Clinical Utility and Analysis of the Run-Roll-Aim Task: Informing Return-to-Duty Readiness Decisions in Active-Duty Service Members

Julianna H. Prim; Oleg V. Favorov†; Amy S. Cecchini‡; Matthew R. Scherer§; Margaret M. Weightman ‖; Karen L. McCulloch¶*

ABSTRACT Introduction: The Assessment of Military Multitasking Performance (AMMP[1]) consists of six dual-task and multitask military-relevant performance-based assessments which were developed to provide assistance in making return-to-duty decisions after concussion or mild traumatic brain injury (mTBI.) The Run-Roll-Aim (RRA) task, one component of the AMMP, was developed to target vulnerabilities following mTBI including attention, visual function, dynamic stability, rapid transition, and vestibular function. One aim of this study was to assess the known-group and construct validity of the RRA, and additionally to further explore reliability limitations reported previously. Materials and Methods: A cross-sectional study consisting of 84 Active Duty service members in two groups (healthy control – HC and individuals experiencing persistent mTBI symptoms) completed neurocognitive tests and the RRA. The RRA task requires a high level of mobility and resembles military training activities in a maneuver that includes combat rolls, fast transitions, obstacle avoidance, and visual search. Observational and inertial sensor data were compared between groups and performance across four trial times was compared within groups. Correlations between RRA results and neurocognitive test scores were analyzed. Results: Simple observational measures (time, errors) did not differ between groups. Spectral power analysis of the inertial sensor data showed significant differences in motor performance between groups. Within group one-way ANOVAs showed that in HC trial 1, time was significantly different than trials 2,3 and 4 ($F(3,47) = 4.60$, $p < 0.01$, Tukey HSD $p < 0.05$) while the mTBI group showed no significant difference in time between trials. During testing individuals with mTBI were less likely to complete the multiple test trials or required additional rest between trials than HCs ($\chi^2 = 10.78$, $p < 0.01$). Small but significant correlations were seen with two neurocognitive tests of attention and RRA performance time. Conclusion: While observational scores were not sensitive to group differences, inertial sensor data showed motor performance on the forward run, combat roll, and backward run differed significantly between groups. The RRA task appeared challenging and provoked symptoms in the mTBI group, causing 8 of 33 mTBI participants to stop the task or require additional rest between trials while none of the HC participants had to stop. Individuals with mTBI demonstrated slower learning of the complex motor sequence compared to HCs who had significant improvement after one trial of RRA. Complex novel training maneuvers like RRA may aid clinicians in informing return to duty decisions.

INTRODUCTION

Since 2000, over 383,000 Department of Defense (DoD) service members (SM) have sustained traumatic brain injury (TBI) with 82.4% of these cases classified as mild (mTBI) or concussion.[1] Although post-concussive symptoms typically resolve within 10–14 days following injury,[2,3] persistent deficits, that may affect complex, duty-relevant task performance has not been extensively studied. Therapists in military medical facilities are challenged to objectively evaluate a range of neurocognitive, sensorimotor, and somatic impairments associated with mTBI when making return to duty (RTD) recommendations.[4]

Following sports-related concussion, return to play decisions are made based on comparison of pre-injury balance, cognition, and symptom reporting to post-injury performance.[5–7] While some smaller, specialized military units have adopted this baseline testing approach, baseline testing for all service members is not feasible given the time, personnel and resource demands associated with obtaining such measures in Brigade sized units ranging in size from 4,000 to 5,000 personnel. Across MTFs (or across the DoD), clinicians assess duty readiness using validated subjective and objective measures. Clinical measures are prone to ceiling effects in pre-morbidly high functioning military personnel and are often validated in civilian populations that may include adolescents or older adults.[8,9] Self-report symptom

*Curriculum in Human Movement Science, Department of Allied Health Sciences, University of North Carolina at Chapel Hill, 3070 Bondurant Hall, 321 S. Columbia St. Chapel Hill, NC 27599-7135.

†Department of Biomedical Engineering, University of North Carolina at Chapel Hill, NC 27599-7575.

‡Intrepid Spirit Clinic, Contractor for The Geneva Foundation, Fort Bragg, NC 28310.

§Clinical and Rehabilitative Medicine Research Program, U.S. Army Medical Research and Materiel Command, Fort Detrick, MD 21702-5012.

‖Courage Kenny Research Center, Minneapolis, MN 55407.

¶Division of Physical Therapy, Department of Allied Health Sciences, School of Medicine, University of North Carolina at Chapel Hill, NC 27599-7135.

The views expressed are solely those of the authors and do not reflect the official policy or position of the US Army, US Navy, US Air Force, the Department of Defense, or the US Government.

doi: 10.1093/milmed/usy425

reporting as a measure is known to be somewhat unreliable due to under or over reporting, based on operational needs, command pressure, or other aspects of warrior culture and demands that result in stressors unique to a military population.[10] Symptom minimization is an especially concerning situation given the potential for further harm to the SM and others in complex, dangerous, and kinetic environments.

The use of military-relevant complex tasks targeting multiple domains of function in RTD assessment shows promise for improving prognostic accuracy by minimizing ceiling effects associated with single domain measures.[4] While postural and dynamic instability are typically observed in the acute stages post-concussion,[6,7] these may be less evident sub-acutely once gross sensorimotor performance has normalized. Efforts to improve the sensitivity of dynamic stability assessments include the use of dual task walking paradigms to challenge available brain resources.[11–16] Detection of subtle differences in gait and postural sway post-concussion have been demonstrated in laboratory settings.[17,18] However, some technology dependent approaches lack clinical feasibility sufficient for widespread use. Similarly, isolated measures of postural stability may have limited utility in detecting movement dysfunction beyond the acute post-concussive phase without the benefit of a baseline assessment or operationally feasible instrumentation.[6,19] Use of challenging tasks required for tactical maneuvers (e.g., running, obstacle avoidance, diving, and rolling) offer an alternative paradigm with clear face validity for SM.[4,8,20]

In recent years, a multidisciplinary group of military and civilian clinician scientists developed novel dual and multitask test components that integrate SM competencies to challenge known mTBI-related vulnerabilities.[21,22] The Assessment of Military Multitasking Performance (AMMP) was developed to assist in military RTD decision-making following concussion by challenging common mTBI impairments in military relevant dual- and multitask scenarios.[20,23] Six AMMP test items were developed in an iterative manner to assure that each task could be tested and scored reliably.[23] Results of the AMMP study have been summarized for the global test battery,[20,21] but results of individual test components are shared in separate publications.[21,22] The purpose of this paper is to report the construct and known-group validity findings related to the Run Roll Aim (RRA) AMMP component, and to further analyze reliability of scoring limitations shared previously.[23]

METHODS

This assessment development, known-group study was conducted at Fort Bragg, NC. The study received approval from the Womack Army Medical Center (WAMC) Institutional Review Board and all participants provided informed consent.

Participants

Participants consisted of two groups: healthy controls (HC) and patients with mTBI. All participants were active duty service members (ADSM) aged 18–42 years stationed at Fort Bragg. Participants with persistent post-concussive symptoms from a mTBI occurring between 2 weeks to 2 years prior to testing were recruited from a clinical population receiving outpatient rehabilitation services at the WAMC TBI Clinic. HC participants were recruited via briefings or flyers. All HC participants were eligible to deploy and were excluded if they reported a concussion within the 12 months preceding enrollment. All participants were able to perform everyday activities that required moderate exertion (Borg Rating of Perceived Exertion between 12 and 14)[24] and all reported an ability to tolerate a 3-hour testing session with breaks if needed. Participants were excluded if they had a duty-limiting medical condition that prevented continuous activity for up to 30 minutes; a history of psychiatric disorder; moderate or severe brain injury; penetrating head injury; or visual or hearing deficits that prevented participation in testing.

Measures and Procedures

Participants completed a single test session lasting up to 3 hours that began with an intake questionnaire followed by neurocognitive tests. All AMMP subtests were administered by a physical or occupational therapist examiners in a counterbalanced sequence in an effort to minimize bias from order effects.

Intake questionnaires included demographic information (age, ethnicity, education level, first language, and learning disabilities) and military history (pay grade, length in military service, current military occupational specialty, and number and duration of deployments) as well as symptom self-report questionnaires. The Post Traumatic Stress Disorder Checklist-Civilian (PCL-C)[25] measured stress-related symptoms while the Neurobehavioral Symptom Inventory (NSI)[26] measured common concussion related symptoms. Current pain and energy level, other injury and behavioral health history (recent sleep history, hearing impairments), and a question about perceived readiness to be deployed to a combat zone in 72 hours were also collected.

The neurocognitive tests administered were the *Neuropsychological Assessment Battery (NAB)*[27] (digits forward, digits backward, numbers, and letters), Comprehensive Trail-Making Test (CTMT),[28] the Test of Memory Malingering (TOMM),[29] Simple Reaction Time (SRT),[30] Tower of Hanoi, and the Wide Range Achievement Test Version 4 (WRAT-4) Reading Test[31] as an estimate of educational background and intelligence. All neurocognitive measures used have known sensitivity to cognitive vulnerabilities associated with mTBI[21] and could be administered in a timely manner by study examiners. In order to avoid repeated testing and to limit test burden for mTBI participants, previously completed cognitive tests (NAB numbers and letters, DF/DB, CTMT, TOMM) were obtained from the medical record with testing done in the preceding weeks to months, while all HC participants completed neurocognitive tests during their AMMP testing session.

Run-Roll-Aim Task

The RRA task (outlined in Supplementary Fig. 1) is a high level mobility and agility task designed to challenge dynamic stability, target acquisition, and tolerance for rolling in an operationally relevant test condition while carrying a simulated weapon (Bluegun). Prior to each of four trials, participants were cued as to which visual targets (odd or even numbers) to attend to on a computer screen on the floor and visible from the RRA course. These numbers could only be viewed by using a near focus scope (BARSKA Blueline 10 × 40 Monocular) mounted on the mock weapon. The computer display was advanced by the examiner with a remote to guide the task sequence. Initial combat roll direction was cued on the computer screen with a large letter (R or L) and an arrow. Participants were instructed to roll in the direction of the letter, a less automatic cue than an arrow, intending to induce a Stroop effect during incongruent conditions (arrow pointing left with displayed "R"). Congruent and incongruent cues were counterbalanced in each direction during the four trials. Subjects were given one practice trial which included only congruent cues.

The SM walked through the RRA course with verbal and computer screen instructions prior to completing a practice trial to ensure that all the components of the task were performed correctly. If the participant demonstrated more than one error during the practice trial, an additional congruent practice trial was performed. Four test trials followed, with a brief rest between trials to allow for rater scoring of the trial. If requested, additional rest was allowed, as needed. In most instances, the SM made the decision whether to continue testing, but the examiner discontinued testing when neccessary (i.e., participant demonstrated an increasing pattern of symptoms such as degradation in balance, slowed movement, observable discomfort with the task, report of increased visual blurring). Observational data for each trial included errors in course completion, Stroop effects (errors in following directional arrow), errors in visual target identification and time for task completion,. The Stroop effect was characterized by three possible errors: hesitation (a delay in response of one second or more), self-correction (initation of roll in wrong direction that was self-corrected), and rolls wrong direction. In order to accurately judge these responses, examiners had to carefully observe test performance.

STATISTICAL ANALYSIS

Data were entered and verified using an on-line Research Electronic Data Capture 209 (REDCap)[32] and password protected Excel database. All statistical analyses used SPSS V22.0 (IBM, Inc) or R (R Foundation for Statistical Computing, Vienna, Austria). Descriptive analyses were performed on demographic and military history characteristics. A subset of our sample, evaluated by two examiners ($n = 26$), allowed for evaluation of inter-rater reliability using the Kappa statistic. Previously reported findings showed inter-rater reliability was acceptable (ICC > 0.93) for course completion time and number of correct and incorrect odd/even numbers, but task error ratings and judgment of responses to the directional Stroop effect were below acceptable reliability standards (ICC = 0.64, 95% confidence interval 0.13–0.92).[23] Construct validity was assessed using Pearson correlation coefficients that included RRA metrics and neurocognitive test scores. Construct validity analysis included both HC and mTBI participants. A sample size over 80 provided 80% power to detect a correlation for expected convergence at a minimum of 0.30 at a two-sided alpha of 0.05. Known-group validity was evaluated by comparing RRA scores for time, numbers identified and errors between HC and those with mTBI. Independent *t*-tests were used to test for significant differences between groups for continuous data if it was normally distributed. If non-normal distribution was determined based on the Wilks–Sharpiro test, a non-parametric Mann–Whitney *U* test with alpha of 0.05 was used to evaluate between group differences. Post hoc analyses investigated practice effects analyzing the relationship between trial 1 time and subsequent trial times within groups (one-way ANOVA, Tukey-HSD, $p < 0.05$). Post hoc, a Chi-square test was used to test for differences in task completion (all four trials with no need for extended rest) with an alpha of 0.05. Extended rest was defined by participant requesting extra time between trials and confirmed by the recorded start and end time of RRA.

RRA Motor Performance Analysis

During the RRA test, each subject wore lumbar and forehead triaxial accelerometry sensors (*NexGen Ergonomics Inc.*) attached using adjustable waist- and headbands. The continuous time series output values (100 Hz sampling rate) were used as objective quantitative measures of each subject's motor performance. Analysis was performed on the magnitudes (Euclidean norms) of the torso and head triaxial acceleration vectors, converted from the time domain to the frequency domain using the Fast Fourier Transform (FFT) procedure and expressed as power spectra.

Although the entire power spectrum can be used to characterize a subject's performance on the RRA test, a large majority of the constituent frequencies carry very little power and can be discarded without any significant loss of information. Therefore, the statistical analysis was performed on a subset of frequencies that exhibited the highest average power in the dataset of all the subjects in the study. Each of the chosen frequencies was autoscaled by subtracting its mean and dividing by its standard deviation. This set of normalized frequencies, considered as a "performance" vector, offered a distilled quantitative description of the subject's motor performance. Accordingly, the performance of any given subject was treated as a point in the "performance space" defined by the selected subset of frequencies. The null hypothesis, that the population means of performance

vectors of the HC and mTBI samples are equal, was tested using Hotelling's *T*-square multivariate test.[33]

RESULTS

Descriptive Analyses

Eighty four active duty SM (51 HC, 33 mTBI) were enrolled in this component of the study with one participant being excluded from analysis due to incomplete assessments. The mean service time was 7.1 years (SD = 5.6), with 58 SMs having deployment history to Iraq or Afghanistan. The average number of lifetime mTBI was 4.0 (SD = 7.4). Participants in the mTBI group had a higher prevalence of pain (mTBI: 79.4%, $n = 27$; HC: 41.5%, $n = 15$) and self-reported Post Traumatic Stress (PTS) (mTBI: 20.5%, $n = 7$; HC: 5.8%, $n = 3$). Other demographic characteristics are listed in Table I. Participants with mTBI were significantly younger, had fewer years of education and military service, lower reading levels, and reported more stress than HC participants.

Inter-rater Reliability

A subset of 26 participants (19 SM with mTBI, 7 HC SM) that completed the RRA scored by the same two examiners were included in the post hoc analysis investigating inter-rater reliability (Table II). The score for "rolled in the wrong direction" had acceptable IRR (mean kappa 0.89, 0.78–1.0), while "Hesitate" (0.28) and "Self-Correct" (0.35) error scores had unacceptably low IRR. In further analyses, these two errors were combined into one category (Table II). Although the IRR improved, it was still inadequate.

Known-Group Validity

There were significant group differences in the ability to complete the task. ($\chi^2 = 10.78$, $p < 0.01$). Eight of 33 (24%) mTBI participants were either unable to complete all four trials or required additional rest between trials due to symptom provocation. This response did not occur with any of the HC participants. There were no significant group differences in summated trial performance time, number of correct visual targets identified, or Stroop effects committed in any of the four trials (Table III). An analysis of variance (ANOVA) within group on task completion time revealed significant variation in how trial 1 related to trial 2, 3, and 4 (Fig. 1) for the HC group only ($F(3,47) = 4.60$, $p < 0.01$). A post hoc Tukey HSD showed that trial 1 differed from trials 2, 3, and 4 while the final three trials did not differ, indicating a possible rapid learning effect with practice. A practice effect would be expected in both groups as participants became more familiar with the task sequence, however the mTBI group did not have significant differences between any of the trial times ($F(3,29) = 1.57$, $p > 0.2$). Only mTBI participants who completed all four trials were included in this analysis.

TABLE I. Demographic Characteristics of Study Participants

Characteristic	Healthy Controls $n = 50$	mTBI $n = 33$	*p*-Value
Age in years	**Mean (SD)**	**Mean (SD)**	0.001[a]
	30.2 (6.1)	26.2 (5.2)	
Sex	***n* (%)**	***n* (%)**	0.112[b]
Women	10 (20)	2 (6.1)	
Men	40 (80)	31 (93.9)	
Race/ethnicity			0.273[b]
Caucasian	25 (50)	21 (63.6)	
African American	15 (30)	4 (12.1)	
Hispanic	6 (12)	3 (9.1)	
Asian	3 (6)	3 (9.1)	
Other	1 (2)	2 (6.1)	
Education			0.008[b]
High school	6 (12)	6 (18.2)	
Trade school	1 (2)	2 (6.1)	
Some college	20 (40)	22 (66.7)	
Bachelor's degree	17 (34)	3 (9.1)	
Advanced degree	6 (12)	0 (0.0)	
Years in military	**Mean (SD)**	**Mean (SD)**	0.004[a]
	8.4 (5.5)	5.2 (4.6)	
Reading level:	61.1 (5.5)	58.1 (6.0)	0.018[a]
WRAT-4			
(raw reading)			
Stress symptoms:	22.2 (8.2)	34 (14.7)	<0.001[c]
PCL-C sum	**Median (range)**	**Median (range)**	
	19 (17–63)	32 (17–73)	

Note: for PCL-C, (HC $n = 50$, mTBI $n = 31$). [a]*t*-Test, [b]Chi-Square, [c]Mann–Whitney *U*.

TABLE II. Inter-rater Reliability (IRR) Analysis for Stroop Effect and Other Errors in the Run Roll Aim

Scoring Item (Metrics)	Trial	Inter-rater Reliability
Hesitate[a]	1	0.669
A 1 second or longer delay on Stroop	2	0.904
response	3	NA
	4	0.28
Self-correct[a]	1	0.345
Started to roll in wrong direction then	2	NA
self-corrected to the right direction	3	NA
	4	0.882
Rolls wrong[a]	1	0.898
A roll in the wrong direction as indicated	2	1.0
by the Stroop task	3	NA
	4	0.778
Other errors	1	0.660
	2	1.0
	3	0.686
	4	1.0
Hesitate and self-correct[b]	1	0.639
	2	0.905
	3	NA
	4	0.407

IRR values for subset population: $n = 26$ (19 mTBI, 7 HC), all trials had two raters.
[a]Specific Stroop error. [b]Exploratory combination of error categories. All calculations used the Kappa statistic, (NA – neither rater judged the presence of the error in any participants and Kappa statistic could not be calculated.)

TABLE III. Run-Roll Aim Multitask Known-Group Analysis

Metrics	HC $n = 50$ Mean (SD) Median (Range)	mTBI $n = 30$ Mean (SD) Median (Range)	p-Value
Trial 1 time (minutes)	0.80 (0.24) 0.74 (0.54–1.8)	0.80 (0.21) 0.74 (0.57–1.4)	0.893
Trial 1 correct	13.5 (0.70) 14 (11–14)	13.2 (1.0) 13 (9–14)	0.159
Trial 1 errors	1.9 (1.3) 2 (0–6)	2.4 (2.3) 2 (0–11)	0.317
Trial 1 Stroop effect	n, error (n, no error)	n, error (n, no error)	0.234[b]
Hesitation	33 (18)	17 (16)	
Self-correction	11 (40)	5 (28)	0.575[b]
Rolls wrong direction	15 (36)	9 (24)	0.706[b]
Total time (minutes) (four trials)	2.9 (0.59) 2.8 (1.9–4.2)	3.0 (0.57) 3.0 (2.2–4.3)	0.515[a]
Total correct (four trials)	4.9 (4.6) 4 (0–25)	5.4 (4.4) 4 (0–16)	0.438
Total errors (four trials)	4.9 (4.6) 4 (0–25)	5.4 (4.4) 4 (0–16)	0.438
Total (four trials) Stroop effect	n, error (n, no error)	n, error (n, no error)	0.459[c]
Hesitation	59 (145)	41 (84)	
Self-correction	22 (182)	9 (116)	0.281[c]
Rolls wrong direction	22 (182)	11 (114)	0.562[c]

Analyses were Mann–Whitney U unless otherwise noted, [a]t-test, [b]Fisher's exact test, [c]Chi-square test. These observational metrics were not significantly different between Healthy Control (HC) and mTBI groups.

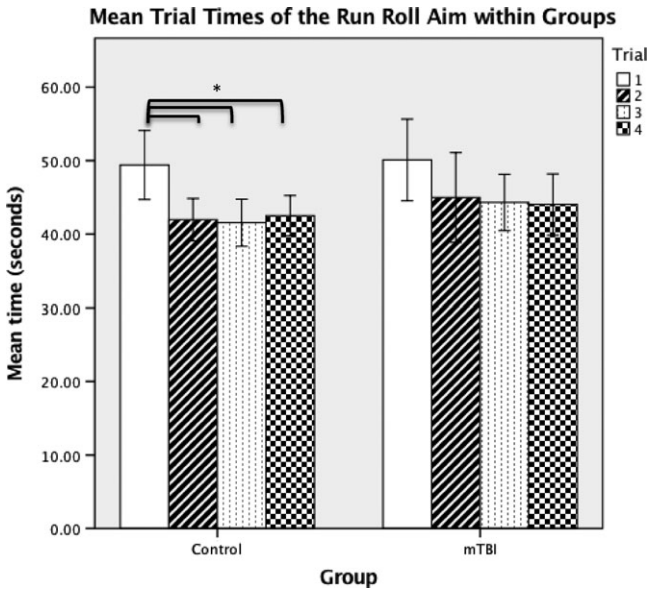

FIGURE 1. Mean perforamnce times on the RRA by Trials. Error Bars: ±2 SE. Analysis used one-way ANOVA and Tukey *post hoc* for comparisions between trials. *$p < 0.05$.

HC and mTBI Group Differences in RRA Motor Performance

Each sensor was attached to the head or trunk in a standardized approach, with one axis aligned with the front-back direction of the body, the second axis aligned with the lateral direction, and the third axis aligned vertically. While the 3D directional acceleration is likely to be a rich source of movement information, this paper confines the analysis to acceleration vector magnitude (i.e., the Euclidean length) (Fig. 2). As Figure 2A shows, the execution of the RRA task involved three periods of motor activity (forward running and combat roll; lateral shuffle; and combat roll and backward running) separated by two periods of near complete immobility (during which the subject searched for visual targets through a scope). Each period of movement was analyzed separately. The analysis of the forward run period was performed on the first 512 time bins (covering 5.12 seconds) of that period. The analysis of the lateral shuffle period was performed on 512 time bins centered on the midpoint of that period. The analysis of the backward run period was performed on the last 512 time bins of that period.

Each series was converted from the time domain to the frequency domain using the Fast Fourier Transform (FFT) procedure and expressed as a power spectrum (Fig. 2B). For each movement period, the power spectra of all four RRA trials for all subjects were averaged and the frequency bins were sorted in descending order (Fig. 2C). In Figure 2C it appears that for each plot, the first 3–5 frequency bins with the highest power form an outstanding group. Therefore, all but the first four frequency bins with the highest power were discarded, and the remaining bins were autoscaled. The four autoscaled bins from the head accelerometer and the four autoscaled bins from the torso accelerometer were combined into an 8D "performance" vector. To determine whether such performance vectors are sensitive to mTBI, Hotelling's

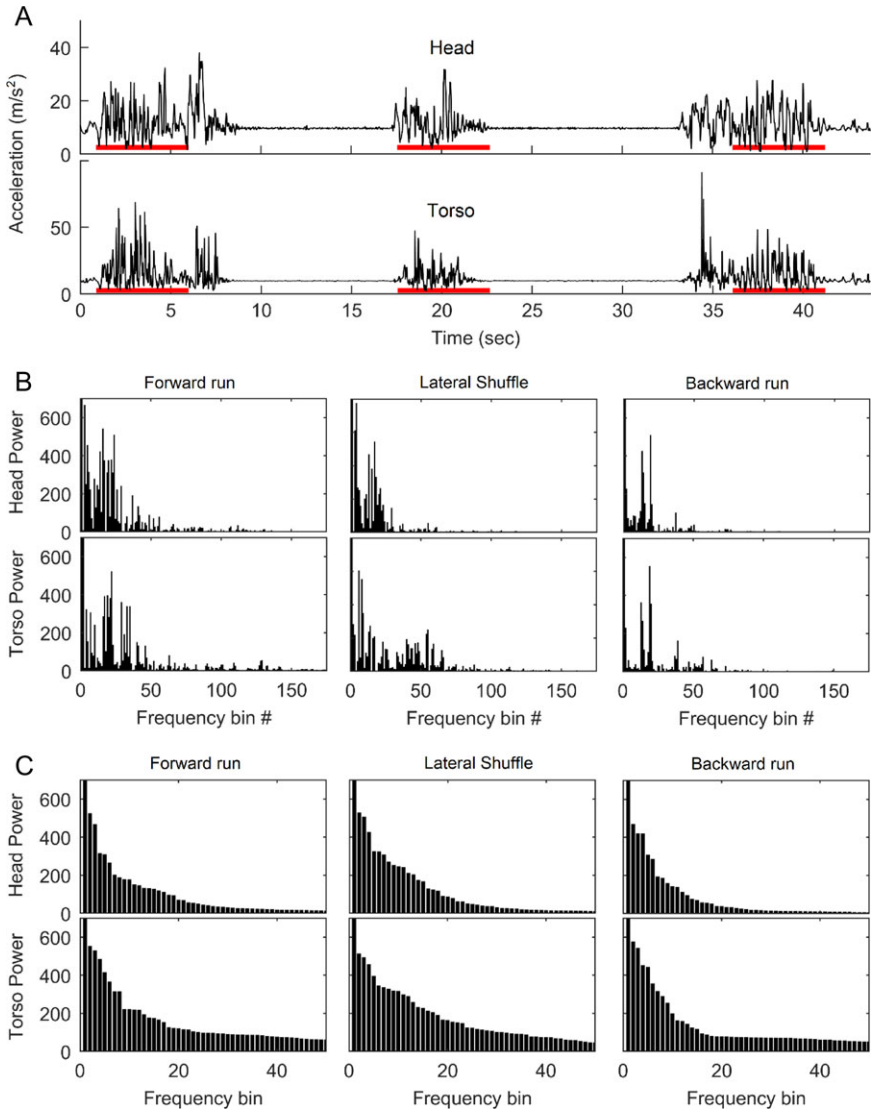

FIGURE 2. Motor performance on the RRA test. (A) An exemplary time series of the head and torso 3-axial accelerometer readings (plotted as the length of the 3D acceleration vector) recorded while a particular subject was going through the test. Red horizontal bars indicate 512-bin time periods selected for the frequency analysis. (B) FFT power spectra of the time periods selected in panel A. (C) Average spectral power, with the frequency bins sorted in the descending order.

multivariate T-square statistic[33] was used to test the null hypothesis that the population means of performance vectors of the HC and mTBI samples are equal at $\alpha = 00.05$. The T-square statistics of the forward run period [$T^2 = 2.91$ ($p = 0.0037$)], the lateral shuffle period [$T^2 = 5.45$ ($p = 0$)], and the backward run period, [$T^2 = 2.81$ ($p = 0.0051$)] all had $p < 0.01$ and after Bonferroni correction for multiple testing the null hypothesis is rejected for all three periods. Therefore the tested population of mTBI subjects was significantly different in their performance on the RRA test from the HC population.

Construct Validity

The RRA task demonstrated a small but significant correlation with NAB Numbers-Letters (a measure of memory/attention) for total time for each trials, as well as aggregate time (sum of trials 1 to 4) (Table IV). The RRA task correlated with the CTMT (measure of executive function and attention) on total time for trials 2 and 4, and aggregate time. RRA was correlated with NAB-Digits Backward (a measure of memory/attention), but only for trial 4 time and post assessment SRT.

DISCUSSION

The AMMP was designed to challenge SM performance to reveal post-concussion functional deficits inconsistent with readiness to RTD.[2,4,5] Findings revealed significant task tolerance difference between HC and mTBI groups with 4 mTBI participants (of 33 total) requiring extended rest between conditions and 4 more unable to complete all trials

TABLE IV. Run Roll Aim Pearson Correlation Coefficients With Neurocognitive Measures

	Trial 1 Total Time	Trial 2 Total Time	Trial 3 Total Time	Trial 4 Total Time	RRA Aggregate Time	RRA Aggregate Correct
NAB numbers-letters: Part D	−0.31 (−0.5, −0.08)	−0.4 (−0.58, −0.19)	−0.37 (−0.56, −0.16)	−0.45 (−0.62, −0.25)	−0.44 (−0.61, −0.23)	0.01 (−0.22, 0.24)
	$n = 73, p = 0.009$*	$n = 73, p < 0.001$*	$n = 72, p = 0.001$*	$n = 73, p < 0.001$*	$n = 73, p < 0.001$*	$n = 73$ $p = 0.956$
NAB Digit: backward	−0.15 (−0.37, 0.09)	−0.13 (−0.35, 0.11)	−0.11 (−0.33, 0.13)	−0.3 (−0.5, −0.07)	−0.19 (−0.4, 0.05)	−0.01 (−0.24, 0.23)
	$n = 71$ $p = 0.215$	$n = 71$ $p = 0.288$	$n = 70$ $p = 0.378$	$n = 71, p = 0.011$*	$n = 71$ $p = 0.118$	$n = 71$ $p = 0.956$
CTMT: composite index	−0.28 (−0.48, −0.05)	−0.31 (−0.5, −0.08)	−0.23 (−0.44, 0)	−0.41 (−0.58, −0.2)	−0.35 (−0.54, −0.13)	0.07 (−0.16, 0.3)
	$n = 73, p = 0.018$*	$n = 73, p = 0.008$*	$n = 72, p = 0.049$*	$n = 73, p < 0.001$*	$n = 73, p = 0.002$*	$n = 73$ $p = 0.528$
SRT: baseline	0.06 (−0.16, 0.28)	0.2 (−0.02, 0.41)	0.11 (−0.12, 0.32)	0.06 (−0.16, 0.28)	0.11 (−0.11, 0.31)	0.2 (−0.02, 0.41)
	$n = 80$ $p = 0.587$	$n = 80$ $p = 0.069$	$n = 79$ $p = 0.347$	$n = 80$ $p = 0.587$	$n = 84$ $p = 0.333$	$n = 80$ $p = 0.069$
SRT: end of testing	0.07 (−0.15, 0.29)	0.29 (0.07, 0.48)	0.19 (−0.03, 0.4)	0.07 (−0.15, 0.29)	0.23 (0.01, 0.42)	0.29 (0.07, 0.48)
	$n = 78$ $p = 0.528$	$n = 78, p = 0.011$*	$n = 77$ $p = 0.092$	$n = 78$ $p = 0.528$	$n = 81, p = 0.043$*	$n = 78, p = 0.011$*
WRAT: standardized score	−0.19 (−0.39, 0.03)	−0.15 (−0.36, 0.07)	−0.13 (−0.34, 0.1)	−0.16 (−0.37, 0.06)	−0.15 (−0.36, 0.07)	−0.2 (−0.4, 0.02)
	$n = 80$ $p = 0.09$	$n = 80$ $p = 0.178$	$n = 79$ $p = 0.259$	$n = 80$ $p = 0.158$	$n = 80$ $p = 0.184$	$n = 80$ $p = 0.078$

Correlations of RRA times and errors with select neurocognitive tests (other tests had non-significant correlations). Values represent r values, (95% CI lower, upper), n = number of participants included in analysis, p-value (*<0.05).

due to obvious symptom exaccerbation. Extended rest was an observational measure defined by participant requesting extra time between trials and represents an important clinical consideration although not defined a priori. In order to detect differences in those with mTBI and HC, inertial sensor measurement was required. Inertial sensor analysis using FFT demonstrated significant differences between groups in participant "performance vector" scores. These scores characterize raw performance within specified kinematic time domains. Duty readiness may be represented as "the vector-sum of relevant military competenies"[4] and as rehabilitation progresses, we hypothesize the performance of mTBI group would move toward the healthy control range.

Although the combination of cognitive and motor challenges in the RRA approximated high level physical performance required of ADSM, simple observational data (trial time, errors) did not distinguish between groups. Human elements of observation are likely contributors to these limitations. The emphasis on rapid motor response that is inherent in military training may explain why the directional Stroop cue did not differentiate between those with mTBI and the HC participants. While the directional Stroop effect povided a method for assessing difficulty with inhibition, expected following mTBI, there are limited real-life situations where such an artificial effect occurs. Although experimental tasks of cognitive control are hypothesized to be sensitive to mTBI related impairments,[34] several studies suggest that individuals with chronic mTBI do not show performance deficits on these tasks.[35–38] In addition, there were difficulties in reliably judging hesitations and self-corrections related to the directional Stroop effect, and since this aspect of the test did not differentiate between groups (Table III), retention of this element is not warranted. Obstacle avoidance was also rarely problematic for participants (1 error on 320 trials), so its inclusion does not add specific value for this population.

The practice effect difference between groups may have important implications for RTD, given there is often a need to rapidly master soldiering skills. The HC group showed a significant decrease in trial time (trial 1 to trial 2) while the mTBI group did not. Previous studies have found a novelty effect in individuals with mTBI[39] meaning the learning of a new task requires more practice to master. Our results support this learning delay as the mTBI group showed a trend toward improvement with successive trials (Fig. 1), but statistically trial 1 did not differ from trials 2–4. Evidence from this study shows that components of the RRA has reasonable psychometric properties and may have clinical value. Inter-rater reliability was adequate based on predetermined levels of acceptability (>0.85) in all measures except observational error ratings.[23]

Weak to moderate correlations (i.e., 0.3–0.5) between RRA sub-scores and neurocognitive domains involving attention and reaction time were expected based on construct validity of other AMMP multitask subtests[20,21] and the

specific requirements of the RRA. Correlations between the RRA and NAB Numbers-Letters, NAB Digit: Backward, and CTMT were confirmed (Table IV), but the clinical importance is questionable. The NAB Numbers-Letters and CTMT involved psychomotor speed, information processing speed, selective attention, and resistance to distraction which were required during the RRA. The NAB-Digits- Backward involves working memory and attention which is not as taxed during the RRA reflected by the weaker and insignificant correlations. The significant but relatively small correlations were expected between the RRA and domain-specific neurocognitive measures, given the combined nature of multiple factors in RRA versus the discrete nature of neurocognitive measures. In future studies, construct validity of the RRA could be measured via comparisons to other multitask assessments.

The RRA task described in this report required specialty equipment (computer display and examiner remote to drive task components, relatively expensive laboratory grade high-quality wireless inertial sensors) and complex analyses to detect movement differences between mTBI and HC participants. A simpler task design that may be easier to administer clinically while capitalizing on the elements of the test that appear to appropriately challenge individuals with mTBI, including dynamic movement transitions (stand to prone, combat roll) and visual search elements, would be a logical next step. An alternative visual search task could be a horizontal strip of random letters and numbers posted in view of the mat used for combat rolls. Prior to each of the four trials, the direction of the combat roll and a visual target (odd or even numbers, vowels or consonants) could be provided. Pilot testing by our group with this simpler version suggests it can be easily administered to detect potential problems with similarly challenging military training activities. While movement differences were identified during forward and backward running as well as the combat roll maneuver, the use of inertial sensors was required. A more clinically feasible means to collect and interpret this data is the focus of a currently funded Department of Defense study (Grants.gov ID: GRANT12296682).

LIMITATIONS

Significant between-group differences for years of education, military service, and Wide Range Achievement Test scores (used as a measure of intelligence) may have contributed to bias in study findings, limiting interpretation of results although none of these attributes likely contributed to an ability to complete this novel task. The concurrent validity findings related to the neurocognitive tests may have been limited by the variability in the testing timeframe. Not all tests were completed on the same day for subjects with mTBI. Convenience sampling and examiners not being blind to each subject's mTBI status may have also introduced bias. Participants with mTBI also had more significant self-reported symptoms of post-traumatic stress and pain, which may have influenced performance on the RRA. All participants with mTBI were at least 2 months post injury with chronic symptom complaints and were being followed in the Fort Bragg TBI Pipeline. Specific information regarding the focus of mTBI rehabilitation services for participants was not collected, therefore it is possible that the impairments that were present in this group of SMs with mTBI were not the deficits (i.e., vestibular complaints) targeted in RRA.

CONCLUSION

Military service requires superior sensorimotor control under complex conditions. The multi-modal design of the AMMP offers assessments that increases the relevance of required SM performance that may improve the ability for therapists to estimate real world functioning over self-report or single domain concussion assessment metrics. The novel multitask approach of the RRA has strong face validity, it challenges SMs with relevant task elements. Removal of measurement components that did not differentiate between groups will simplify scoring and potentially improve utility in RTD decision making. Although observational measures did not differentiate between groups, the finding that 8 of 33 subjects with mTBI were unable to or had difficulty completing multiple trials supports the notion that further testing and scoring refinement is warranted. Specifically, capturing activity intolerance and learning curve differences may enhance the relevance of a streamlined RRA for duty readiness decisions. Further research is indicated to explore the sensitivity of performance vector analysis as a method for assessing duty readiness and measuring sensorimotor performance. Movement differences between groups, detected by inertial sensors, may provide a valuable means to evaluate SM performance, but will require additional work to facilitate clinical implementation.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Military Medicine* online.

PREVIOUS PRESENTATIONS

American Congress of Rehabilitation Medicine 2015, International Brain Injury Association Congress 2016, American Physical Therapy Assocation IVSTEP Conference 2016, American Physical Therapy Association Combined Sections Meeting 2016, 2017.

FUNDING

Support for the Assessment of Military Multitasking Performance validation study was provided by the United States Army Medical Research Materiel Command (award no. W81XWH-12-2-0070).

ACKNOWLEDGMENTS

The authors would like to thank CAPT Henry McMillan, for his work as site principal investigator at Fort Bragg for his dedication to make this

project a success; Ms. Caroline Cleveland for her help with data acquisition and for serving as research assistant at Fort Bragg, NC. We would like to thank the service members of USARIEM in Natick, MA and Fort Bragg, NC for their willingness to participate in this research study. We thank Dr Arthur Estrada of USAARL, Fort Rucker, AL and Dr Edward Zambraski of USARIEM, Natick, MA for their support of this project.

REFERENCES

1. DoD Worldwide Numbers for TBI | DVBIC. Available at http://dvbic.dcoe.mil/dod-worldwide-numbers-tbi; accessed August 16, 2018. Updated August 9, 2018.
2. McCrory P, Meeuwisse WH, Dvořák J, et al: 5th international conference on concussion in sport (Berlin). Br J Sports Med 2017; 51(11): 837. doi:10.1136/bjsports-2017-097878.
3. Belanger HG, Curtiss G, Demery JA, Lebowitz BK, Vanderploeg RD: Factors moderating neuropsychological outcomes following mild traumatic brain injury: a meta-analysis. J Int Neuropsychol Soc 2005; 11 (3): 215–27. doi:10.1017/S1355617705050277.
4. Scherer MR, Weightman MM, Radomski MV, Davidson LF, McCulloch KL: Returning service members to duty following mild traumatic brain injury: exploring the use of dual-task and multitask assessment methods. Phys Ther 2013; 93(9): 1254–67. doi:10.2522/ptj.20120143.
5. Leddy JJ, Baker JG, Willer B: Active rehabilitation of concussion and post-concussion syndrome. Phys Med Rehabil Clin N Am 2016; 27(2): 437–54. doi:10.1016/j.pmr.2015.12.003.
6. Guskiewicz KM: Postural stability assessment following concussion: one piece of the puzzle. Clin J Sport Med 2001; 11(3): 182–9.
7. King LA, Horak FB, Mancini M, et al: Instrumenting the balance error scoring system for use with patients reporting persistent balance problems after mild traumatic brain injury. Arch Phys Med Rehabil 2014; 95(2): 353–9. doi:10.1016/j.apmr.2013.10.015.
8. McCulloch KL, Goldman S, Lowe L, et al: Development of clinical recommendations for progressive return to activity after military mild traumatic brain injury: guidance for rehabilitation providers. J Head Trauma Rehabil 2015; 30(1): 56–67. doi:10.1097/HTR.0000000000000104.
9. Gottshall KR, Gray NL, Drake AI, Tejidor R, Hoffer ME, McDonald EC: To Investigate the influence of acute vestibular impairment following mild traumatic brain injury on subsequent ability to remain on activity duty 12 months later. Mil Med 2007; 172(8): 852–7. doi:10.7205/MILMED.172.8.852.
10. Hettich T, Whitfield E, Kratz K, Frament C: Case report: use of the Immediate Post Concussion Assessment and Cognitive Testing (ImPACT) to assist with return to duty determination of special operations soldiers who sustained mild traumatic brain injury. J Spec Oper Med 2010; 10(4): 48–55.
11. Catena RD, van Donkelaar P, Chou L-S: Different gait tasks distinguish immediate vs. long-term effects of concussion on balance control. J Neuroeng Rehabil 2009; 6: 25. doi:10.1186/1743-0003-6-25.
12. Catena RD, van Donkelaar P, Chou L-S: The effects of attention capacity on dynamic balance control following concussion. J Neuroeng Rehabil 2011; 8: 8. doi:10.1186/1743-0003-8-8.
13. Vallée M, McFadyen BJ, Swaine B, Doyon J, Cantin J-F, Dumas D: Effects of environmental demands on locomotion after traumatic brain injury. Arch Phys Med Rehabil 2006; 87(6): 806–813. doi:10.1016/j.apmr.2006.02.031.
14. Howell DR, Osternig LR, Chou L-S: Dual-task effect on gait balance control in adolescents with concussion. Arch Phys Med Rehabil 2013; 94(8): 1513–20. doi:10.1016/j.apmr.2013.04.015.
15. Howell DR, Osternig LR, Chou L-S: Return to activity after concussion affects dual-task gait balance control recovery. Med Sci Sports Exerc 2015; 47(4): 673–80. doi:10.1249/MSS.0000000000000462.
16. Kleiner M, Wong L, Dubé A, Wnuk K, Hunter SW, Graham LJ: Dual-task assessment protocols in concussion assessment: a systematic literature review. J Orthop Sports Phys Ther 2018; 48(2): 87–103. doi:10.2519/jospt.2018.7432.
17. Alberts JL, Thota A, Hirsch J, et al: Quantification of the balance error scoring system with mobile technology. Med Sci Sports Exerc 2015; 47 (10): 2233–40. doi:10.1249/MSS.0000000000000656.
18. Fino PC, Parrington L, Pitt W, et al: Detecting gait abnormalities after concussion or mild traumatic brain injury: a systematic review of single-task, dual-task, and complex gait. Gait Posture 2018; 62: 157–66. doi:10.1016/j.gaitpost.2018.03.021.
19. Brown HJ, Siegmund GP, Guskiewicz KM, Van Den Doel K, Cretu E, Blouin J-S: Development and validation of an objective balance error scoring system. Med Sci Sports Exerc 2014; 46(8): 1610–6. doi:10.1249/MSS.0000000000000263.
20. Radomski MV, Weightman MM, Davidson LF, et al: Development of a measure to inform return-to-duty decision making after mild traumatic brain injury. Mil Med 2013; 178(3): 246–53. doi:10.7205/MILMED-D-12-00144.
21. Radomski MV, Davidson LF, Smith L, et al: Toward return to duty decision-making after military mild traumatic brain injury: preliminary validation of the charge of quarters duty test. Mil Med 2018; 183(7–8): e214–22. doi:10.1093/milmed/usx045.
22. Scherer MR, Weightman MM, Radomski MV, et al: Measuring soldier performance during the patrol-exertion multitask: preliminary validation of a postconcussive functional return-to-duty metric. Arch Phys Med Rehabil 2018; 99(2S): S79–S85. doi:10.1016/j.apmr.2017.04.012.
23. Weightman MM, McCulloch KL, Radomski MV, et al: Further development of the assessment of military multitasking performance: iterative reliability testing. PLoS One 2017; 12(1): e0169104. doi:10.1371/journal.pone.0169104.
24. Borg GA: Psychophysical bases of perceived exertion. Med Sci Sports Exerc 1982; 14(5): 377–81.
25. Weathers FW, Litz BT, Herman DS, Huska JA, Keane TM: The PTSD Checklist (PCL): Reliability, validity, and diagnostic utility. Annual convention of the international society for traumatic stress studies; October 1993; San Antonio, TX.
26. Cicerone KD, Kalmar K: Persistent postconcussion syndrome: the structure of subjective complaints after mild traumatic brain injury. J Head Trauma Rehabil 1995; 10: 1–17.
27. Zgaljardic DJ, Temple RO: Neuropsychological Assessment Battery (NAB): performance in a sample of patients with moderate-to-severe traumatic brain injury. Appl Neuropsychol 2010; 17(4): 283–8. doi:10.1080/09084282.2010.525118.
28. Reynolds CR: Comprehensive Trail Making Test (CTMT). *Austin*, TX: Pro-Ed. 2001.
29. Tombaugh TN: Test of Memory Malingering:(TOMM). New York, Mulit Health Systems, Inc, 1996.
30. Kane RL, Roebuck-Spencer T, Short P, Kabat M, Wilken J: Identifying and monitoring cognitive deficits in clinical populations using Automated Neuropsychological Assessment Metrics (ANAM) tests. Arch Clin Neuropsychol 2007; 22(Suppl 1): S115–26. doi:10.1016/j.acn.2006.10.006.
31. Wilkinson GS, Robertson GJ: Wide Range Achievement Test 4 (WRAT4). Lutz, FL, Psychological Assessment Resources, 2006.
32. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG: Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009; 42(2): 377–81. doi:10.1016/j.jbi.2008.08.010.
33. Hotelling H: The generalization of student's ratio. In: Breakthroughs in Statistics, pp 54–65. Edited by Kotz S, Johnson NL New York, Springer, 1992. doi:10.1007/978-1-4612-0919-5_4; Springer Series in Statistics. New York, NY.

34. Pontifex MB, O'Connor PM, Broglio SP, Hillman CH: The association between mild traumatic brain injury history and cognitive control. Neuropsychologia 2009; 47(14): 3210–6. doi:10.1016/j.neuropsychologia.2009.07.021.
35. Larson MJ, Clayson PE, Farrer TJ: Performance monitoring and cognitive control in individuals with mild traumatic brain injury. J Int Neuropsychol Soc 2012; 18(2): 323–33. doi:10.1017/S1355617711001779.
36. Larson MJ, Farrer TJ, Clayson PE: Cognitive control in mild traumatic brain injury: conflict monitoring and conflict adaptation. Int J Psychophysiol 2011; 82(1): 69–78. doi:10.1016/j.ijpsycho.2011.02.018.
37. Larson MJ, Perlstein WM, Demery JA, Stigge-Kaufman DA: Cognitive control impairments in traumatic brain injury. J Clin Exp Neuropsychol 2006; 28(6): 968–86. doi:10.1080/13803390600646860.
38. Killam C, Cautin RL, Santucci AC: Assessing the enduring residual neuropsychological effects of head trauma in college athletes who participate in contact sports. Arch Clin Neuropsychol 2005; 20(5): 599–611. doi:10.1016/j.acn.2005.02.001.
39. Suchy Y, Euler M, Eastvold A: Exaggerated reaction to novelty as a subclinical consequence of mild traumatic brain injury. Brain Inj 2014; 28(7): 972–9. doi:10.3109/02699052.2014.888766.