AFRL-RH-WP-TR-2022-XXXX

Evaluation of U.S. Air Force Performance Assessment Tool to Detect the Cognitive Performance Effects of Operator Dehydration

Jeffrey. B Phillips, Madison C. McInnis, Kody Coleman, Roger Carff, Adrien Moucheboeuf, Allison Bew, Andrew Dorsey

The Florida Institute for Human and Machine Cognition

April 6, 2022

Final Report

NOTICE AND SIGNATURE PAGE

AFRL-SA-WP-TR-201x-xxxx HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//SIGNED//	//SIGNED//
SIGNATURE BLOCK OF UNIT CHIEF	SIGNATURE BLOCK OF DEPT HEAD

REPORT DOCUMENTATION PAGE		*Form Approved* *OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**		
1. REPORT DATE *(DD-MM-YYYY)* 04-06-2022	**2. REPORT TYPE** Final	**3. DATES COVERED *(From – To)***
4. TITLE AND SUBTITLE Evaluation of the U.S. Air Force Performance Assessment Tool to Detect the Cognitive Performance Effects of Operator Dehydration.		**5a. CONTRACT NUMBER**
		5b. GRANT NUMBER 2020_029_01
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Jeffery B. Phillips, Madison C. McInnis, Kody Coleman, Roger Carff, Adrien Moucheboeuf, Allison Bew, Andrew Dorsey		**5d. PROJECT NUMBER**
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The Florida Institute for Human and Machine Cognition 40 S. Alcaniz Street Pensacola, FL, 32502		**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Airman Biosciences Division Dept 2729 R Street Wright-Patterson AFB, OH 45433-7913		**10. SPONSORING/MONITOR'S ACRONYM(S)**
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION / AVAILABILITY STATEMENT Distribution A: Approved for public release; distribution is unlimited. Case Number: 88ABW-201x-xxxx, xx Mon 201x		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT The United States Air Force School of Aerospace Medicine (USAFSAM) has recently developed a psychometric test called the Performance Assessment Tool (PAT). The PAT possesses several qualities that make it an ideal tool for assessing operational stress on human performance. Previous studies have shown that PAT maintains high test-retest reliability, operational face validity, and criterion validity for a wide range of stressors that are specific to tactical aviation. Another advantage of PAT is the minimal time, less than an hour, required for training and administration. PAT's ease of us is further evidenced by its ability to automatically identify each participant's optimal difficulty level for stress testing. Despite the initial success of PAT, studies such as Phillips, 2019 have identified areas in need of improvement if the PAT is to become a mainstay in aviation research. The purpose of the reported effort was to alter PAT according to the suggestions made by Phillips, 2019. This report describes how the PAT software was re-developed and will detail the changes made to PAT. Additionally, this report presents data from a pilot study that was conducted to validate the re-developed software. Results from the investigation show that the updated PAT maintains good test-rest reliability and provides the ability to manipulate workload by adding or subtracting tasks.		
15. SUBJECT TERMS Dehydration, PAT, tactical aviation, workload, validity, reliability, NASA TLX		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U	SAR	18	19b. TELEPHONE NUMBER *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

This page intentionally left blank

TABLE OF CONTENTS

Section

LIST OF FIGURES

Page

LIST OF TABLES

Page

1.0 SUMMARY

The United States Air Force School of Aerospace Medicine (USAFSAM) has recently developed a psychometric test called the Performance Assessment Tool (PAT). The PAT possesses several qualities that make it an ideal tool for assessing operational stress on human performance. Previous studies have shown that PAT maintains high test-retest reliability, operational face validity, and criterion validity for a wide range of stressors that are specific to tactical aviation. Another advantage of PAT is the minimal time, less than an hour, required for training and administration. PAT's ease of us is further evidenced by its ability to automatically identify each participant's optimal difficulty level for stress testing. Despite the initial success of PAT, studies such as Phillips, 2019 have identified areas in need of improvement if the PAT is to become a mainstay in aviation research. The purpose of the reported effort was to alter PAT according to the suggestions made by Phillips, 2019. This report describes how the PAT software was re-developed and will detail the changes made to PAT. Additionally, this report presents data from a pilot study that was conducted to validate the re-developed software. Results from the investigation show that the updated PAT maintains good test-rest reliability and provides the ability to manipulate workload by adding or subtracting tasks.

2.0 INTRODUCTION

The United States Air Force School of Aerospace Medicine (USAFSAM) is leading a series of studies to evaluate the effects of operational stressors on human cognitive and perceptual performance. The goals of these studies are to 1) better understand and quantify the effects of operational stress on human performance and 2) develop effective mitigation strategies. Common occupational stressors that have been identified to significantly degrade human cognitive and perceptual performance include hypoxia, hypocapnia, thermal stress, fatigue, motion sickness, spatial disorientation, dehydration, and high cognitive workload. Unfortunately, many psychometric tools used to identify hypoxia-related performance deficits are limited in their utility due to low test-retest reliability. Low test-retest reliability results in high error variance, which makes detection of hypoxia-related performance deficits improbable (Kerlinger, 1973).

To develop a cognitive-perceptual task that is sensitive to the effects of stressors while also having high operational face validity, the USFSAM, in cooperation with the University of Notre Dame, developed the Performance Assessment Tool (PAT). The PAT is a computerized multi-tasking performance assessment involving one primary task of tracking and three additional tasks including math, memory, and mannequin (see PAT subsection below for further explanation). Importantly, the PAT requires many of the same cognitive processes as flying. Initial validation studies revealed that the PAT possesses high test-retest reliability. Phillips and colleagues (2019) further investigated the reliability and criterion validity of PAT and found excellent test-retest reliability and sensitivity to the performance effects of hypoxia. However, building upon the suggestions of Phillips and colleagues (2019), the PAT was recently altered to provide a more powerful performance assessment.

The current study consisted of two phases: Phase 1) PAT software re-development. Phase 2) Validation of re-developed software. The specific aim of Phase 1 was to improve the PAT's design and increase sensitivity to performance deficits associated with operational stressors in aviation. Following completion of Phase 1, Phase 2 aimed to 1) Provide an evaluation of the updated PAT's (PAT 2.6b) test-retest reliability and 2) To determine that PAT 2.6b provides the ability to manipulate workload by adding or subtracting tasks.

3.0 METHODS

3.1 Phase 1

3.1.1 Changes Made to PAT 2.6

In PAT 2.6, the algorithm for calculating the difficulty level for subtasks (math, mannequin, and memory) was tied to momentary tracking accuracy. Due to the overlapping visual demands imposed by both the tracking task and the additional subtasks, it was challenging for participants to increase or maintain subtask difficulty level. The subtask difficulty levels in PAT 2.6b no longer depend on tracking accuracy. This change allows participants to achieve higher difficulty levels.

While changing the algorithm for how task difficulty was set allowed participants to achieve higher math levels, the difficulty of the math tasks did not correspond with these levels. Previously, in PAT 2.6, the number of distractors that could be presented to the user ranged from two to five distractors depending on difficulty level. Therefore, the user could easily memorize the math tasks without doing any computation. To enhance the individual difficulty levels of the math subtask in PAT 2.6b, we increased the range of distractors presented to the user to be between two to ten distractors depending on difficulty level. The number of distractors per difficulty level is further outlined elsewhere (Villano & Crowell, 2021).

The memory items in PAT 2.6b are now presented in a red square in the multi-functional display (MFD) (Figure 1) on the user's computer screen as opposed to the red distractor that was in PAT 2.6. This change provides consistent areas of interest for math, mannequin, and memory to allow better computation of eye-tracking metrics like saccadic velocity.

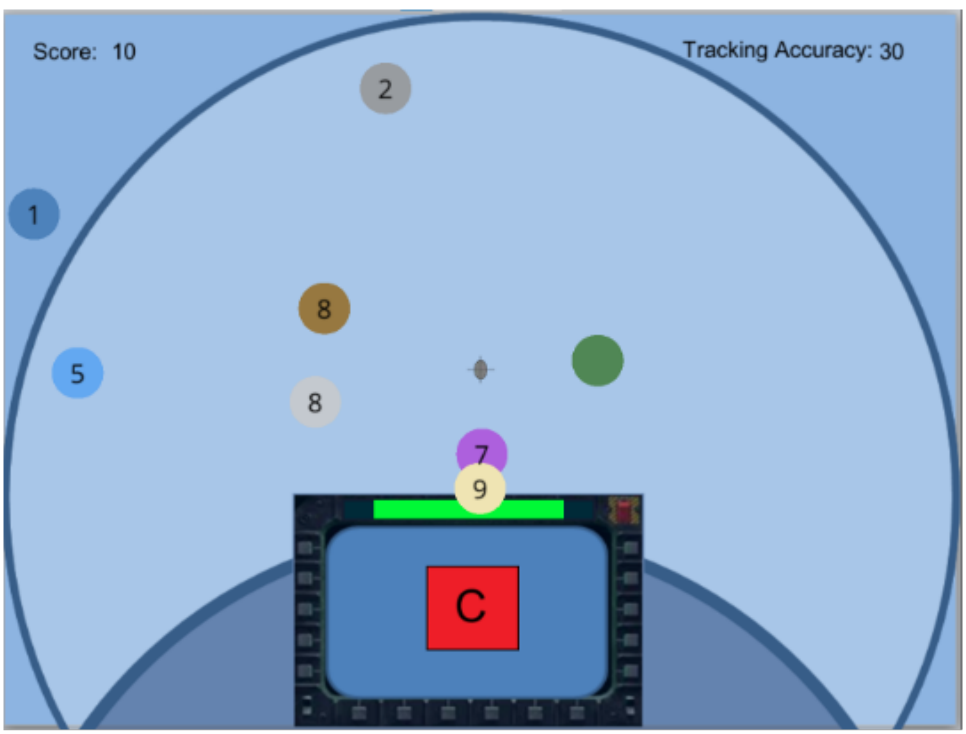


Figure 1. Display of memory subtask with time bar

In PAT 2.6b, a lime green time bar has been implemented into the three subtasks: math, mannequin, and memory (Figure 2). Located at the top of the MFD, this allows the user to view how much time they have left to complete the subtask. The length of the time bar is proportional to how much time the user has left to complete the task, i.e., as the user runs out of time the length of the time bar decreases.

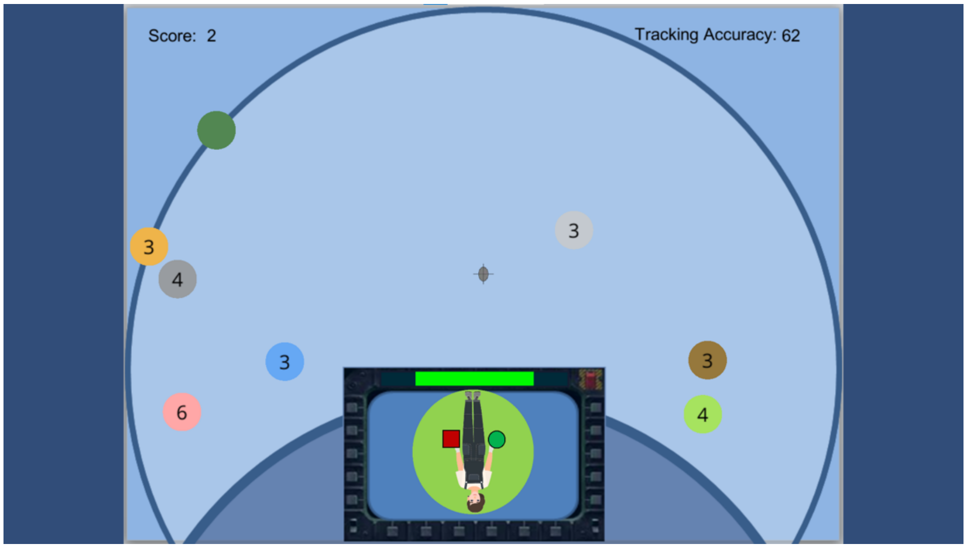


Figure 2. Display of mannequin subtask with time bar

Additionally, the scoring formulas were adjusted by multiplying the percentage correct by 100 before multiplying it to produce larger scores for tracking and secondary task scores produced by the PAT 2.6 scoring algorithm. The formulas for the tracking, math, memory, and mannequin subtasks multiply the percentage of correct responses by the difficulty level of the subtask. This value is multiplied by 100 to convert the percent to a whole number score (Section 3.1.3). The scaling factor of 100 was added across performance metrics in a hope to maximize experimental variance as described in Kerlinger, 1986. Increasing variance in this manner could have a negative effect on test-retest reliability. Negative effects associated with the expansion in variance across will be apparent by comparing inter-class correlation coefficients across the two scoring methods.

3.1.2 PAT 2.6b Description

The PAT is a performance assessment tool developed by USFSAM that is based on Mission Essential Competencies. As a result of the Air Force Research Laboratory's (AFRL) effort to establish Mission Essential Competencies in aviation (Alliger et al., 2007; Symons et al., 2007), seven psychometric constructs have been deemed critical to aviation. These constructs include cognitive proficiency, visual perception, attention-vigilance, spatial processing, memory, reasoning, and psychomotor processing. The PAT is designed to be sensitive to changes in these psychometric constructs that occur in extreme environments such as those encountered by tactical aviators (I.e., high altitude, high workload, thermal stress, increased G-forces, etc.) The PAT is composed of a manual tracking task, a working memory task, an addition task, and a mannequin (spatial processing) task previously described by Villano & Crowell, 2021. The PAT can be administered with one, two, three or all four of the tasks simultaneously.

3.1.3 PAT 2.6b Performance Metrics

For the participant to advance in difficulty levels for the math, memory, and mannequin subtasks, they need to answer three questions in a row correctly. For the participant to advance in difficulty for tracking, they must maintain an average accuracy of at least 85% throughout the iteration. Tracking difficulty will decrease if a participant has an average accuracy of 65% or less at the end of an iteration. If the participant's average accuracy is between 65% and 85% then the difficulty level will not change. The maximum amount of time for each task is five seconds. As the difficulty level increases the time allotted to execute the subtask decreases by 10%.

Formulas for scoring subcomponents of PAT 2.6b:

Tracking = % On Target x 100 x Final Tracking Difficulty Level

Math = % Correct Math Items *x* 100 x Math Final difficulty Level

Mannequin = % Correct Mannequin Items x 100 x Mannequin Final Difficulty Level

Memory = % Correct Memory Items **x** 100 x Memory Final Difficulty Level

PCOLAC = Tracking + Math + Mannequin + Memory

PATC = Dual Task Total Score (combination of tracking and math score) + Mannequin Total Score + Memory Total Score

The Pensacola Composite score takes the sum of the new values for tracking, math, memory, and mannequin scores because it factors in the participant's individual difficulty levels per task. The PAT Composite score adds the combined values of all the tasks before considering each individual difficulty level. The previous formulas for PAT 2.6 can be found in Phillips 2019.

3.2 Phase 2

3.2.1 Participants

*** Ten participants (5 Female, 5 Male) were recruited from The Institute for Human and Machine Cognition (IHMC) and were administered the PAT and Nasa Task Load Index (TLX). The study was determined to be an exempt protocol IRB-2021-0015 by the IHMC IRB Review Board. The average age of participants was 26.8 years with a range of 19-32 years (SD = 4.6). No efforts were made to recruit individuals of a specific ethnic background.

3.2.2 Instrumentation

The PAT is presented on a computer monitor and the user controls inputs using a joystick or computer mouse; a HOTAS Warthog (Thrustmaster, Carentoir, France) joystick was used in the current study. We will continue to refer to the use of a joystick instead of a mouse. The NASA TLX is a subjective workload assessment tool that allows users to perform subjective workload assessments on operators working with human-machine interface systems (Hart & Staveland, 1988; Hart, 2006). The NASA TLX provides an overall workload score based on the weighted average of ratings on six scales: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, Frustration. Further description of the scales can be found elsewhere (Appendix). Participants completed the NASA TLX on an iPad mini 2.

3.2.3 Procedure

Participants were familiarized with PAT by executing one iteration of each PAT subtask individually and then completing the tasks combinations listed in Table 1. Participants were instructed to fill out the NASA TLX after every iteration they completed during familiarization. Iterations were two minutes long with the Minimum Number of Tasks set to two. After completing the individual subtasks and subtask combinations, participants played the full PAT (Tracking + Math + Mannequin + Memory) 10 times, answering the TLX only after the first iteration. All subtask combinations and the first seven full PAT playthroughs were done in practice mode, while the last three were done in acquisition mode. The practice mode allows participants to view the subtask instructions and adjusts the difficulty level of the subtasks based on the accuracy of the answers given by the user. The acquisition mode does not contain the instruction screens, but the difficulty level of the subtasks will adjust based on the accuracy of the answers given. The order, mode, tasks, and NASA TLX status can be found in Tables 1 & 2.

3.2.4 Statistical Analysis

For each of the PAT task combinations NASA TLX ratings, means and standard deviations were obtained across all participants. A series of inter-class correlation coefficients were conducted over the last 10 full PAT sessions for each of the tracking and composite scoring approaches.

4.0 RESULTS

The mean and standard deviation of NASA TLX total workload rating for each of the different PAT task combinations is presented in Table 1. Individual means and standard deviations of workload ratings for each TLX subscale during the different PAT task combinations can be seen in Table 2.

Table 1. NASA TLX Total Workload Ratings for PAT Task Combinations

Task	Mean	SD
Tracking	54.3	21.3
Tracking + Mannequin	67.6	27.3
Tracking + Math	59.4	22.0
Tracking + Memory	42.3	18.4
Tracking + Mannequin + Memory	65.2	27.3
Tracking + Math + Memory	70.2	28.4
Tracking + Math + Mannequin	65.2	25.9
Tracking + Math + Mannequin + Memory (Full PAT)	70.7	23.1

Table 2. NASA TLX Subscale Workload Ratings for PAT Task Combinations

PAT Task and TLX Subscale	Mean	SD
Tracking Mental	6	5.27
Tracking Physical	4.9	6.28
Tracking Temporal	9.4	4.70
Tracking Performance	14.7	4.55
Tracking Effort	10.7	6.13
Tracking Frustration	8.6	4.53
Tracking + Mannequin Mental	11.8	4.61
Tracking + Mannequin Physical	4.9	5.26
Tracking + Mannequin Temporal	12.1	5.09

Tracking + Mannequin Performance	15.3	4.79
Tracking + Mannequin Effort	11.7	5.25
Tracking + Mannequin Frustration	11.8	6.43
Tracking + Math Mental	10.5	4.86
Tracking + Math Physical	4.4	5.4
Tracking + Math Temporal	10.8	5.25
Tracking + Math Performance	13.1	3.51
Tracking + Math Effort	10.8	4.37
Tracking + Math Frustration	9.8	4.67
Tracking + Memory Mental	9	3.97
Tracking + Memory Physical	4	3.65
Tracking + Memory Temporal	8.7	5.14
Tracking + Memory Performance	7.3	4.95
Tracking + Memory Effort	7.9	4.46
Tracking + Memory Frustration	5.4	5.02
Tracking + Mannequin Memory + Mental	13.3	4.64
Tracking + Mannequin + Memory Physical	5.6	5.44
Tracking + Mannequin + Memory Temporal	11.8	4.34
Tracking + Mannequin + Memory Performance	12.9	5.07
Tracking + Mannequin + Memory Effort	11.9	5.20
Tracking + Mannequin + Memory Frustration	9.8	6.92
Tracking + Math + Memory Mental	13.8	4.92
Tracking + Math + Memory Physical	5.7	6.33
Tracking + Math + Memory Temporal	12.8	4.77
Tracking + Math + Memory Performance	13.6	4.86
Tracking + Math + Memory Effort	13.3	5.27
Tracking + Math + Memory Frustration	11	6.75
Tracking + Math + Mannequin Mental	13.1	5.00
Tracking + Math + Mannequin Physical	5.4	5.74
Tracking + Math + Mannequin Temporal	12.8	5.87
Tracking + Math + Mannequin Performance	12.1	4.15
Tracking + Math + Mannequin Effort	12.3	4.99
Tracking + Math + Mannequin Frustration	9.5	5.84
Tracking + Math + Mannequin + Memory Mental	14.6	4.06
Tracking + Math + Mannequin + Memory Physical	5.8	5.87
Tracking + Math + Mannequin + Memory Temporal	14	3.83
Tracking + Math + Mannequin + Memory Performance	13.3	4.22
Tracking + Math + Mannequin + Memory Effort	12	4.69
Tracking + Math + Mannequin + Memory Frustration	11	5.98

PCOLA Composite scores and Tracking scores were calculated across the last 10 iterations of full PAT that participants performed. To examine the scoring effects of the old PAT formulas (Phillips, 2019) versus the new PAT formulas, PCOLA Composite and Tracking scores were calculated using both formulas. PCOLA Composite 1 and Tracking 1 represent the previous formulas presented in Phillips, 2019. PCOLA Composite 2 and Tracking 2 represent the updated formulas from the current investigation. A series of inter-class correlation coefficients

(ICCs) were calculated across these scores as a measure of test-retest reliability. ICCs were performed on PCOLA Composite 1, PCOLA Composite 2, Tracking 1, and Tracking 2 (Table 3).

Table 3. PAT ICCs

Measure	ICC	*p*-Values
PCOLA Composite 1	0.87	< 0.001
PCOLA Composite 2	0.82	< 0.001
Tracking 1	0.89	<0.001
Tracking 2	0.88	<0.001

The NASA-TLX scores showed a 16.4-point increase in subjective mental workload between tracking only and the full PAT. NASA TLX subscale scores suggest that the biggest contributors to perceived workload were Mental Demand and Temporal Demand (Table 4).

Table 4. NASA TLX Subscale Workload Ratings for Tracking Only and Full PAT

TLX Subscale	Tracking Mean (SD)	Full PAT Mean (SD)	Difference
Mental Demand	6 (5.27)	14.6 (4.06)	8.6
Physical Demand	4.9 (6.28)	5.8 (5.87)	1.1
Temporal Demand	9.4 (4.7)	14 (3.83)	4.6
Performance	14.7 (4.55)	13.3 (4.22)	1.4
Effort	10.7 (6.13)	12 (4.69)	1.3
Frustration	8.6 (4.53)	11 (5.98)	2.4
Total	54.3 (21.3)	70.7 (23.1)	16.4

5.0 DISCUSSION

The current effort was conducted to make improvements to PAT 2.6's design and increase its sensitivity to performance deficits associated with operational stressors in aviation. In addition, the investigation aimed to provide an evaluation of PAT 2.6b's test-retest reliability and to determine if PAT 2.6b can provide workload manipulation by performing tracking only versus full PAT.

Changes were made to the PAT 2.6, as described in section 3.4 of this report, to increase task difficulty and correct shortcomings that were identified in Phillips, 2019. Once the agreed upon changes were made, a convenance sample was recruited and asked to execute PAT 2.6b and its components in a sequence described in Table 2.

The observed ICCs for PAT tracking and composite metrics were above .80 which is considered good according to established psychometric standards (Ponterotto & Ruckdeschel, 2007). Increasing the scale of the performance metrics by 10 times did have a slight negative effect on the reliability of those metrics. However, the reliability of the metrics remains well within the good range (>.80) according to psychometric standards. The PCOLA composite 2 score gives the subtasks (math, memory, and mannequin) significantly more impact on the score than PCOLA composite 1 and partially explains the drop in the test-retest reliability coefficient from 0.87 to 0.82 thereby increasing experimental variance while keeping error variance low (Kerlinger, 1986). This quality should make PAT even more sensitive to performance disruptions associated with environmental stress.

The NASA-TLX results suggest that investigators can measure psychometric performance on at least two broad levels of workload with tracking only representing a low workload condition and full PAT representing a high workload condition. This is an important capability as it allows investigators to address the effects of stress at varying levels of cognitive workload to determine if there is stress by cognitive workload interaction effects. More work should be done to classify workload across different PAT task combinations to enable investigators to manipulate workload across a broad continuum as opposed to high versus low.

6.0 CONCLUSION

The PAT is a capable tool for quantifying the negative performance effects of operational stress on psychometric performance. Specifically, the PAT is targeted towards performance in tactical aviation. Initial studies have shown that PAT possesses high face validity as an aviation relevant task. The PAT's high face validity can be attributed to its development around mission essential competencies previously identified by AFRL. In addition to high face validity, a study conducted by Phillips, 2019 demonstrated that PAT possesses high sensitivity in measuring the psychometric effects of hypoxia. However, the PAT is a flexible tool that can accommodate a broad range of operational stressors, besides hypoxia, through the modification of session and trial duration. For example, a hypoxia experiment may require two, two-minute iterations whereas a fatigue experiment may require five, six-minute iterations. Despite the aviation related stressor that is in question, PAT can provide standard metrics to quantify and document the performance effects of that stressor and its interactions with other stressors. A further advantage of PAT is its ease of use -- participants can be trained to use PAT in less than an hour and PAT will automatically identify each participant's optimal difficulty level for stress testing. Additionally, PAT presents an engaging interface and is executed like a videogame rather than a classic psychometric test.

More work must be completed to establish PAT's sensitivity to a broader variety of stressors to include dehydration, fatigue, G-stress, motion sickness and other common aeromedical stressors. Additional studies should be conducted to establish PAT's reliability across a broad range of iteration times to establish the psychometric properties of PAT when played for sustained periods longer than two minutes. The manual tracking component is physically fatiguing and may lead to a drop off in performance over longer sustained periods of play. Previous studies employing PAT have been conducted using standard 22-inch computer screens. However, the visual search area of the tracking task and math tasks will increase if larger, non-standard screens are used. Future studies should also be conducted to establish if difficulty increases with screen size.

7.0 REFERENCES

Alliger, G.M., Beard, R., Bennett, W., Colegrove, C.M., Garrity, M. (2007). Understanding mission essential competencies as a work analysis method. Technical Report Air Force Research Laboratory Human Effectiveness Directorate Warfighter Readiness Research Division.

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 Years Later. Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting, 904-908. Santa Monica: HFES.

Hart, S. G. & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.) Human Mental Workload. Amsterdam: North Holland Press.

Kerlinger, F. N. (1986). *Foundations of behavioral research* (2nd ed.). New York: Holt, Rinehart and Winston.

Loesch, L. C., & Vacc, N. A. (Eds.) (1997). Research in counseling & therapy. Greensboro, NC: ERIC Counseling and Student Services Clearinghouse.

Phillips, J.B. (2019). Evaluation of the reliability and criterion validity of the performance assessment tool. Report to Sponsor (Air Force Research Laboratory).

Ponterotto, J.G. & Ruckdeschel, D.E (2007). An overview of coefficient alpha and a reliability for estimating adequacy of internal consistency coefficients with psychological research measures. *Perceptual Motor Skills*, 105, 10023-7478.

Symons, S., France, M., Bell, J., Bennett, W. (2007). Linking knowledge and skills to mission essential competency-based syllabus development for distributed mission operations. Technical Report Air Force Research Laboratory Human Effectiveness Directorate Warfighter Readiness Research Division.

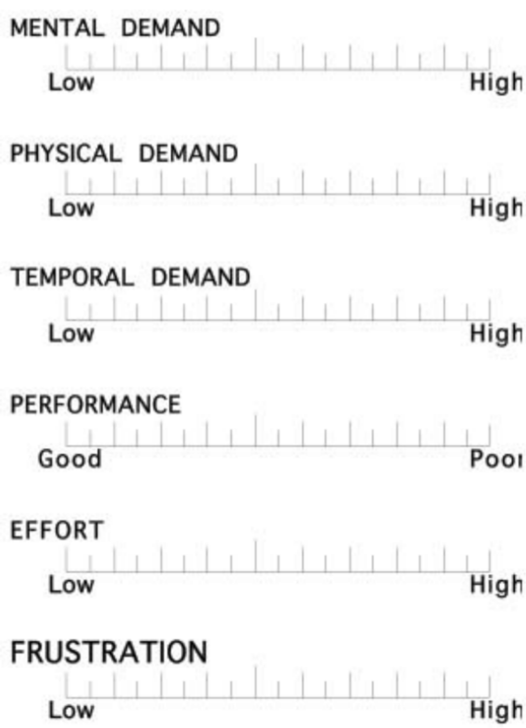
Villano, M. & Crowell, C.R. (2021). Performance assessment tool (PAT) for PAD4P2CP user documentation version 2.6b. Air Force Research Laboratory 711th Human Performance Wing.

LIST OF ABBREVIATIONS AND ACRONYMS

AFRL – Air Force Research Laboratory
PAT - Performance Assessment Tool
PCOLAC - Pensacola Composite
PATC - Performance Assessment Tool Composite
IHMC - Institute for Human and Machine Cognition
ICC - Inter-class Correlation Coefficients
NASA-TLX - Nasa Task Load Index
USAFSAM - United States Air Force School of Aerospace Medicine

APPENDIX

NASA TLX Rating Scale and Definition
(Retrieved from Hart, 2006)



RATING SCALE DEFINITIONS		
Title	Endpoints	Descriptions
MENTAL DEMAND	*Low/High*	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	*Low/High*	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	*Low/High*	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
EFFORT	*Low/High*	How hard did you have to work (mentally and physically) to accomplish your level of performance?
PERFORMANCE	*Good/Poor*	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
FRUSTRATION LEVEL	*Low/High*	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?