Check for updates

Article

Assessment
2019, Vol. 26(1) 125–143
© The Author(s) 2017
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1073191116689820
journals.sagepub.com/home/asm
SAGE

Development, Administration, and Structural Validity of a Brief, Computerized Neurocognitive Battery: Results From the Army Study to Assess Risk and Resilience in Servicemembers

Tyler M. Moore[1], Ruben C. Gur[1,2], Michael L. Thomas[3], Gregory G. Brown[3,4], Matthew K. Nock[5], Adam P. Savitt[1], John G. Keilp[6,7], Steven Heeringa[8], Robert J. Ursano[9], Murray B. Stein[3], and On behalf of the Army STARRS Collaborators[*]

Abstract

The Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS) is a research project aimed at identifying risk and protective factors for suicide and related mental health outcomes among Army Soldiers. The New Soldier Study component of Army STARRS included the assessment of a range of cognitive- and emotion-processing domains linked to brain systems related to suicidal behavior including posttraumatic stress disorder, mood disorders, substance use disorders, and impulsivity. We describe the design and application of the Army STARRS neurocognitive test battery to a sample of 56,824 soldiers. We investigate its structural and concurrent validity through factor analysis and correlation of scores with demographics. We conclude that, in addition to being composed of previously well-validated measures, the Army STARRS neurocognitive battery *as a whole* demonstrates good psychometric properties. Correlations of scores with age and sex differences mostly replicate previously published findings, highlighting moderate to large effect sizes even within this restricted age range. Factor structures of scores conform to theoretical expectations. This neurocognitive battery provides a brief, valid measurement of neurocognition that may be helpful in predicting mental health and military performance. These measures can be integrated with neuroimaging to offer a powerful tool for assessing neurocognition in Servicemembers.

Keywords

Army STARRS, Penn Computerized Neurocognitive Battery, neurocognitive assessment, posttraumatic stress disorder, psychometrics

The Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS, 2012) is a research project aiming primarily to investigate the recent increase in suicide rates among Army Soldiers (Kessler et al., 2013; Ursano et al., 2014). The study includes retrieval of historical data, prospective data collection, and biological sample collection across a number of substudies. One of these substudies, the New Soldier Study (NSS), involves the administration of computerized psychiatric symptom inventories, personality assessments, and neurocognitive tests to Army Soldiers at the onset of basic training. The final sample for the NSS comprises over 50,000 participants. Note that the NSS is only one of several studies within the Army STARRS project, and is the focus of the present study due to the assessment goals (neurocognitive and clinical) that motivated it.

[1]University of Pennsylvania, Philadelphia, PA, USA
[2]Philadelphia Veterans Administration Medical Center, Philadelphia, PA, USA
[3]University of California, San Diego, La Jolla, CA, USA
[4]VA San Diego Healthcare System, San Diego, CA, USA
[5]Harvard University, Cambridge, MA, USA
[6]New York State Psychiatric Institute, New York, NY, USA
[7]Columbia University, New York, NY, USA
[8]University of Michigan, Ann Arbor, MI, USA
[9]Uniformed Services University of the Health Sciences, Bethesda, MD, USA
*The complete list of Army STARRS Collaborators in contributing authors' section at the end of the article.

Corresponding Author:
Tyler M. Moore, Brain Behavior Laboratory, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA.
Email: tymoore@upenn.edu

The neurocognitive tests selected for inclusion in the Army STARRS battery are designed to assess a broad range of cognitive and emotion processing domains that have been related to disorders and problems of interest in Army STARRS, including suicidal behavior, posttraumatic stress disorder (PTSD), mood disorders, substance and alcohol use disorders, and impulsive behavior. Most tests in this battery are from the Penn Computerized Neurocognitive Battery (CNB; Gur et al., 2001; Gur et al., 2010), and were included because they are based on functional neuroimaging (Gur, Erwin, & Gur, 1992; Gur et al., 2001; Roalf et al., 2014), normed on large samples (Gur et al., 2012; Gur et al., 2014), and adaptable for minimally proctored group administration. Other tasks chosen for this battery—specifically, the Go/No-Go (GNG) task and the Emotional Stroop (ESTROOP) task, were not part of the original CNB but were added to augment the Army STARRS battery by additional established suicide-related behavioral measures of impulsivity (Keilp, Sackeim, & Mann, 2005; Nock et al., 2010). That is, tests selected from the CNB were chosen because of the CNB's well-established validity and history of use; however, the coverage of neurocognitive domains by the CNB is not without gaps, and the ESTROOP and GNG were selected to fill those gaps.

There is evidence that neurocognitive correlates of suicidal behavior include primarily deficits in executive control related to frontal lobe functioning, such as problems with abstraction and mental flexibility, attention, impulse control, and decision making (reviewed in Jollant, Lawrence, Olié, Guillaume, & Courtet, 2011; Richard-Devantoy, Orsat, Dumais, Turecki, & Jollant, 2014). Related mental health problems such as PTSD and traumatic brain injury (TBI) also have been associated with deficits in executive functions that include sustained attention, working memory, but also episodic memory (e.g., Brenner et al., 2010; Leskin & White, 2007; Uddo, Vasterling, Brailey, & Sutker, 1993; Vasterling et al., 2002; review in Pitman, Shalev, & Orr, 2000). Notably, alcohol and substance abuse exacerbate these deficits specifically in the areas of verbal memory, attention, and processing speed performance (Samuelson et al., 2006), and face memory (Samuelson et al., 2009). These domains are also implicated in depression (reviews in Kurtz & Gerraty, 2009; McClintock, Husain, Greer, & Cullum, 2010). Deficits in affect processing are more specifically linked to depression (e.g., Gur, Erwin, & Gur, 1992; Naranjo et al., 2011), as well as to proneness to aggression (Weiss et al., 2006).

Our aims in selecting a neurocognitive battery for Army STARRS were as follows: (a) to sample behavioral measures that are sensitive to the integrity of frontotemporal brain systems, which are implicated in conditions that enhance proneness to suicide and (b) measure both cognitive and emotion-processing (social cognition) domains of functioning that have been documented in these conditions and are relevant to vocational and social adjustment. Because of time constrains (two sessions of about 20 minutes each were allotted by the protocol), we had to forego administration of other tests such as measures of verbal and spatial reasoning, additional episodic memory and social cognition domains, and motor speed. Data showing psychometric properties of individual tests in initial subsamples have been reported (Thomas et al., 2013; Thomas et al., 2015).

With these aims in mind, the following tests were selected for inclusion in the battery: Penn Conditional Exclusion Test (PCET), Penn Continuous Performance Test (PCPT), Short Letter-N-Back (SLNB), GNG, Penn Face Memory Test (PFMT), Penn Emotion Identification Test (ER40), and ESTROOP-Style Test. The PCET was chosen because deficits in abstraction, problem solving, and mental flexibility have been associated with suicidal behavior (Keilp et al., 2001; Neuringer, 1964; Schotte & Clum, 1987; Schotte, Cools, & Payvar, 1990; see also LeGris & van Reekum, 2006, for a review). Deficits in mental flexibility, abstract reasoning, and problem solving are also associated with several psychiatric conditions that confer high risk for suicide, including borderline personality disorder (Fertuck, Lenzenweger, Clarkin, Hoermann, & Stanley, 2006), PTSD (Danckwerts & Leathem, 2003), major depression (Mialet, Pope, & Yurgelun-Todd, 1996; Paelecke-Habermann, Pohl, & Leplow, 2005), and alcohol abuse (Noël, Bechara, Dan, Hanak, & Verbanck, 2007), as well as schizophrenia and spectrum disorders (Saykin et al., 1991). Impairments in executive functioning may be of particular concern for military personnel who suffer the combined effects of mild TBI and PTSD, as deficits in these areas of cognition are seen with high frequency.

The PCPT was chosen because lapses in executive control of attention and vigilance contribute to impairments in declarative memory (Takashima et al., 2006) and complex problem solving. Many psychiatric conditions that are associated with attentional deficits (including PTSD, major depressive disorder, bipolar disorder, and psychosis) are known to contribute to suicidal behavior (Arsenault-Lapierre, Kim, & Turecki, 2004; Keilp, Gorlyn, Oquendo, Burke, & Mann, 2008).

The SLNB was chosen because the ability to actively maintain and refresh goal-related information is a major executive domain (Baddeley & Della Sala, 1996) that relates to dorsolateral prefrontal structures in healthy people (Ragland et al., 1997, Ragland et al., 2002) and is sensitive to effects of TBI (Vallat-Azouvi et al., 2007), depressive disorders (Christopher & MacDonald, 2005), and PTSD (Shaw et al., 2009).

The GNG was chosen because poor GNG performance has been found in attention deficit disorder (Barkley, 1997; Durston et al., 2007), those at genetic risk for attention deficit disorder (Durston, Mulder, Casey, Ziermans, & van

Engeland, 2006; Wood et al., 2011), drug abusers (Verdejo-García, Bechara, Recknor, & Pérez-García, 2006), bipolar disorder with suicidal behavior (Harkavy-Friedman et al., 2006), subjects who had experienced childhood abuse (Navalta, Polcari, Webster, Boghossian, & Teicher, 2006), subjects undergoing tryptophan depletion (LeMarquand et al., 1998; Robinson & Sahakian, 2009), and after administration of alcohol (Ostling & Fillmore, 2010). Poor performance has also been associated with self-ratings of impulsiveness in healthy volunteers (Keilp et al., 2005), and with more violent suicidal behavior. The GNG has been used extensively in electroencephalography and functional imaging studies to produce reliable activation of ventral prefrontal and striatal brain regions in both healthy people (Durston, Thomas, Worden, Yang, & Casey, 2002; Horn, Dolan, Elliott, Deakin, & Woodruff, 2003) and patients with attention deficit hyperactivity disorder (Casey et al., 1997).

The PFMT (Gur et al., 1997) was chosen because the memory system is to some extent domain specific, with greater left hemispheric involvement in verbal memory and greater right hemispheric involvement in face and shape memory. Memory for faces is related to emotional processing, and may be sensitive to the effects of PTSD and depressive disorders (Gur, Erwin, & Gur, 1992; Naranjo et al., 2011). In short, detection of memory deficits is of obvious importance because they not only could impair work performance but, in the case of deficits in memory for faces, may also indicate damage to limbic structures involved in affect regulation.

The ER40 was chosen because emotion recognition is a critical aspect of social information processing and social problem solving. Difficulties in decoding facial affect lead to misjudgment of intentions of peers or foes, and can fuel social isolation, alienation, and hostility (e.g., Weiss et al., 2006). Various psychiatric conditions modify emotional information processing; for example, individuals with PTSD have a heightened sensitivity to fearful faces (Masten et al., 2008), whereas individuals with borderline personality disorder are quick to categorize emotional expressions (Fertuck et al., 2009). Impairment in affect processing is also linked to depression (e.g., Gur, Erwin, Gur, Zwil, et al., 1992; Gur, Erwin, & Gur, 1992; Naranjo et al., 2011), proneness to aggression (Weiss et al., 2006), as well as to schizophrenia (Heimberg, Gur, Erwin, Shtasel, & Gur, 1992; Kohler et al., 2003; Kohler, Hanson, & March, 2013).

The ESTROOP was chosen because Stroop-style tasks using pathology-specific words have demonstrated a relationship between psychopathology and attentional bias in depression (Williams, Mathews, & MacLeod, 1996), anxiety (Foa, Feske, Murdock, Kozak, & McCarthy, 1991; McNally, Kaspi, Riemann, & Zeitlin, 1990; Teachman, Smith-Janik, & Saporito, 2007), PTSD (Kaspi, McNally, & Amir, 1995), and substance use (Cox, Fadardi, & Pothos, 2006). Suicide-specific ESTROOP tasks have found significant interference in suicide-specific trials in recent suicide attempters (Williams & Broadbent, 1986) over and above bias to generally negative or neutral words (Becker, Strohbach, & Rinck, 1999). Recent work has also demonstrated the utility of suicide-specific ESTROOP scores as behavior markers for future suicide attempts (Cha, Najmi, Park, Finn, & Nock, 2010).

The present analysis examined the psychometric structure of the neurocognitive battery in an effort to derive useful indices of performance that can help link clinical parameters to neuroimaging and genomic measures in a translational context. We took advantage of the unusually large sample to obtain estimates of factorial structure on half the sample, and replicated in the other half.

Method

Participants and Administration

Army soldiers were recruited to volunteer without compensation for the Army STARRS NSS at the start of basic training. Due to the potential danger of soldiers feeling compelled to participate due to the fear of disapproval (or worse) from their commanding officers, extra emphasis was given to the fact that this was a voluntary procedure. Those who did not participate were explicitly offered the opportunity for recreational activities of their choosing (as opposed to, e.g., fitness training or other less desirable activities). All participants described below were recruited specifically for the NSS.

The current sample comprises 56,824 participants (82.3% male) from three Army bases in the United States tested between February of 2011 and November of 2012. Mean age was 21.0 (*SD* = 3.6) with only 2% age 32+, and racial breakdown was as follows: 69% White; 20% Black; 2.8% Asian; 1.4% American Indian; 0.8% Pacific Islander; and 5.8% other. All soldiers were asked to provide informed, written consent prior to participation in research. Army commanders provided sufficient time to complete all surveys and tests, which were administered in a group format using laptop computers. Research proctors monitored the testing environment and assisted with questions and technical difficulties. Surveys and tests were administered in a fixed order in 90-minute sessions over 2 days of testing. The neurocognitive part of the computerized assessment was administered in the last 20 minutes of the 90-minute assessment session.

Tests Administered

Penn Conditional Exclusion Test. The PCET (Kurtz, Ragland, Moberg, & Gur, 2004) is designed to test a participant's ability to learn rules and principles, recognize unexpected changes in those rules, and adjust accordingly. It is based on the “Odd Man Out” paradigm, participants are asked to determine which particular object does not belong to a group of other objects. In the case of the PCET version administered in the Army STARRS battery, the objects vary on three characteristics: size, shape, and the thickness of the lines composing them. For example, if three of the objects are stars, and one of the objects is a square, it might be the case that the square is the “odd man out” (and therefore the correct answer), because it is the only nonstar. On the other hand, if the square and two of the stars are large and one of the stars is small, the small star might be the “odd man out,” because it is the only small object.

On each trial, participants select the object they believe to be the “odd man out,” and are immediately told if they were correct or incorrect. Participants are given 48 trials to learn which characteristic (size, shape, or line thickness) is determining the “odd man out,” and then must get 10 consecutive correct answers. After those correct answers, the characteristic is changed—for example, the participant might have correctly learned that size is the important characteristic, but after the 10 consecutive correct answers selecting the odd size, the important characteristic will change (perhaps to shape). The participant must then recognize that the rule has changed, determine what the new rule is, and again respond with 10 consecutive correct answers. Finally, after those 10 correct answers, the rule is changed a third time, and the participant must again determine the new rule (and respond accordingly).

The PCET is scored based on a composite of total correct responses and the number of rules/principles the participant learned. Specifically, a performance composite score is calculated by multiplying the number of principles learned (plus 1 to accommodate those who do not learn a single rule) by proportion of correct responses (i.e., correct responses/total responses).

Penn Continuous Performance Test. The PCPT (Kurtz, Ragland, Bilker, Gur, & Gur, 2001) is a test of vigilance and visual attention. Participants are shown a series of configurations of red seven-segment displays (as on a digital clock display), and asked to press a space bar when the stimulus is a number (first half) or letter (second half). Each trial lasts 1 second, during which the stimulus is displayed for 300 ms followed by a blank screen displayed for 700 ms. Total test time is 3 minutes (1.5 for numbers and 1.5 for letters).

Short Letter-N-Back. In the SLNB, participants are asked to pay attention to letters that flash on the computer screen one at a time, and to press the spacebar according to a specified principle. In the Army STARRS implementation, the participant was instructed to press the spacebar whenever the letter on the screen is the same as the one before the previous letter (2-back). In all trials, the participant has 2.5 seconds to press the spacebar, and is given a practice session before beginning. This task is scored based on the total number of true positives.

Go/No-Go. The GNG task is a measure of impulse control that requires subjects to respond to either a single designated target or a series of targets, and to inhibit responding to a particular low-frequency nontarget. The goal of the task is to induce subjects to develop a tendency to respond, and then to interrupt that tendency with an intermittent nontarget. In their simplest form, GNG tasks use a series of letters or symbols as targets, and a single letter or figure as a nontarget. In the Army STARRS GNG task, participants see a series of Xs and Ys quickly displayed at different positions on the screen. Each stimulus is shown for 300 ms, followed by a uniform black screen for 900 ms. Participants are instructed to respond (press the spacebar) if and only if an X appears in the upper half of the screen. Thus, participants must inhibit the impulse to respond to both Xs in the lower half of the screen and Ys generally.

Penn Face Memory Test. The PFMT presents examinees 20 faces that they will be asked to identify later. Faces are shown in succession for an encoding period of 5 seconds each. After this initial learning period, examinees are immediately shown a series of 40 faces—20 targets and 20 distractors—and are asked to decide whether they have seen each face before by choosing 1 of 4 ordered categorical response options: “definitely yes,” “probably yes,” “probably no,” or “definitely no.” Stimuli consist of black-and-white photographs of faces presented on a black background. All faces were rated as having neutral expressions and were balanced for gender and age (Gur et al., 1993; Gur et al., 2001). Responses and response times are recorded during test administration; however, there are no time limits during recognition testing or explicit instructions to work quickly.

Penn Emotion Identification Test. The ER40 (Carter et al., 2009; Erwin et al., 1992; Habel et al., 2000; Kohler et al., 2003; Mathersul et al., 2009) measures the ability of an individual to recognize the specific emotion being expressed by a poser. Participants are shown a series of 40 faces, and asked to choose (among five options) which emotion the person in the photograph is expressing. The five options are Happy, Sad, Anger, Fear, or No Emotion. There are four male and four female faces for each emotion, for a total of 40 faces (8 actor photos × 5 emotions = 40).

Emotion Stroop-Style Test. The traditional Stroop (1935) paradigm measures the degree to which semantic processing interferes with color identification. In the classic case, color words (e.g., GREEN, RED, BLUE) are displayed in potentially incongruous font colors. Participants are required to name the font colors while ignoring the semantic content of the color words. Response latencies on incongruous words (interference trials) are thought to capture an effort to inhibit a prepotent bias to ignore font color when reading. The ESTROOP adds another layer to this paradigm by displaying emotionally valenced words in addition to color words. These valenced words are either generally negative (e.g., *alone*, *rejected, stupid*) or specific to suicide (e.g., *suicide*, *dead*, *funeral*) and have been previously used in other suicide-related behavioral measures (Nock et al., 2010). The ESTROOP measures interference due to attentional bias by subtracting response latencies to neutral words from those for negative or suicide-specific words.

Data Analysis

Data Cleaning. Flags were assigned to test sessions with response patterns consistent with hardware/software malfunction and/or subject inattention, misunderstanding, or noncompliance. First, histograms for all measures (accuracy and speed) were examined visually for impossible results (e.g., negative response times), suspicious patterns (e.g., unusually high frequency of an exact millisecond resolution response time), or suspicious distributions (e.g., bimodal), any of which would suggest possible software/hardware failure. No such problems were found. Next, response patterns for individual tests were examined for subject-related problems (e.g., noncompliance); that is, thresholds for whether to flag a test session varied by test. For example, sessions for the Continuous Performance Test (CPT; a rapid task) were flagged if there were 10 consecutive responses (presses) or 20 consecutive nonresponses. Sessions for the ER40 (a deliberative task) were flagged if the same emotion was selected ≥7 times in a row and/or if there was at least 1 response time ≤250 ms. Such rules were established for each of the seven tests based on post hoc examination of the data, such that a flagging rule was applied only if it described a situation where a participant-related problem almost certainly existed. These flagged test sessions were excluded from analysis unless otherwise indicated. Note that only the flagged test was excluded, not the entire battery; thus, it was possible for some participants to have data for only some of the individual tests. Thus, missing data were handled using pairwise deletion in all analyses described below. However, as an added precaution, we also reran all analyses using more stringent quality assurance criteria—specifically, participants with any missing data were removed (listwise), and outliers >3 standard deviations from the mean on each score were removed. Supplementary Table S1 shows the percentages of scores on each test that fell outside this range of ±3 standard deviations. All results using the more stringent criteria are shown in the supplement as indicated below in the Results section for each analysis.

Concurrent Validity. To assess the concurrent validity of the individual tests composing the Army STARRS battery, we examined gender differences within each test using t tests, and plotted each test score's (accuracy and speed) relationship with age. The tests' relationships with age were tested statistically via robust linear regression (Maronna & Yohai, 2000) including both age and age squared to account for nonlinearity. Robust linear regression was used due to our suspicion that the assumption of homoscedasticity would be violated. This suspicion was tested using the Breusch and Pagan (1979) method, which tests the likelihood (given the sample size) of the linear relationship between the independent variables and regression residuals; the test statistic is distributed as a χ^2, and a statistically significant value indicates a violation of the homoscedasticity assumption. All analyses were performed using the *stats*, *car* (Fox & Weisberg, 2011), and *robustbase* (Rousseeuw et al., 2016) packages in R (v3.2.0; R Core Team, 2016), and plots were created using SPSS (version 21). Additionally, a large number of participants (N = 3,949) reported having ever experienced a TBI, here defined as having resulted in, (a) a perforated eardrum, (b) loss of consciousness greater than 30 minutes, or both. Thus, because TBI is an obvious potential confounder, the above analyses were performed again after excluding participants who reported TBI.

Factor Analysis. To assess the latent structure of the Army STARRS battery, we first estimated unidimensional and two-factor exploratory factor solutions (EFAs). Note that a major step of most EFAs—that is, judging, empirically, or theoretically, the appropriate number of factors to extract—was not necessary in this case due to the small number of variables (seven). Extracting three or more factors would guarantee that at least one of the factors was indicated by fewer than three variables, making those factors not properly identified. We thus chose to estimate only unidimensional and two-factor solutions.

Next, based on the exploratory results and the theory that motivated test selection, we estimated a confirmatory bifactor model of the efficiency scores with two specific factors and one general factor. Bifactor modeling is a way to estimate the contribution of a test to an overall dimension (performance in this case) after controlling for its specific factor, and vice versa. Bifactor models are similar to higher order models (in which one general factor comprises the lower order factors, which themselves comprise the individual tests), except that, in a bifactor model, there are direct effects of the general factor on the individual tests. For more information on strengths and weaknesses of bifactor modeling, see Reise (2012) and Reise, Moore, and Haviland (2010). Note, however, that because of the brevity of the battery, a higher order model was not feasible (without mathematical constraints), as the higher order factor needs at least three lower order factors to be identified.

Table 1. Gender Effects on the Army STARRS Neurocognitive Test Battery.

	Gender differences						
	Male (M)			Female (F)			Difference
Test	*M*	*SD*	*N*	*M*	*SD*	*N*	(M − F)
ESTROOP_Accuracy	−0.02	1.00	46,788	0.15	0.90	10,036	−0.17*
CPT_Accuracy	0.00	1.00	43,282	0.04	0.97	9,321	−0.04
ER40_Accuracy	−0.02	1.01	43,893	0.10	0.92	9,544	−0.12*
GNG_Accuracy	0.00	1.00	43,283	0.04	0.98	9,216	−0.04*
SLNB_Accuracy	0.02	1.00	35,744	−0.06	1.00	7,231	0.08*
PCET_Accuracy	0.04	0.99	37,293	−0.17	1.03	7,538	0.21*
PFMT_Accuracy	−0.01	1.00	40,281	0.06	1.00	8,915	−0.07*
ESTROOP_RT	−0.03	1.00	46,788	0.09	0.97	10,036	−0.12*
CPT_RT	0.01	1.01	43,282	−0.05	0.90	9,321	0.06*
ER40_RT	0.05	1.01	43,893	−0.27	0.87	9,544	0.32*
GNG_RT	−0.06	0.97	43,283	0.29	1.08	9,216	−0.35*
SLNB_RT	−0.01	0.99	35,744	0.02	1.04	7,231	−0.03
PCET_RT	0.01	1.01	37,293	−0.06	0.94	7,538	0.07*
PFMT_RT	0.01	1.01	40,281	−0.03	0.96	8,915	0.04
ESTROOP_Efficiency	0.00	1.00	46,788	0.05	0.94	10,036	−0.05*
CPT_Efficiency	0.00	1.00	43,282	0.05	0.94	9,321	−0.05*
ER40_Efficiency	−0.05	1.00	43,893	0.25	0.91	9,544	−0.30*
GNG_Efficiency	0.04	0.98	43,283	−0.18	1.04	9,216	0.22*
SLNB_Efficiency	0.02	0.99	35,744	−0.06	1.03	7,231	0.08*
PCET_Efficiency	0.02	1.00	37,293	−0.07	0.99	7,538	0.09*
PFMT_Efficiency	−0.01	1.00	40,281	0.07	0.98	8,915	−0.08*

Note. RT = response time; *SD* = standard deviation; CPT = Continuous Performance Task; ER40 = Emotion Recognition; GNG = Go/No-Go; SLNB = Short Letter-N-Back; PCET = Penn Conditional Exclusion Test; PFMT = Penn Face Memory Test; ESTROOP = Emotional Stroop Task. Accuracy, response times, and efficiency scores are in z score units such that a difference of 0.05 indicates a 0.05 standard deviation difference.
*$p < .001$.

Here, we have only two lower order factors. Also, note that we do not compare the bifactor model with a standard correlated-traits model because, even if the latter had better fit indices and lower (better) information criteria, it would not be the model of choice because one of the purposes of the confirmatory model is to generate one overall score, something not possible with a correlated-traits model. When subfactor scores are desired, we use and recommend the two-factor exploratory model shown below.

All EFAs were performed using least squares extraction and oblimin rotation in the *psych* package (Revelle, 2013) in R, and the confirmatory model was estimated using the robust maximum likelihood estimator in M*plus* (v6; Muthén & Muthén, 1998-2013). Also, due to a minor estimation problem for the CFA, the residual variance of ER40 had to be constrained to be >0. Note that this value was not fixed (specified in the model), but was simply constrained using the MODEL CONSTRAINT command in M*plus*, which removes a problematic portion of the maximum likelihood estimation search space.

Though there are a number of ways to evaluate the fit of a model (and many corresponding "thresholds" for acceptable fit), we follow the recommendations of Hu and Bentler (1998, 1999) throughout this article. Missing data were handled using pairwise deletion (but see below), and the confirmatory model was identified by setting one loading to 1.0 per factor. Additionally, to achieve some level of cross-validation and avoid sample-specific solutions, the total sample was randomly split into an exploratory group ($N = 26{,}050$) and confirmatory group ($N = 25{,}000$). All EFA results reported below are based on the exploratory group, and the CFA on the confirmatory group. Note, however, that one of the potential hazards of random-split cross-validation is that the two groups, by chance, might differ in some consequential way. Here, the most important way they might differ is in the variances of the test scores (accuracy, response time, and efficiency). We therefore tested for equal variance between groups using F tests, and compared the groups on age and sex.

Results

Sex Differences

Table 1 shows the results of the analyses examining sex differences. Because scores were z scores standardized to the global mean, values in the rightmost column of Table 1

Table 2. Associations of Age With Accuracy and Speed on Seven Neurocognitive Tests (Full Sample).

	Age (standardized)		Age squared	
Score	β	*p*	β	*p*
ESTROOP_Accuracy	**.118**	**<.001**	**−.015**	**<.001**
CPT_Accuracy	**.190**	**<.001**	**−.030**	**<.001**
ER40_Accuracy	−.002	.702	−.003	.052
GNG_Accuracy	**.125**	**<.001**	**−.014**	**<.001**
SLNB_Accuracy	**.043**	**<.001**	**−.010**	**<.001**
PCET_Accuracy	**−.019**	**.029**	−.005	.168
PFMT_Accuracy	**.099**	**<.001**	**−.009**	**<.001**
ESTROOP_Speed	**−.110**	**<.001**	**.007**	**<.001**
CPT_Speed	**.038**	**<.001**	**−.004**	**<.001**
ER40_Speed	**−.175**	**<.001**	**.012**	**<.001**
GNG_Speed	**−.097**	**<.001**	**.005**	**.022**
SLNB_Speed	**−.043**	**<.001**	**.006**	**<.001**
PCET_Speed	**−.113**	**<.001**	−.001	.620
PFMT_Speed	**−.079**	**<.001**	**−.001**	**<.001**

Note. CPT = Continuous Performance Task; ER40 = Emotion Recognition; GNG = Go/No-Go; SLNB = Short Letter-N-Back; PCET = Penn Conditional Exclusion Test; PFMT = Penn Face Memory Test; ESTROOP = Emotional Stroop Task. Age squared is the square of standardized age. Significant effects are bolded.

("Difference") can be interpreted as effect sizes. The sex differences in accuracy are mostly consistent with previous findings using the same tests (Gur et al., 2012). Specifically, for accuracy females outperform males on attention (CPT), impulse control (GNG and ESTROOP), face memory (PFMT), and emotion identification (ER40), while males outperform females on mental flexibility (PCET), and working memory (SLNB). In terms of speed, females are faster (lower response time) on the PCET, PFMT, and ER40. We found no significant sex differences on the SLNB, and the slower performance of males on the CPT is not consistent with previous findings. Finally, males perform faster on the ESTROOP and much faster on the GNG, both measures of impulsivity. This finding is consistent with their poorer accuracy reported above and a speed/accuracy trade-off characteristic of the GNG paradigm (Trommer, Hoeppner, Lorber, & Armstrong, 1988).

The bottom portion of Table 1 shows sex differences in efficiency, which is the average of an individual's accuracy and speed scores. All six gender differences in efficiency reported here are highly significant. Specifically, females outperform males on the CPT, ESTROOP, and PFMT, and quite substantially on the ER40. Males outperform females on the GNG, SLNB, and PCET.

Supplementary Table S2 shows the results of the above analyses performed using listwise deletion and outlier removal. All significant sex differenced remain significant, and no previously nonsignificant effect becomes significant. In addition, when the above analyses were conducted on the full sample after removing participants reporting a TBI, almost all results remained. The only exception was that the difference between males and females on PFMT RT became significant (new M − F = 0.6; $p < .001$).

Age Effects

Tests of heteroscedasticity were confirmed ($p < .05$) for all variables, and we thus proceeded with robust regression. All linear effects of age were significant ($p < .05$) except for ER40 accuracy ($p = .70$). Additionally, many nonlinear terms (age squared) were significant, and this information is shown in Table 2. To further explore the nonlinear relationships, these associations are presented graphically. Figure 1 shows the relationships between age and overall accuracy and speed, and Figure 2 shows each individual test score's relationship with age. The age trends in Figure 1 are clear: Accuracy increases with age until approximately 27 years, and speed correspondingly decreases with age at enlistment.

Both age-related trends (in accuracy and speed) are further examined in Figure 2 by reducing the summary scores to their individual component test scores; Figure 2a shows the four executive-/frontal lobe-related tests (CPT, SLNB, GNG, and ESTROOP), and Figure 2b shows the three reasoning-/memory-related tasks (PCET, PFMT, and ER40).

Finally, Supplementary Table S3 shows the results of the above analyses performed using listwise deletion and outlier removal. Relationships with age remained largely consistent with the full sample, with the following exceptions: The nonlinear association with SLNB accuracy became nonsignificant, the linear association with PCET accuracy became nonsignificant, the nonlinear association with PCET accuracy became significant, the nonlinear association with GNG speed became nonsignificant, the nonlinear association with PCET speed became significant, and the nonlinear association with PFMT speed became nonsignificant. In addition, when the above analyses were conducted on the full sample after removing participants reporting a TBI, almost all results remained. The only exceptions were (a) the age-squared term for PFMT speed became nonsignificant, (b) the linear age term for PCET accuracy became nonsignificant, and (c) the age-squared term for ER40 accuracy became significant ($p < .05$).

Exploratory Factor Analysis

Results of the comparison of the exploratory (E) and confirmatory (C) samples were mixed. They did not differ significantly by age (mean = 21.01 years for E and 21.02 for C) or sex (17.76% female for E and 17.38% for C). However, the variance of 7 out of the 21 scores did differ significantly between groups, even after correcting for multiple comparisons. Table 3 lists the adjusted p values for the tests of equality of variances; unequal variance was detected for

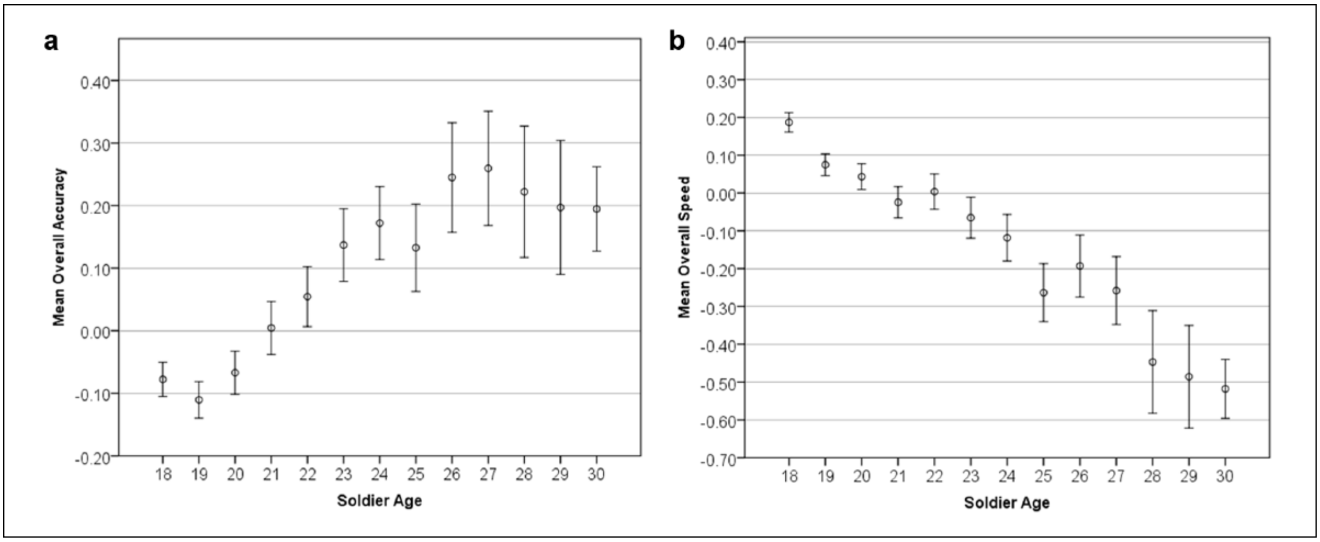


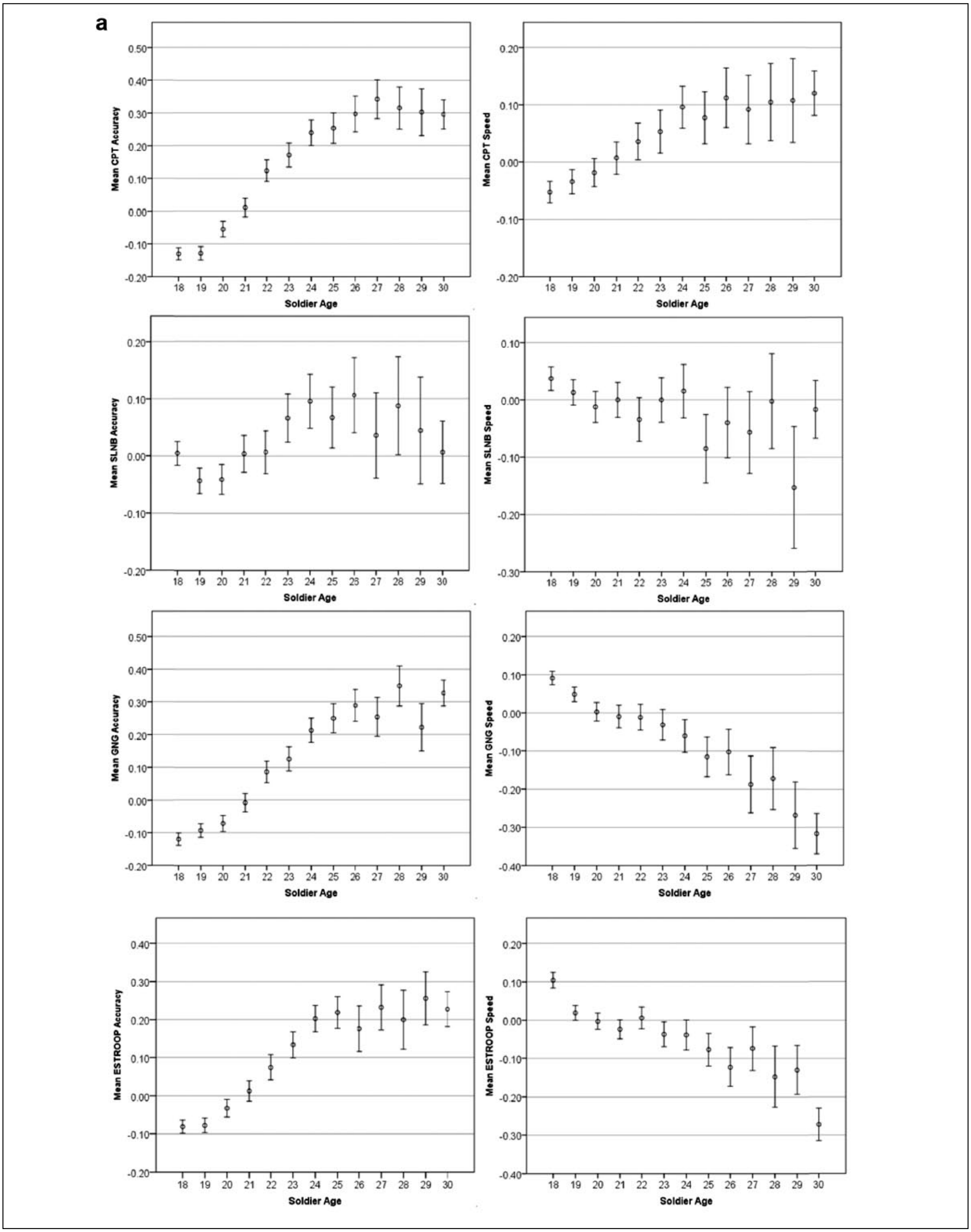
Figure 1. Age trends in standardized overall (a) accuracy and (b) speed (z scores) for the STARRS battery, with 95% confidence intervals.
Note. STARRS = Study to Assess Risk and Resilience in Servicemembers.

ESTROOP RT, GNG RT, SLNB RT, PFMT RT, ESTROOP Efficiency, CPT Efficiency, and ER40 Efficiency. Thus, cross-validation of the EFA with the CFA below should be interpreted with some caution.

Table 4 shows the unidimensional (one-factor) and two-factor exploratory solutions of the Army STARRS efficiency, accuracy, and speed scores. The fit of the unidimensional models for all three score types was moderate-to-poor. Specifically, the root mean square error of approximation (RMSEA) for efficiency, accuracy, and speed were 0.082 (±0.003), 0.065 (±0.003), and 0.096 (±0.003), respectively; and the degrees of freedom–corrected root mean square residuals (RMSRs) were 0.07, 0.05, and 0.08, respectively. The borderline fit of the unidimensional accuracy model indicates that one might be justified in calculating an "overall accuracy" score while ignoring multidimensionality, but the same could not be said of the speed and efficiency models, which are clearly multidimensional.

The two-factor models from Table 4 mostly confirm the hypothesis that the Army STARRS battery scores are multidimensional. The fit of the efficiency model is excellent (RMSEA = 0.038 ± 0.003; RMSR = 0.03), and mostly conforms to theory. Factor 1 comprises the four tests designed to functions of attention, working memory, impulse control, and management of emotional interference (CPT, LNB, GNG, and ESTROOP, respectively). By contrast, Factor 2 comprises the three tests that require more complex cognition involving additional temporoparietal functions of abstraction and mental flexibility, episodic memory, and emotion recognition (PCET, PFMT, and ER40, respectively). In light of age group effects in Figure 1 and Table 2, it is notable that the tests included in Factor 1 show better scores in the older cohorts while those of Factor 2 remain stable or get lower with increased cohort age. Note also that the moderate interfactor correlation (.45) between Factors 1 and 2 suggests that, despite the multidimensional structure of efficiency scores, an underlying (general performance) factor explaining covariance among all six tests does exist. Finally, the factor pattern shown in Table 4, in which "rapid" tests load on Factor 1 and "deliberative" tests load on Factor 2, is consistent with the idea that there are two "modes" of thinking (fast and slow) that recruit different brain regions. This phenomenon is discussed in an influential book (Kahneman, 2011).

To further examine the structure of the efficiency scores, we analyzed each component of efficiency (accuracy and speed) separately. The rationale is that the structure of efficiency could be, (a) the result of accuracy and speed having the same structure as each other, inevitably resulting in the same structure for efficiency; or (b) the result of some combination of unique accuracy and speed structures, which combine in such a way as to yield the efficiency structure in Table 4. The two-factor model for accuracy in Table 4 suggests the latter, because the structure of accuracy deviates somewhat from the structure of efficiency and maintains moderate-to-good fit (RMSEA = 0.051 ± 0.003; RMSR = 0.03). Specifically, Factor 1 now comprises only two tests (GNG and ESTROOP) plus two cross-loadings[1] (CPT and SLNB), and Factor 2 comprises five tests (PCET, CPT, LNB, PFMT, and ER40). Essentially, the attention and working memory tasks (CPT and LNB) switch from Factor 1 (in the efficiency model) to Factor 2 (in the accuracy model), though both retain cross-loadings on Factor 1. The


a
Mean CPT Accuracy
Mean CPT Speed
Mean SLNB Accuracy
Mean SLNB Speed
Mean GNG Accuracy
Mean GNG Speed
Mean ESTROOP Accuracy
Mean ESTROOP Speed
Soldier Age
18 19 20 21 22 23 24 25 26 27 28 29 30

(continued)

Figure 2. (continued)

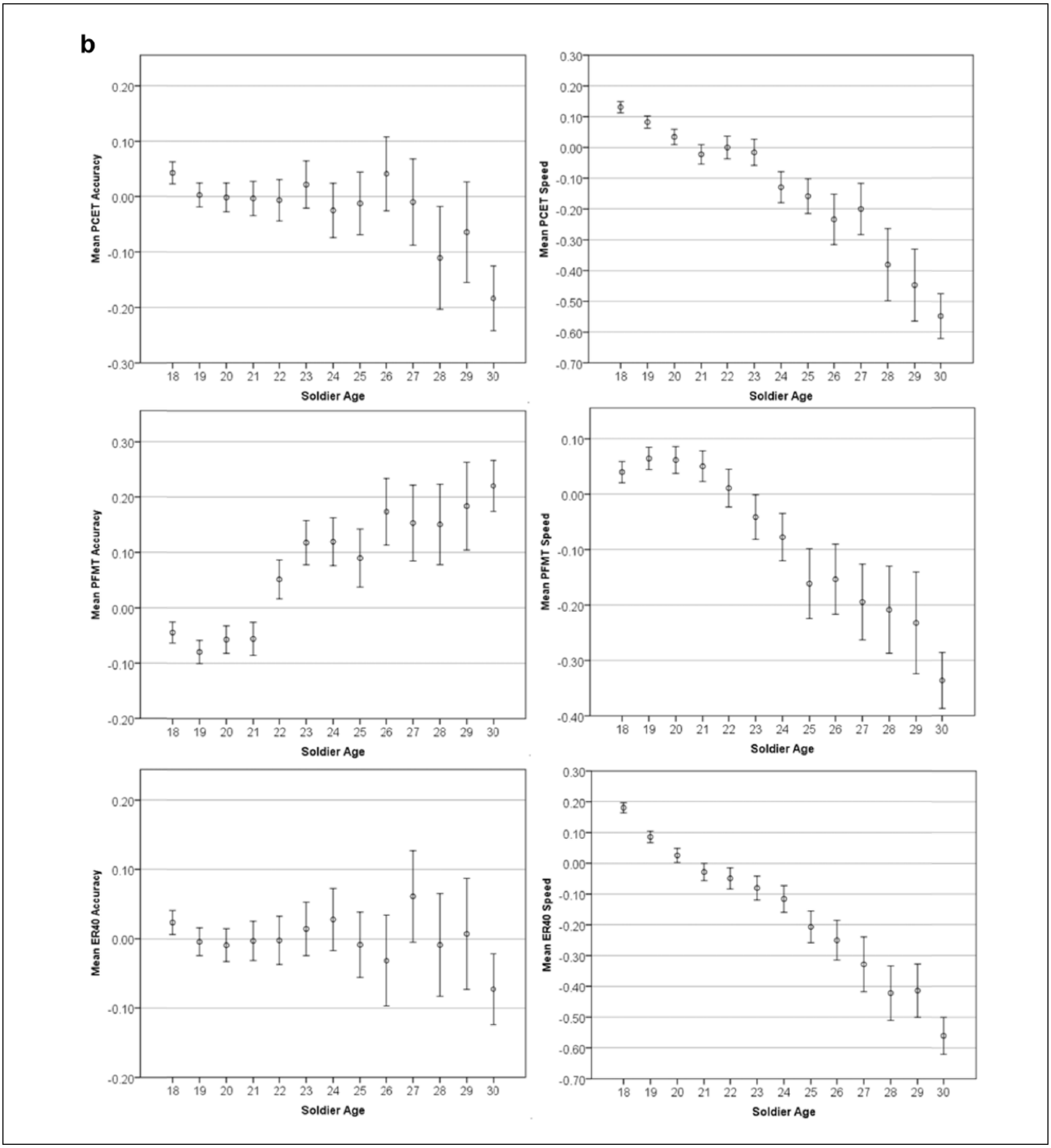


Figure 2. Age trends in standardized accuracy and speed (*z* scores) for the STARRS battery, with 95% confidence intervals. (a) Four executive-/frontal lobe–related tests (CPT, SLNB, GNG, and ESTROOP) and (b) Three reasoning-/memory-related tasks (PCET, PFMT, and ER40).

Note. STARRS = Study to Assess Risk and Resilience in Servicemembers; CPT = Continuous Performance Task; SLNB = SLNB = Short Letter-N-Back; GNG = Go/No-Go; ESTROOP = Emotional Stroop Task; PCET = Penn Conditional Exclusion Test; PFMT = Penn Face Memory Test; ER40 = Emotion Recognition.

Table 3. *p* Values for *F* Tests of Equal Variance Between Exploratory and Confirmatory Samples Used for Factor Analyses.

Score	*p*
ESTROOP_Accuracy	.118
CPT_Accuracy	1.000
ER40_Accuracy	.095
GNG_Accuracy	1.000
SLNB_Accuracy	1.000
PCET_Accuracy	1.000
PFMT_Accuracy	1.000
ESTROOP_RT	**<.001**
CPT_RT	.422
ER40_RT	.118
GNG_RT	**.048**
SLNB_RT	**.050**
PCET_RT	1.000
PFMT_RT	**<.001**
ESTROOP_Efficiency	**<.001**
CPT_Efficiency	**<.001**
ER40_Efficiency	**.008**
GNG_Efficiency	1.000
SLNB_Efficiency	1.000
PCET_Efficiency	1.000
PFMT_Efficiency	.486

Note. Significant values bolded. CPT = Continuous Performance Task; ER40 = Emotion Recognition; GNG = Go/No-Go; SLNB = Short Letter-N-Back; PCET = Penn Conditional Exclusion Test; PFMT = Penn Face Memory Test; ESTROOP = Emotional Stroop Task. *p* Values are corrected for multiple comparisons using the Holm (1979) method.

reason for the shift of the LNB is unclear, but it might be due to task difficulty that creates a distribution of accuracy scores more similar to the complex cognition tasks. The nearly equal loadings of the CPT on Factor 1 and Factor 2 (0.32 and 0.33, respectively) make interpretation difficult.

Finally, the two-factor model for speed in Table 4 very closely matches that for efficiency, and has good fit (RMSEA = 0.031 ± 0.003; RMSR = 0.02). Factor 1 comprises four tests that emphasize vigilant, rapid responses, while Factor 2 comprises three tests that put less emphasis on speed. Note that soldiers were asked to work as quickly as possible for all seven tests, but the three tests composing Factor 2 (PCET, PFMT, and ER40) require at least a momentary pause to contemplate the response. Thus, the answer to the question posed above—that is, how do the structures of accuracy and speed combine to result in the clean, two-factor structure for efficiency?—appears to be that speed exerts enough influence over the somewhat complex structure of accuracy to result in a structure of efficiency that more nearly mimics that for speed.

Finally, Supplementary Table S4 shows the results of the above analyses performed using listwise deletion and outlier removal. Results are consistent with those of the full sample, except that, for accuracy, the CPT and LNB have no cross-loadings—that is, they cleanly load on Factor 1 and Factor 2, respectively. In the full sample, they both cross-loaded on Factor 1 and Factor 2.

Confirmatory Factor Analysis

The exploratory analyses described above provided clear guidance about which tests should compose which factors in a confirmatory model; these specifications were consistent with neurocognitive theory. Figure 3 shows the two-factor confirmatory bifactor model of the Army STARRS battery efficiency scores based on results obtained in the above exploratory analyses. We chose to model efficiency, because it combines both types of performance information (accuracy and speed), and there is a published bifactor model of the CNB (Moore, Reise, Gur, Hakonarson, & Gur, 2015) using five of the seven tests used in the STARRS battery as part of a larger battery. The fit of the model in the second half of the sample, seen in Figure 3, is excellent (comparative fit index = 0.98; RMSEA = 0.026 ± 0.003; SRMR = 0.015), and some important characteristics are notable. First, as suggested by the moderate (0.53) interfactor correlation in the two-factor exploratory model (see Note in Table 4), the overall efficiency factor underlying all tests (right side of Figure 3) is also moderate (mean loading = 0.44). This supports the idea that a general efficiency score can be calculated from all seven test scores, but that one should not ignore the multidimensionality (as one could if the overall efficiency dimension was very strong).

Second, it is worth noting that the overall efficiency factor is slightly more determined by the cognition and memory tests (mean loading = 0.57) than by the executive attention tests (mean loading = 0.35), which means that the neuropsychological phenomena determining soldiers' overall efficiency are measured more precisely by the former (though only moderately). By contrast, the individual group factors (left side of Figure 3) show the opposite effect—that is, the executive/attention factor (mean loading = 0.47) is stronger than the reasoning/memory factor (mean absolute loading = 0.28). This means that, even after controlling for general performance efficiency, a moderate amount of the covariance among tests is explained by neuropsychological processes related specifically to executive/attention abilities. The same cannot be said of the reasoning/memory tests, because most of the covariance among those three tests seems to be explained almost entirely by general efficiency ability. Indeed, after controlling for general efficiency, the group factor loading of the PFMT becomes negative (−0.48), though not significantly so. Such a weak group factor indicates that although the reasoning/memory tests are good measures of the neuropsychological phenomena controlling general efficiency ability, one should use extreme caution if trying to create subscale scores designed

Table 4. Unidimensional and Two-Factor Solutions of Efficiency, Accuracy, and Speed Scores From the Army STARRS Battery.

	Efficiency			Accuracy			Speed		
		Two-factor			Two-factor			Two-factor	
Test	Uni	F1	F2	Uni	F1	F2	Uni	F1	F2
PCET	.31		.41	.23		.45	.50		.55
CPT	.49	.39		.59	.32	.33	.31	.39	
SLNB	.53	.50		.47	.22	.32	.25	.40	
GNG	.66	.77		.76	.90		.43	.69	
PFMT	.37		.31	.42		.30	.40		.37
ER40	.40		.68	.30		.39	.60		.70
ESTROOP	.46	.45		.52	.45		.25	.23	

Note. Uni = unidimensional; CPT = Continuous Performance Task; ER40 = Emotion Recognition; GNG = Go/No-Go; SLNB = Short Letter-N-Back; PCET = Penn Conditional Exclusion Test; PFMT = Penn Face Memory Test; ESTROOP = Emotional Stroop Task; STARRS = Study to Assess Risk and Resilience in Servicemembers. Rotation = oblimin. Interfactor correlations for efficiency, accuracy, and speed are .45, .58, and .41, respectively; loadings with absolute value less than .20 not shown.

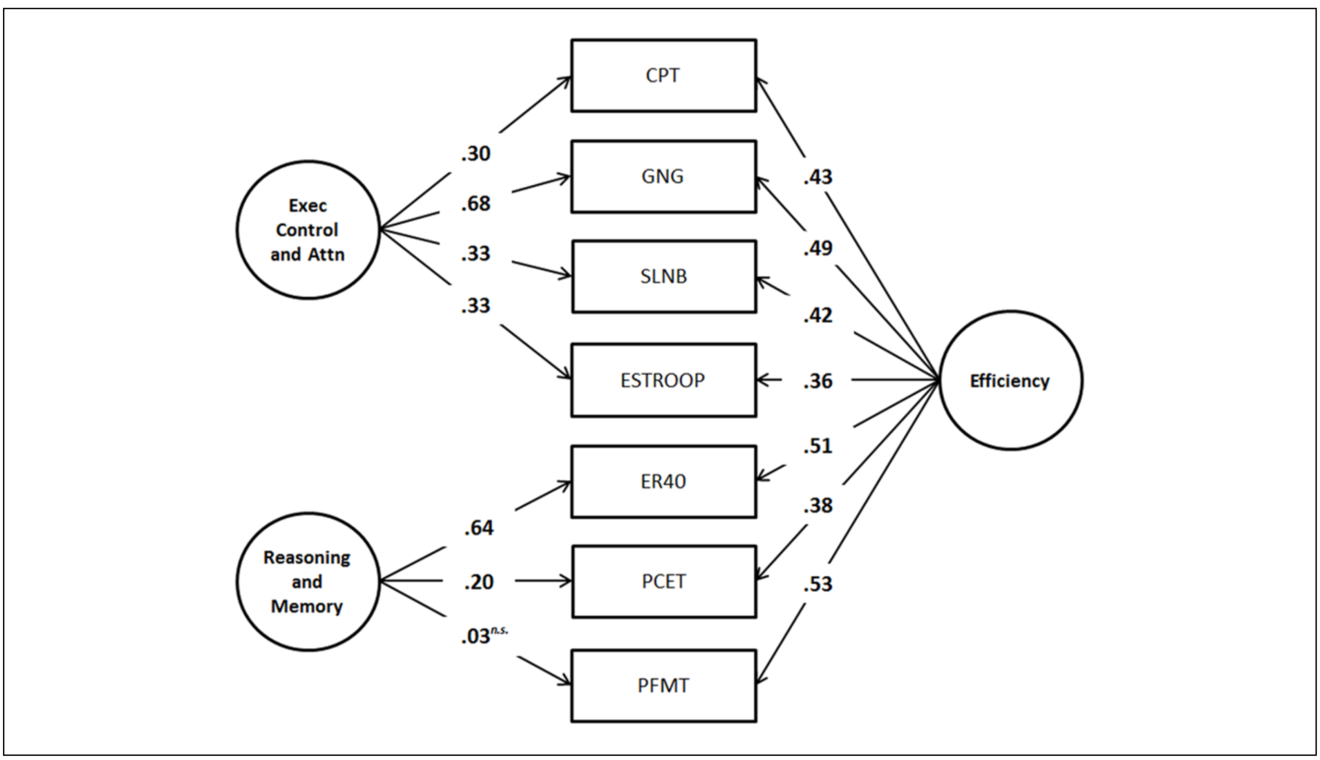


Figure 3. Confirmatory bifactor analysis of the Army STARRS neurocognitive battery efficiency scores.
Note. Exec = Executive; Attn = Attention; CPT = Continuous Performance Task; ER40 = Emotion Recognition; GNG = Go/No-Go; SLNB = Short Letter-N-Back; PCET = Penn Conditional Exclusion Test; PFMT = Penn Face Memory Test; ESTROOP = Emotional Stroop Task; *ns* = not significant. Results are standardized such that the variance of the latent variables is 1.00. All coefficient estimates are significant at the 0.005 level unless indicated otherwise.

to test reasoning and memory uniquely with the current combination of tests. Instead, one could use the general efficiency score as a close proxy to reasoning and memory.

Finally, Supplementary Figure S1 shows the results of the above analysis performed using listwise deletion and outlier removal. Fit of the model remains excellent (comparative fit index = 0.98; RMSEA = 0.036 ± 0.005; SRMR = 0.014), and relative loadings remain mostly unchanged. For example, the Executive Control factor remains dominated by GNG, Reading/Memory remains dominated by ER40, and the general efficiency factor remains dominated by PFMT and GNG. The only exception is that the loading of the ER40 on the general factor is somewhat higher in the full sample (0.51) compared with the more limited sample (0.40).

Discussion

With the recent increase in soldier suicides, as well as the constant hazard of TBI in battle, rapid, and efficient neurocognitive testing of executive control functioning is becoming important for the military. Abnormalities or changes in neurocognitive test scores have been implicated in myriad problems related to the military, including suicidal behavior, PTSD, mood disorders, substance and alcohol use disorders, and impulsive behavior. The CNB used in the Army STARRS research project is an efficient, easily administered battery that could be used for research (as in this case) or for more applied needs, such as in the war theater.

We have applied this battery to a large sample of Servicemembers with minimal complications and have obtained data on over 50,000 soldiers within a short period of time. In the present study, we examined the factorial and concurrent validity of this battery by testing its sensitivity to sex differences and age effects on performance and by evaluating its factorial structure. Robust sex differences were documented on measures that have revealed such differences in earlier studies. For example, Gur et al. (2012) found that females outperformed males on ER40 and PFMT accuracy and on ER40 speed; Longenecker, Dickinson, Weinberger, and Elvevåg (2010) found that males outperformed females on N-Back accuracy; von Kluge (1992) found that females outperformed males on Stroop task accuracy, and males outperformed females on Stroop task speed[2]; while age effects were either novel or consistent with previous findings. With respect to aggregate (cross-battery) results, the improvement in accuracy (Figure 1a) until age 27 is consistent with findings reported by Schaie (1994) and Whitley et al. (2016), and is further supported by evidence for continuing brain development until that age (Sowell, Thompson, Holmes, Jernigan, & Toga, 1999; Yakovlev & Lecours, 1967). The steady decrease in speed (Figure 1b) is also fully consistent with previous literature, which shows greater adverse effects of age on speed than on memory (e.g., Irani et al., 2012). Additionally, Salthouse (2000) found uniform decreases in performance speed with age on multiple neurocognitive tests.

Three of the five significant positive correlations of accuracy with age (SLNB, CPT, and PFMT) are consistent with a previous study using the same tests in a large civilian sample (Gur et al., 2012). The fourth significant positive correlation (GNG150) is consistent with previous findings that accuracy increases with age in response inhibition tasks (e.g., Kramer, Humphrey, Larish, Logan, & Strayer, 1994). To summarize, the face memory test (PFMT) and two of the faster paced tests that more overtly require executive functions like attention and cognitive control (CPT and GNG) correlate positively with age, whereas two of the three tests that require more contemplation (ER40 and PCET) correlate negatively with age (at least after age 18). To our knowledge, there has been no previous finding that ESTROOP accuracy increases with age until age ~25, at which time it plateaus; it is therefore unclear whether the ESTROOP accuracy results speak to the concurrent validity of the Army STARRS battery. Overall, however, it appears that older recruits have better attention skills (CPT), are less impulsive (GNG), and have better memory for faces (PFMT). On the other hand, they are less sensitive to emotions (ER40) and have less mental flexibility (PCET). These age effects are quite surprising in their consistency and magnitude. Age of recruits makes a difference; older recruits are demonstrably more accurate but also slower across nearly all tests. When specific domains are examined, older recruits are less impulsive although, after age 26, they tend to become more rigid.

Notably, trends in speed (all negative except CPT, indicating higher RT or slower responding with increased age) are opposite to those reported in Gur et al. (2012), who found almost all positive correlations between age and speed. This is because of the differing age cohorts—18+ years here, compared with 8 to 21 years in Gur et al. (2012). The Gur et al (2012) study showed in a cohort of 3,500 children annually faster response speed from age 8 to 17, where it flattens through age 21. The present results indicate that within the age range of 18 to 30 years response speed is generally lower with increased age of cohorts, consistent with several previous studies (e.g., Deary & Der, 2005; Fozard, Vercruyssen, Reynolds, Hancock, & Quilter, 1994; Myerson, Hale, Wagstaff, Poon, & Smith, 1990; Salthouse, Hambrick, & McGuthry, 1998). On the other hand, accuracy on most tasks does continue to improve until approximately age 27. The more pronounced age group effects evident in the GNG, CPT, ESTROOP, and PFMT are consistent with the hypothesized association of these measures with frontal lobe functioning. Frontal lobe maturation is protracted in humans, reaching its apex in the early 20s, whereas the motor cortex (responsible for immediate response times) matures by late adolescence (e.g., Filipek, Richelme, Kennedy, & Caviness, 1994; Gogtay et al., 2004; Huttenlocher 1979; Matsuzawa et al., 2001; Pfefferbaum et al., 1994; Sowell et al., 1999; Yakovlev & Lecours, 1967). These results indicate that the battery is sensitive to demographic parameters that affect neurocognitive performance.

The factor analysis supported the theory that motivated the construction of the STARRS battery, which was to emphasize the measurement of executive functioning and sample to a more limited extent the domains of memory, reasoning, and social cognition. Based on the results of the present study, in which concurrent and structural validity were largely confirmed, we are comfortable recommending the STARRS battery for research and applied purposes. Of theoretical interest, the factor analyses also provide some evidence in favor of a dual-process model of cognition (Chaiken & Trope, 1999). Specifically, when performance efficiency is separated into accuracy and speed components, the factor structures of the latter two do not perfectly match the structure of efficiency scores (see Table 4). A dual-process framework would predict such a phenomenon, because dual-process models posit separate cognitive processes (one fast, one slower) for arriving at an accurate response (see Kahneman, 2011). Thus, the neural activation required to arrive at a correct answer on two different tasks might involve different proportions of rapid versus slow neurocognitive processing. The result could be two different patterns of covariance among speed scores and among accuracy scores that combine to form a nonetheless theoretically sound covariance structure for efficiency. Evidence from neuroimaging (Goel, Buchel, Frith, & Dolan, 2000; Goel & Dolan, 2003) lends additional support to dual-process models of cognition.

Two other popular test batteries that have similar goals deserve mention for brief comparison: the Automated Neuropsychological Assessment Metrics (ANAM; Reeves, Kane, & Winter, 1995) and the NIH Toolbox (Gershon et al., 2010). The ANAM, designed by the U.S. Department of Defense, includes 22 tests designed to measure accuracy and speed in the cognitive domains of executive function, episodic memory, and decision making. Similarly, the NIH Toolbox includes approximately 55 tests designed to measure cognition, emotion, motor, and sensation (nihtoolbox.org). The Army STARRS battery is more focused on executive control and uniquely examines social cognition (via the ER40) and nonconscious emotional processing (via the ESTROOP). It is also unique in including tests that have demonstrated links to regional brain function (Gur, Erwin, & Gur, 1992; Gur et al., 2010; Roalf et al., 2014). These two advantages are notable for a couple of reasons. First, by definition, the psychopathologies of interest here (including those associated with suicide risk, such as depression) are associated not only with changes in executive control but also changes in emotional and social processing phenomena such as those tapped by the ER40 and ESTROOP. Second, to demonstrate a link between brain function and measurement outcome is to meet the highest standard of validity, clearly articulated by Borsboom (2005). Five of the seven tests on the Army STARRS battery (SLNB, ER40, PFMT, PCET, and CPT) meet this standard (Roalf et al., 2014), which is more than can be said for either the ANAM or NIH Toolbox as of 2016.

An important limitation of the present study is that the sample is likely unique in which it comprised individuals who chose to join the U.S. Army and are therefore likely more novelty seeking and less risk averse than the general population (see Rademaker, Vermetten, Geuze, Muilwijk, & Kleber, 2008). An implication is that the effects found here with respect to age and sex may in fact be underestimated due to the narrow sampling—that is, if the sample comprised the whole spectrum of impulsivity and risk aversion found in the general population, the increased variance in tests of executive control might result in larger effects. Further research is needed to determine the extent to which U.S. Army personnel, especially those in noncombat roles, resemble the general population psychologically. A second limitation of the present study is that the optimal evaluation of valid score interpretation—namely, tests of whether the scores are sensitive to outcomes of interest such as psychopathology and suicide risk—has not been presented here. This work is currently underway by multiple Army STARRS collaborators, and information about the progress of said research is available here: http://starrs-ls.org/#/list/publications. To be clear, only when this further work is completed will the validity of the Army STARRS battery (as intended) be established. The present study established only what may be considered preliminary validity.

Notwithstanding the sample's limited scope and cross-sectional nature, the present analysis supports the psychometric validity of the Army STARRS neurocognitive battery. The availability of this battery can have great potential for augmenting the set of assessment tools that can help in the early detection and intervention of vulnerability to neurocognitive deficits. Neuroimaging has increasingly been used for early diagnosis, and the present battery can be administered in the scanner to confirm linkage between dysfunction and regional brain activation. Out of the scanner, the battery can provide a more affordable step for identifying individuals with regional brain dysfunction. Arguably, the future of neuropsychology would likely involve such a combined use of neuroimaging-validated cognitive tests with in-scanner verification in smaller subsamples where specific hypotheses can be pursued. The present study is a step in that direction.

Authors' Note

The contents are solely the responsibility of the authors and do not necessarily represent the views of the Department of Health and Human Services, NIMH, the Veterans Administration, Department of the Army, or the Department of Defense.

Contributing Authors

The Army STARRS Team consists of Coprincipal Investigators: Robert J. Ursano, MD (Uniformed Services University of the Health Sciences) and Murray B. Stein, MD, MPH (University of California, San Diego, and VA San Diego Healthcare System); Site Principal Investigators: Steven Heeringa, PhD (University of Michigan) and Ronald C. Kessler, PhD (Harvard Medical School); National Institute of Mental Health (NIMH) collaborating scientists: Lisa J. Colpe, PhD, MPH and Michael Schoenbaum,

PhD; Army liaisons/consultants: COL Steven Cersovsky, MD, MPH (USAPHC) and Kenneth Cox, MD, MPH (USAPHC); Neurocognitive Working Group Cochairs: Gregory G. Brown, PhD (University of California, San Diego); Ruben C. Gur, PhD (University of Pennsylvania); Matthew K. Nock, PhD (Harvard University); Neurocognitive Working Group: Robert B. Baron, MSE (University of Pennsylvania); Colleen M. Brensinger, MS (University of Pennsylvania); Margaret L. Hudson, MPH (University of Michigan); Devin Hunt (NIMH); Chad Jackson, MSCE (University of Pennsylvania); Adam Jaroszewski, BS (Harvard University); Tyler M. Moore, PhD, MSc (University of Pennsylvania); Allison Mott, BA (University of Pennsylvania); James A. Naifeh, PhD (Uniformed Services University of the Health Sciences); Megan Quarmley, BS (University of Pennsylvania); Victoria Risbrough, PhD (University of California, San Diego); Adam Savitt, BA (University of Pennsylvania), Murray B. Stein, MD, MPH (University of California, San Diego, and VA San Diego Healthcare System); Michael L. Thomas, PhD (University of California, San Diego); Virginie M. Patt (University of California, San Diego), and Robert J. Ursano, MD (Uniformed Services University of the Health Sciences); Other team members: Pablo A. Aliaga, MA (Uniformed Services University of the Health Sciences); COL David M. Benedek, MD (Uniformed Services University of the Health Sciences); Susan Borja, PhD (NIMH); Laura Campbell-Sills, PhD (University of California, San Diego); Catherine L. Dempsey, PhD, MPH (Uniformed Services University of the Health Sciences); Richard Frank, PhD (Harvard Medical School); Carol S. Fullerton, PhD (Uniformed Services University of the Health Sciences); Nancy Gebler, MA (University of Michigan); Robert K. Gifford, PhD (Uniformed Services University of the Health Sciences); Stephen E. Gilman, ScD (Harvard School of Public Health); Marjan G. Holloway, PhD (Uniformed Services University of the Health Sciences); Paul E. Hurwitz, MPH (Uniformed Services University of the Health Sciences); Sonia Jain, PhD (University of California, San Diego); Tzu-Cheg Kao, PhD (Uniformed Services University of the Health Sciences); Karestan C. Koenen, PhD (Columbia University); Lisa Lewandowski-Romps, PhD (University of Michigan); Holly Herberman Mash, PhD (Uniformed Services University of the Health Sciences); James E. McCarroll, PhD, MPH (Uniformed Services University of the Health Sciences); Katie A. McLaughlin, PhD (Harvard Medical School); Rema Raman, PhD (University of California, San Diego); Sherri Rose, PhD (Harvard Medical School); Anthony Joseph Rosellini, PhD (Harvard Medical School); Nancy A. Sampson, BA (Harvard Medical School); LCDR Patcho Santiago, MD, MPH (Uniformed Services University of the Health Sciences); Michaelle Scanlon, MBA (National Institute of Mental Health); Jordan W. Smoller, MD, ScD (Harvard Medical School); Patti L. Vegella, MS, MA (Uniformed Services University of the Health Sciences); Christina Wassel, PhD (University of Pittsburgh); and Alan M. Zaslavsky, PhD (Harvard Medical School).

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Army STARRS was sponsored by the Department of the Army and funded under cooperative agreement number U01MH087981 with the U.S. Department of Health and Human Services, National Institutes of Health, National Institute of Mental Health (NIH/NIMH).

Notes

1. Technically, the 0.22 cross-loading of the SLNB on Factor 1 does not meet the conventional cutoff of 0.30 for evaluating factor loading salience, but we selected 0.20 as the cutoff for discussing factor loadings here. See Kline (2014) for a nuanced discussion of factor loadings and their meanings.
2. Note, however, that the task used by von Kluge (1992) was a traditional Stroop task, whereas ours was an ESTROOP task. Concurrent validity support provided by von Kluge (1992), therefore, is only partial.

Supplemental Material

Supplementary material for this article is available online.

References

Army STARRS. (2012). Retrieved from http://www.armystarrs.org/

Arsenault-Lapierre, G., Kim, C., & Turecki, G. (2004). Psychiatric diagnoses in 3275 suicides: A meta-analysis. *BMC Psychiatry*, *4*, 37. doi:10.1186/1471-244X-4-37

Baddeley, A., & Della Sala, S. (1996). Working memory and executive control. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *351*, 1397-1404.

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin*, *121*, 65-94.

Becker, E. S., Strohbach, D., & Rinck, M. (1999). A specific attentional bias in suicide attempters. *Journal of Nervous and Mental Disease*, *187*, 730-735.

Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics*. Cambridge, England: Cambridge University Press.

Brenner, L. A., Terrio, H., Homaifar, B. Y., Gutierrez, P. M., Staves, P. J., Harwood, J. E., . . . Warden, D. (2010). Neuropsychological test performance in soldiers with blast-related mild TBI. *Neuropsychology*, *24*, 160-167.

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, *47*, 1287-1294.

Carter, C. S., Barch, D. M., Gur, R., Gur, R., Pinkham, A., & Ochsner, K. (2009). CNTRICS final task selection: Social cognitive and affective neuroscience–based measures. *Schizophrenia Bulletin*, *35*, 153-162.

Casey, B. J., Castellanos, F. X., Giedd, J. N., Marsh, W. L., Hamburger, S. D., Schubert, A. B., . . . Rapoport, J. L. (1997). Implication of right frontostriatal circuitry in response inhibition and attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*, 374-383.

Cha, C. B., Najmi, S., Park, J. M., Finn, C. T., & Nock, M. K. (2010). Attentional bias toward suicide-related stimuli predicts suicidal behavior. *Journal of Abnormal Psychology*, *119*, 616-622.

Chaiken, S., & Trope, Y. (Eds.). (1999). *Dual-process theories in social psychology*. New York, NY: Guilford Press.

Christopher, G., & MacDonald, J. (2005). The impact of clinical depression on working memory. *Cognitive Neuropsychiatry*, *10*, 379-399.

Cox, W. M., Fadardi, J. S., & Pothos, E. M. (2006). The addiction-Stroop test: Theoretical considerations and procedural recommendations. *Psychological Bulletin*, *132*, 443-476.

Danckwerts, A., & Leathem, J. (2003). Questioning the link between PTSD and cognitive dysfunction. *Neuropsychology Review*, *13*, 221-235.

Deary, I. J., & Der, G. (2005). Reaction time, age, and cognitive ability: Longitudinal findings from age 16 to 63 years in representative population samples. *Aging, Neuropsychology, and Cognition*, *12*, 187-215.

Durston, S., Davidson, M. C., Mulder, M. J., Spicer, J. A., Galvan, A., Tottenham, N., . . . Casey, B. J. (2007). Neural and behavioral correlates of expectancy violations in attention-deficit hyperactivity disorder. *Journal of Child Psychology and Psychiatry*, *48*, 881-889.

Durston, S., Mulder, M., Casey, B. J., Ziermans, T., & van Engeland, H. (2006). Activation in ventral prefrontal cortex is sensitive to genetic vulnerability for attention-deficit hyperactivity disorder. *Biological Psychiatry*, *60*, 1062-1070.

Durston, S., Thomas, K. M., Worden, M. S., Yang, Y., & Casey, B. J. (2002). The effect of preceding context on inhibition: An event-related fMRI study. *NeuroImage*, *16*, 449-453.

Erwin, R. J., Gur, R. C., Gur, R. E., Skolnick, B., Mawhinney-Hee, M., & Smailis, J. (1992). Facial emotion discrimination: I. Task construction and behavioral findings in normal subjects. *Psychiatry Research*, *42*, 231-240.

Fertuck, E. A., Jekal, A., Song, I., Wyman, B., Morris, M. C., Wilson, S. T., . . . Stanley, B. (2009). Enhanced "Reading the Mind in the Eyes" in borderline personality disorder compared to healthy controls. *Psychological Medicine*, *39*, 1979-1988.

Fertuck, E. A., Lenzenweger, M. F., Clarkin, J. F., Hoermann, S., & Stanley, B. (2006). Executive neurocognition, memory systems, and borderline personality disorder. *Clinical Psychology Review*, *26*, 346-375.

Filipek, P. A., Richelme, C., Kennedy, D. N., & Caviness, V. S. (1994). The young adult human brain: An MRI-based morphometric analysis. *Cerebral Cortex*, *4*, 344-360.

Foa, E. B., Feske, U., Murdock, T. B., Kozak, M. J., & McCarthy, P. R. (1991). Processing of threat-related information in rape victims. *Journal of Abnormal Psychology*, *100*, 156-162.

Fox, J., & Weisberg, S. (2011). *An R Companion to applied regression* (2nd ed.). Thousand Oaks, CA: Sage. Retrieved from http://socserv.socsci.mcmaster.ca/jfox/Books/Companion/

Fozard, J. L., Vercruyssen, M., Reynolds, S. L., Hancock, P. A., & Quilter, R. E. (1994). Age differences and changes in reaction time: The Baltimore Longitudinal Study of Aging. *Journal of Gerontology*, *49*, P179-P189.

Gershon, R. C., Cella, D., Fox, N. A., Havlik, R. J., Hendrie, H. C., & Wagster, M. V. (2010). Assessment of neurological and behavioural function: The NIH Toolbox. *Lancet Neurology*, *9*, 138-139.

Goel, V., Buchel, C., Frith, C., & Dolan, R. J. (2000). Dissociation of mechanisms underlying syllogistic reasoning. *NeuroImage*, *12*, 504-514.

Goel, V., & Dolan, R. J. (2003). Explaining modulation of reasoning by belief. *Cognition*, *87*, B11-B22.

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., . . . Thompson, P. M. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 8174-8179.

Gur, R. C., Calkins, M. E., Satterthwaite, T. D., Ruparel, K., Bilker, W. B., Moore, T. M., . . . Gur, R. E. (2014). Neurocognitive growth charting in psychosis spectrum youths. *JAMA Psychiatry*, *71*, 366-374.

Gur, R. C., Erwin, R. J., & Gur, R. E. (1992). Neurobehavioral probes for physiologic neuroimaging studies. *Archives of General Psychiatry*, *49*, 409-414.

Gur, R. C., Erwin, R. J., Gur, R. E., Zwil, A. S., Heimberg, C., & Kraemer, H. C. (1992). Facial emotion discrimination: II. Behavioral findings in depression. *Psychiatry Research*, *42*, 241-251.

Gur, R. C., Jaggi, J. L., Ragland, J. D., Resnick, S. M., Shtasel, D., Muenz, L., & Gur, R. E. (1993). Effects of memory processing on regional brain activation: Cerebral blood flow in normal subjects. *International Journal of Neuroscience*, *72*, 31-44.

Gur, R. C., Ragland, J. D., Moberg, P. J., Turner, T. H., Bilker, W. B., Kohler, C., . . . Gur, R. E. (2001). Computerized neurocognitive scanning: I. Methodology and validation in healthy people. *Neuropsychopharmacology*, *25*, 766-776.

Gur, R. C., Ragland, J. D., Mozley, L. H., Mozley, P. D., Smith, R., Alavi, A., . . . Gur, R. E. (1997). Lateralized changes in regional cerebral blood flow during performance of verbal and facial recognition tasks: Correlations with performance and "effort." *Brain and Cognition*, *33*, 388-414.

Gur, R. C., Richard, J., Calkins, M. E., Chiavacci, R., Hansen, J. A., Bilker, W. B., . . . Gur, R. E. (2012). Age group and sex differences in performance on a computerized neurocognitive battery in children age 8-21. *Neuropsychology*, *26*, 251-265.

Gur, R. C., Richard, J., Hughett, P., Calkins, M. E., Macy, L., Bilker, W. B., . . . Gur, R. E. (2010). A cognitive neuroscience-based computerized battery for efficient measurement of individual differences: Standardization and initial construct validation. *Journal of Neuroscience Methods*, *187*, 254-262.

Habel, U., Gur, R. C., Mandal, M. K., Salloum, J. B., Gur, R. E., & Schneider, F. (2000). Emotional processing in schizophrenia across cultures: Standardized measures of discrimination and experience. *Schizophrenia Research*, *42*, 57-66.

Harkavy-Friedman, J. M., Keilp, J. G., Grunebaum, M. F., Sher, L., Printz, D., Burke, A. K., . . . Oquendo, M. (2006). Are BPI and BPII suicide attempters distinct neuropsychologically? *Journal of Affective Disorders*, *94*, 255-259.

Heimberg, C., Gur, R. E., Erwin, R. J., Shtasel, D. L., & Gur, R. C. (1992). Facial emotion discrimination: III. Behavioral findings in schizophrenia. *Psychiatry Research*, *42*, 253-265.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70.

Horn, N. R., Dolan, M., Elliott, R., Deakin, J. F. W., & Woodruff, P. W. R. (2003). Response inhibition and impulsivity: An fMRI study. *Neuropsychologia*, *41*, 1959-1966.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure analysis: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*, 424-453.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Huttenlocher, P. R. (1979). Synaptic density in human frontal cortex: Developmental changes and effects of aging. *Brain Research*, *163*, 195-205.

Irani, F., Brensinger, C. M., Richard, J., Calkins, M. E., Moberg, P. J., Bilker, W., . . . Gur, R. C. (2012). Computerized neurocognitive test performance in schizophrenia: A lifespan analysis. *American Journal of Geriatric Psychiatry*, *20*, 41-52.

Jollant, F., Lawrence, N. L., Olié, E., Guillaume, S., & Courtet, P. (2011). The suicidal mind and brain: A review of neuropsychological and neuroimaging studies. *World Journal of Biological Psychiatry*, *12*, 319-339.

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Strauss, and Giroux.

Kaspi, S. P., McNally, R. J., & Amir, N. (1995). Cognitive processing of emotional information in posttraumatic stress disorder. *Cognitive Therapy and Research*, *19*, 433-444.

Keilp, J. G., Gorlyn, M., Oquendo, M. A., Burke, A. K., & Mann, J. J. (2008). Attention deficit in depressed suicide attempters. *Psychiatry Research*, *159*, 7-17.

Keilp, J. G., Sackeim, H. A., Brodsky, B. S., Oquendo, M. A., Malone, K. M., & Mann, J. J. (2001). Neuropsychological dysfunction in depressed suicide attempters. *American Journal of Psychiatry*, *158*, 735-741.

Keilp, J. G., Sackeim, H. A., & Mann, J. J. (2005). Correlates of trait impulsiveness in performance measures and neuropsychological tests. *Psychiatry Research*, *135*, 191-201.

Kessler, R. C., Colpe, L. J., Fullerton, C. S., Gebler, N., Naifeh, J. A., Nock, M. K., . . . Heeringa, S. G. (2013). Design of the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *International Journal of Methods in Psychiatric Research*, *22*, 267-275.

Kline, P. (2014). *An easy guide to factor analysis*. New York, NY: Routledge.

Kohler, C. G., Hanson, E., & March, M. E. (2013). Emotion processing in schizophrenia. In D. L. Roberts & D. L. Penn (Eds.), *Social cognition in schizophrenia: From evidence to treatment* (pp. 173-195). New York, NY: Oxford University Press.

Kohler, C. G., Turner, T. H., Bilker, W. B., Brensinger, C. M., Siegel, S. J., Kanes, S. J., . . . Gur, R. C. (2003). Facial emotion recognition in schizophrenia: Intensity effects and error pattern. *American Journal of Psychiatry*, *160*, 1768-1774.

Kramer, A. F., Humphrey, D. G., Larish, J. F., Logan, G. D., & Strayer, D. L. (1994). Aging and inhibition: Beyond a unitary view of inhibitory processing in attention. *Psychology and Aging*, *9*, 491-512.

Kurtz, M. M., & Gerraty, R. T. (2009). A meta-analytic investigation of neurocognitive deficits in bipolar illness: Profile and effects of clinical state. *Neuropsychology*, *23*, 551-562.

Kurtz, M. M., Ragland, J. D., Bilker, W., Gur, R. C., & Gur, R. E. (2001). Comparison of the continuous performance test with and without working memory demands in healthy controls and patients with schizophrenia. *Schizophrenia Research*, *48*, 307-316.

Kurtz, M. M., Ragland, J. D., Moberg, P. J., & Gur, R. C. (2004). The Penn Conditional Exclusion Test: A new measure of executive-function with alternate forms for repeat administration. *Archives of Clinical Neuropsychology*, *19*, 191-201.

LeGris, B. N., & van Reekum, M. D. (2006). The neuropsychological correlates of borderline personality disorder and suicidal behaviour. *Canadian Journal of Psychiatry*, *51*, 131-142.

LeMarquand, D. G., Pihl, R. O., Young, S. N., Tremblay, R. E., Séguin, J. R., Palmour, R. M., & Benkelfat, C. (1998). Tryptophan depletion, executive functions, and disinhibition in aggressive, adolescent males. *Neuropsychopharmacology*, *19*, 333-341.

Leskin, L. P., & White, P. M. (2007). Attentional networks reveal executive function deficits in posttraumatic stress disorder. *Neuropsychology*, *21*, 275-284.

Longenecker, J., Dickinson, D., Weinberger, D. R., & Elvevåg, B. (2010). Cognitive differences between men and women: A comparison of patients with schizophrenia and healthy volunteers. *Schizophrenia Research*, *120*, 234-235.

Maronna, R. A., & Yohai, V. J. (2000). Robust regression with both continuous and categorical predictors. *Journal of Statistical Planning and Inference*, *89*, 197-214.

Masten, C. L., Guyer, A. E., Hodgdon, H. B., McClure, E. B., Charney, D. S., Ernst, M., . . . Monk, C. S. (2008). Recognition of facial emotions among maltreated children with high rates of post-traumatic stress disorder. *Child Abuse & Neglect*, *32*, 139-153.

Mathersul, D., Palmer, D. M., Gur, R. C., Gur, R. E., Cooper, N., Gordon, E., & Williams, L. M. (2009). Explicit identification and implicit recognition of facial emotions: II. Core domains and relationships with general cognition. *Journal of Clinical and Experimental Neuropsychology*, *31*, 278-291.

Matsuzawa, J., Matsui, M., Konishi, T., Noguchi, K., Gur, R. C., Bilker, W., & Miyawaki, T. (2001). Age-related volumetric changes of brain gray and white matter in healthy infants and children. *Cerebral Cortex*, *11*, 335-342.

McClintock, S. M., Husain, M. M., Greer, T. L., & Cullum, C. M. (2010). Association between depression severity and neurocognitive function in major depressive disorder: A review and synthesis. *Neuropsychology*, *24*, 9-34.

McNally, R. J., Kaspi, S. P., Riemann, B. C., & Zeitlin, S. B. (1990). Selective processing of threat cues in posttraumatic stress disorder. *Journal of Abnormal Psychology*, *99*, 398-402.

Mialet, J. P., Pope, H. G., & Yurgelun-Todd, D. (1996). Impaired attention in depressive states: A non-specific deficit? *Psychological Medicine*, *26*, 1009-1020.

Moore, T. M., Reise, S. P., Gur, R. E., Hakonarson, H., & Gur, R. C. (2015). Psychometric properties of the Penn Computerized Neurocognitive Battery. *Neuropsychology*, *29*, 235-246.

Muthén, L. K., & Muthén, B. O. (1998-2013). *Mplus user's guide* (7th ed.). Los Angeles, CA: Author.

Myerson, J., Hale, S., Wagstaff, D., Poon, L. W., & Smith, G. A. (1990). The information-loss model: A mathematical theory of age-related cognitive slowing. *Psychological Review*, *97*, 475-487.

Naranjo, C., Kornreich, C., Campanella, S., Noël, X., Vandriette, Y., Gillain, B., . . . Constant, E. (2011). Major depression is associated with impaired processing of emotion in music as well as in facial and vocal stimuli. *Journal of Affective Disorders*, *128*, 243-251.

Navalta, C. P., Polcari, A., Webster, D. M., Boghossian, A., & Teicher, M. H. (2006). Effects of childhood sexual abuse on neuropsychological and cognitive function in college women. *Journal of Neuropsychiatry and Clinical Neurosciences*, *18*, 45-53.

Neuringer, C. (1964). Rigid thinking in suicidal individuals. *Journal of Consulting Psychology*, *28*, 54-58.

Nock, M. K., Park, J. M., Finn, C. T., Deliberto, T. L., Dour, H. J., & Banaji, M. R. (2010). Measuring the "suicidal mind": Implicit cognition predicts suicidal behavior. *Psychological Science*, *21*, 511-517.

Noël, X., Bechara, A., Dan, B., Hanak, C., & Verbanck, P. (2007). Response inhibition deficit is involved in poor decision making under risk in nonamnesic individuals with alcoholism. *Neuropsychology*, *21*, 778-786.

Ostling, E. W., & Fillmore, M. T. (2010). Tolerance to the impairing effects of alcohol on the inhibition and activation of behavior. *Psychopharmacology*, *212*, 465-473.

Paelecke-Habermann, Y., Pohl, J., & Leplow, B. (2005). Attention and executive functions in remitted major depression patients. *Journal of Affective Disorders*, *89*, 125-135.

Pfefferbaum, A., Mathalon, D. H., Sullivan, E. V., Rawles, J. M., Zipursky, R. B., & Lim, K. O. (1994). A quantitative magnetic resonance imaging study of changes in brain morphology from infancy to late adulthood. *Archives of Neurology*, *51*, 874-887.

Pitman, R. L., Shalev, A. Y., & Orr, S. P. (2000). Posttraumatic stress disorder: Emotion, conditioning, and memory. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 1133-1147). Cambridge: MIT Press.

R Core Team. (2016). *R: A language and environment for statistical computing*. Retrieved from http://www.R-project.org/

Rademaker, A. R., Vermetten, E., Geuze, E., Muilwijk, A., & Kleber, R. J. (2008). Self-reported early trauma as a predictor of adult personality: A study in a military sample. *Journal of Clinical Psychology*, *64*, 863-875.

Ragland, J. D., Glahn, D. C., Gur, R. C., Censits, D. M., Smith, R. J., Mozley, P. D., . . . Gur, R. E. (1997). PET regional cerebral blood flow during working and declarative memory: Relationship with task performance. *Neuropsychology*, *11*, 222-231.

Ragland, J. D., Turetsky, B. I., Gur, R. C., Gunning-Dixon, F., Turner, T., Schroeder, L., . . . Gur, R. E. (2002). Working memory for complex figures: An fMRI comparison of letter and fractal n-back tasks. *Neuropsychology*, *16*, 370-379.

Reeves, D., Kane, R., & Winter, K. (1995). *Automated Neuropsychological Assessment Metrics (ANAM): Test administrators guide Version 3.11* (Report No. NCRF-95-01). San Diego, CA: National Cognitive Recovery Foundation.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*, 667-696.

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, *92*, 544-559.

Revelle, W. (2013). "*psych": Procedures for personality and psychological research* (Ver. 1.6.12). Retrieved from https://cran.r-project.org/web/packages/psych/index.html

Richard-Devantoy, S., Orsat, M., Dumais, A., Turecki, G., & Jollant, F. (2014). Neurocognitive vulnerability: Suicidal and homicidal behaviours in patients with schizophrenia. *Canadian Journal of Psychiatry*, *59*, 18-25.

Roalf, D. R., Ruparel, K., Gur, R. E., Bilker, W., Gerraty, R., Elliott, M. A., . . . Gur, R. C. (2014). Neuroimaging predictors of cognitive performance across a standardized neurocognitive battery. *Neuropsychology*, *28*, 161-179.

Robinson, O. J., & Sahakian, B. J. (2009). A double dissociation in the roles of serotonin and mood in healthy subjects. *Biological Psychiatry*, *65*, 89-92.

Rousseeuw, P. J., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., Verbeke, T., & Maechler, M. (2016). *"robustbase": Basic robust statistics* (Ver. 0.4-5). Retrieved from http://CRAN.R-project.org/package=robustbase

Salthouse, T. A. (2000). Aging and measures of processing speed. *Biological Psychology*, *54*, 35-54.

Salthouse, T. A., Hambrick, D. Z., & McGuthry, K. E. (1998). Shared age-related influences on cognitive and noncognitive variables. *Psychology and Aging*, *13*, 486-500.

Samuelson, K. W., Neylan, T. C., Lenoci, M., Metzler, T. J., Cardenas, V., Weiner, M. W., & Marmar, C. R. (2009). Longitudinal effects of PTSD on memory functioning. *Journal of the International Neuropsychological Society*, *15*, 853-861.

Samuelson, K. W., Neylan, T. C., Metzler, T. J., Lenoci, M., Rothlind, J., Henn-Haase, C., . . . Marmar, C. R. (2006). Neuropsychological functioning in posttraumatic stress disorder and alcohol abuse. *Neuropsychology*, *20*, 716-726.

Saykin, A. J., Gur, R. C., Gur, R. E., Mozley, P. D., Mozley, L. H., Resnick, S. M., . . . Stafiniak, P. (1991). Neuropsychological function in schizophrenia: Selective impairment in memory and learning. *Archives of General Psychiatry*, *48*, 618-624.

Schaie, K. W. (1994). The course of adult intellectual development. *American Psychologist*, *49*, 304-313.

Schotte, D. E., & Clum, G. A. (1987). Problem-solving skills in suicidal psychiatric patients. *Journal of Consulting and Clinical Psychology*, *55*, 49-54.

Schotte, D. E., Cools, J., & Payvar, S. (1990). Problem-solving deficits in suicidal patients: Trait vulnerability or state phenomenon? *Journal of Consulting and Clinical Psychology*, *58*, 562-564.

Shaw, M. E., Moores, K. A., Clark, R. C., McFarlane, A. C., Strother, S. C., Bryant, R. A., . . . Taylor, J. D. (2009). Functional connectivity reveals inefficient working memory systems in post-traumatic stress disorder. *Psychiatry Research*, *172*, 235-241.

Sowell, E. R., Thompson, P. M., Holmes, C. J., Jernigan, T. L., & Toga, A. W. (1999). *In vivo* evidence for post-adolescent brain maturation in frontal and striatal regions. *Nature Neuroscience*, *2*, 859-861.

Stroop, J. R. (1935). Studies of interference in verbal reactions. *Journal of Experimental Psychology*, *12*, 242-248.

Takashima, A., Jensen, O., Oostenveld, R., Maris, E., van de Coevering, M., & Fernandez, G. (2006). Successful declarative memory formation is associated with ongoing activity during encoding in a distributed neocortical network related to working memory: A magnetoencephalography study. *Neuroscience*, *139*, 291-297.

Teachman, B. A., Smith-Janik, S. B., & Saporito, J. (2007). Information processing biases and panic disorder: Relationships among cognitive and symptom measures. *Behavior Research and Therapy*, *45*, 1791-1811.

Thomas, M. L., Brown, G. G., Gur, R. C., Hansen, J. A., Nock, M. K., Heeringa, S., . . . Stein, M. B. (2013). Parallel psychometric and cognitive modeling analyses of the Penn Face Memory Test in the Army Study to Assess Risk and Resilience in Servicemembers. *Journal of Clinical and Experimental Neuropsychology*, *35*, 225-245.

Thomas, M. L., Brown, G. G., Gur, R. C., Moore, T. M., Patt, V. M., Nock, M. K., . . . Stein, M. B. (2015). Measurement of latent cognitive abilities involved in concept identification learning. *Journal of Clinical and Experimental Neuropsychology*, *37*, 653-669.

Trommer, B. L., Hoeppner, J. A. B., Lorber, R., & Armstrong, K. J. (1988). The go-no-go paradigm in attention deficit disorder. *Annals of Neurology*, *24*, 610-614.

Uddo, M., Vasterling, J. J., Brailey, K., & Sutker, P. B. (1993). Memory and attention in combat-related post-traumatic stress disorder (PTSD). *Journal of Psychopathology and Behavioral Assessment*, *15*, 43-52.

Ursano, R. J., Colpe, L. J., Heeringa, S. G., Kessler, R. C., Schoenbaum, M., & Stein, M. B. (2014). The Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *Psychiatry*, *77*, 107-119.

Vallat-Azouvi, C., Weber, T., Legrand, L., Azouvi, P., Allain, P., Etcharry-Bouyx, F., . . . Withaar, F. K. (2007). Working memory after severe traumatic brain injury. *Journal of the International Neuropsychological Society*, *13*, 770-780.

Vasterling, J. J., Duke, L. M., Brailey, K., Constans, J. I., Allain, A. N., & Sutker, P. B. (2002). Attention, learning, and memory performances and intellectual resources in Vietnam veterans: PTSD and no disorder comparisons. *Neuropsychology*, *16*, 5-14.

Verdejo-García, A., Bechara, A., Recknor, E. C., & Pérez-García, M. (2006). Executive dysfunction in substance dependent individuals during drug use and abstinence: An examination of the behavioral, cognitive and emotional correlates of addiction. *Journal of the International Neuropsychological Society*, *12*, 405-415.

von Kluge, S. (1992). Trading accuracy for speed: Gender differences on a Stroop task under mild performance anxiety. *Perceptual and Motor Skills*, *75*, 651-657.

Weiss, E. M., Kohler, C. G., Nolan, K. A., Czobor, P., Volavka, J., Platt, M. M., . . . Gur, R. C. (2006). The relationship between history of violent and criminal behavior and recognition of facial expression of emotions in men with schizophrenia and schizoaffective disorder. *Aggressive Behavior*, *32*, 187-194.

Whitley, E., Deary, I. J., Ritchie, S. J., Batty, G. D., Kumari, M., & Benzeval, M. (2016). Variations in cognitive abilities across the life course: Cross-sectional evidence from *Understanding Society*: The UK Household Longitudinal Study. *Intelligence*, *59*, 39-50.

Williams, J. M. G., & Broadbent, K. (1986). Distraction by emotional stimuli: Use of a Stroop task with suicide attempters. *British Journal of Clinical Psychology*, *25*, 101-110.

Williams, J. M. G., Mathews, A., & MacLeod, C. (1996). The emotional Stroop task and psychopathology. *Psychological Bulletin*, *120*, 3-24.

Wood, A. C., Rijsdijk, F., Johnson, K. A., Andreou, P., Albrecht, B., Arias-Vasquez, A., . . . Kuntsi, J. (2011). The relationship between ADHD and key cognitive phenotypes is not mediated by shared familial effects with IQ. *Psychological Medicine*, *41*, 861-871.

Yakovlev, P. L., & Lecours, A. R. (1967). The myelogenetic cycles of regional maturation of the brain. In A. Minkowski (Ed.), *Regional development of the brain in early life* (pp. 3-70). Philadelphia, PA: F.A. Davis.