Routledge
Taylor & Francis Group

Virtual reality Stroop task for assessment of supervisory attentional processing

Thomas D. Parsons[1], Christopher G. Courtney[2], and Michael E. Dawson[2]

[1]Department of Psychology, University of North Texas, Denton, TX, USA
[2]Department of Psychology, University of Southern California, Los Angeles, CA, USA

The current project is a preliminary attempt at assessing the potential of a three-dimensional virtual reality Stroop task (VRST) for investigation of supervisory attentional processing. The VRST was compared with paper-and-pencil and computer automated (two-dimensional) versions of the Stroop. Psychophysiological measures were also used to assess varying levels of user arousal. The three versions of the Stroop task were conducted with a sample of 50 healthy university students, between the ages of 18 and 28 years, which included equivalent distributions of men and women from ethnically diverse populations. No significant differences were found for age, gender, or education. Results indicated that the typical Stroop effect pattern found in the Delis–Kaplan Executive Function System (D-KEFS) and Automated Neuropsychological Assessment Metrics (ANAM) occurs in the VRST; the complexity condition of the VRST offers potential for assessment of exogenous and endogenous attentional processing; the two single-item presentation formats of the Stroop are comparable; there is a psychophysiological difference in arousal between low- and high-threat zones; and while there is a learning effect present in the VRST across conditions, it is limited by the arousal conditions in the high-threat zones. Additionally, no negative side effects were associated with use of the virtual environment. We conclude that the VRST has the potential to offer a novel approach to assessment of supervisory attentional processing within an ecologically relevant environment.

Keywords: Neuropsychological assessment; Psychophysiology; Ecological validity; Virtual environment.

Neurocognitive assessment of executive functioning represents a principal objective of neuropsychological evaluations. The executive functions revolve around supervisory attentional networks that include: selective attention, inhibitory control, planning, problem solving, and some aspects of short-term memory (Baddeley 1996; Baddeley & Hitch, 1974; Norman & Shallice, 1986). The Stroop (Stroop, 1935) is a test to evaluate the executive aspects of attention control and is one of the most widely used tasks in cognitive neuroscience (Gazzaniga, Ivry, & Mangun, 1998; MacLeod, 1992). Although there are variations in Stroop stimuli across these studies (e.g., number of items), they all measure freedom from distractibility, selective attention, response conflict, and response inhibition.

In a typical neuropsychological assessment, the Stroop test is a paper-and-pencil measure used to assess executive functioning and attentional control through the presentation of blocks of multiple stimuli on a card (multi-item presentation; Uttl & Graf, 1997). An example of a multi-item paper-and-pencil Stroop is the Delis–Kaplan Executive Function System (D-KEFS) Color–Word Interference Test (Delis, Kaplan, Kramer, & Ober, 1997). During the D-KEFS Color–Word Interference Test, a participant is to perform the following: (a) word reading: read words that are the names of colors, (b) color naming: name the color of ink patches, (c) inhibition: name the color ink in which incongruent color words are printed, and (d) inhibition/switching: shift between color naming, word reading, and inhibition. While multi-item presentations of rows of Stroop stimuli have been well validated in a number of psychometric studies, there are situations in which a single-item presentation is preferable (see Davidson, Zacks, & Williams, 2003). Examples

include instances in which researchers are interested in: (a) analyzing reaction times, number correct, and/or impact of errors for individual stimuli; and (b) randomized counterbalancing of various trial types (e.g., neutral, interference, cued, facilitated). Multi-item presentations may result in confounds from or interactions with visual distractor interference. Further, multi-item presentations expose the participant to all the stimuli at once, which may enhance the training/learning curve resulting in greater practice effects when compared to single-item procedures (see Lemay, Bedard, Roulea, & Tremblay, 2004).

In the past decade, a number of computerized Stroop-like tasks have been developed: CogSport card recognition task (CogSport, 1999), Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT Color Match; Lovell, Collins, Podell, Powell, & Maroon, 2000), Automated Neuropsychological Assessment Metrics (ANAM) Stroop (ANAM; Johnson, Vincent, Johnson, Gilliland, & Schlegel, 2008), and HeadMinder picture recognition (Erlanger, Feldman, & Kutner, 1999). Of these computerized batteries, the ANAM single-item Stroop presentation is a close approximation of the traditional Stroop test, in which subjects press one of three computer keys to identify each of three colors (i.e., red, green, or blue; see Reeves, Winter, Bleiberg, & Lane, 2007).

Although computer automated versions of the Stroop task offer a number of advantages over traditional paper-and-pencil testing (e.g., increased standardization of administration; increased accuracy of timing presentation and response latencies; ease of administration and data collection; and reliable and randomized presentation of stimuli for repeat administrations; see Schatz & Browndyke, 2002), the ANAM Stroop lacks the extensive normative data supporting the paper-and-pencil versions of the Stroop, and stimuli are always presented in the middle of the screen against a blank background. A more enhanced assessment would involve a more flexible presentation format (Bugg, Jacoby, & Toth, 2008), in which stimulus complexity may be adjusted to manipulate differing aspects of exogenous and endogenous attentional processing. Exogenous attention refers to the impact of external physical events upon automatic attention. Endogenous attention refers to active direction of attention to something deemed important by the participant (Posner, 1980). For example, automatic processing and exogenous attention may be engaged by having a subject stare at a computer screen that has stimuli consistently presented in a fixed central location. Contrariwise, an example of controlled processing and endogenous attention

is reflected in a scenario in which the stimuli appear randomly throughout the computer screen. Assessment of the extent to which attentional selectivity can be inferred is relative to whether the stimuli are in fixed (expected) or variable (unexpected) spatial locations of the monitor screen (see Posner, 1980).

Automatic and controlled processing using the Stroop

Much of the research into the Stroop has involved assessment of automatic and controlled processing of attentional stimuli (Fisk & Schneider, 1981). Whilst automaticity of processing (e.g., word reading) is considered as parallel, requiring little effort, and not under the participant's direct control, controlled processing (e.g., interference) is understood as serial, effortful, and under a participant's direct control (Schneider & Shiffrin, 1977; see MacLeod, 1991, for a review). Although Shiffrin and Schneider (1977) use the decreased processing speed in controlled versus automatic processing (assuming automaticity to be an all-or-none happening), alternative conceptions have emerged. MacLeod and Dunbar (1988) view the features of automaticity as developing relative to the degree of training on respective tasks. This learned association may increase the practice-related decline in Stroop interference reaction time as color naming becomes increasingly automatic, and an automatic reading suppression response develops (Cohen, Dunbar & McClelland, 1990; Dulaney & Rogers, 1994).

The automatic and controlled processing of attentional demands requires varying levels of effort (Kahneman & Treisman, 1984). Although some high-information-load tasks may be relatively automatic in controlled circumstances, the total available processing capacities may be increased or decreased by other factors such as arousal. The increased difficulty found in the Stroop interference task has been shown to directly evoke autonomic changes in cardiovascular arousal (Hoshikawa & Yamamoto, 1997). Seeley et al. (2007) identified an "executive control" network and a "salience" network. The executive network links dorsolateral frontal and parietal neocortices. The "salience" network is anchored by dorsal anterior cingulate and orbital frontoinsular cortices that connect to subcortical and limbic areas. Whilst the executive network is frequently coactivated with the salience network in tasks of attention, working memory, and response selection (Kerns et al., 2004; Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis,

2004), the salience network also activates in response to threats (Craig, 2002). Increased arousal may impact the processing of salient information and enhance the contrast between stimuli with different levels of salience (Critchley, 2005). This issue of intensity and attentional effort may be described as a top-down supervisory attentional network, in which threats to attention result in recruitment and integration of performance monitoring (prefrontal–anterior cingulate) and motivational (mesolimbic) systems (Sarter, Gehring, & Kozak, 2006). Further, the involvement of the anterior cingulate cortex (ACC) in the Stroop color interference task has been delineated (Liotti, Woldorff, Perez, & Mayberg, 2000; Pardo, Pardo, Janer, & Raichle, 1990).

While both paper-and-pencil and computerized versions of the Stroop have been widely used, their ecological validity may diminish predictions about real-world functioning (Chaytor & Schmitter-Edgecombe, 2003). Traditional Stroop measures do not replicate the diverse environment in which persons live. The functional and predictive relationship between a person's performance on a Stroop test and behavior in a variety of real-world settings can be limited by the fact that traditional Stroop tests do not manipulate the environment to look at the impact of a person's level of arousal on their performance. Supervisory attentional processing has been found to be impacted by influences from many factors. One of these factors is arousal level (Critchley, 2005; Seeley et al., 2007). As mentioned above, there is a need for assessment measures that assess whether arousal impacts automatic and controlled processing.

In summary, paper-and-pencil (multi-item presentation) versions of the Stroop do not allow for studies of differentiation of automatic and controlled processing relative to individual stimuli. While computerized (single-item presentation) versions offer some advance in this regard, they are often not well validated and have limited generalizability. An example of a computerized Stroop assessment is the ANAM. The ANAM Stroop presents stimuli at a fixed point in the middle of the screen. As a result, the ANAM Stroop does not assess for information related to the impact of adjustments to stimulus complexity (in which the complexity of presentation is manipulated). Finally, the assessment of supervisory attention should reflect the impact of training and varying levels of arousal found in real-world situations. While paper-and-pencil and computerized Stroop tests offer psychometric rigor, there is need for methodological techniques that allow for assessments that better capture the subject's performance in a real-world setting.

Virtual reality Stroop task

We address these issues using a virtual reality Stroop task (VRST). The VRST involves the subject being immersed in a virtual high-mobility multipurpose wheeled vehicle (HMMWV) as Stroop stimuli appear on the windshield (see Figure 1). The VRST assesses automatic processing and exogenous attention by presenting stimuli consistently in a fixed central location in the simple interference condition. The VRST assesses controlled processing and endogenous attention during the complex interference condition by presenting stimuli at random positions throughout the computer screen. Hence, the adjustments to stimulus complexity on the VRST can be used to assess attention at both exogenous (automatic processing) and exogenous (controlled processing) levels (Posner, 1980).

The VRST was designed to be a measure of supervisory attentional processing (executive functioning) and emulates aspects of both the paper-and-pencil and ANAM versions of the Stroop test. Like the paper-and-pencil Stroop, the VRST uses Stroop stimuli (though in single-item presentation) to assess 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

Figure 1. A virtual reality Stroop task in which the subject is immersed in a virtual high-mobility multipurpose wheeled vehicle (HMMWV) as Stroop stimuli appear on the windshield.

assessment of reaction time to single-item presentations of Stroop stimuli. The VRST goes beyond the ANAM and paper-and-pencil versions of the Stroop by assessing for the impact of training, stimulus intensity (arousal from threatening stimuli), and stimulus complexity (endogenous and exogenous presentations) in a virtual environment that replicates real-world activities. The VRST adds a simulation environment with military relevant stimuli in high- and low-threat settings. The VRST includes both simple (presented in the middle of the screen) and complex (presented in different locations on the screen) presentation of the interference stimuli.

In the following, we review results related to the following four primary questions:

(1) For the comparison among the Stroop types:
   (1a) Does the typical Stroop effect (decreased processing speed in controlled versus automatic processing; Shiffrin & Schneider, 1977) pattern found in the D-KEFS and (ANAM) occur in the VRST?
   (1b) Does the complexity condition of the VRST offer potential for assessment of exogenous and endogenous attentional processing (Posner, 1980)?
(2) Is there a psychophysiological difference in arousal between low- and high-threat zones?
(3) For the VRST, is there a learning effect (see MacLeod & Dunbar, 1988) present across the Stroop conditions?
(4) Is there a difference in learning across high- and low-threat conditions (Critchley, 2005; Seeley et al., 2007)?

METHOD

This study was designed to compare the VRST with paper-and-pencil (D-KEFS) and computer automated (ANAM) versions of the Stroop on behavioral measures. We utilized psychophysiological measures to assess varying levels of threat on the user. We aimed to assess different aspects of supervisory attentional control within the VRST. This study expands considerably upon a previously published initial validation study (Parsons, Courtney, Arizmendi, & Dawson, 2011).

Participants

The University of Southern California's Institutional Review Board approved the study. A total of 50 college-aged subjects participated in the study. The age range of participants was 18 to 28 years of age (age: $M = 19.71$; $SD = 2.82$). Participants were 75% female. The education range of participants was 12 to 15 years (education: $M = 13.6$; $SD = 1.07$). No significant differences were found for age, gender, or education. After informed consent was obtained, basic demographic information was recorded. Participants were also given a medical health history form to assess the presence of any mental or physical disorders that may have hindered their performance. No participants were excluded for responses given on this form.

Design and measures

Participants completed three versions of the Stroop Interference Task: (a) the multi-item presentation of the D-KEFS (Delis et al., 1997); (b) the single-item presentation of the ANAM Stroop task (ANAM, 2007); and (c) the single-item presentation of the VRST. Testing occurred in a quiet, climate-controlled environment in a university-owned computer lab. Psychophysiological assessment occurred during the VRST version of the Stroop task. The order in which the various Stroop tests were administered was counterbalanced across subjects. Participants completed the simulator sickness questionnaire, which includes a pre- and post-VR exposure symptom checklist (Kennedy, Fowlkes, Berbaum, & Lilienthal, 1992).

Paper-and-pencil Stroop task

The D-KEFS (Color–Word Interference Test) was used for the multi-item paper-and-pencil presentation of the Stroop test. We followed the D-KEFS manual's prescribed approach to administration and the D-KEFS's "Scoring Assistant" software for scoring of the Color–Word Interference Test. The subject was seated at a desk and was presented with the following stimuli from the D-KEFS: (a) "word reading" stimulus card with 50 color words printed in black ink; (b) "color naming" card with 50 colored blocks; (c) "color–word interference" card with 50 color names printed in a discrepant ink color; and (d) "color–word interference/switching" card, in which the subject performed the interference task if and only if the words (50 total words) did not have a box drawn around them. Participants were timed for completion of each condition, and all incorrect responses were recorded.

Computerized Stroop task

The ANAM (single-item presentation) Stroop task requires the subject to press a computer key labeled red, green, or blue to identify each color stimulus presented. We followed the ANAM (2007) manual's approach to administration and scoring of the ANAM Stroop test. Each participant used the keyboard to take the ANAM Stroop test. A short practice session (10 stimuli) was provided just prior to the actual testing of each Stroop condition (i.e., color-naming, word-reading, and interferences tasks). There are three possible blocks of 50 trials for this test. In the first block, the words RED, GREEN, and BLUE are presented individually in black type on the display. The user is instructed to read each word aloud and to press a corresponding key for each word ("red" = 1; "green" = 2; and "blue" = 3). In the second block, a series of XXXXs is presented on the display in one of three colors (red XXXXs, green XXXXs, or blue XXXXs). The user is instructed to say the color of the XXXXs aloud and to press the corresponding key based on color. In the third block, a series of individual words ("RED," "GREEN," or "BLUE") are presented in a color that does not match the name of the color depicted by the word. The user is instructed to say the color of the word aloud rather than reading the actual word and to press the response key assigned to that color. The participants were asked to respond to each individual color stimulus as quickly as possible without making mistakes. Speed and duration of single-item stimulus presentations were user defined— meaning that a new stimulus appeared only after the user correctly identified the previous stimulus. The ANAM (2007) scores the Stroop as follows:

- *Color–word score* was calculated by (a) multiplying the number of correct colors named by the number of correct words named; and (b) dividing the product by the sum of number of correct colors named plus the number of correct words named.
- *Interference score* was calculated by taking the number correct on the interference task and subtracting the color–word score.

Virtual reality Stroop task

Two monitors were used: (a) one for displaying the Launcher application, which is used by the examiner administering the test; and (b) another for displaying the participant's view of the virtual environment in the head-mounted display (HMD; eMagin Z800 with an InterSense InteriaCube 2+

attached for tracking). 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 × 2 in stereo mode.

Animation software was utilized for development of the virtual Iraqi/Afghani and HMMWV VRST environment. The environments were rendered in real time using a graphics engine with a fully customizable rendering pipeline, including vertex and pixel shaders, shadows, bump maps, and screenspace geometric primitives (Parsons & Reinebold, 2012). A MATLAB scoring program (Wu et al., 2010) and human–computer interface (Clinical Neuropsychology and Simulation Interface; CNS-I) was employed for data acquisition, to guide stimulus presentation, and for psychophysiological monitoring (Wu, Lance, & Parsons, 2013). The CNS-I also allowed for key events in the environment to be logged and time stamped with millisecond temporal accuracy.

Participants rode in the simulated HMMWV through alternating zones of low threat and high threat. Amongst low-threat zones, little activity in the virtual environment was presented aside from driving down a desert road. During the high threat zones, gunfire, explosions, and shouting amongst other stressors were presented throughout. The participants experienced four low-threat and four high-threat zones designed to manipulate levels of arousal. The order of threat level presentation was counterbalanced across participants. The VRST was completed during exposure to the high- and low-threat zones. The VRST consisted of four conditions: (a) word reading, (b) color naming, (c) simple interference (stimuli presented in the middle of windshield), and (d) complex interference (stimuli presented in variable locations on windshield). Each Stroop condition was experienced once in a high-threat zone and once in a low-threat zone. As with the ANAM Stroop test, the presentation speed of individual stimuli was user defined, meaning that the subsequent stimulus did not appear until the appropriate key was pressed for the stimulus currently being viewed. The ANAM (2007) manual's approach to scoring was adopted for comparison:

- *Color–word score* was calculated by (a) multiplying the number of correct colors named by the number of correct words named; and (b) dividing the product by the sum of number of correct colors named plus the number of correct words named.

- *Simple interference score* was calculated by taking the number correct on the Interference task and subtracting the color–word score—during the simple presentation of stimuli.
- *Complex interference score* was calculated by taking the number correct on the interference task and subtracting the color–word score—during the complex presentation of stimuli.

A weighted score out of 50 was used for comparison to the D-KEFS and VRST. The D-KEFS has a total of 50 stimuli. The VRST is a timed measure with a maximum of 50 stimuli available per zone. To maintain counterbalancing of threat stimuli, there were times (e.g., a participant's response speed was not fast enough to allow for all 50 stimuli to be presented) that the number of available VRST stimuli was below 50.

Psychophysiological responses were recorded using the Biopac MP150 system throughout the participants' period of immersion within the VRST. Electrocardiogram (ECG), electrodermal activity (EDA), and respiration were recorded as participants rode in the simulated HMMWV through alternating zones of low threat and high threat. ECG, EDA, and respiration were recorded simultaneously throughout exposure to the VRST task. Data were logged with use of an in-house designed program called the NeuroSim Interface, which recorded raw psychophysiological signals as well as behavioral responses and allowed for the logging of events in the virtual environment. 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 nondominant hand. Electrodes were filled with a 0.05-molar isotonic sodium chloride (NaCl) paste to provide a continuous connection between the electrodes and the skin. ECG was used to record rate and regularity of heart rate using 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. Respiration was recorded with a transducer belt placed around the rib cage.

Feature extraction and optimal response classification for VRST and psychophysiological responses were examined using a MATLAB scoring program (i.e., CNS-I) designed specifically for this study (Wu et al., 2010). This allowed for assessment of performance validity (suboptimal effort) and screening for outliers to establish data integrity: (a) identification of outliers as observations exceeding three standard deviations from the median reaction time; (b) exclusion of observations that are in both the top 1% in speed and simultaneously in the bottom 1% of accuracy; and (c) filtering and pattern recognition assessment for establishing feature sets using support vector machine classifiers (Wu et al., 2013).

Interbeat intervals (IBIs) were scored as the time difference in milliseconds between successive R waves in the ECG signal. Median IBIs for each zone experienced in the VRST served as the features for analysis. Tonic skin conductance response levels (SCLs) were also calculated for each zone and analyzed, as were frequencies of nonspecific responses (NSRs) in the electrodermal recording (Dawson, Schell, & Filion, 2007). NSRs were defined as an increase of greater than 0.01 μS, with a peak amplitude coming within a window of 1 to 3 s following the initial increase in the signal. The number of NSRs occurring in each zone was then calculated for analysis. Interbreath interval (IBrI) was recorded from the respiration data and was scored for each zone by calculating the median interval in seconds between peak amplitudes in the raw data signal, signifying one full breath cycle. All psychophysiological features extracted were checked visually for accuracy and were scored manually if artifacts impeded the scoring program from correctly assessing the signal. Because psychophysiological response recording was utilized to assess changes in arousal caused by the threat level manipulation, paired-samples *t* tests were used to compare responses for each psychophysiological feature during the high-threat and low-threat zones.

RESULTS

For all analyses of variance (ANOVAs) that included a variable with more than one degree of freedom, paired-samples *t* tests were used to identify the precise nature of any main effects. Additionally, all significant main effects and interactions were followed up with paired-samples *t* tests in order to determine the nature of these effects. Effect sizes were proffered for each analysis. Correlational analyses were conducted to assess for similarities across modalities. A sequentially rejective test procedure based on a modified Bonferroni inequality was used on significant *t*-test results to prevent inflation of Type I error rates (Rom, 1990). Additionally, a Greenhouse–Geisser correction was used for all reported main effects and interactions with greater than one degree of freedom.

Behavioral results of the three Stroop types: Does the typical Stroop effect pattern found in the D-KEFS and ANAM occur in the VRST?

Behavioral dependent variables consisted of reaction time and number of correct responses. Correlational analyses on these variables were conducted to assess the similarities in color–word results across modalities (e.g., D-KEFS, ANAM, and VRST): D-KEFS and ANAM: $r = .24$, $p = .14$; D-KEFS and VRST: $r = .21$, $p = .16$; VRST and ANAM: $r = .493$, $p < .001$. In order to assess the differences between the three Stroop types, a 3 (Stroop type: D-KEFS, ANAM, VRST) by 2 (Stroop condition: color–word and interference conditions) repeated measures ANOVA was employed. Due to the lack of a complex interference condition in the ANAM Stroop, only simple interference was used in this ANOVA for the VRST and the D-KEFS Stroop. In order to compare the three Stroop types directly, color–word scores and interference scores were calculated, which make up the two-level Stroop condition variable. Color–word scores were calculated by multiplying the number of correct responses during the color-naming and word-reading conditions and dividing the product by the sum of the number of correct color-naming and word-reading responses. The interference scores were calculated by subtracting the color–word score from the number of correct responses during the interference condition. The VRST presented a maximum of 50 stimuli in each condition. However, participants only experienced each condition for 1 minute, leading to variability in the number of stimuli presented based upon the speed with which responses were given. The timing of each condition was constrained to ensure that each participant experienced the exact same environmental stimuli in the VE at exactly the same time. Due to the fact that the different Stroop tests have a different number of trials for each condition, a weighted score out of 50 was created for each, which created an adjusted score depending on the number of trials that each participant experienced. The D-KEFS Stroop displayed 50 trials for each condition, while the average number of trials experienced in VRST was 47 ($SD = 2.96$; min = 26, max = 50), and the average number of ANAM trials experienced was 41 ($SD = 5.16$; min = 25, max = 52).

Comparison of color–word and interference scores across all three Stroop types

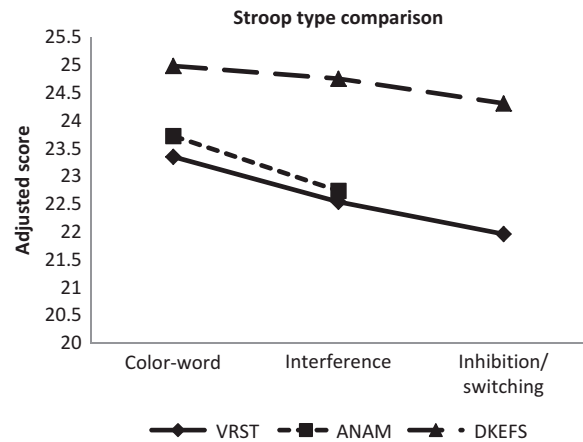Correlational analyses were conducted on these variables to assess the similarities in interference

Figure 2. Adjusted color–word and interference scores for the three Stroop types. VRST = virtual reality Stroop task; ANAM = Automated Neuropsychological Assessment Metrics; D-KEFS = Delis–Kaplan Executive Function System. It should be noted that the ANAM Stroop and the D-KEFS do not include a complex interference condition. Given the additional task demands of the D-KEFS inhibition/switching, we compared it to the VRST complex condition. The adjusted scores are calculated by creating a proportion for each individual's color–word and interference scores based on the number of trials actually experienced and adjusting that score to be out of the total possible number of trials, which were 50.

"accuracy" scores across modalities (e.g., D-KEFS and ANAM: $r = .28$, $p = .086$; D-KEFS and VRST: $r = .45$, $p < .01$; VRST and ANAM: $r = .38$, $p < .01$). In terms of reaction time, the D-KEFS and ANAM were significantly correlated for color naming ($r = .36$, $p < .01$); word reading ($r = .70$, $p < .001$); and interference ($r = .55$, $p < .001$).

The repeated measures ANOVA involving all three Stroop types revealed a main effect of both Stroop type, $F(2, 48) = 41.42$, $p < .001$, $\eta^2 = .522$, and Stroop condition, $F(1, 49) = 15.78$, $p < .001$, $\eta^2 = .293$. In all cases, participants had lower scores in the interference condition than they did in the color–word condition (see Figure 2). The Stroop type main effect was the result of higher scores on the D-KEFS than on both the VRST, $t(49) = 8.20$, $p < .001$, $d = 2.34$, and the ANAM Stroop, $t(49) = 8.11$, $p < .01$, $d = 2.31$. The difference between the VRST and ANAM Stroop conditions was not significant, nor was there a significant interaction between Stroop type and Stroop condition.

Reaction time differences between VRST and ANAM: Are the two single-item presentation formats (i.e., VRST and ANAM) of the Stroop comparable?

To assess differences in processing speed associated with the VRST and ANAM, which provide

reaction times for each individual stimulus, a separate 2 (Stroop type) by 3 (Stroop condition) repeated measures ANOVA was employed. The median reaction times for each condition were used, which represent the amount of time it took for the participant to answer with the correct response. The D-KEFS Stroop test was not included in these analyses because the nature of the multiple-item presentation format does not allow for recording of reaction times to individual stimuli.

The reaction-time-related analyses in the VRST and ANAM (see Figure 3) revealed a main effect of Stroop type, $F(1, 49) = 1378.08$, $p < .001$, $\eta^2 = .968$, such that it took participants longer to respond correctly to VRST stimuli than to ANAM Stroop stimuli. Additionally, there was a main effect of Stroop condition, $F(2, 48) = 25.65$, $p < .001$, $\eta^2 = .533$, being that interference reaction times in both Stroop types were slower than color-naming reaction times, $t(49) = 8.83$, $p < .001$, $d = 2.52$, and word-reading reaction times, $t(49) = 7.67$, $p < .001$, $d = 2.19$. Reaction times between color naming and word reading did not differ significantly. An interaction between Stroop type and Stroop condition was also uncovered, $F(2, 48) = 25.56$, $p < .001$, $\eta^2 = .532$. The interaction was a result of a greater difference between interference reaction times and the other two conditions in the VRST than was present in the ANAM Stroop, $t(49) = 5.28$, $p < .001$, $d = 1.50$. This t-score was calculated with use of a paired-samples t test between the difference of interference and the average reaction times of the color-naming and word-reading conditions for both VRST and ANAM Stroop.
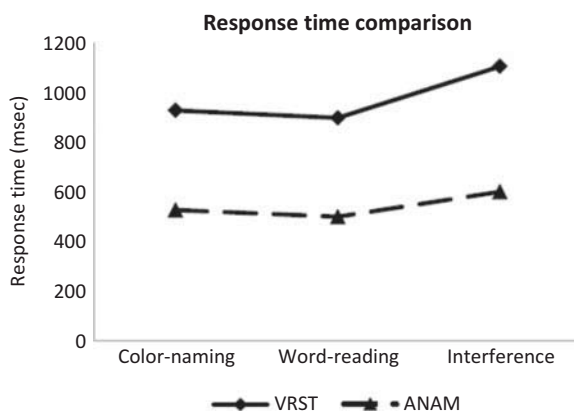

Figure 3. Reaction times for the virtual reality Stroop task (VRST) and Automated Neuropsychological Assessment Metrics (ANAM) Stroop reported in milliseconds.

Does the complexity condition of the VRST offer potential for assessment of exogenous and endogenous attentional processing (Posner, 1980)?

Herein there was a comparison between the VRST simple/complex conditions and the D-KEFS interference/switching conditions. Scores were calculated for the VRST complex presentation of the interference stimuli. Scores were also calculated for the D-KEFS interference/switching. A separate 2 (Stroop type) by 3 (Stroop condition) repeated measures ANOVA was utilized to examine differences in the D-KEFS and VRST for color–word, interference, and the complexity condition of the VRST related to the D-KEFS interference/switching.

The results of the ANOVA involving the VRST and D-KEFS Stroop tests to assess VRET complex interference and the interference/switching (D-KEFS) effects resulted in a main effect of Stroop type, which again was the result of higher scores on the D-KEFS, $F(1, 49) = 57.64$, $p < .001$, $\eta^2 = .573$ (see Figure 2). A main effect of Stroop condition was also revealed, $F(2, 48) = 17.34$, $p < .001$, $\eta^2 = .287$, such that performance was better in color–word tasks than in either the simple interference condition, $t(49) = 4.99$, $p < .001$, $d = 1.42$, or the complex interference condition, $t(49) = 6.49$, $p < .001$, $d = 1.85$, and the complex interference condition resulted in significantly poorer performance than did the simple interference condition, $t(49) = 2.87$, $p < .01$, $d = 0.82$. The interaction between Stroop type and Stroop condition was not significant. In further analyses, constrained to correlation of the accuracy scores for D-KEFS inhibition/switching and VRST complex interference, the results were not significant (D-KEFS and VRST: $r = .20$, $p = .19$).

Effects of threat level in the VRST

Is there a behavioral performance difference between high- and low-threat zones?

Analyses of the effect of threat level on the color–word and interference scores for both simple and complex interference resulted in a main effect of threat level, $F(1, 49) = 6.81$, $p < .05$, $\eta^2 = .118$, and condition, $F(2, 48) = 45.97$, $p < .001$, $\eta^2 = .648$ (see Figure 4). The interaction between threat level and condition was also significant, $F(2, 48) = 3.71$, $p < .05$, $\eta^2 = .129$. The interaction was a result of a significant difference in performance
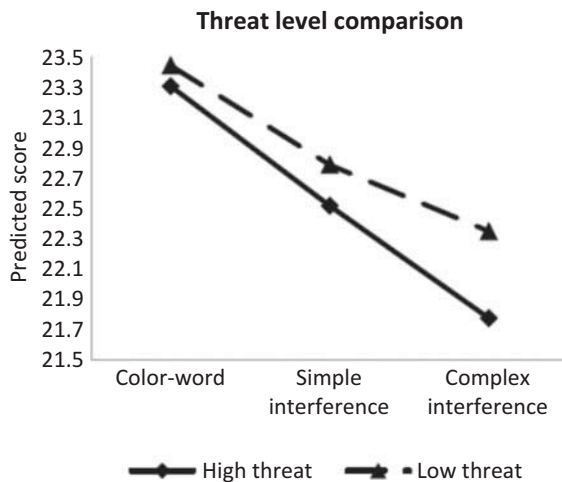
Threat level comparison

Figure 4. A comparison of the adjusted scores for each of the virtual reality Stroop task (VRST) conditions, broken down between high- and low-threat zones.

only during the complex interference condition, $t(49) = 2.23$, $p < .05$, $d = 0.63$, such that the high-threat zones resulted in lower performance scores than the low-threat zones. The other two conditions did not differ as a result of the threat manipulation, though performance in the simple interference condition evidenced a trend toward lower scores in the high-threat zones, $t(49) = 1.87$, $p = .07$, $d = 0.54$.

For the VRST, is there a learning effect present across the Stroop conditions relative to arousal conditions?

Due to the fact that each Stroop condition in the VRST was experienced in both a high-threat and a low-threat zone, an additional pair of ANOVAs were utilized to assess the possibility of learning effects based on the order in which the high- and low-threat zones were experienced. A total of

26 participants experienced the low-threat zones first, and 24 participants began in a high-threat zone. The same 2 (threat level) by 3 (Stroop condition) ANOVAs were utilized to test for order effects for both accuracy performance (i.e., color–word and interference scores) and processing speed (i.e., reaction times).

Assessment of learning effects on processing speed yielded a different set of information. When testing for order effects regarding whether the first zone experienced was high threat or low threat, we observed that reaction time performance revealed an interaction between threat level and Stroop condition for both the low-threat first participants, $F(2, 24) = 6.80$, $p < .01$, $\eta^2 = .118$, and the participants that experienced high threat first, $F(2, 22) = 4.41$, $p < .05$, $d = 4.46$. However, the nature of these interactions differed between the two groups.

Although participants who experienced the low-threat zones first demonstrated a main effect of threat level, $F(2, 24) = 12.56$, $p < .01$, $d = 2.76$, such that reaction times in the low-threat zone were slower than those in the high-threat zone, which was experienced second (descriptive values can be seen in Table 1), this group of participants evidenced no significant differences in reaction times between high- and low-threat zones on the color-naming and word-reading tasks and rather only evidenced significant differences based on threat level during interference tasks. Reaction times during simple interference were significantly slower than those in the high-threat simple interference, $t(25) = 3.63$, $p < .01$, $d = 1.45$, and the same pattern held for the complex interference condition, $t(25) = 2.86$, $p < .01$, $d = 1.14$.

The opposite was true for participants who experienced the high-threat zones first, as this group of participants evidenced faster reaction times in the low-threat zones, $F(2, 22) = 13.13$, $p < .01$,

TABLE 1
Reaction times by order of high- and low-threat zones

Participants	Condition	High-threat zones		Low-threat zones	
		Median	SD	Median	SD
Experiencing low-threat zones first	Color naming	886.80	101.04	884.47	105.56
	Word reading	904.76	88.16	907.12	115.85
	Simple interference	1025.80	142.31	1132.71	212.99
	Complex interference	1036.89	149.65	1121.36	205.67
Experiencing high-threat zones first	Color naming	947.56	101.06	916.43	101.17
	Word reading	958.72	120.63	961.59	106.83
	Simple interference	1199.89	281.44	1094.62	238.19
	Complex interference	1156.13	203.35	1110.78	237.35

Note. Reaction times in ms.

TABLE 2
Accuracy scores by order of high- and low-threat zones

Participants	Score	High-threat zones		Low-threat zones	
		Mean	SD	Mean	SD
Experiencing low-threat zones first	Color–word score	23.24	1.30	23.48	1.27
	Simple interference score	23.02	1.92	22.49	3.08
	Complex interference score	22.15	2.35	22.18	3.33
Experiencing high-threat zones first	Color–word score	23.39	0.99	23.40	1.17
	Simple interference score	21.93	2.56	23.13	2.42
	Complex interference score	21.34	3.62	22.54	2.41

$d = 2.58$. Similar to the pattern revealed by the low-threat-first group, this group of participants showed no significant difference in reaction times between high- and low-threat zones on the color-naming and word-reading tasks, only showing differences based on threat level during interference tasks. The participants who experienced the high-threat zones first had slower reaction times in the high-threat zones during the simple interference task, $t(23) = 3.37$, $p < .01$, $d = 1.40$.

Learning effects also had an effect on participant accuracy performance; however, the effects were more selective than those found in reaction time analyses (see Table 2 for a descriptive statistical summary). In participants that experienced the low-threat zones first, there was no main effect of threat level, nor was there an interaction between threat level and Stroop condition. These participants did display the expected Stroop condition main effect, with lower scores on the interference tasks than on the color–word tasks, $F(2, 24) = 6.04$, $p < .01$, $d = 3.98$. Contrariwise, order effects did appear to affect participants who experienced the high-threat zones first. Participants in this group responded less accurately when in the high-threat zones than when in the low-threat zones, $F(2, 22) = 7.89$, $p < .01$, $d = 3.33$. And, as was the case in the group that experienced the low-threat zones first, participants who first experienced the high-threat zones were less accurate as the task became more difficult, resulting in a significant Stroop condition main effect, $F(2, 22) = 6.73$, $p < .01$, $d = 3.61$. The interaction between threat level and Stroop condition among the high-threat-first participants was not significant.

Are there psychophysiological differences between high- and low-threat zones?

Assessment of the effects of the threat level manipulation in the VRST called for a separate set of analyses to examine both behavioral

performance and psychophysiological response differences in both high- and low-threat zones. Behavioral performance accuracy differences were assessed using the color–word and interference scores with a 2 (threat level) by 3 (Stroop condition) repeated measures ANOVA, which included the color–word score as well as both the simple and complex (relative to VRST) interference scores as the three Stroop conditions. Reaction times were also compared between the high- and low-threat zones and were subjected to the same repeated measures ANOVA.

Across all psychophysiological features that were analyzed, responses during the high-threat zones were higher than responses during the low-threat zones. Differences between IBIs, as recorded by ECG, in the high- versus low-threat zones were significant, with shorter IBIs (i.e., faster heart rate) in the high-threat zones, $t(49) = 3.07$, $p < .01$, $d = 0.87$. SCLs were also higher during high-threat zones, $t(49) = 4.36$, $p < .001$, $d = 1.24$. Participants responded with an increased number of nonspecific responses during the high-threat zones, $t(49) = 3.26$, $p < .01$, $d = 0.93$, and participants had a faster breathing rates when experiencing the high-threat zones as well, $t(49) = 3.82$, $p < .001$, $d = 1.09$ (see Figure 5).

DISCUSSION

Our goal was to compare the VRST with paper-and-pencil (D-KEFS) and computer automated (ANAM) versions of the Stroop on behavioral measures. We also utilized psychophysiological measures to assess the effects of threat on arousal. Finally, we aimed to assess different aspects of supervisory attentional control within the VRST. The primary results were: (a) the typical Stroop effect pattern found in the D-KEFS and ANAM was replicated in the VRST; (b) the complexity condition of the VRST offers potential for assessment of exogenous and endogenous attentional
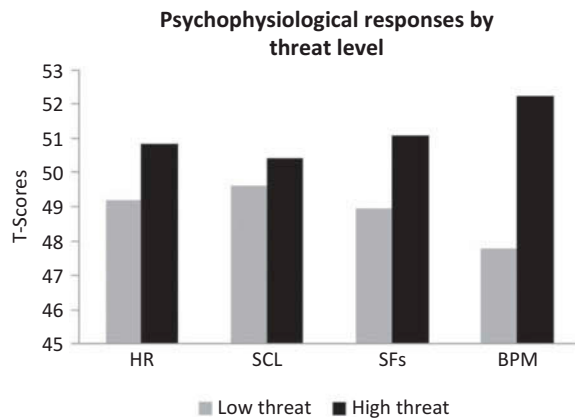
Psychophysiological responses by threat level

Figure 5. Psychophysiological responses shown here as *t*-scores in order to set all response measures to the same scale for display purposes. Interbeat and interbreath intervals were first converted to beats and breaths per minute, respectively, in order to avoid confusion related to lower bars representing faster heart rates or breathing rates. In this figure, HR stands for heart rate, SCL stands for skin conductance level, SFs stands for spontaneous fluctuations, and BPM stands for breaths per minute.

processing; (c) the two single-item presentation formats (i.e., VRST and ANAM) of the Stroop produced similar effects; (d) there was greater psychophysiological arousal in the VRST high-threat zones than in the low-threat zones; and (e) while there was a learning effect present in the VRST across conditions, it was limited to the high-threat zones. No negative side effects were associated with use of the VRST.

Stroop effect pattern found in the D-KEFS, ANAM, and VRST

The typical Stroop effect pattern found in the D-KEFS and ANAM was replicated in the VRST. In all three types of Stroop tests, participants exhibited lower performance scores in the interference condition than in the color–word condition. The Stroop type main effect was the result of higher scores on the D-KEFS. The difference between the VRST and ANAM Stroop conditions was not significant, nor was there a significant interaction between Stroop type and Stroop condition. This difference is expected given the fact that the D-KEFS version uses a multiple-item Stroop task rather than a single-item method found in the ANAM and VRST. The single-item versions are viewed to be preferable given the fact that they provide performance measures (reaction times and errors) for individual stimuli, thus making it possible to eliminate error trials from the reaction time analyses.

Exogenous and endogenous attention in VRST

The complexity condition of the VRST offers potential for assessment of exogenous and endogenous attentional processing. The results support the use of the VRST to assess exogenous and endogenous attentional processing (Posner, 1980). This conclusion is supported by the fact that there was not a significant correlation between the complex interference condition of the VRST and interference/switching found in the D-KEFS. While the D-KEFS interference/switching condition is a Stroop interference plus rule/task switching paradigm, the complexity condition of the VRST more closely reflects attentional selectivity (found in endogenous attention tasks) that can be inferred relative to whether the stimuli are in fixed (expected) or variable (unexpected) spatial locations of the monitor screen (see Posner, 1980). Future studies are needed to assess the impact of these findings. For example, it would be of import to assess the convergent validity of these measures using correlational analyses.

Comparability of two single-item presentation formats (VRST and ANAM) of the Stroop

The comparison of the two single-item presentation formats (i.e., VRST and ANAM) of the Stroop revealed similar result patterns. It is important to note that it took participants longer to respond correctly to VRST stimuli than to ANAM Stroop stimuli. Further, there was a greater difference between interference reaction times and the other two conditions in the VRST than was present in the ANAM Stroop. This is as expected given the fact that the VRST version is a military-relevant virtual environment that presents both high- and low-threat zones, which creates the additional task of trying to ignore potentially distracting and threatening environmental stimuli.

Impact of arousal upon performance

Analyses of the effect of threat level on the color–word and interference scores for both simple and complex interference resulted in a main effect of threat level and condition. A significant interaction resulted from a significant difference in performance only during the complex interference condition, such that the high-threat zones resulted in poorer performance than the low-threat zones.

These findings reflect the perspective that high-information-load tasks may be relatively automatic in controlled circumstances—for example, in low-threat zones with little activity. However, the total available processing capacities may be decreased by other factors such as arousal (e.g., threat zones with a great deal of activity). Maintenance of attention can be under voluntary control but not attentional intensity (Kahneman & Treisman, 1984). Further, these findings appear to reflect Seeley et al.'s (2007) differentiation of "executive control" and "salience" networks. For the VRST, the executive network may be activated during low-threat conditions (Kerns et al., 2004; Ridderinkhof et al., 2004), and the salience network may activate in response to threats found in the high-threat zones (Craig, 2002).

Psychophysiological difference in arousal between low- and high-threat zones

All psychophysiological measures utilized in this research support the notion that the high-threat zones created a greater level of arousal than did low-threat zones. Heart rate increased, which is indicative of defensive responding (Courtney, Dawson, Schell, Iyer, & Parsons, 2010). Skin conductance levels increased, which is consistent with previous research suggesting that increased levels of anxiety or arousal result in increased SCLs and increased number of nonspecific responses (Dawson et al., 2007; Orr & Pitman, 1987). Respiration rate also increased during high-threat zones, which has been shown to be associated with increased arousal and stress (Wientjes, 1992). The virtual environment allows for participants to be placed in a highly threatening scenario, which proffers the ability to examine how participants will respond when stressors similar to those found in real-world situations are present.

Learning effect and processing speed in the VRST across conditions

Assessment of learning effects on processing speed revealed that participants who experienced the low-threat zones first had slower reaction times in the low-threat zone than in the high-threat zone. Although this may seem counterintuitive given the fact that low-threat zones seem to be easier than high-threat zones, it actually comports well with the idea that automaticity develops relative to the degree of practice (or learned associations) on respective tasks. Learned associations during the

first half (low-threat zone) may result in faster reaction time in the second half even though it is a threat zone (see Cohen et al., 1990; MacLeod & Dunbar, 1988). This idea is also reflected in the fact that the reaction times of persons experiencing the low-threat zone first performed lower on simple interference than on the high-threat simple interference, and the same pattern held for the complex interference condition. It is important to note that the opposite was true for participants who experienced the high-threat zone first, as those participants evidenced faster reaction times in the low-threat zones. This reflects findings of practice-related decline in Stroop interference reaction time as color naming becomes increasingly automatic, and the development of an automatic reading suppression response occurs (Dulaney & Rogers, 1994; Lemay et al., 2004).

Learning effect and accuracy in the VRST across conditions

Learning effects were also apparent in participant accuracy performance, though the effects were more selective than those found in reaction time analyses. In participants that experienced the low-threat zones first, there was no main effect of threat level, nor was there an interaction between threat level and Stroop condition. These participants did display the expected Stroop condition main effect, with lower scores on the interference tasks than on the color–word tasks. Contrariwise, order effects did appear to affect participants who experienced the high-threat zones first. Participants in this group responded less accurately in the high-threat zones than in the low-threat zones. It is perhaps the case that experiencing the high-threat zone first limited learning to the extent that the practice effect did not occur when moving to the low-threat zone. This comports well with the view that high-information-load tasks might be relatively automatic in controlled circumstances (e.g., low-threat zones with little activity; Seeley et al., 2007). The total available processing capacities may be decreased by other factors such as arousal (e.g., threat zones with a great deal of activity). These findings seem to support an arousal model over a learned association view of automaticity. Perhaps the difference between VRST accuracy and reaction times can be used in future studies to differentiate the potential differences found in automaticity relative to arousal models and learned association views. As such, performance asymmetries observed in the Stroop task can be accounted for by differences in practice when the learning condition

occurs during low threat, but not when the threat is high.

Limitations and projections for future studies

Our findings should be understood in the context of some limitations. First, although the current study counterbalances order of presentation and threat level, future studies are needed to assess the mediating effects of psychophysiological difference in arousal between low- and high-threat zones. Future studies should use mediator analyses to assess the current pattern of results as they suggest possible interactions with arousal levels, akin to conflict adaptation effects. These, if confirmed, would be interesting findings for research on interference control. Further, the high-ambient distractors versus low-ambient distractors could be separated in future studies to include a high-ambient-distractor–low-threat condition in addition to the current high-ambient-distractor–high-threat and low-ambient-distractor–low-threat conditions.

Second, recent research indicates that healthy, undergraduate research subjects in a "neuropsychological experiment" may put forth suboptimal effort (An, Zakzanis, & Joordens, 2012). For the VRST, we examine performance using scoring algorithms to screen for outliers and assess data integrity (Wu et al., 2010, 2013). It is important to note that the data are not free of potential confounds due to poor effort. Screening was done at the time of testing and also during the data analysis process to eliminate obvious poor effort, but data may still contain some individuals who provided less than optimal effort. This situation is true for most normative neuropsychological data and particularly true for computerized testing given the reduced interaction with the examiner. Future studies should make attempts at including well-standardized effort tests with other cohorts (healthy and clinical) to assess the sensitivity and specificity of the VRST. Furthermore, these findings are based on a relatively small sample size. As a necessary next step, the reliability and validity of the VRST needs to be established using a larger sample of participants. This will ensure that the current findings are not an anomaly due to sample size. Although our findings need to be replicated to assess potential impact of sample size, effect size analysis revealed strong effects for all analyses, which reduces the likelihood of anomalous results relative to sample size. Additionally, the ability of the VRST to accurately differentiate learning and arousal factors related to automatic and controlled attentional

processing needs to be based on carefully established critical values using specific cutoff points for classifying (a) a positive (learning effect likely) or negative (learning effect unlikely) finding; and (b) a positive (arousal effect likely) or negative (arousal effect unlikely) finding.

Whilst the VRST needs to be further validated, current findings provide data regarding the potential of the VRST for measurement of supervisory attentional processing. The VRST replicated and extended the Stroop effects found in the D-KEFS and flatscreen ANAM versions of the Stroop. Furthermore, the increased difficulty of the VRST may be useful in creating a greater amount of variability in performance levels of participants for assessment of questions of automatic and controlled processing. Psychophysiological data gleaned from the study support the notion that the high-threat zones created a greater level of arousal. We propose that this added threat manipulation creates greater ecological validity for this cognitive task. The VRST in a threat zone presents a very different and more stressful situation than responding to stimuli that are printed on a piece of paper or are presented on a blank computer screen.

REFERENCES

An, K. Y., Zakzanis, K. K., & Joordens, S. (2012). Conducting research with non-clinical healthy undergraduates: Does effort play a role in neuropsychological test performance? *Archives of Clinical Neuropsychology*, *27*, 849–857.

Automated Neuropsychological Assessment Metrics (Version 4) [Computer software]. (2007). Norman, OK: C-SHOP.

Baddeley, A. D. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology*, *49*, 5–28.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–90). San Diego, CA: Academic Press.

Bugg, J. M., Jacoby, L. L., & Toth, J. P. (2008). Multiple levels of control in the Stroop task. *Memory & Cognition*, *36*, 1484–1494.

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review*, *13*, 181–197.

CogSport. (1999). Parkville: CogState.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, *97*, 332–361.

Courtney, C. G., Dawson, M. E., Schell, A. M., Iyer, A., & Parsons, T. D. (2010). Better than the real thing: Eliciting fear with moving and static computer-generated stimuli. *International Journal of Psychophysiology, 78*, 107–114.

Craig, A. D. (2002). How do you feel? Interoception: The sense of the physiological condition of the body. *Nature Reviews Neuroscience, 3*, 655–666.

Critchley, H. D. (2005). Neural mechanisms of autonomic, affective, and cognitive integration. *Journal of Comparative Neurology, 493*, 154–166.

Davidson, D. J., Zacks, R. T., & Williams, C. (2003). Stroop interference, practice, and aging. *Aging, Neuropsychology and Cognition, 10*, 85–98.

Dawson, M. E., Schell, A. M., & Filion, D. L. (2007). The electrodermal system. In T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (3rd ed., pp. 159–181). New York, NY: Cambridge University Press.

Delis, D. C., Kaplan, E., Kramer, J. H., & Ober, B. A. (1997). *Delis–Kaplan Executive Function Scale*. San Antonio, TX: The Psychological Corporation.

Dulaney, C. L., & Rogers, W. A. (1994). Mechanisms underlying reduction in Stroop interference with practice for young and old adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 470–484.

Erlanger, D. M., Feldman, D. J., & Kutner, K. (1999). *Concussion resolution index*. New York, NY: HeadMinder.

Fisk, A. D., & Schneider, W. (1981). Control and automatic processing during tasks requiring sustained attention: A new approach to vigilance. *Human Factors, 23*, 737–750.

Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (1998). *Cognitive neuroscience: The biology of the mind*. New York, NY: Norton.

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

Johnson, D. R., Vincent, A. S., Johnson, A. E., Gilliland, K., & Schlegel, R. E. (2008). Reliability and construct validity of the Automated Neuropsychological Assessment Metrics (ANAM) mood scale. *Archives of Clinical Neuropsychology, 1*, 73–85.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman, D. R. Davies, & J. Beatty (Eds.), *Varieties of attention* (pp. 29–61). New York, NY: Academic Press.

Kennedy, R. S., Fowlkes, J. E., Berbaum, K. S., & Lilienthal, M. G. (1992). Use of a motion sickness history questionnaire for prediction of simulator sickness. *Aviation, Space, and Environmental Medicine, 63*, 588–593.

Kerns, J. G., Cohen, J. D., MacDonald, A. W., III, Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science, 303*, 1023–1026.

Lemay, S., Bedard, M. A., Roulea, I., & Tremblay, P. L. G. (2004). Practice effect and test–retest reliability of attentional and executive tests in middle-aged to elderly subjects. *The Clinical Neuropsychologist, 18*, 284–302.

Liotti, M., Woldorff, M. G., Perez, R., & Mayberg, H. S. (2000). An ERP study of the temporal course of the Stroop color–word interference effect. *Neuropsychologia, 38*, 701–711.

Lovell, M. R., Collins, M. W., Podell, K., Powell, J., & Maroon, J. (2000). *ImPACT: Immediate Post-Concussion Assessment and Cognitive Testing*. Pittsburgh, PA: NeuroHealth Systems, LLC.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*, 163–203.

MacLeod, C. M. (1992). The Stroop task: The "gold standard" of attentional measures. *Journal of Experimental Psychology: General, 121*, 12–14.

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 126–135.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research and theory* (Vol. 4, pp. 1–18). New York, NY: Plenum.

Orr, S. P., & Pitman, R. K. (1987). Electrodermal psychophysiology of anxiety disorder: Orienting response and spontaneous fluctuations. *Biological Psychiatry, 22*, 653–656.

Pardo, J. V., Pardo, P. J., Janer, K. W., & Raichle, M. E. (1990). The anterior cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm. *Proceedings of the National Academy of Sciences, 87*, 256–259.

Parsons, T. D., Courtney, C., Arizmendi, B., & Dawson, M. (2011). Virtual Reality Stroop Task for neurocognitive assessment. *Studies in Health Technology & Informatics, 143*, 433–439.

Parsons, T. D., & Reinebold, J. (2012). Adaptive virtual environments for neuropsychological assessment in serious games. *IEEE Transactions on Consumer Electronics, 58*, 197–204.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*, 3–25.

Reeves, D. L., Winter, K. P., Bleiberg, J., & Lane, R. L. (2007). ANAM Genogram: Historical perspectives, description, and current endeavors. *Archives of Clinical Neuropsychology, 22S*, S15–S37.

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science, 306*, 443–447.

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

Sarter, M., Gehring, W. J., & Kozak, R. (2006). More attention must be paid: The neurobiology of attentional effort. *Brain Research Reviews, 51*, 145–160.

Schatz, P., & Browndyke, J. (2002). Applications of computer-based neuropsychological assessment. *Journal of Head Trauma Rehabilitation, 17*, 395–410.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information-processing: I. Detection, search, and attention. *Psychological Review, 84*, 1–66.

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G., Kenna, H., . . . Greicius, M. D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *Journal of Neuroscience, 27*, 2349–2356.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*, 127–190.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662.

Uttl, B., & Graf, P. (1997). Color–Word Stroop test performance across the adult life span. *Journal of Clinical and Experimental Neuropsychology*, *19*, 405–420.

Wientjes, C. J. E. (1992). Respiration in psychophysiology: Methods and applications. *Biological Psychology*, *34*, 179–203.

Wu, D., Courtney, C., Lance, B., Narayanan, S. S., Dawson, M., Oie, K., & Parsons, T. D. (2010). Optimal arousal identification and classification for affective computing: Virtual reality Stroop task. *IEEE Transactions on Affective Computing*, *1*, 109–118.

Wu, D., Lance, B., & Parsons, T. D. (2013). Collaborative filtering for brain-computer interaction using transfer learning and active class selection. *PLoS ONE*, 1–18.