible] | [illegible] |
| [illegible] | [illegible] | [illegible] | [illegible] |
| | [illegible] | [illegible] | [illegible] |
| | [illegible] | [illegible] | [illegible] |
| [illegible] | [illegible] | [illegible] | [illegible] |
| | [illegible] | [illegible] | [illegible] |
| | [illegible] | [illegible] | [illegible] |
| [illegible] | [illegible] | [illegible] | [illegible] |
| | [illegible] | [illegible] | [illegible] |
| | [illegible] | [illegible] | [illegible] |

[illegible]
[illegible]

**Table 2. Correlations of Intelligence and Stress Variables with Performance**

The best predictors of performance appear to be the subjective stress states and coping style measures. Most notably, pre- and post- task engagement scores correlate highly ($p<.001$) with performance on the simultaneous task, while post-task engagement correlates highly with performance on the successive task. **Table 2** also suggests that, while simultaneous and successive tasks have some common predictors, the optimal set of predictors for each type of task may differ somewhat.

Conclusions

This preliminary study confirmed that the modified versions of the original SVT developed by Temple, *et al*. (2000) show the vigilance decrement characteristic of performance of longer monitoring tasks. Thus, these tasks appear to qualify as the performance-sampling component of the predictive battery. The data also confirm previous findings suggesting that personality traits are, at best, no more than modest predictors of vigilance. However, additional preliminary research has shown some personality factors have the capacity to predict stress states and coping during vigilance, which may contribute to their utility in the context of performance of a longer, sustained monitoring task. While previous vigilance studies report equivocal findings regarding the intelligence – vigilance association, the present data support inclusion of short intelligence tests in the predictive battery. Finally, consistent with previous findings, both stress states assessed using the DSSQ and coping scales correlated with performance, supporting inclusion of these measures in the vigilance battery. It was found that post-task state measures were correlated more strongly with performance than pre-task measures. This finding may be a consequence of the post-task measures being more representative of states experienced during the task itself. Indeed, the stress response to the SVT resembles that seen with longer-duration tasks. If so, post-task measures taken from the initial SVT may be predictive of a subsequent, longer criterion task. This hypothesis is being tested in the validation study for the predictive battery.

The analyses from the preliminary study allowed us to reduce both the number of tests and the number of items on some tests. **Table 3** shows the tests selected for the validation study and the number of items selected from each test. These tests are being administered along with the SVT and a specially constructed criterion measure in the validation study described below. The DSSQ subscales are administered both as a pre-test and post-test in the validation study.

| **Test Name** | **No. of Items** |
|---|---|
| O-LIFE Unusual Experiences subscale | 8 |
| Impulsiveness scale | 17 |
| UPPS Sensation Seeking subscale | All |
| Cognitive Failures Questionnaire Distractibility subscale | All |
| Pittsburgh Sleep Quality Index | All |
| ETS Advanced Vocabulary test (Crystallized Intelligence) | All |
| ETS Letter Sets test (Fluid Intelligence) | All |
| DSSQ Mood State | 29 |
| DSSQ Motivation State Subscale | 15 |
| DSSQ Thinking Style Subscale | 30 |
| DSSQ Thinking Content Subscale | 16 |
| DSSQ General State of Mind Subscale | 30 |
| NASA-TLX Workload Scales | 6 |
| Opinion of Task questionnaire | 20 |
| Task Coping (Dealing With Problems) questionnaire | 21 |

**Table 3. List of Tests and Items Retained in the Validation Study**

## CRITERION TASKS

Two concurrent research efforts are being pursued to validate the ability of the Vigilant Warrior battery to predict vigilance performance using different criterion tasks. In the field validation effort we are using an actual Improvised Explosive Device (IED) detection performance test normally used for training and operational evaluation of combat threat-detection skills in an outdoor setting. We are collecting data from Soldiers and Marines as they engage IED-detection problems embedded in several performance lanes implemented by the US Army and Marine Corps. Solvig (2006), for example, provides an unclassified description of these performance lanes, which is available on the Internet (see References section.

The more extensive primary validation research is being conducted in a controlled laboratory simulation environment. The criterion task developed for this work simulates a real-world, electronic-display-monitoring environment with high vigilance requirements similar to those to which Army operators might be exposed in operational settings. This battlefield situation-monitoring task presents a tactical situation display on a computer monitor, which portrays a two-dimensional plan-view map of a geographical area within which the positions of military combat vehicles are represented. Static components of the display include terrain features and reference grid lines. The dynamic components of the display are moving combat vehicles, the positions of which change with each display update. The symbolic combat vehicles appear in three columns that move from left to right across the screen and return. The movement of all three columns is patterned. **Figure 3** shows an example of the symbols and background imagery. The center column of combat vehicles is led by a combat tank with two gun barrels. The display is updated every second with the gun barrels displayed for 50 msec. Participants are asked to respond whenever the gun barrels are of different lengths (simultaneous condition) or are both longer than the standard length (successive condition). In **Figure 3**, the left panel shows the lead tank in the middle column with its gun barrels of the standard length (the non-signal event). The right panel in **Figure 3** shows a “simultaneous condition” signal event (unequal-length gun barrels). “Successive condition" signal events occur when the gun barrels are both as long as the long gun barrel in the right panel of **Figure 3**.

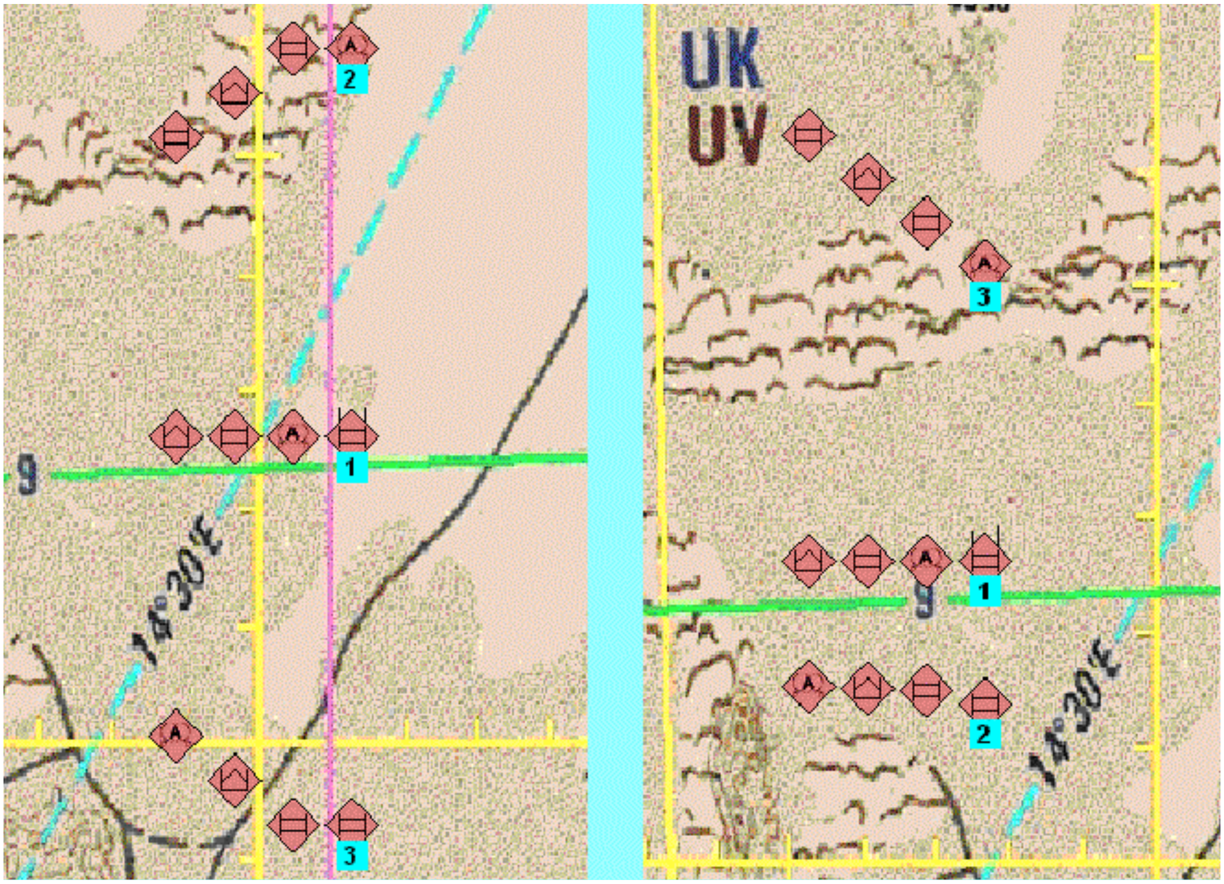



**Figure 3. Partial Screen Displays From the Battlefield Situation-Monitoring Task Showing a Non-Signal Event (Left Panel) and a Simultaneous Condition Signal Event (Right Panel)**

Two versions of the battlefield situation-monitoring task were created to examine the battery's capacity to predict performance under special task conditions and concurrent demands of Army vigilance tasks. The cued version (CUE) captures Army vigilance tasks that are augmented by probabilistic information about potential signal sources. At the heart of the Army's vision for the future soldier is a vast communications and sensor network that will focus on providing both commanders and individual soldiers with near real-time tactical and strategic information to enhance performance and situation awareness. On initial consideration, such a technologically sophisticated information system would appear to resolve fully the classic vigilance problem for the future soldier. Infrequent, perceptually vague signals requiring rapid and accurate responses should be largely eliminated by using combined sensor and intelligence data to predict such events and display them in a salient manner, which should obviate any decrement associated with traditional watch-keeping behavior. However, closer scrutiny of this new perceptual environment reveals that, rather than eliminating the vigilance problem, the augmented perception offered by advanced systems may simply shift the problem to a more cognitive domain. The unavoidable problem of enhancing the soldier's information resources is that this information will always be imperfect. Brief (but operationally significant) delays in information arrival, distorted intelligence, data pre-processing that is not always transparent to the receiver, and spoofing of sensors are all probable flaws in any such system that may complicate detection of significant events. Thus, under some potential circumstances, the soldier's vigilance abilities may be compromised by inaccurate information intended to augment detection. The preceding analysis suggests that the vigilant future soldier will be one who can make effective use of the augmented information to detect significant events, but also maintain an ability to detect rare events that are not spatially or temporally cued by the system. In the cued version of the battlefield situation-monitoring task, display updates include predictors of future signals. These cues are accurate predictions for a majority of signal occurrences. Dependent measures focus on cued vs. uncued signals, and analyses examine the predictive capabilities of the battery for both types of event.

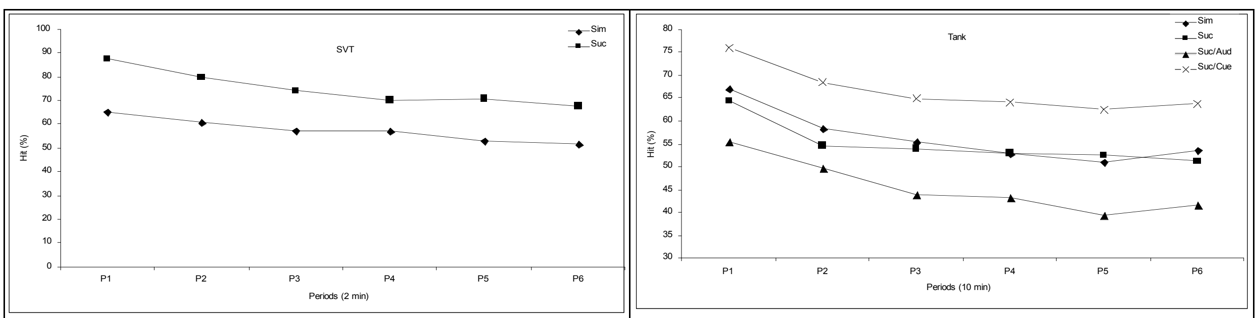
The second version of the battlefield situation-monitoring task (Auditory Secondary Task – AST) represents the common vigilance condition in which the soldier is engaged in tasks in addition to the one of monitoring a visual display for infrequent signals. This condition requires the subject to respond to a secondary auditory display source in order to answer queries about the location of specific vehicles on the map—a task akin to responding to radio messages from collaborators. This additional task increases the mental resource demands imposed upon the subject and permits testing the ability of the battery to predict vigilance performance under multitasking conditions.

**VALIDATION STUDIES**

Two validation studies using the Vigilant Warrior battery are underway. We are conducting the first at the University of Cincinnati under the direction of Professors Joel Warm and Gerald Matthews. The second validation study is being conducted at three military installations located in the United States under the supervision of Dr. Jennifer Murphy of the Army Research Institute (ARI). We describe the two validation studies separately below.

Laboratory Validation at University of Cincinnati

This study is examining the efficacy of the personality variables and the SVT in predicting performance on the simulated battlefield situation-monitoring task described above. Four hundred participants were assigned to one of four conditions as follows: (1) Simultaneous condition, (2) Successive condition, (3) Successive condition with AST, and (4) Successive condition with CUE. The signal rate was 5% in all conditions. Following a 2-minute practice version of the criterion task, which provided auditory feedback on hits, misses, and false alarms, six 10-minute consecutive trial blocks of the criterion task were presented, and presentation order was fully randomized within each condition. We completed data collection in March 2008, and analysis of the data is well underway. The initial analyses show a strong "Vigilance Decrement" in both the SVT and the simulated battlefield situation-monitoring task (see **Figure 4**). Panel 1 of **Figure 4** shows the cell means for each SVT condition (Simultaneous and Successive) and each 2-minute trial block. Panel 2 shows the cell means for four conditions (Simultaneous, Successive, Successive with AST, and Successive with CUE) and each 10-minute trial block. In both tasks, percent hits showed main effects for condition (SVT: $F_{(1,398)} = 4980.81$, $p<.001$; Criterion Task: $F_{(3,396)} = 2559.82$, $p<.001$) and trial blocks (SVT: $F_{(5,1990)} = 68.63$, $p<.001$; Criterion Task: $F_{(5,1980)} = 65.87$).




**Panel 1: SVT 2-Min. Trial Blocks (SIM & SUC conditions)** | **Panel 2: Criterion Task 10-Min. Trial Blocks (four conditions)**

**Figure 4. Plot of Cell Means for SVT and Laboratory-Focused Criterion Task in Validation Study**

The condition x trial blocks interaction for percent hits was not reliable for the Criterion Task ($F_{(15,1980)} = 0.88$, n.r.), but was reliable for the SVT ($F_{(5,1990)} = 3.80$, $p<.01$). The latter effect was due to the larger difference in the means between Simultaneous and Successive in Period 1 (see Panel 1 of **Figure 4**).

In a preliminary test of the efficacy of the SVT in predicting the simulated battlefield situation-monitoring task, we created two derived measures from each. The first was provisionally called "Vigilance Skill," and consisted of the average of the hits across all six trial blocks. High scores on this measure indicate consistent success on each task. The second derived measure was provisionally called "Vigilance Decrement," and was constructed by subtracting the score on Trial Block 6 from Trial Block 1. High scores on this measure indicate a larger Vigilance Decrement over time. We then tested the hypothesis that high SVT Vigilance Skill would predict simulated battlefield situation-monitoring task Vigilance Skill using the participants from the Simultaneous and Successive conditions in both tasks. For the Vigilance Skill derived measure, the overall correlation was 0.52 ($R^2$ = .27, $p<.01$). For this derived variable, participants who did well on the SVT were also moderately likely to do well on the simulated battlefield situation-monitoring task.

A similar test using the Vigilance Decrement measure failed to show the same pattern. In this case, there was no relationship (R = -0.01) between SVT Vigilance Decrement and simulated battlefield situation-monitoring task Vigilance Decrement. We are currently performing additional analyses to determine why. We are also beginning the analysis of the personality variables and their relationship to Criterion Task performance.

These results provide the first concrete evidence to indicate that we have developed a vigilance task with two attractive aspects: (a) the simulated battlefield situation-monitoring task has strong face validity and (b) the findings with this criterion task duplicate those obtained with more abstract laboratory versions. The Successive Condition manipulation is especially interesting because of the findings for the AST and CUE conditions. Our results indicate that even unreliable cueing improved performance while the AST reduced performance as expected by increasing the multitasking workload.

Field Validation at Selected Military Installations

We conducted the first pilot study during March 2008. Using an abbreviated version of the Vigilant Warrior battery, we collected data from 47 US Army personnel and obtained IED "detection lane" performance data on each participant. We are in the initial stages of data analysis at this time. One very preliminary finding is that the SVT component of the battery is sensitive both to the ambient light in the testing room and to the level of brightness and contrast settings of the laptop computers used to present the task. Quick adjustments to the laptop computers' brightness and contrast settings alleviated part of the problem, but the ambient lighting (very high luminance) of the field setting made the SVT more difficult than the same task conducted in our university laboratory. Nevertheless, we were able to get stable data from three days of testing.

Our preliminary analyses show the following:

- Mean hits were lower and false alarms were higher in all conditions than those obtained in the laboratory. We attributed this to the luminance effects described above.
- Mean hits were higher in the Successive condition compared to the Simultaneous condition. This replicated the laboratory findings for the SVT.
- Mean hits in each of the six 2-minute trial blocks showed an increase over the first four trial blocks, followed by a decrease in the final two trial blocks. This pattern suggested the participants in this first field pilot study were still adjusting to the task during the first few trial blocks in spite of having a 2-minute practice block with feedback prior to performing the SVT. This differs from the university laboratory where similar adjustment completes by the end of the first 2-minute trial block of the actual task. The reason for this difference is unclear and must await the analysis of the participants' responses to the post-task workload, task, and coping questionnaires.

**FUTURE DIRECTIONS**

We continue to analyze the data from the laboratory validation at the University Of Cincinnati using the Simulated Battlefield Situation Display Monitoring Criterion Task. Our next step is to determine if the measured personality, task, workload, and coping variables contribute reliably to predicting the criterion task. If so, there is reason to expect that we can provide a reliable and powerful selection tool to the military and to civilian activities, including air-traffic control, airport baggage inspection, industrial quality control, robotic manufacturing, medical monitoring, and public-safety assurance. Because individuals vary significantly in their ability to perform vigilance tasks effectively, the availability of a personnel-selection tool that reliably assigns individuals with exceptional vigilance abilities to tasks requiring sustained attention represents a significant advance in overall operational human effectiveness in current and future combat environments, as well as civilian environments requiring sustained attention. Furthermore, the development of the Simulated Battlefield Situation Display Monitoring Criterion Task, which duplicates all the characteristics of those obtained with more abstract laboratory versions like the SVT, suggests we can build a selection battery with strong face validity. If it is found that a shortened 12-minute version of the Simulated Battlefield Situation Display Monitoring Criterion Task exhibits the same characteristics, then such a task may easily substitute for the SVT, which is hypersensitive to physical display conditions, including ambient light and display brightness and contrast. If this is the case, we will have developed a selection battery that will be well received by both test administrators and personnel being assessed.

**REFERENCES**

Davies, D.R., & Parasuraman, R. (1982). The psychology of vigilance. London: Academic Press.

Dember, W.N. & Warm, J.S. (1979). *Psychology of Perception, 2nd Edition*, New York: Holt, Rinehart, & Winston.

Eysenck, H.J. & Eysenck, M. W. (1985). *Personality and individual differences: A natural science approach.* New York: Plenum.

Hancock, P.A, & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors*, *31*, 519-537.

Helton, W.S., Dember, W.N., Warm, J.S., & Matthews, G. (1999). Optimism, pessimism, and false failure feedback: Effects on vigilance. *Current Psychology, 18*, 311-325

Johnson, A., & Proctor, R.W. (2003). *Attention: Theory and practice.* Thousand Oaks, CA: Sage.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Mackworth, N.M. (1948). The breakdown of vigilance during prolonged search. *Quarterly Journal of Experimental Psychology*, **1**, 6-21.

Matthews, G., & Campbell, S.E. (1998). Task-induced stress and individual differences in coping. In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, pp. 821-825. Santa Monica, CA: Human Factors and Ergonomics Society.

Matthews, G., & Davies, D.R. (2001). Individual differences in energetic arousal and sustained attention: A dual-task study. *Personality and Individual Differences, 31,* 575-589.

Matthews, G., Davies, D.R., & Lees, J.L. (1990) Arousal, extraversion, and individual differences in resource availability. *Journal of Personality and Social Psychology, 59*, 150-168.

Matthews, G., Davies, D.R., Westerman, S.J., & Stammers, R.B. (2000) *Human performance*: *Cognition, stress and individual differences*. London: Psychology Press.

Robertson, I.H., Manly, T., Andrade, J. Baddeley, B.T., & Yiend, J. (1997).Oops! Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia, 35,* 747-758.

Solvig, E. (4/22/2006). Marine base simulates Iraq conditions. *USA Today (via The (Palm Springs, Calif.) Desert Sun*. see http://www.usatoday.com/news/nation/2006-04-22-marinetraining_x.htm.

Temple, J. G., Warm, J. S., Dember, W. N., Jones, K. S., LaGrange, C. M., & Matthews, G. (1997). The effects of caffeine and signal salience on performance, workload, and stress in an abbreviated vigilance task. *Human Factors, 42*, 183-194.

Warm, J.S. (1993). Vigilance and target detection. In B.M. Huey & C.D. Wickens (Eds.), *Workload transition: Implications for individual and team performance* (pp. 139-170). Washington, DC: National Academy Press.

Warm, J.S., Dember, W.N., & Hancock, P.A. (1996). Vigilance and workload in automated systems. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 183-200). Mahwah, NJ: Erlbaum.