Experimental Results (2022), 3, e15, 1–11
doi:10.1017/exp.2022.11

PSYCHOLOGY AND PSYCHIATRY
NOVEL-RESULT

The development, and day-to-day variation, of a Military-Specific Auditory N-Back Task and Shoot-/Don't -Shoot Task

Christopher Vine[1,*], Sarah Coakley[1,2], Stephen Myers[1], Sam Blacker[1] and Oliver Runswick[1,3]

[1]Occupational Performance Research Group, Institute of Sport, Nursing and Allied Health, University of Chichester, Chichester, United Kingdom, [2]Faculty of Sport, Allied Health and Performance Science, St Mary's University, Twickenham, United Kingdom, and [3]Institute of Psychiatry, Psychology & Neuroscience, Kings College London, London, United Kingdom
*Corresponding author. Email: c.vine@chi.ac.uk

(Received 16 May 2022; Revised 02 August 2022; Accepted 15 August 2022)

Abstract
During military operations, soldiers are required to successfully complete numerous physical and cognitive tasks concurrently. Understanding the typical variance in research tools that may be used to provide insight into the interrelationship between physical and cognitive performance is therefore highly important. This study assessed the inter-day variability of two military-specific cognitive assessments: a Military-Specific Auditory N-Back Task (MSANT) and a Shoot-/Don't-Shoot Task (SDST) in 28 participants. Limits of agreement ±95% confidence intervals, standard error of the mean, and smallest detectable change were calculated to quantify the typical variance in task performance. All parameters within the MSANT and SDST demonstrated no mean difference for trial visit in either the seated or walking condition, with equivalency demonstrated for the majority of comparisons. Collectively, these data provided an indication of the typical variance in MSANT and SDST performance, while demonstrating that both assessments can be used during seated and walking conditions.

Key words: decision making; external validity; occupational; performance

Introduction

During military operations, personnel are required to maintain performance in both their role-specific physical tasks (e.g., load carriage) and corresponding cognitive tasks (e.g., decision making and communication) (Crawford et al., 2017; Scribner, 2016). Failure to maintain performance, in either domain, can result in reduced combat readiness and decreased operational performance (Crawford et al., 2017; Vrijkotte et al., 2016). Consequently, there is growing interest in the relationship between military-specific physical activity and cognitive performance within military operators (Armstrong et al., 2022; Bhattacharyya et al., 2017; Eddy et al., 2015; Giles et al., 2019; Kobus et al., 2010; Nibbeling et al., 2014; Son et al., 2019, 2022; Vine et al., 2021). Despite this interest, the methodologies and approaches used to investigate this relationship have differed considerably, particularly concerning the assessment of cognitive performance.

Based on the assessment tools used to date, and the visual and auditory requirements of soldiers, two assessment tools were developed: A Military Specific Auditory N-Back Task (MSANT), and a

Shoot-/Don't-Shoot (SDST). The former used phonetically described pairs of letters and represented aspects of military radio communications, while the latter represented aspects of any military scenario where visual search and inhibition are required (e.g., assaulting an enemy position or operations in builtup areas). The current study, therefore, aimed to detail the methodology of the MSANT and SDST, along with quantifying the typical day-to-day variability of both assessment tools under seated and walking condition. The investigation did not seek to investigate the influence of physical fatigue or dual tasking on the performance of these assessment tools.

Methods

The full methods for this project are available in the Supplementary Material, with the raw data available at: https://osf.io/jekv8/. Briefly, the study comprised of two elements. First, the day-to-day variability of the MSANT and SDST was assessed in a seated condition on three separate occasions (Part 1). This was chosen due to the large variability in potential application of these assessment tools in future projects. Second, within a sub-sample of the study population, the day-to-day variability of the MSANT and SDST was assessed during a 10-min walking activity, on three separate occasions (Part 2). While a matched study population for this part of the study would have been optimal, given the time required for this portion of the study (a result of the necessity to reach a physiological steady state before conducting the test, and the recovery period required between each walking bout to prevent the onset of fatigue), a sub-sample approach was instead chosen. Physiological steady state refers to the stabilization in the physiological responses to exercise (e.g., increases in heart rate). Without this stability, variability in cognitive performance could be induced as a consequence of adapting to the exercise stimulus opposed to just reflecting the typical variation in test performance.

All laboratory visits were separated by a minimum of 48 hr, and participants were required to arrive in a fed and hydrated state having avoided caffeine for a minimum of 3 hr. Study visits were completed at approximately the same time of day (±2 hr) to control for the potential effect of circadian rhythm on test performance. All participants were recruited from the university population (all were students or from academic positions), spoke fluent English, and had self-declared normal or corrected to normal vision.

Twenty-eight participants volunteered for Part 1 of the study (14 male, 14 female, age [mean ± SD] 27.3 ± 4.3 year) and 12 participants for Part 2 (6 male, 6 female, age 28.4 ± 3.5 year). Sample size for Part 1 was calculated using an a priori power calculation (G Power; version 3.1.9.4) (Prajapati et al., 2010). For the seated portion of the investigation, 28 participants were required to a moderate effect size ($f = 0.25$), with a statistical power of 80% and an alpha level of 0.05, based upon a correlation coefficient of $r = 0.5$ (identified from initial pilot testing). A moderate effect size (Cohen, 1988) was selected based on the combination of effect sizes reported in previous investigations, utilizing similar cognitive assessment tools (Eddy et al., 2015), and the anticipated smallest effect size of interest to military policymakers. The sub-sample size was designed to represent the typical size (and therefore likely variation) of study populations within this research area (e.g., Bhattacharyya et al., 2017; Crowell et al., 1999; Eddy et al., 2015). Ethical approval was provided by the Institution's Research Ethics Committee, with written consent obtained from all participants.

Cognitive assessments

The MSANT involved identifying a pair of phonetically described letters two previous to an auditory tone (i.e., 2-back). During the seated condition, participants recorded their answers, while during walking trials, participants were required to relay their answers verbally which were recorded on their behalf. The SDST was designed to be a visual search and inhibition task similar to those tasks previously employed within the literature (Armstrong et al., 2022; Eddy et al., 2015; Kobus et al., 2010). The assessment involved responding appropriately to targets and non-targets. Participants were instructed to place equal

https://doi.org/10.1017/exp.2022.11 Published online by Cambridge University Press

importance on both response time and accuracy. For the SDST, there was a 2:1 ratio between targets and non-targets.

For Part 1, during the first visit, participants were familiarized (two full trial completions of each assessment) with the MSANT and SDST, in a randomized counterbalanced order. For the second, and third visits, participants completed the MSANT and SDST in the same randomized counterbalanced order. For Part 2, a sub-sample of 12 participants completed 3 additional laboratory visits completing the SDST and MSANT while walking on a treadmill. Again the MSANT and SDST were completed in a randomized order. All tests were completed with 10 min of seated rest between trials to negate the influence of physical fatigue. To enable a physiological steady state to occur, participants completed 5 min of walking before the commencement of the cognitive assessments. For all walking trials, participants walked on a motorized treadmill (6.5 km·h^{-1}, 1% gradient) at a load carriage speed representing a typical "enemy contact" speed (Armstrong et al., 2019).

Statistical analysis

Data were principally analyzed using JASP (JASP, 2020; v0.14.1). For normally distributed data, a one-way ANOVA was employed to identify whether a likely main effect of assessment time point was apparent. Effect sizes are presented as Omega squared (ω^2) (Levine & Hullett, 2002). For non-normally distributed data, a Friedman's test was employed with effect sizes presented using Kendall's *W*. Holm-Bonferroni adjusted pairwise comparisons, and pairwise comparisons using Conover's test were made post-hoc as appropriate. For key assessment variables, equivalency between trials was calculated using the two one-sided test approach (Lakens et al., 2018). Based upon the a priori sample size calculation, $d =$ 0.5 was employed as the smallest effect size of interest. To describe the typical variation in assessment parameters between trials, Limits of agreement ± 95% confidence intervals, standard error of the mean, and smallest detectable change values were calculated (Hopkins, 2000; Ludbrook, 2010; van Kampen et al., 2013).

Results

Descriptive statistics are presented in Table 1, with day-to-day variation descriptors reported in Table 2. One participant was removed from the analysis, due to being more than two SDs outside the remainder of the data set.

Seated performance

Military-Specific Auditory N-Back Task

There was no likely main effect for time for total correct response ($\chi^2_{(4)} = 4.531, p = .361$, Kendall's $W =$ 0.492) or combined correct responses ($\chi^2_{(4)} = 3.856, p = .426$, Kendall's $W = 0.488$); however, a likely main effect for time was evident for partial correct responses ($\chi^2_{(4)} = 11.846, p = .019$, Kendall's $W =$ 0.426). For the key variable of combined correct responses, the comparison between trial 1 versus trial 2 was both statistically equivalent ($W_{(25)} = 64, p = .002$) and not statistically different ($W_{(25)} = 70, p =$ .938). Similarly, trial 2 versus trial 3 were both statistically equivalent ($W_{(25)} = 20$, $p = .06$) and not statistically different. Likewise, trial 1 versus trial 3 were also both statistically equivalent ($W_{(25)} = 50, p =$ .032) and not statistically different.

Shoot-/Don't-Shoot Task

There was no likely main effect for time on either shoot correct ($\chi^2_{(4)} = 4.00, p = .406$, Kendall's $W =$ 0.175), don't-shoot correct ($\chi^2_{(4)} = 3.069, p = .546$, Kendall's $W = 0.482$), total correct ($\chi^2_{(4)} = 3.375, p =$ .497, Kendall's $W = 0.471$), and average response time ($F_{(2.981,77.515)} = 1.035$, $p = .382$, $\omega^2 = 0.001$).

https://doi.org/10.1017/exp.2022.11 Published online by Cambridge University Press

Table 1. Descriptive statistics for cognitive assessments (mean ± SD [range]) during seated (S) and walking (W) conditions

Task (condition)	Parameter	Experimental trial				
		FAM 1	FAM 2	Trial 1	Trial 2	Trial 3
SDST (S)	Total correct (%)	96.4 ± 3.3	97.3 ± 2.7	96.5 ± 3.0	96.3 ± 3.2	97.1 ± 3.2
		[86.1–100.0]	[91.7–100.0]	[88.9–100.0]	[88.9–100.0]	[83.3–100.0]
	RT (ms)	579 ± 58	574 ± 57	562 ± 57	550 ± 51	528 ± 43
		[490–684]	[478–683]	[472–704]	[450–639]	[433–655]
	ASTO (ms/CR)	16.7 ± 1.6	16.4 ± 1.7	16.2 ± 1.8	15.9 ± 1.4	15.1 ± 1.4
		[14.1–20.4]	[13.7–19.2]	[13.7–19.9]	[13.2–18.3]	[12–18.7]
SDST (W)	Total correct (%)			94.9 ± 5.3	96.1 ± 3.8	96.5 ± 5.4
				[80.6–100]	[88.9–100]	[80.6–100]
	RT (ms)			594 ± 70	575 ± 69	566 ± 69
				[496–678]	[457–661]	[451–666]
	ASTO (ms/CR)			17.4 ± 1.4	16.6 ± 1.6	16.3 ± 1.9
				[15–19.4]	[13.9–18.4]	[13.3–19]
MSANT (S)	Total correct (%)	87.7 ± 15	88.5 ± 16.7	90.4 ± 14.6	90.8 ± 16	94.2 ± 9.5
		[50–100]	[30–100]	[40–100]	[30–100]	[60–100]
	Combined score (%)	91 ± 11.3	91.4 ± 12.8	92.9 ± 10.8	92.7 ± 12.3	95.1 ± 7.7
		[60–100]	[46.7–100]	[56.7–100]	[46.7–100]	[70–100]
MSANT (W)	Total correct (%)			93.3 ± 8.9	95 ± 10	94.2 ± 9
				[70–100]	[70–100]	[80–100]
	Combined score (%)			95.3 ± 6.7	96.1 ± 7.9	95.8 ± 6.5
				[76.7–100]	[76.7–100]	[83.3–100]

Note. Blank cells denote data that were not collected due to the seated condition acting as the familiarization for the walking condition.
Abbreviations: ASTO, accuracy-speed trade-off; CR, correct response; FAM, familiarization; MSANT, Military-Specific Auditory N-Back Task; RT, response time; S, seated; SDST, Shoot-/Don't-Shoot Task; W, walking.

There was, however, a main effect for time in the accuracy-speed trade-off (ASTO) parameter ($F_{(4,104)} = 7.037$, $p < .001$, $G\eta^2 = 0.089$). Importantly, the sole noteworthy difference occurred between familiarization 1 and trial 3 ($t_{(26)} = 4.855$, $p < .001$, $d = 0.756$) suggesting no discernible difference was likely between performances in the three experimental trials, following two familiarization trials. For the key variable of total correct responses, trial 1 versus trial 2, trial 1 versus trial 3, and trial 2 versus trial 3 were both statistically equivalent (1 vs. 2: $W_{(26)} = 93, p = .011$; 1 vs. 3: $W_{(26)} = 61, p = .047$; 2 vs. 3: $W_{(26)} = 41, p = .040$) and not statistically different. For the other key variable of ASTO, all comparisons were likely neither statistically equivalent (1 vs. 2: $t_{(26)} = -1.701, p = .050$; 2 vs. 3: $t_{(26)} = -0.127, p = .45$; 1 vs. 3: $t_{(26)} = 0.287$, $p = .612$) nor statistically different.

Walking performance

Military-Specific Auditory N-Back Task

As with seated MSANT performance, there was no likely effect of time on total correct responses ($\chi^2_{(2)} = 1.000$, $p = .607$, Kendall's $W = 0.568$), partial correct responses ($\chi^2_{(2)} = 1.280$, $p = .527$, Kendall's $W = 0.541$), and combined correct responses ($\chi^2_{(2)} = 1.000$, $p = .607$, Kendall's $W = 0.582$). For the key variable of combined correct responses, trials 1 versus 2 and trials 2 versus 3 were statistically equivalent

https://doi.org/10.1017/exp.2022.11 Published online by Cambridge University Press

https://doi.org/10.1017/exp.2022.11 Published online by Cambridge University Press

Table 2. Descriptive statistics for cognitive assessments (mean ± SD) during seated (S) and walking (W) conditions

		Trial 1 versus 2			Trial 2 versus 3			Trial 1 versus 3		
Task (condition)	Parameter	Mean Bias ± 95% CI	SEM	SDC	Mean Bias ± 95% CI	SEM	SDC	Mean Bias ± 95% CI	SEM	SDC
SDST (S)	Total correct (%)	0.2 ± 5.1	2.4	6.7	−0.8 ± 4.8	2.3	6.4	−0.6 ± 6.1	2.9	8.0
	RT (ms)	12 ± 88	42	116	22 ± 65	31	87	34 ± 84	40	111
	ASTO ($ms{\cdot}CR^{-1}$)	0.3 ± 2.8	1.3	3.7	0.7 ± 2.3	1.1	3.1	1.1 ± 2.9	1.4	3.8
SDST (W)	Total correct (%)	−1.2 ± 6.2	2.7	7.5	−0.5 ± 8.3	3.6	10.1	−1.6 ± 4.1	1.8	4.9
	RT (ms)	19 ± 60	26	73	9 ± 66	29	80	28 ± 75	33	91
	ASTO ($ms{\cdot}CR^{-1}$)	0.8 ± 2.3	1.0	2.8	0.3 ± 3.1	1.4	3.8	1.0 ± 2.3	1.0	2.8
MSANT (S)	Total correct (%)	−0.4 ± 18.5	8.8	24.5	−3.5 ± 13.2	6.3	17.5	−3.8 ± 15.2	7.2	20.0
	Combined score (%)	0.3 ± 14.5	6.9	19.1	−2.4 ± 11.2	5.4	14.8	−2.2 ± 11.9	5.7	15.7
MSANT (W)	Total correct (%)	−1.7 ± 13.5	5.9	16.4	0.8 ± 16.1	7.0	19.5	−0.8 ± 21.2	9.3	25.7
	Combined score (%)	−0.8 ± 10.3	4.5	12.5	0.3 ± 11.8	5.2	14.3	−0.6 ± 15.8	6.9	19.1

Abbreviations: ASTO, accuracy-speed trade-off; CI, confidence intervals; CR, correct response; MSANT, Military-Specific Auditory N-Back Task; RT, response time; S, seated; SDC, smallest detectable change; SDST, Shoot-/Don't-Shoot Task; SEM, standard error of the mean; W, walking.

(1 vs. 2: $W_{(11)} = 12$, $p = .017$; 2 vs. 3: $W_{(11)} = 13$, $p = .020$) and not statistically different. Conversely, trial 1 versus 3 was neither statistically equivalent nor statistically different.

Shoot-/Don't-Shoot Task

Again there were no likely effects of time on shoot correct responses ($\chi^2_{(2)} = 4.800$, $p = .091$, Kendall's $W = 0.449$), don't-shoot correct responses ($\chi^2_{(2)} = 2.480$, $p = .289$, Kendall's $W = 0.672$), total correct responses ($\chi^2_{(2)} = 3.161$, $p = .206$, Kendall's $W = 0.741$), response times ($F_{(2,22)} = 2.880$, $p = .077$, $\eta_G^2 = 0.018$), and ASTO ($F_{(2,22)} = 2.713$, $p = .088$, $\eta_G^2 = 0.042$). For the key variable of total correct responses, all comparisons were neither statistically equivalent (1 vs. 2: $W_{(11)} = 6$, $p = .096$; 1 vs. 3: $W_{(11)} = 0$, $p = .093$; 2 vs. 3: $W_{(11)} = 14$, $p = .084$) nor statistically different. Similarly, for the other key variable of ASTO all comparisons were likely neither statistically equivalent (1 vs. 2: $t_{(11)} = 0.127$, $p = .549$; 2 vs. 3: $t_{(11)} = -1.205$, $p = .127$; 1 vs. 3: $t_{(11)} = 0.787$, $p = .776$) nor statistically different.

Discussion

This study has described the methods of two military-specific cognitive assessment tools (MSANT and SDST) and quantified their typical day-to-day variability. These data provide typical magnitudes of variance for the key assessment parameters. While no likely performance differences were observed across the experimental measurement points, not all walking comparisons were statistically equivalent, suggesting additional data are required before this assertion is made, for the given equivalency bounds. It should however be noted that borderline statistically significant results may become non-significant where correction for multiple testing is utilized. The current investigation has also demonstrated the suitability of these assessment tools for use during military-specific physical activity within a laboratory setting.

Before this investigation, the day-to-day performance variation in any military-specific cognitive assessments had not been quantified. This is an issue for several reasons, including the translational ability of research findings to the "real world" (Close et al., 2019) and also for methodological decision making (e.g., sample size calculations). Moreover, with military operations rarely conducted in isolation, information on inter-test performance is highly relevant to research investigating sequential or repeated bout performance. The comparison between seated and walking performance was not a research question of interest in the current study, particularly given that deficits in cognitive performance are typically observed after ~30 min of military activity (e.g., Eddy et al., 2015; Giles et al., 2019). However, observationally, the typical variation in performance between trials appears similar between seated and walking conditions.

Familiarizing participants with assessment tools is critical for research, particularly when time limitations may inhibit access to study participants (e.g., military populations). Collectively, the current study's data demonstrate that beyond two full seated trials, a continued improvement in performance was not likely apparent, suggesting this familiarization length is sufficient to minimize possible learning effects.

Several limitations exist with the current investigation, including the use of a civilian population, and the limited walking sub-sample size. As acknowledged previously, the smaller sub-sample size was chosen for largely practical reasons, although it does match many studies within this area, highlighting issues with underpowered investigations. Future research should attempt to pair reliable and applied cognitive tasks (such as those described herein) with operationally relevant and appropriate physical activity. This in turn will support enhanced applied research as well as enabling a greater focus to be placed on developing mitigation strategies where the greatest mission impact can be obtained.

Acknowledgment. The authors would like to thank the participants for their participation in the current project.

Supplementary Materials. To view supplementary material for this article, please visit http://doi.org/10.1017/exp.2022.11.

Authorship contributions. Research concept and study design: C.V., S.C., S.M., S.B., O.R; Data collection, data and statistical analyses, and writing of the initial manuscript: C.V.; Reviewing/editing final manuscript: C.V., S.C., S.M., S.B., and O.R.

Data availability statement. The raw data for this project can be viewed at: https://osf.io/jekv8/.

Funding statement. The authors declare no funding was received for this research.

Conflict of interest. The authors have no conflicts of interest to declare.

References

Armstrong, N., Ward, A., Lomax, M., Tipton, M. J., & House, J. R. (2019). Wearing body armour and backpack loads increase the likelihood of expiratory flow limitation and respiratory muscle fatigue during marching. *Ergonomics*, **62**, 1181–1192.

Armstrong, N. C., Smith, S. J. R., Risius, D., Doyle, D., Wardle, S. L., Greeves, J. P., House, J. R., Tipton, M., & Lomax, M. (2022). Cognitive performance of military men and women during prolonged load carriage. *BMJ Military Health*, e002000.

Bhattacharyya, D., Pal, M., Chatterjee, T., & Majumdar, D. (2017). Effect of load carriage and natural terrain conditions on cognitive performance in desert environments. *Physiology & Behavior*, **179**, 253–261.

Close, G. L., Kasper, A. M., & Morton, J. P. (2019). From paper to podium: quantifying the translational potential of performance nutrition research. *Sports Medicine*, **49**, 25–37.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Earlbaum Associates.

Crawford, C., Teo, L., Lafferty, L., Drake, A., Bingham, J. J., Gallon, M. D., O'connell, M. L., Chittum, H. K., Arzola, S. M., & Berry, K. (2017). Caffeine to optimize cognitive function for military mission-readiness: a systematic review and recommendations for the field. *Nutrition Reviews*, **75**, 17–35.

Crowell, H. P., Krausman, A. S., Harper, W. H., Faughn, J. A., & Sharp, M. A. (1999). *Cognitive and physiological performance of soldiers while they carry loads over various terrains*. Army Research Laboratory.

Eddy, M. D., Hasselquist, L., Giles, G., Hayes, J. F., Howe, J., Rourke, J., Coyne, M., O'Donovan, M., Batty, J., & Brunyé, T. T. (2015). The effects of load carriage and physical fatigue on cognitive performance. *PloS One*, **10**, e0130817.

Giles, G. E., Hasselquist, L., Caruso, C., & Eddy, M. D. (2019). Load carriage and physical exertion influence cognitive control in military scenarios. *Medicine and Science in Sports and Exercise*, **51**, 2540–2546.

Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Sports Medicine*, **30**, 1–15.

JASP. (2020). *JASP (Version 0.14.1)[Computer software].*

Kobus, D. A., Brown, C. M., Wu, L., Robusto, K., & Bartlett, J. (2010). *Cognitive performance and physiological changes under heavy load carriage.* Pacific Science and Engineering Group Inc.

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, **1**, 259–269.

Levine, T. R., & Hullett, C. R. (2002). Eta squared, partial eta squared, and misreporting of effect size in communication research. *Human Communication Research*, **28**, 612–625.

Ludbrook, J. (2010). Confidence in Altman-Bland plots: A critical review of the method of differences. *Clinical and Experimental Pharmacology and Physiology*, **37**, 143–149.

Nibbeling, N., Oudejans, R. R. D., Ubink, E. M., & Daanen, H. A. M. (2014). The effects of anxiety and exercise-induced fatigue on shooting accuracy and cognitive performance in infantry soldiers. *Ergonomics*, **57**, 1366–1379.

Prajapati, B., Dunne, M., & Armstrong, R. (2010). Sample size estimation and power analysis. *Optometry Today*, **16**, 123–132.

Scribner, D. R. (2016). Predictors of shoot–don't shoot decision-making performance: An examination of cognitive and emotional factors. *Journal of Cognitive Engineering and Decision Making*, **10**, 3–13.

Son, M., Hyun, S., Beck, D., Jung, J., & Park, W. (2019). Effects of backpack weight on the performance of basic short-term/working memory tasks during flat-surface standing. *Ergonomics*, **62**, 548–564.

Son, M., Jung, J., Hwang, D., Beck, D., & Park, W. (2022). The effect of backpack weight on the performance of basic short-term/working memory tasks while walking along a pre-determined route. *Ergonomics*, 1–23.

van Kampen, D. A., Willems, W. J., van Beers, L. W. A. H., Castelein, R. M., Scholtes, V. A. B., & Terwee, C. B. (2013). Determination and comparison of the smallest detectable change (SDC) and the minimal important change (MIC) of four-shoulder patient-reported outcome measures (PROMs). *Journal of Orthopaedic Surgery and Research*, **8**, 40.

Vine, C. A. J., Myers, S. D., Coakley, S. L., Blacker, S. D., & Runswick, O. R. (2021). Transferability of military-specific cognitive research to military training and operations. *Frontiers in Psychology*, **12**, 386.

Vrijkotte, S., Roelands, B., Meeusen, R., & Pattyn, N. (2016). Sustained military operations and cognitive performance. *Aerospace Medicine and Human Performance*, **87**, 718–727.

Cite this article: Vine C, Coakley S, Myers S, Blacker S, Runswick O (2022). The development, and day-to-day variation, of a Military-Specific Auditory N-Back Task and Shoot-/Don't -Shoot Task. *Experimental Results,* 3, e15, 1–11. https://doi.org/10.1017/exp.2022.11

Peer Reviews

Reviewing editor: Dr. Gregory Postal
Uniformed Services University of the Health Sciences, F Edward Hebert School of Medicine, Psychiatry, 130 South Churchill Drive, Fayetteville, North Carolina, United States, 28303-5065

Minor revisions requested.

doi:10.1017/exp.2022.11.pr1

Review 1: The Development, and Day-to-Day Variation, of a Military-Specific Auditory N-Back Task and Shoot-/Don't-Shoot Task

Reviewer: Dr. Yu Kume

Date of review: 23 May 2022

Conflict of interest statement. Reviewer declares none.

Comment

Comments to the Author: Dear Authors,

I understand that the current study will be focused on effect of cognitive trials according to two conditions (Part 1 or Part 2) on three separate occasions. There are, however, serious issues regarding extreme small samples, different sample-size for each part's experiment, poor information on definition of three separate occasions, how to select statistical analyses to clarify the objective of the present study and so on. Especially, I would like to suggest kindly that selection of statistical methodology would be reconsidered to clarify effect of two cognitive trials according to two experimental conditions (Part 1 and Part 2). For instance, if the interactive effect between conditions and occasions for each cognitive trial was examined statistically according to the objective of this study, I would suggest application of the repeated two-way ANOVA to analyze the data collected in this study. Considering the above concerns carefully, I would like to suggest kindly that the manuscript might need to be rewritten after reanalyzing the data.

Score Card

Presentation

2.2 /5

Is the article written in clear and proper English? (30%)	4/5
Is the data presented in the most useful manner? (40%)	1/5
Does the paper cite relevant and related articles appropriately? (30%)	2/5

Context

1.5 /5

Does the title suitably represent the article? (25%)	2/5
Does the abstract correctly embody the content of the article? (25%)	1/5
Does the introduction give appropriate context? (25%)	1/5
Is the objective of the experiment clearly defined? (25%)	2/5

Analysis

1.2 /5

Does the discussion adequately interpret the results presented? (40%)	1/5
Is the conclusion consistent with the results and discussion? (40%)	1/5
Are the limitations of the experiment as well as the contributions of the experiment clearly outlined? (20%)	2/5

https://doi.org/10.1017/exp.2022.11 Published online by Cambridge University Press

doi:10.1017/exp.2022.11.pr2

Review 2: The Development, and Day-to-Day Variation, of a Military-Specific Auditory N-Back Task and Shoot-/Don't-Shoot Task

Reviewer: Dr. Ryota Sakurai

Date of review: 15 June 2022

Conflict of interest statement. Reviewer declares none.

Comment

Comments to the Author: This study examined the reliability of Military Specific Auditory N-Back Task and Shoot-/Don't-Shoot Task. The present study yielded interesting findings. Please find below some specific comments that need to be addressed.

1.The Part 2 study assessed the day-to-day variability of the MSANT and SDST during a 10-minute walking activity. What was the aim of the Part 2 study? Did the authors want to examine the effect of fatigue on the day-to-day variability of tests or the effects of dual-tasking on the variability?

2.How did you recruit participants. Did the study have inclusion and exclusion criteria regarding study participants?

3.Although previous studies may have described, please explain the methodology of MSANT and SDST in more detail.

4.I was not sure of the meaning of a physiological steady-state. Does five minutes of walking result in a physiological steady-state among participants? A five-minute break would probably bring more physiological stability.

5.Why did the authors calculate the sample size only in the Part 1 study?

6.What were the intervals between trials (i.e., trial 1 to trial 3)?

7.Education level may confound the results of MSANT and SDST. Did you consider the effect?

Score Card

Presentation

3.9 /5

Is the article written in clear and proper English? (30%)	5/5
Is the data presented in the most useful manner? (40%)	3/5
Does the paper cite relevant and related articles appropriately? (30%)	4/5

Context

3.5 /5

Does the title suitably represent the article? (25%)	3/5
Does the abstract correctly embody the content of the article? (25%)	4/5
Does the introduction give appropriate context? (25%)	4/5
Is the objective of the experiment clearly defined? (25%)	3/5

Analysis

3.6 /5

Does the discussion adequately interpret the results presented? (40%)	4/5
Is the conclusion consistent with the results and discussion? (40%)	3/5
Are the limitations of the experiment as well as the contributions of the experiment clearly outlined? (20%)	4/5

https://doi.org/10.1017/exp.2022.11 Published online by Cambridge University Press