Intersession Reliability and Within-Session Stability of a Novel Perception-Action Coupling Task

Christopher Connaboy; Caleb D. Johnson; Alice D. LaGoy; Gert-Jan Pepping; Richard J. Simpson; Zhigang Deng; Luming Ma; Joanne L. Bower; Shawn R. Eagle; Shawn D. Flanagan; Candice A. Alfano

BACKGROUND: The perception-action coupling task (PACT) was designed as a more ecologically valid measure of alertness/reaction times compared to currently used measures by aerospace researchers. The purpose of this study was to assess the reliability, within-subject variability, and systematic bias associated with the PACT.

METHODS: There were 16 subjects (men/women = 9 / 7; age = 27.8 ± 3.6 yr) who completed 4 identical testing sessions. The PACT requires subjects to make judgements on whether a virtual ball could fit into an aperture. For each session, subjects completed nine cycles of the PACT, with each cycle lasting 5 min. Judgement accuracy and reaction time parameters were calculated for each cycle. Systematic bias was assessed with repeated-measures ANOVA, reliability with intraclass correlation coefficients (ICC), and within-subject variability with coefficients of variation (CV_{TE}).

RESULTS: Initiation time (Mean = 0.1065 s) showed the largest systematic bias, requiring the elimination of three cycles to reduce bias, with all other variables requiring, at the most, one. All variables showed acceptable reliability ($ICC > 0.70$) and within-subject variability ($CV_{TE} < 20\%$) with only one cycle after elimination of the first three cycles.

CONCLUSIONS: With a three-cycle familiarization period, the PACT was found to be reliable and stable.

KEYWORDS: reaction time, response time, action boundary, perceptual-motor.

Connaboy C, Johnson CD, LaGoy AD, Pepping G-J, Simpson RJ, Deng Z, Ma L, Bower JL, Eagle SR, Flanagan SD, Alfano CA. *Intersession reliability and within-session stability of a novel perception-action coupling task*. Aerosp Med Hum Perform. 2019; 90(2):1–7.

The perception-action coupling task (PACT) was designed as a more ecologically valid measure of alertness and reaction time, in terms of being a measure that follows motor control theory regarding how humans move and interact with their environments. This is proposed in comparison to a number of measures, but specifically to currently used measures of reaction time by aerospace researchers and clinicians working with aerospace professionals. This study provides evidence that the PACT can be used as a reliable measure, showing consistent performance across multiple testing sessions.

Perception-action coupling describes the inextricable link between perceiving and acting, whereby action both informs and regulates perception, and what is perceived is simultaneously informed and regulated by the action.[12] Gap closure and the accuracy of action-boundary and action-capability perception are the behaviors most commonly analyzed to understand how perception-action behavior is regulated.[11,18,23] These behaviors are often analyzed in response to changes in the task at hand (e.g., changes in rules, load, control interface sensitivity, stimulus regularity, etc.),[18,29] changes in the organism (e.g., force production capacity, postural regulation, visual acuity, anxiety, fatigue, etc.),[13,18,27] or changes in environmental constraints (e.g., altitude, temperature, etc.).

Gap closure refers to goal directed activities that involve intercepting or avoiding objects or events within the environment. Good examples of such goal directed activities relevant to the current study include catching or hitting a ball, jumping a gap, or steering to avoid a collision.[18,21,28] The accuracy of

---

From the Neuromuscular Research Laboratory, Department of Sports Medicine and Nutrition, University of Pittsburgh, Pittsburgh, PA; the School of Exercise Science, Australian Catholic University, Brisbane Campus, Banyo, Australia; the Department of Nutritional Sciences, The University of Arizona, Tuscon, AZ; and the Departments of Computer Science and Psychology, University of Houston, Houston, TX.

This manuscript was received for review in June 2018. It was accepted for publication in November 2018.

Address correspondence to: Christopher Connaboy, 3860 S. Water St., Pittsburgh, PA, 15217; connaboy@pitt.edu.

DOI: https://doi.org/10.3357/AMHP.5190.2019

action-boundary and action-capability perception relates to the concept of affordances, whereby the identification of 'opportunities for action' (i.e., affordances) are regulated by an individual's accuracy in relating their own capabilities that inform on actions (leg strength, body/object dimensions, finger dexterity/fine motor control) to an action-boundary.[11] An action-boundary, as described by Fajen et al.,[11] is the critical point at which the limitations of a particular action are met, necessitating a different action in order to maintain a successful motoric response. Some examples of action boundaries, based on the previous examples of capabilities, would be: maximal distance for jumping (leg strength), smallest opening one could fit through (body/object dimensions), quickest time one is able to manipulate an object with a joystick (finger dexterity/fine motor control).

Accuracy of perceptual-motor judgements, or the ability of an individual to recognize his/her action capabilities and action boundaries, has broad implications for successful control and decision-making during movement tasks. Inadequate attunement to these capabilities and boundaries has been shown to result in altered postural control and movement patterns, increased latency in reactionary measures, and decreases in task performance.[10,13,22] Summarily, it would seem that the dynamic integration of perception and action is key to a number of variables related to behavioral risk and human performance. This would seem to be highly applicable to aerospace professionals (pilots, astronauts, and other military personnel) who operate in challenging environments (e.g., zero gravity, extreme temperatures, adverse weather, hypoxia/hyperoxia)[1,4] and under a range of psychological/physiological stressors (e.g., altered sleep patterns, cognitive/physical fatigue, motion sickness).[20,26] Furthermore, these individuals often operate under these conditions for prolonged periods of time.[33] It follows, then, that the dynamic integration of perception and action would be a key feature in measures meant to capture changes in alertness/reaction time in response to these stressors.

However, traditional measures of alertness or reaction time[6,17,19] mostly require the individual to respond to a given stimulus as quickly as possible, termed simple reaction time measures.[19] Other measures do require an individual to make a quick decision between responding and not responding based on the type of stimulus, often referred to as "go, no-go" tasks or choice/complex reaction time measures.[6,17] Both of these traditional measures, simple and choice reaction time, have been used often in past research in aerospace professionals.[9,31,33] However, even choice measures do not require the individual to make a perceptual judgement, based on the spatial or dynamic properties of the presented stimulus. Therefore, they do not encapsulate the types of decisions that must be made when judging affordances for a given movement behavior. That is, the dynamic integration of perception and action is not fully captured by these traditional measures.

However, research that has incorporated perceptual-motor judgement has shown that successful movement behaviors can be maintained even when an individual reacts slower (delayed reaction times) and movement is initiated at an extended interval from a stimulus signal.[28] A movement solution to a defined task may change in response to changes in organismic or task constraints (i.e., fatigue, sleep disruption, wakefulness) and, while initial reaction time increases or remains consistent, other compensatory strategies may be employed to maintain successful overall performance. Therefore, a delayed reaction time, in and of itself, may not be indicative of unsuccessful performance or translate to disturbed motor planning. This is especially true for astronauts, pilots, and other military personnel, given that they are required to maintain a high level of performance under prolonged stressors coming from a wide-range of sources (mentioned above).[20,26,33] It is possible that only when a specific series of organismic/task constraints goes beyond a key threshold, successful motor performance can no longer be maintained, irrespective of any accommodations achieved through alterations in motor coordination to solve a specific movement task. To enable the identification of these thresholds that may induce perceptual deficits, more sensitive, ecologically valid and robust/usable measures are required.

Based on this need, a novel PACT software was developed following a task first described by Smith and Pepping.[28] In their study, a computer-based task was described in which subjects were asked to make judgements on whether a fixed-size ball could fit through apertures of varying sizes.[28] Subjects were asked to respond in two different conditions, either by moving a mouse cursor toward/away from the aperture or pressing a "yes/no" key.[28] In developing the PACT, the general task requirements and one of the two response types (moving toward/away from the aperture) were maintained. However, the task was moved to a tablet-based application to make it more portable, automated, and user-friendly. Second, where only aperture size was changed between trials in the Smith and Pepping[28] study, the use of a tablet-based application allowed for both ball and aperture size to vary from trial to trial. This, in effect, served to increase the ecological validity of the task, requiring subjects to perceive whether a fit is possible with properties of both the object and target changing from trial to trial.

With these alterations, and the previous work by Smith and Pepping,[28] the PACT would seem to present a perceptual-motor measure that is highly usable and holds good construct validity as an ecologically valid assessment. However, before it can be established whether this software can provide an ecologically valid measure of an individual's ability to accurately identify possibilities for action, it must be shown to be a reliable and stable measure. Therefore, the purpose of this study was threefold: to 1) establish the extent of any systematic bias between testing sessions; 2) examine the test-retest reliability; and 3) determine the within-subject variability associated with the PACT outcome data. The term systematic bias is used based on the definition outlined by Hopkins,[15] as mean changes in performance within and between sessions, which can indicate learning, fatigue, boredom, etc.

METHODS

Subjects

An observational, test-retest design was employed for the current study. There were 16 subjects (men/women = 9/7, age = 27.8 ± 3.6 yr) who reported to the lab for 4 testing sessions. Subjects were asked to refrain from consuming any caffeine in the 4 h prior to testing and to arrive in a well-fed, well-rested, and alert state. To be included in the study subjects had to meet the criteria of: having corrected 20/20 vision, being free from any visual impairments, and having no need to take medications that would have impaired cognitive processes, alertness, or vision. The study protocol was approved in advance by the University of Pittsburgh Institutional Review Board. Each subject provided written informed consent before participating.

Procedures

During each of the four testing sessions, subjects completed nine cycles of the PACT in a quiet environment, with minimal distractions. The PACT requires the subject to make determinations as to whether a series of virtual balls (diameter ranging from 10–60 mm) presented on an iPad (Apple Inc., Cupertino, CA) can fit through a series of virtual apertures (diameter ranging from 18–44 mm). Eight ratios of ball-to-aperture size were presented, ranging from 0.2 to 1.8, depicted in **Fig. 1**. Ball-to-aperture size ratios were presented in a randomized order and each ratio was presented 16 times across each cycle. To perform the PACT, subjects started with their index or middle finger of their dominant hand on the starting button (depicted in Fig. 1). At a randomized interval, between 0.01 and 0.70 s, the ball and aperture appeared on the screen. If the subject determined that the ball could fit through the aperture, they moved their finger from the starting position to a virtual joystick (depicted in Fig. 1), swiping upwards (forward) to direct the ball toward the hole. If they determined that the ball did not fit, the subject moved their finger to the joystick, swiping downwards (backward) to direct the ball away from the aperture. As soon as the action was completed, the subjects moved their finger back to the start button. Subjects were instructed to respond as quickly and accurately as possible and were not provided feedback about their performance throughout testing.

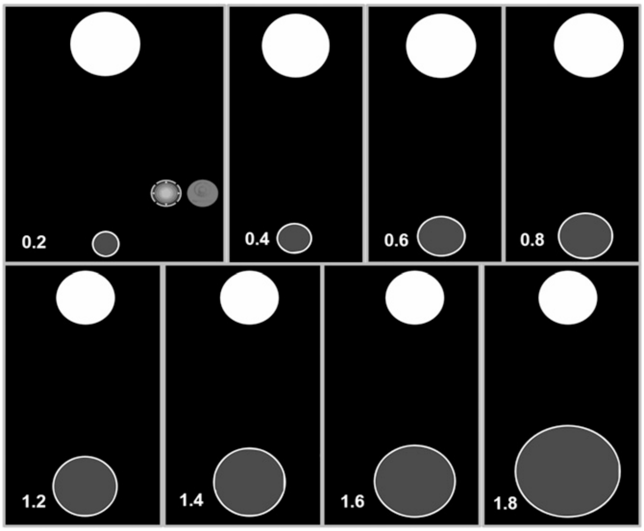

Fig. 1. Depiction of PACT interface and example ball to aperture ratios. Ball = circle at bottom of screen, aperture = circle at top of screen, start button = button on the bottom right, joystick = circle to the left of the start button. Depicted ball and aperture sizes are examples of each ratio; actual ball (10–60 mm) and aperture (18–44 mm) sizes varied randomly within each ratio.

Each cycle of the PACT lasted approximately 5 min, depending on how quickly subjects moved their finger back to the start button after each movement. Subjects were given a 15-min break after completing each set of three consecutive cycles. During breaks, subjects were allowed to move around and perform any activities they wished in order to relax. Testing sessions were separated by at least 6 d to allow for washout, and the mean number of days in between sessions was 9.67 ± 3.4. Finally, across sessions subjects were scheduled for the same general time of day (morning, afternoon, evening).

To assess the accuracy of action boundary judgements (ACC), the ratio of correct to incorrect responses was calculated for each cycle and expressed as a percentage based on the following formula:

$$\textit{Correct response ratio} = \frac{\textit{Number of correct responses}}{\textit{Number of pairings}} \times 100$$

A correct response was considered one where either: 1) the ball fit and the subjects swiped forward on the joystick, or 2) ball did not fit and the subjects swiped backward on the joystick. The reactive component of the PACT was analyzed for only correct responses by dividing the total time between the presentation of the stimulus (the ball-aperture pairing) and the response into different phases. Reaction Time (RT) was calculated as the time interval between the presentation of the stimulus and the subjects lifting their finger off the start button. Movement time (MT) was calculated as the time interval between the subjects lifting their finger and moving it to the joystick to respond. Finally, initiation time (IT) was calculated as the time interval between the subjects initiating a response with the joystick and completing the response.

Statistical Analyses

All statistics were performed using IBM SPSS 23.0 (IBM, Armonk, NY). An iterative approach was taken to the data analyses in an attempt to not only identify the inherent stability of the PACT data, but the relevant testing parameters (i.e., familiarization period, number of testing cycles) necessary to achieve stable measures with the PACT. To assess the presence of any systematic bias within each variable (ACC, RT, MT, IT), 4 × 9 repeated measures analysis of variance (ANOVA) were first calculated, with session (4 levels) and cycle (9 levels) as the two within-subject factors. Further, time-series plots were formulated for visual assessment. Sphericity was assessed with Mauchly's test of sphericity and a Greenhouse-Geiser correction (GG) was applied to P-values as appropriate. The interaction factor (session x cycle) was assessed for each ANOVA and, given the presence of a significant interaction, cycles were eliminated until the systematic bias was eliminated (i.e., 4 × 8, 4 × 7, etc…). Next, the main effects of session and cycle were examined and marginal comparisons were performed with paired *t*-tests, using Bonferroni-corrected *P*-values, when main effects were found to be significant.

After cycles had been removed to eliminate systematic bias, intraclass correlation coefficients [ICC (3,1)] were calculated in an iterative manner, averaging variables across all remaining cycles first and systematically eliminating cycles and recalculating coefficients. This was done on a case by case basis, dependent on the results of the ANOVA for each variable (i.e., if the first three cycles were eliminated, then the first cycle included was cycle 4). Finally, the mean coefficients of variation were calculated for each variable using this same process to assess within-subject variability across testing sessions. Coefficients were calculated using the typical error of the measure (CV_{TE}), as described by Hopkins.[16] In the case of MT, which showed significant departures from normality across cycle averages, log transformations were applied before calculating CV_{TE}, also as described by Hopkins.[16] An alpha level of 0.05 was used for all inferential statistics.

RESULTS

Time-series plots depicting the variability of all variables across cycles and sessions are located in **Fig. 2**. Results of the inferential statistics are summarized below.

The results of the repeated-measures, 4 × 9 ANOVA, including all cycles, showed no presence of systematic bias for RT. Examination of the interaction term showed no significant interaction of session x cycle [F (24, 360) = 0.408, P (GG) = 0.910]. A significant main effect of cycle [F (8, 120) = 2.802, P = 0.007, partial η^2 = 0.157] was observed; however, marginal comparisons did not show any significant differences in RT averaged across cycles (mean difference = 0.001–0.008 s, P = 0.225–1.00). All cycles were included for calculation of ICCs and CV_{TE}.

The results of the repeated-measures, 4 × 9 ANOVA for MT mirrored those for RT, with a nonsignificant interaction term [F (24, 360) = 1.729, P (GG) = 0.167] and main effects for both session [F (3, 45) = 2.329, P (GG) = 0.672] and cycle [F (8, 120) = 0.391, P (GG) = 0.128]. All cycles were included for calculation of ICCs and CV_{TE}.

Initiation time was found to have the most variability, requiring the elimination of the first three cycles from testing sessions before the presence of systematic bias was removed. The results of the repeated-measures, 4 × 6 ANOVA showed the interaction term to be nonsignificant [F (15, 225) = 2.417, P (GG) = 0.056]. Examination of the main effects showed a significant main effect of session [F (3, 45) = 4.491, P = 0.008, partial η^2 = 0.230]; however, examination of the marginal comparisons revealed no significant differences in IT averaged across sessions (mean difference = 0.001–0.043 s, P = 0.063–1.00). Results of the ICCs and CV_{TE} for all variables can be found in **Table I**. For IT, cycles 4–9 were included in the analysis.

Judgement accuracy required one cycle be eliminated from each session to remove the presence of systematic bias. The

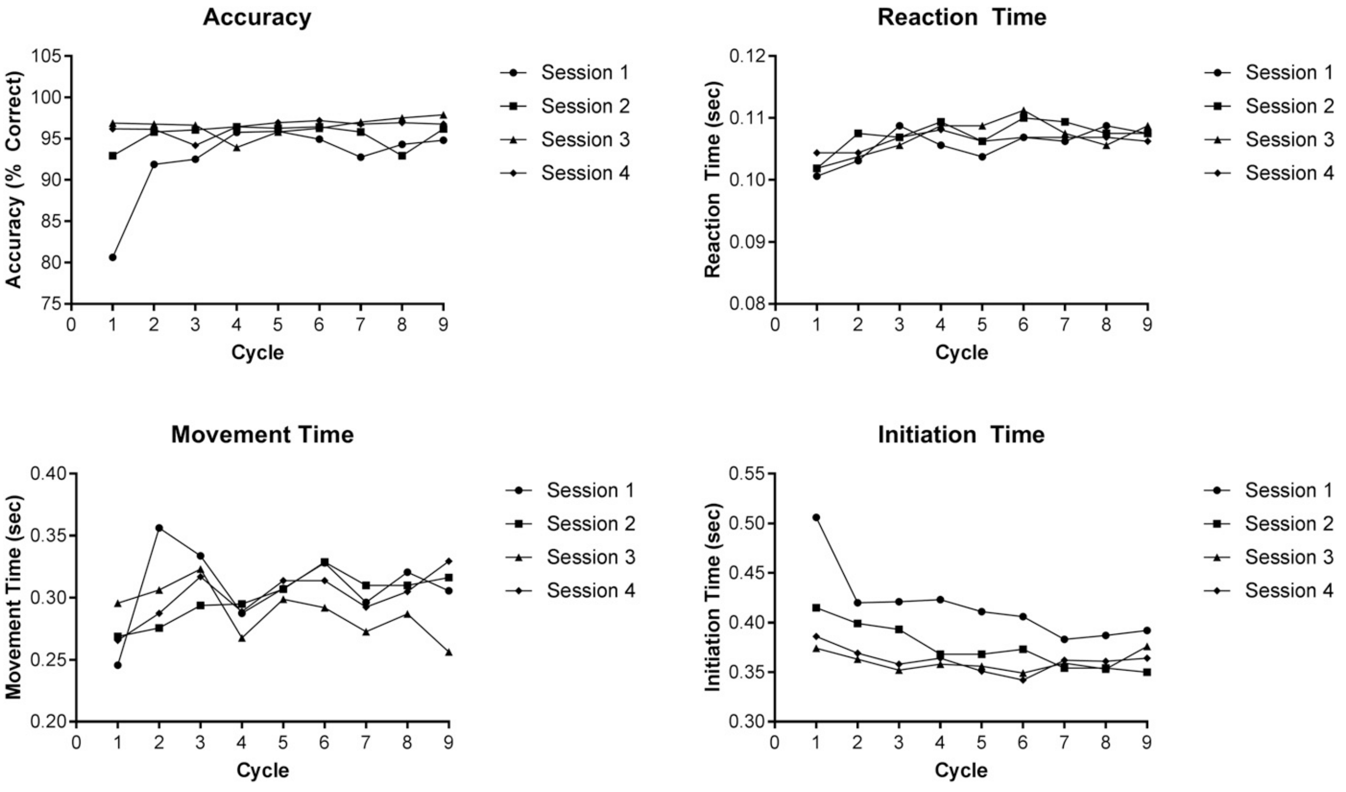

Fig. 2. Means for initiation time, movement time, reaction time, and accuracy across cycles and sessions.

Table I. Intraclass Correlation Coefficients and Coefficients of Variation by Cycle.

	ICC	95% CI	CV_{TE}	95% CI
Reaction Time				
9 Cycles	0.884	0.750–0.955	9.24%	7.51–12.07%
8 Cycles	0.882	0.745–0.954	9.68%	7.86–12.64%
7 Cycles	0.883	0.748–0.955	9.72%	7.90–12.70%
6 Cycles	0.874	0.728–0.951	10.13%	8.23–13.24%
5 Cycles	0.872	0.725–0.950	9.90%	8.04–12.93%
4 Cycles	0.866	0.711–0.948	10.11%	8.21–13.21%
3 Cycles	0.853	0.683–0.943	11.35%	9.22–14.83%
2 Cycles	0.870	0.721–0.950	9.22%	7.49–12.05%
1 Cycle	0.865	0.710–0.947	10.35%	8.41–13.53%
Movement Time				
9 Cycles	0.980	0.956–0.992	12.33%	9.90–16.40%
8 Cycles	0.979	0.955–0.992	12.94%	10.39–17.24%
7 Cycles	0.977	0.950–0.991	13.55%	10.87–18.06%
6 Cycles	0.975	0.947–0.990	13.90%	11.15–18.53%
5 Cycles	0.971	0.938–0.989	14.97%	12.00–20.44%
4 Cycles	0.964	0.923–0.986	15.30%	12.26–20.44%
3 Cycles	0.956	0.906–0.983	17.27%	13.82–23.14%
2 Cycles	0.943	0.879–0.978	20.26%	16.17–27.25%
1 Cycle	0.877	0.739–0.952	25.33%	20.12–34.30%
Initiation Time				
6 Cycles	0.547	0.060–0.821	43.28%	35.16–56.54%
5 Cycles	0.642	0.244–0.860	50.72%	41.19–66.25%
4 Cycles	0.972	0.931–0.990	7.20%	5.85–9.40%
3 Cycles	0.974	0.946–0.990	5.95%	4.83–7.77%
2 Cycles	0.992	0.983–0.997	4.07%	3.31–5.32%
1 Cycle	0.906	0.785–0.964	11.88%	9.65–15.52%
Accuracy				
8 Cycles	0.787	0.519–0.918	2.42%	1.96–3.16%
7 Cycles	0.795	0.514–0.923	2.39%	1.94–3.13%
6 Cycles	0.786	0.504–0.918	2.00%	1.62–2.61%
5 Cycles	0.767	0.474–0.911	2.13%	1.73–2.79%
4 Cycles	0.700	0.359–0.882	2.87%	2.33–3.75%
3 Cycles	0.602	0.204–0.837	3.53%	2.86–4.61%
2 Cycles	0.545	0.127–0.809	4.13%	3.36–5.40%
1 Cycle	0.362	−0.265–0.741	4.58%	3.72–5.98%

ICC = Intraclass correlation coefficient, 95% CI = 95% confidence interval, CV_{TE} = coefficient of variation using typical error.
Initiation time: begins with the fourth cycle; movement time and reaction time: begins with the first cycle; accuracy: begins with the second cycle; Movement time: CV_{TE} result of log transformed values.

Table II. Intraclass Correlation Coefficients and Coefficients of Variation for Movement Time, Reaction Time, and Accuracy with First Three Trials Removed.

	ICC	95% CI	CV_{TE}	95% CI
Reaction Time				
6 Cycles	0.873	0.726–0.951	9.11%	7.40–11.90%
5 Cycles	0.799	0.537–0.923	9.07%	7.37–11.85%
4 Cycles	0.869	0.718–0.949	9.30%	7.56–12.15%
3 Cycles	0.849	0.674–0.941	11.10%	9.01–14.50%
2 Cycles	0.849	0.675–0.941	12.04%	9.78–15.72%
1 Cycle	0.830	0.634–0.934	12.72%	10.33–16.61%
Movement Time				
6 Cycles	0.979	0.955–0.992	14.06%	11.27–18.75%
5 Cycles	0.972	0.935–0.990	13.55%	10.87–18.06%
4 Cycles	0.978	0.953–0.991	14.64%	11.74–19.54%
3 Cycles	0.979	0.955–0.992	15.34%	12.29–20.49%
2 Cycles	0.972	0.940–0.989	16.23%	12.99–21.71%
1 Cycle	0.943	0.878–0.978	19.28%	15.82–26.64%
Accuracy				
6 Cycles	0.766	0.511–0.908	2.62%	2.13–3.42%
5 Cycles	0.749	0.473–0.901	2.87%	2.33–3.75%
4 Cycles	0.815	0.612–0.927	2.02%	1.64–2.64%
3 Cycles	0.809	0.598–0.925	2.19%	1.78–2.86%
2 Cycles	0.820	0.621–0.929	2.05%	1.66–2.67%
1 Cycle	0.707	0.391–0.884	3.28%	2.66–4.28%

ICC = intraclass correlation coefficient, 95% CI = 95% confidence interval, CV_{TE} = coefficient of variation using typical error.
Initiation time: begins with the fourth cycle; movement time and reaction time: begins with the first cycle; accuracy: begins with the second cycle; movement time: CV_{TE} result of log transformed values.

results of the repeated-measures, 4 × 8 ANOVA revealed a nonsignificant interaction term [$F(21, 315) = 1.449$, $P(GG) = 0.226$]. The main effect of session was found to be significant [$F(3, 45) = 3.246$, $P = 0.031$]; however, marginal comparisons showed no significant differences in ACC averaged across session (mean difference = 0.055–2.367%, $P = 0.168$–1.00). Cycles 2–9 were included for calculation of ICCs and CV_{TE}.

Finally, because IT was found to require the removal of three cycles to eliminate systematic bias, a follow-up analysis was conducted where ICCs and CV_{TE} were recalculated for all other variables starting with the fourth cycle. The results of these tests can be found in **Table II**.

DISCUSSION

The current study was undertaken to investigate the reliability and stability of a novel measure of reaction time and accuracy: the PACT. The PACT was developed based on the original work of Smith and Pepping,[28] with slight alterations made to increase the usability and portability of the task. Comparisons with the previous study are difficult, as the authors report on only one variable for the condition most closely matching the current measure (computer-based, mouse movement), which they termed initiation time.[28] However, their definition of initiation time is essentially the equivalent of combining all response time variables from the PACT. In these terms, the mean initiation times reported by Smith and Pepping[28] (687 ms) is very close to the means found in the current study (mean RT+MT+IT = 785 ms). The slightly higher mean times in the current study are most likely related to the higher perceptual demands of the PACT, related to subjects having to attune to both a changing ball and aperture from trial to trial.

The first purpose of the current study was to investigate the presence of systematic bias over repeated sessions and testing cycles of the PACT. Results of repeated measures ANOVAs demonstrated that ACC only required the removal of one cycle, and RT and MT did not require the removal of any cycles to eliminate significant between-session, within-session, or interaction effects. However, IT required the removal of the first three cycles before the session by cycle interaction term became nonsignificant ($P > 0.05$). These results indicate the need for three cycles (approximately 15 min) to obtain a baseline familiarity with the PACT and stable measures for all variables assessed by the PACT. Following this familiarization session, no significant systematic bias was detected for any of the variables.

Previous literature on reaction time measures of a similar nature have generally failed to report analyses for systematic bias, making comparisons difficult. One study by Ayala et al.[3] found no systematic bias in hamstrings reaction time based on latency between a stimulus and muscle activation. However, the authors only report the results of the trial by session interaction effect, with no information on between- or within-session effects. This is especially troublesome given that several studies assessing the reliability of motor pattern or performance metrics (i.e., kinematics, single-leg squat performance) have reported systematic bias due to a between-sessions effect.[8,25,32] These studies have all shown significant differences between the first session and all following sessions, indicating the need for a full familiarization session.

The second aim of the current study was to investigate the test-retest reliability of the PACT. In this effort, cycles were eliminated for each variable to remove the presence of systematic bias and ICCs were calculated in an iterative manner to provide estimates of improvement in reliability with the addition of multiple testing cycles. The interpretation of reliability statistics is variable, with recommendations for acceptable reliability ranging from an ICC of 0.60 to 0.90.[2,14,24] However, Heaton et al.[14] reviewed studies assessing the reliability of neuropsychological measures and reported a range of 0.70 to 0.90 as "generally good." Based on this criteria, IT, MT, and RT were all found to have acceptable reliability with only one cycle of testing, with ACC requiring four cycles (Table I). In examining the trend in ACC (Fig. 2), this effect is evident, with the first three cycles of the first session showing marked improvement, and then leveling off for the remaining cycles and sessions. Further, when the first three cycles were removed due to the systematic bias present in IT (Table II), only one cycle was required to reach adequate reliability for all variables. Overall, the PACT demonstrated superior reliability compared to similar measures, with previously reported ICCs on choice reaction time tasks ranging from 0.26–0.69, and the majority in the range of 0.46–0.52.[5,7,30]

The final aim was to investigate the within-subject variability inherent in the PACT using the same iterative process as for the test-retest reliability. Like reliability, the interpretation of CV_{TE} is variable and dependent on the type of measure being assessed, as well as the expected magnitude of change a researcher or clinician wishes to detect. Also, like systematic bias, a lack of studies reporting within-subject variability for similar measures makes it hard to, at the very least, formulate an expected value for the CV_{TE}. Two studies have included the standard difference of the error for choice reaction time tasks, reporting values of 12.77% and 19.38%. While there are slight differences in the calculation of these metrics (CV and standard error of the difference), these studies provide the best comparison to the current one.

Initiation time and MT demonstrated the greatest effect of additive testing cycles on the CV_{TE}, with RT and ACC showing no significant change in the CV_{TE} beyond the first cycle (Table I). Initiation time required two cycles to achieve a stable CV_{TE} (4.07%), where the addition of cycles produced only a minimal change in the coefficient, and MT required four cycles (15.30%). However, related to previously reported values discussed above, IT showed a consistent CV_{TE} with only one cycle (11.88%) and MT with three cycles (17.27%). Further, when interpreting the coefficients for MT after removal of the first three cycles (Table II), the CV_{TE} was stable with two cycles (16.23%) and, consistent with previous studies, with only one (19.28%). Overall, the results demonstrate that, with a three-cycle familiarization period, the within-subject variability of the PACT, across all variables, is consistent with previous literature on complex reaction times with only one cycle of testing. However, two cycles may be required to achieve a stable CV_{TE}, where additional added cycles yield minimal reductions in the coefficient.

In summary, the results of the current study demonstrate that, with a three-cycle familiarization period, the PACT demonstrates no systematic bias, good reliability, and within-subject variability that is consistent with expected values, requiring only one 5-min cycle of testing. Further, as discussed in length in the introduction, the PACT would seem to hold good construct validity as an ecological assessment, requiring subjects to make perceptual judgements based on scaling the spatial properties of an object to an action boundary. However, in cases where investigators or clinicians working with astronauts or pilots, such as physicians, sports medicine professionals, or psychologists, require a highly reliable measure or are interested in variables that may elicit smaller changes in PACT performance, two cycles of testing may be required (10 min). Finally, these are general recommendations and we urge individuals to consider the results for themselves and make a decision on the necessary familiarization and testing periods based on their specific needs.

ACKNOWLEDGMENTS

This study was funded by the National Aeronautics and Space Administration; grant number NNX15AC13G. The opinions or assertions contained herein are the views of the authors and are not to be constructed as official or reflecting the views of NASA. Further, the funding body did not have a role in the design, analysis, or interpretation of the current study.

Authors and affiliations: Christopher Connaboy, M.Sc., Ph.D., Alice D. LaGoy, B.S., M.S., Shawn R. Eagle, ATC, MAT, and Shawn D. Flanagan, M.S., Ph.D., Neuromuscular Research Laboratory, Department of Sports Medicine and Nutrition, University of Pittsburgh, Pittsburgh, PA; Caleb D. Johnson, M.S., Ph.D., Spaulding National Running Center, Harvard Medical School, Department of Physical Medicine and Rehabilitation, Cambridge, MA; Gert-Jan Pepping, M.Sc., Ph.D., School of Exercise Science, Australian Catholic University, Brisbane Campus, Banyo, Queensland, Australia; Richard J. Simpson, B.Sc., Ph.D., Department of Nutritional Sciences, The University of Arizona, Tuscon, AZ; Zhigang Deng, M.S., Ph.D., and Luming Ma, B.S., M.S., Department of Computer Science, and Candice A. Alfano, Ph.D., Department of Psychology, University of Houston, Houston, TX; Joanne L. Bower, M.Sc., Ph.D., School of Applied Social Sciences, De Montfort University, Leicester, United Kingdom;

REFERENCES

1. Aherne BB, Zhang C, Chen WS, Newman DG. Pilot decision making in weather-related night fatal helicopter Emergency Medical Service accidents. Aerosp Med Hum Perform. 2018; 89(9):830–836.

2. Anastasi A. Psychological testing. New York: Collier Macmillan; 1988.
3. Ayala F, De Ste Croix M, Sainz de Baranda P, Santonja F. Inter-session reliability and sex-related differences in hamstrings total reaction time, pre-motor time and motor time during eccentric isokinetic contractions in recreational athlete. J Electromyogr Kinesiol. 2014; 24(2):200–206.
4. Bouak F, Vartanian O, Hofer K, Cheung B. Acute mild hypoxic hypoxia effects on cognitive and simulated aircraft pilot performance. Aerosp Med Hum Perform. 2018; 89(6):526–535.
5. Broglio SP, Ferrara MS, Macciocchi SN, Baumgartner TA, Elliott R. Test-retest reliability of computerized concussion assessment programs. J Athl Train. 2007; 42(4):509–514.
6. Chmura J, Krysztofiak H, Ziemba AW, Nazar K, Kaciuba-Uscilko H. Psychomotor performance during prolonged exercise above and below the blood lactate threshold. Eur J Appl Physiol Occup Physiol. 1998; 77(1–2):77–80.
7. Collie A, Maruff P, Makdissi M, McCrory P, McStephen M, Darby D. CogSport: reliability and correlation with conventional cognitive tests used in postconcussion medical evaluations. Clin J Sport Med. 2003; 13(1):28–32.
8. Connaboy C, Coleman S, Moir G, Sanders R. Measures of reliability in the kinematics of maximal undulatory underwater swimming. Med Sci Sports Exerc. 2010; 42(4):762–770.
9. Dart T, Gallo M, Beer J, Fischer J, Morgan T, Pilmanis A. Hyperoxia and hypoxic hypoxia effects on simple and choice reaction times. Aerosp Med Hum Perform. 2017; 88(12):1073–1080.
10. Daviaux Y, Mignardot J-B, Cornu C, Deschamps T. Effects of total sleep deprivation on the perception of action capabilities. Exp Brain Res. 2014; 232(7):2243–2253.
11. Fajen BR, Riley MA, Turvey MT. Information, affordances, and the control of action in sport. Int J Sport Psychol. 2009; 40(1):79–107.
12. Gibson JJ. The visual perception of objective motion and subjective movement. Psychol Rev. 1954; 61(5):304–314.
13. Graydon MM, Linkenauger SA, Teachman BA, Proffitt DR. Scared stiff: the influence of anxiety on the perception of action capabilities. Cogn Emotion. 2012; 26(7):1301–1315.
14. Heaton RK, Temkin N, Dikmen S, Avitable N, Taylor MJ, et al. Detecting change: a comparison of three neuropsychological methods, using normal and clinical samples. Arch Clin Neuropsychol. 2001; 16(1):75–91.
15. Hopkins WG. Measures of reliability in sports medicine and science. Sports Med. 2000; 30(1):1–15.
16. Q1 ~~Hopkins WG. Spreadsheets for analysis of validity and reliability. A new view of statistics. Sportscience. 2015; 19:36–42.~~
17. Kopp B, Rist F, Mattler U. N200 in the flanker task as a neurobehavioral tool for investigating executive control. Psychophysiology. 1996; 33(3):282–294.
18. Lessard DA, Linkenauger SA, Proffitt DR. Look before you leap: Jumping ability affects distance perception. Perception. 2009; 38(12):1863–1866.
19. Logan GD, Cowan WB, Davis KA. On the ability to inhibit simple and choice reaction time responses: a model and a method. J Exp Psychol Hum Percept Perform. 1984; 10(2):276–291.
20. Morris MB, Wiedbusch MD, Gunzelmann G. Fatigue incident antecedents, consequences, and aviation operational risk management resources. Aerosp Med Hum Perform. 2018; 89(8):708–716.
21. Oudejans RR, Michaels CF, Bakker FC, Dolne MA. The relevance of action in perceiving affordances: perception of catchableness of fly balls. J Exp Psychol Hum Percept Perform. 1996; 22(4):879–891.
22. Palmer CJ, Riccio GE, Van Emmerik RE. Orienting under load: intrinsic dynamics and postural affordances for visual perception. Ecol Psychol. 2012; 24(2):95–121.
23. Pepping G-J, Li Fo-X. The role of haptic exploration of ground surface information in perception of overhead reachability. J Mot Behav. 2008; 40(6):491–498.
24. Portney L, Watkins M. Foundations of clinical research: application to practice. Stamford (CT): Appleton & Lange; 1993.
25. Rawcliffe AJ, Simpson RJ, Graham SM, Psycharakis SG, Moir GL, Connaboy C. Reliability of the kinetics of British Army foot drill in untrained personnel. J Strength Cond Res. 2017; 31(2):435–444.
26. Sallinen M, Akerstedt T, Harma M, Henelius A, Ketola K, et al. Recurrent on-duty sleepiness and alertness management strategies in long-haul airline pilots. Aerosp Med Hum Perform. 2018; 89(7):601–608.
27. Schweigart G, Mergner T, Barnes G. Object motion perception is shaped by the motor control mechanism of ocular pursuit. Exp Brain Res. 2003; 148(3):350–365.
28. Smith J, Pepping G-J. Effects of affordance perception on the initiation and actualization of action. Ecol Psychol. 2010; 22(2):119–149.
29. Stone JA, Maynard I, North JS, Panchuk D, Davids K. Emergent perception–action couplings regulate postural adjustments during performance of externally-timed dynamic interceptive actions. Psychol Res. 2015; 79(5):829–843.
30. Strauss GP, Allen DN, Jorgensen ML, Cramer SL. Test-retest reliability of standard and emotional Stroop tasks: an investigation of color-word and picture-word versions. Assessment. 2005; 12(3):330–337.
31. Takács E, Czigler I, Pató LG, Balázs L. Dissociated components of executive control in acute hypobaric hypoxia. Aerosp Med Hum Perform. 2017; 88(12):1081–1087.
32. Urquhart BG, Moir GL, Graham SM, Connaboy C. Reliability of 1RM split-squat performance and the efficacy of assessing both bilateral squat and split-squat 1RM in a single session for non-resistance-trained recreationally active men. J Strength Cond Res. 2015; 29(7):1991–1998.
33. Vrijkotte S, Roelands B, Meeusen R, Pattyn N. Sustained military operations and cognitive performance. Aerosp Med Hum Perform. 2016; 87(8):718–727.