*Medicine Meets Virtual Reality 18*
*J.D. Westwood et al. (Eds.)*
*IOS Press, 2011*

*© 2011 The authors. All rights reserved.*
*doi:10.3233/978-1-60750-706-2-433*

# Virtual Reality Stroop Task for Neurocognitive Assessment

Thomas D. PARSONS,[a,1] Christopher G. COURTNEY,[a] Brian ARIZMENDI,[a] and Michael DAWSON[a2]
[a] *University of Southern California, Los Angeles, CA*

**Abstract.** Given the prevalence of traumatic brain injury (TBI), and the fact that many mild TBIs have no external marker of injury, there is a pressing need for innovative assessment technology. The demand for assessment that goes beyond traditional paper-and-pencil testing has resulted in the use of automated cognitive testing for increased precision and efficiency; and the use of virtual environment technology for enhanced ecological validity and increased function-based assessment. To address these issues, a Virtual Reality Stroop Task (VRST) that involves the subject being immersed in a virtual Humvee as Stroop stimuli appear on the windshield was developed. This study is an initial validation of the VRST as an assessment of neurocognitive functioning. When compared to the paper-and-pencil, as well as Automated Neuropsychological Assessment Metrics versions of the Stroop, the VRST appears to have enhanced capacity for providing an indication of a participant's reaction time and ability to inhibit a prepotent response while immersed in a military relevant simulation that presents psychophysiologically arousing high and low threat stimuli.

**Keywords.** Neuropsychological assessment, psychophysiology, ecological validity, virtual environment

## Introduction

The assessment of traumatic brain injury (TBI) has become a difficult challenge for the DoD medical health system. The reports are sobering: 12-20% of Service Members report symptoms of TBI in theater or during re-deployment [1-2]; 47% of all blast injuries in war zones affect the head [3]; two out of five injuries during OIF II were head, face, or neck injuries [4]; and blast injuries often produce symptoms similar to classical TBI, thereby complicating detection, diagnosis, and treatment [5]. Additionally, mild TBIs (mTBIs) often go undiagnosed when other life-threatening wounds occurred [6]. While many mTBIs resolve in a matter of days or weeks, some cases develop post-concussive syndrome, which includes a number of persistent behavioral, cognitive, and psychological symptoms. Such symptoms vary greatly in both severity and onset [7]. Given the prevalence of TBIs, and the fact that many mTBIs have no external marker of injury, there is a pressing need for innovative technology for initial assessment, treatment, and rehabilitation. The demand for TBI

1 Corresponding Author: Thomas D. Parsons, Ph.D., Director of NeuroSim Lab, University of Southern California, Institute for Creative Technologies, Los Angeles, CA. E-mail: tparsons@ict.usc.edu.
2 This research is partially supported by the U.S. Army Research Laboratory, Human Research & Engineering Directorate, Translational Neuroscience Branch (Aberdeen Proving Ground, MD).

assessment that goes beyond traditional paper-and-pencil testing has resulted in the use of automated neurocognitive testing for increased precision and efficiency; as well as the use of virtual environment technology for enhanced ecological validity and increased function-based assessment. Computerized testing batteries, such as the Automated Neuropsychological Assessment Metrics (ANAM; [8]), provide increased accuracy and efficiency (over traditional paper-and-pencil versions) for repeated administration [9]. ANAM has been used for TBI assessment with civilians, athletes, and military personnel [10].

While the ANAM has been found to have adequate predictive value, it does not replicate the diverse military environment in which Soldiers function. A study examining active duty military personnel with mTBI found that 46 percent experienced occupational impairment four to eight months after the injury [11]. Because such injuries can greatly interfere with the ability to perform complex cognitive and emotional processing tasks involved in optimal work performance, there is also a need for assessment technology with greater ecological validity. At varying levels of threat, Soldiers must be able to exercise control of executive functions including the ability to direct and maintain attention, organize incoming stimuli, reason about abstractions, problem-solve, self regulate, and coordinate psychomotor performance. For a neurocognitive measure to be relevant to the assessment of Soldier cognitive functioning, it should go beyond paper-and-pencil measures and provide some indication of a Soldier's reaction time as well as the tendency to perseverate in a response despite external feedback within high and low threat settings. While the military has historically evaluated such cognitive abilities and related predictions of performance decrements through observation or as net outcomes (e.g., task or mission completion), much less focus has been applied to assessment of environmental and occupational challenges. Although the paper-and-pencil and ANAM Stroop are both validated neurocognitive assessments, there is a need for militarily-relevant tests.

To address these issues, the Virtual Reality Stroop Task (VRST) was developed. Like the paper-and-pencil Stroop, the VRST assesses simple attention, gross reading speed, divided attentional abilities, and executive functioning. Like the ANAM, the VRST automates the paper-and-pencil Stroop task and allows for assessment of reaction time. The VRST goes beyond the ANAM and paper-and-pencil versions of the Stroop by replicating the diverse military environment in which Soldiers function.

## 1. Methods

This study was designed as an initial validation of the VRST. A further goal of this study was to utilize psychophysiological measures to predict levels of threat and cognitive workload. In addition, we compared paper-and-pencil Stroop, ANAM Stroop, and VRST on behavioral measures such as reaction time (the time from stimulus onset to the first button press), response time (the time it took for the correct response to be made), and number of correct responses in each zone. Note, the paper-and-pencil version does not allow for assessment of reaction time of each individual response to a stimulus, so the analyses reflect percentage correct. The three main questions were: 1) Can threat level in the VRST be predicted using behavioral and physiological data? 2) Can Stroop difficulty in the VRST be predicted using behavioral and physiological data? 3) How do paper-and-pencil, ANAM, and VR versions of the Stroop behavioral results compare?

### 1.1. Participants and Procedure

The University of Southern California's Institutional Review Board approved the study. A total of 20 college-aged subjects participated in the study. After informed consent was obtained, basic demographic information was recorded. Presentation of the ANAM and VRST versions of the Stroop were counterbalanced. While experiencing the VRST, participant psychophysiological responses were recorded using the Biopac system. Electrocardiographic activity (ECG), and Electrodermal activity (EDA), were recorded simultaneously using a Biopac MP150 system and a computer running Acknowledge software. EDA was measured with the use of 8 mm silver-silver chloride electrodes placed on the volar surface of the distal phalanges of the index and middle fingers of the non-dominant hand. Electrodes were filled with a 0.05 molar isotonic NaCl paste to provide a continuous connection between the electrodes and the skin. Skin conductance responses were scored as the largest amplitude response beginning in a window of 1 to 3 s following stimulus onset. A response was defined as having amplitude greater than 0.01 μS. EDA was included because it tends to be sensitive to the presence of startling or threatening stimuli, and positive or negative emotional events [12]. ECG was recorded with use of a Lead 1 electrode placement, with one 8 mm silver-silver chloride electrode placed on the right inner forearm about 2 cm below the elbow and another placed in the same position on the left inner forearm. A third 8 mm silver-silver chloride electrode was placed on the left inner wrist to serve as a ground. Electrode sites were cleaned with alcohol prep pads in order to improve contact. Interbeat intervals (IBIs) were scored as the time difference in seconds between successive R waves in the ECG signal. A median interbeat interval was recorded during each of the same 5 second sampling periods used to assess skin conductance level. ECG was included because cognitive workload has been linked with changes in heart rate [13]. Further, in previous studies [14-16], greater sympathetic predominance in cardiac control in response to Stroop task demands has been observed. Following completion of the VRST protocol, subjects were assessed for simulator sickness. Notably, none of the subjects reported simulator sickness following VRST.

### 1.2. Paper-and-Pencil Stroop Task

The paper-and-pencil Stroop test measures an individual's ability to shift cognitive set and proffer a measure of a subject's ability to inhibit a prepotent (i.e., an overlearned) response in favor of an unusual one. For the Stroop Color and Word Test, the subject was seated at a desk and presented a Word Page with color words printed in black ink, a Color Page with 'Xs' printed in color, and a color-Word Page with words from the first page printed in colors from the second page (the color and the word do not match).

### 1.3. The ANAM Version of the Stroop Task

The ANAM Stroop is a computer automated version that requires the subject to press one of three computer keys to identify each color. Subjects were seated in front of a blank computer screen. The ANAM Stroop began with a computerized trial of practice words, during which time color words were presented on the otherwise blank screen and the subject was asked to respond to the color of these words by pressing an appropriate key on a keypad. For the interference task subjects pressed the key corresponding to the color of the letters rather than the color indicated by the word.

### *1.4. Virtual Reality Stroop Task*

The VRST involves the subject being immersed in a virtual Humvee as Stroop stimuli appear on the windshield. The VRST is a measure of executive functioning and was designed to emulate the paper-and-pencil as well as ANAM version of the Stroop test. The apparatus used for the virtual humvee included a Pentium 4 desktop computer with a 3 GHz Processor; 3 GB of RAM; and an nVidia GeForce 6800. Two monitors were used: 1) one for displaying the Launcher application which is used by the examiner administering the test; and 2) another for displaying the participant's view of the virtual environment in the HMD. Like the ANAM version, the VRST requires an individual to press one of three computer keys to identify each of three colors, (i.e., red, green, or blue). Unlike the ANAM version, the VRST adds a simulation environment with military relevant stimuli in high and low threat settings. Participants wore an eMagin Z800 Head Mounted Display with an InterSense InteriaCube 2+ attached for tracking. A Logitech Driving Force steering wheel was clamped on to the edge of a table in front of the monitors. Accelerator and brake pedals was positioned under the table. To increase the potential for sensory immersion, a tactile transducer was built using a three foot square platform with six Aura bass shaker speakers (AST-2B-04, 4Ω 50W Bass Shaker) attached. The tactile transducer was powered by a Sherwood RX-4105 amplifier with 100 Watts per Channel x 2 in Stereo Mode.

Development of the scenes, levels, and the virtual Iraqi/Afghani environment was done using Maya animation software. The environments were rendered in real time using the Gamebryo 3-D graphics engine with a fully customizable rendering pipeline, including vertex and pixel shaders, shadows, bump maps, and screenspace geometric primitives. The application also utilizes the NeuroSim Interface (NSI) developed in the Neuroscience and Simulation Laboratory (NeuroSim) at the University of Southern California. The NSI was used for data acquisition, stimulus presentation, psychophysiological monitoring, and communication between the psychophysiological recording hardware and the virtual environment. Configuration parameters were saved to files using the NSI and automatically loaded through its control module, allowing the experimenter to rapidly switch configurations in order to perform specific experimental sequences. The NSI also enabled the sending of event markers from the stimulus presentation computer to the data recording device. Finally, the NSI used compiled Matlab scripts to filter the incoming psychophysiological data in real-time. The software runs on Windows XP 32-bit, and requires 5 Gb of free Hard Drive space.

### *1.5 Stimuli and Design*

Participants were immersed in the VRST as psychophysiological responses were recorded. ECG and EDA were collected as participants rode in a simulated Humvee through alternating zones of low threat (i.e., little activity aside from driving down a desert road) and high threat (i.e., gunfire, explosions, and shouting amongst other stressors). The participants experienced 3 low threat and 3 high threat zones designed to manipulate levels of arousal (start section; palm ambush; safe zone; city ambush; safe zone; and bridge ambush). The order of threat levels was counterbalanced across participants. The VRST was employed to manipulate levels of cognitive workload, and was completed during exposure to the high and low threat zones. The VRST consisted of 3 conditions: 1) word-reading, 2) color-naming, and 3) interference. Each Stroop condition was experienced once in a high threat zone and once in a low threat zone.

Stimuli were presented for 1.25 seconds each, and participants were asked to respond as quickly as possible without making mistakes.

### *1.6 Data Analytics*

Two separate stepwise regressions were used to determine the efficacy of the psychophysiological and behavioral data for predicting levels of threat and cognitive workload. Specifically, average skin conductance level, heart rate (recorded as interbeat intervals), reaction time, response time, and number correct for each zone were entered as predictors for each dependent variable. In order to determine differences in participant response to the ANAM, VRST, and the paper-and-pencil Stroop test, a 3 (Stroop condition) by 3 (presentation type; ANAM, VRST, and paper and pencil Stroop) repeated measures ANOVA was employed for the percentage of correct answers in each condition. A Greenhouse-Geisser correction was used for all reported main effects and interactions with greater than one degree of freedom. Additionally, all significant main effects and interactions were followed up with paired-samples t-tests in order to determine the nature of these effects. A sequentially rejective test procedure based on a modified Bonferroni inequality was used on significant t-test results to prevent inflation of type 1 error rates [17].

## 2. Results

<u>Can threat level in the VRST be predicted using behavioral and physiological data?</u> Results revealed that skin conductance level was the only significant predictor of threat level, $\beta = 4.25$, $t(19) = 2.75$, $p < 0.01$, with increased skin conductance levels in high threat zones. Behavioral data did not predict threat level.

<u>Can Stroop difficulty in the VRST be predicted using behavioral and physiological data?</u> Heart rate and behavioral data were reliably predictive of cognitive workload. Response time significantly predicted workload as responses were slowed during the highest difficulty interference condition, $\beta = 0.34$, $t(19) = 5.18$, $p < 0.001$. Heart rate was also predictive of workload, as heart rate increased during the interference task, $\beta = 0.19$, $t(19) = 2.84$, $p < 0.01$. Subjects also gave more correct responses during the low cognitive load tasks of color naming and word-reading than during the high cognitive workload interference task, $\beta = 0.17$, $t(19) = 2.74$, $p < 0.01$. Thus, heart rate and response time increased during the highest difficulty interference condition, while the number of correct responses decreased

<u>How do paper-and-pencil, ANAM, and VRST behavioral results compare</u>? The results of the analyses on the percentage of correct responses in each condition revealed a main effect of Stroop condition, $F\,(2, 18) = 22.26$, $p < 0.001$. This was the result of significantly fewer correct responses in the interference condition than either the color-naming condition, $t\,(19) = 4.35$, $p < 0.001$, or the word-reading condition, $t\,(19) = 5.75$, $p < 0.001$. The color-naming and word-reading conditions did not differ significantly. A main effect of presentation type was also exhibited, $F\,(2, 18) = 47.34$, $p < 0.001$. A significantly greater percentage of correct responses were exhibited in the paper and pencil Stroop in comparison with either the ANAM, $t\,(19) = 3.49$, $p < 0.01$, or the VRST, $t\,(19) = 7.07$, $p < 0.001$. The ANAM also resulted in a greater percentage of correct responses than the VRST, $t\,(19) = 7.24$, $p < 0.001$. Finally, an interaction

between Stroop condition and presentation type was uncovered, $F$ (4, 16) = 7.11, $p < 0.01$, which was due to significant differences between interference and both color-naming, $t$ (19) = 3.67, $p < 0.01$, and word-reading conditions, $t$ (19) = 4.48, $p < 0.001$, during the VRST only. The ANAM and paper and pencil Stroop tests failed to create significant differences in performance between the Stroop conditions.

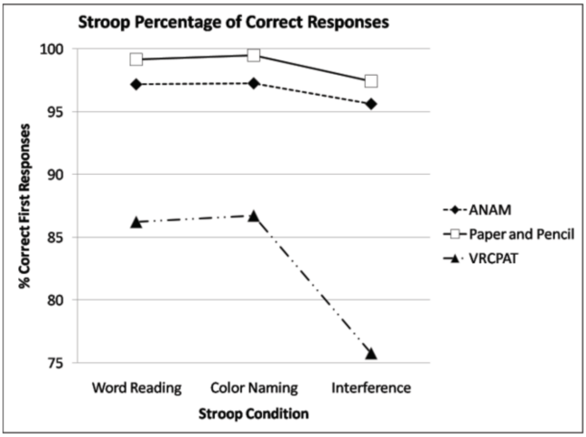


**Figure 1.** Comparison of three versions of the Stroop

## 3 Discussion

Our goal was to conduct an initial pilot study to validate the VRST through comparison with paper-and-pencil and ANAM versions of the Stroop test. We believe that this goal was met. Results revealed that skin conductance level was the only significant predictor of threat level, with increased skin conductance levels in the high threat zones. These results seem to comport well with findings that electrodermal activity tends to be sensitive to the presence of startling or threatening stimuli, and positive or negative emotional events [12]. Heart rate and behavioral data were reliably predictive of cognitive workload. This appears consistent with other work in human computer interaction that has linked cognitive workload with changes in heart rate [13]. Finally, consistent with previous studies [14-16], a shift toward greater sympathetic predominance in cardiac control in response to Stroop task demands was observed, as heart rate and response time increased during the highest difficulty interference condition, while the number of correct responses decreased. As one would expect, high threat zones in the VRST resulted in a significantly smaller proportion of correct responses when compared to the paper-and-pencil and ANAM versions of the Stroop. Interestingly, the difference between ANAM interference and color-naming conditions were not significant. An interaction between Stroop condition and presentation type was uncovered, which is due to significant differences between interference and color-naming in the high threat zones, and the low threat zones. A main effect of presentation type was also demonstrated, which was due to significantly faster responses to the ANAM than the VRST's low threat, and high threat zones.

We recognize that the current findings are only a first step in the development of this tool. More steps are necessary to continue the process of test development and to fully establish the VRST as a measure that contributes to existing assessment procedures for the diagnosis of mTBI. Although the VRST as a measure must be fully validated, current findings provide preliminary data regarding the validity of the virtual environment as a measure of executive functioning. Nevertheless, the fairly small sample size requires that the reliability and validity of the VRST be established using a larger sample of well-matched participants. This will ensure that current findings are not a sample size–related anomaly.

In summary, the findings reported herein provide an initial validation of the VRST as a neurocognitive assessment of Soldier neurocognitive functioning. When compared to ANAM and paper-and-pencil versions of the Stroop, the VRST appears to provide an enhancement in that it has the capacity for providing an indication of a Soldier's reaction time and ability to inhibit a prepotent response while immersed in a military relevant simulation that presents psychophysiologically arousing stimuli.

## References

[1] A.I. Schneiderman, E.R. Braver, & H.K. Kang, Understanding sequelae of injury mechanisms and mild Traumatic Brain Injury incurred during the conflicts in Iraq and Afghanistan: Persistent postconcussive symptoms and Posttraumatic Stress Disorder, *American Journal of Epidemiology* **167** (2008), 1446-1452.

[2] T. Tanielian, & L.H. Jaycox, (Eds.) *Invisible Wounds of War: Psychological and Cognitive Injuries,Their Consequences, & Services to Assist Recovery*. Santa Monica, CA: RAND Corporation (2008).

[3] K.H. Taber, D.L. Warden, & R.A. Hurley, Blast-related traumatic brain injury: what is known? *Journal of Neuropsychiatry and Clinical Neuroscience* **18** (2006), 141-145.

[4] A.L. Wade, J.L. Dye, C. Mohrle, & M.R. Galarneau. Head, face, and neck injuries during Operation Iraqi Freedom II: results from the US Navy-Marine Corps Combat Trauma Registry. *Journal of Trauma* **63** (2007), 836-840.

[5] J.M. Wightman, S.L. Gladish, Explosions and blast injuries, *Annals of Emergency Medicine* **37** (2001), 664-678.

[6] E.M. Martin, W.C. Lu, K. Helmick, L. French, and D.L. Warden, (2008). Traumatic brain injuries sustained in the Afghanistan and Iraq wars, *American Journal of Nursing* **108** (2008), 40-7.

[7] I. Cernak, J. Savic, D. Ignjatovic, M. Jevtik, Blast injury from explosive munitions, *Journal of Trauma* **47** (1999), 96-103.

[8] R.L. Kane & G.G. Kay, Computerized assessment in neuropsychology: A review of test and test batteries. *Neuropsychology Review* **3** (1992), 1–117.

[9] T. Roebuck-Spencer, W. Sun, A.N. Cernich, K. Farmer, & J Bleiberg, Assessing change with the Automated Neuropsychological Assessment Metrics (ANAM): Issues and challenges, *Archives of Clinical Neuropsychology* **22(S1)** (2007), S79-S87.

[10] T.D. Parsons, A. Notebaert, & K. Guskewitz, Application of Reliable Change Indices to Computerized Neuropsychological Measures of Concussion, *The International Journal of Neuroscience* **119** (2009), 492-507.

[11] A.I. Drake, N. Gray, S. Yoder, M. Pramuka, & M. Llewellyn, Factors predicting return to work following mild traumatic brain injury: A discriminant analysis. *Journal of Head Trauma Rehabilitation* **15** (2000) 1103-1112.

[12] T. Butler, H. Pan, O. Tueschent, et al., Human fear-related motor neurocircuitry. *Neuroscience* **150** (2007), 1-7.

[13] Y. Hoshikawa and Y. Yamamoto, Effects of Stroop color-word conflict test on the autonomic nervous system responses, *Heart and Circulatory Physiology* **272** (1997), 1113–1121.

[14] Y.N. Boutcher and S.H. Boutcher, Cardiovascular response to Stroop: effect of verbal response and task difficulty, *Biological Psychology* **73** (2006), 35–241.

[15] J.P.A. Delaney and D.A. Brodie, Effects of short-term psychological stress on the time and frequency domains of heart rate variability, *Perceptual and Motor Skills* **91** (2000), 515–524.

[16] K.J. Mathewson, M.K. Jetha, I.E. Drmic, S.E. Bryson, J.O. Goldberg, G.B. Hall, D.L. Santesso, S.J. Segalowitz, & L.A. Schmidt, Autonomic predictors of Stroop performance in young and middle-aged adults, *International Journal of Psychophysiology* **76** (2010), 123-129.

[17] D.M. Rom. A sequentially rejective test procedure based on a modified Bonferroni inequality. *Biometrika* **77** (1990), 663–665.