

Reliability and Minimal Detectable Change of the Standardized Assessment of Reaction Time

Landon B. Lempke, PhD, ATC*†‡; Eric J. Shumski, MS§||;
Thomas A. Prato, MS, ATC§||; Robert C. Lynall, PhD, ATC§||

*Michigan Concussion Center, School of Kinesiology, University of Michigan, Ann Arbor; †Division of Sports Medicine, Boston Children's Hospital, MA; ‡Micheli Center for Sports Injury Prevention, Waltham, MA; §UGA Concussion Research Laboratory and ||UGA Biomechanics Laboratory, Department of Kinesiology, University of Georgia, Athens

Context: Postconcussion reaction time deficits are common, but existing assessments lack sport-related applicability. We developed the Standardized Assessment of Reaction Time (StART) tool to emulate the simultaneous cognitive and motor function demands in sport, but its reliability is unestablished.

Objectives: To determine the intrarater, interrater, and test-retest reliability of StART and to examine the dual-task effect, time effect, and relationships between StART and computerized and laboratory-based functional reaction time assessments.

Design: Prospective cohort study.

Setting: Clinical laboratory.

Patients or Other Participants: Twenty healthy, physically active individuals (age = 20.3 ± 1.8 years, females = 12, no concussion history = 75%).

Main Outcome Measure(s): Participants completed the StART, computerized reaction time (Stroop task via CNS Vital Signs), and laboratory-based jump landing and cutting reaction time under single-task and dual-task (subtracting by 6s or 7s) cognitive conditions at 2 testing sessions a median of 7 days apart. We used intraclass correlation coefficients (ICCs), repeated-measure analysis of variance, and Pearson *r* correlations to address our aims.

Results: Overall, good to strong interrater (ICC [2,k] range = 0.83–0.97), intrarater (ICC [3,k] range = 0.91–0.98), and test-retest (ICC [3,k] range = 0.69–0.89) reliability were observed. A significant reaction time assessment-by-cognitive condition interaction was present ($P = .018$, $\eta_p^2 = 0.14$), with StART having the largest dual-task effect. Main time effects for dual-task conditions were seen across all reaction time assessments (mean difference = -25 milliseconds, $P = .026$, $\eta_p^2 = 0.08$) with improved performance at the second testing session. No StART outcomes correlated with computerized reaction time ($P > .05$), although some correlated with single-task (r range = 0.42–0.65) and dual-task (r range = 0.19–0.50) laboratory cutting reaction time.

Conclusions: The StART demonstrated overall reliable performance relative to other reaction time measures. Reliability coupled with a strong dual-task effect indicates that StART is a valid measure for examining functional reaction time and may have future utility for sport-related concussion return-to-play decision-making.

Key Words: mild traumatic brain injury, response time, visuomotor function, return to play

Key Points

- Excellent overall reliability with only 3 trials indicates any clinician can accurately and quickly use the Standardized Assessment of Reaction Time (StART).
- Minimal detectable changes were observed and indicate strong metric sensitivity for future clinical use.
- The StART demonstrated a similar or stronger dual-task effect compared with other reaction time measures and thus may better induce cognitive-motor interference clinically.

Concussions are a widespread condition occurring across sport,^{1,2} hospital,³ and military⁴ populations and result in transient dysfunctional neurotransmission throughout the brain.⁵ Concussions present clinically with heightened symptoms, sensorimotor impairment, and dysfunctional neurocognition. Postconcussion deficits typically normalize for most people between 21 and 28 days after injury.^{1,6} To ensure that common deficits can be detected, best clinical practices call for standardized assessments to examine these areas and thus to accurately diagnose concussions.^{7,8} With neurocognitive assessments, clinicians specifically evaluate numerous cognitive domains, with reaction time being one robust and valid domain of focus.⁹ Postconcussion

reaction time deficits are well-established, are considered an important assessment outcome due to deficits observed immediately after injury until 21 to 59 days postinjury,¹⁰ and are robust across all reaction time methods used.^{10–12} Thus, numerous methods ranging widely in complexity exist to ensure that reaction time can be tested in any setting after concussion.^{12–17}

Current reaction time measures are typically evaluated in sports medicine⁷ via either computerized neurocognitive assessments^{12,18} or clinical assessments (eg, drop ruler test).^{14,15} One major concern is that computerized reaction time and clinical reaction time measures do not correlate with functional reaction time determined using whole-body

motion-capture tracking.^{13,19} Increasingly, researchers have also identified lingering impairments on dual-task (ie, simultaneously completing cognitive and motor tasks) gait assessments 2 months after concussion²⁰ and consistently observed a 2-fold increase in musculoskeletal injury for up to 2 years.^{21,22} The reaction time and gait findings^{13,20–22} collectively raise concerns about return-to-sport safety and the lack of functional, sport-like applicability, as current clinical measures²³ are not associated with heightened musculoskeletal injury after concussion,^{21,22} whereas dual-task gait outcomes are.^{24,25} The recommended diagnostic assessments may not be well-suited for determining return-to-sport readiness, as the on-field demands are not emulated.

Clinical assessments occur in a controlled, quiet testing environment^{26,27} and consist of simple finger movements (computerized neurocognitive testing) or static stances (postural stability). These circumstances differ greatly from the dynamic environments of most sport settings and the whole-body movement coordination needed to successfully compete and avoid injury. Such differences between clinical and sport environments may be important considerations for ensuring return-to-sport safety, and some components could perhaps be emulated via dual-task and whole-body functional movement assessments. We have developed a novel reaction time assessment battery called the *Standardized Assessment of Reaction Time (StART)* to address these concerns and potentially improve return-to-sport patient safety.^{28,29} However, before implementation in clinical practice, we must first understand the StART psychometric properties and their relationship with established computerized and laboratory-based functional reaction time assessments to ensure reliability and identify the minimal detectable change in scores.

The purposes of our study were to (1) determine the intrarater, interrater, and test-retest reliability; standard error of measurement; and minimal detectable change of StART; (2) examine the dual-task and time effect of StART compared with other reaction time assessments; and (3) examine the relationship between StART and previously established reaction time measures (computerized¹⁸ and functional movement^{13,30}). We hypothesized that (1) StART would display good³¹ intrarater, interrater, and test-retest reliability (intraclass correlation coefficients [ICCs] ≥ 0.75) and relatively small measurement error and detectable change; (2) StART would display a similar dual-task effect (ie, worse reaction time during the dual-task relative to the single-task condition) and time effect as other reaction time assessments; and (3) StART would not correlate with computerized reaction time but would correlate with functional movement reaction time.

METHODS

Study Design and Participants

An *a priori* power analysis³² was conducted using the test-retest reliability ICCs provided by Lynall et al³⁰ for their laboratory-based jump landing, single-legged hop, and cutting reaction time performance under single and dual tasks. Power analysis was calculated using 2-tailed tests ($\alpha = .05$, $\beta = 0.90$) and the lowest published³⁰ test-retest reliability ICC value of 0.75 relative to the alternative

hypothesis ICC of 0.21 (lowest fair ICC interpretation value).^{32,33} A sample size of 20 participants was needed to maintain adequate power for detecting true effects.

Therefore, a convenience sample of 20 healthy, recreationally active, college-aged participants were enrolled and completed all assessments across 2 testing sessions a median of 7 (interquartile range = 7–7; range = 7–38) days apart as part of the prospective repeated-measures study. All participants were recruited across the university campus through fliers placed in university buildings; interested individuals contacted the research team to be screened and to enroll in the study. Volunteers were included if they were healthy young adults (age = 18–30 years old), English was their primary language, and they were *recreationally active* (performed physical activity for a minimum of 30 minutes, 3 times per week).³⁴ Volunteers were excluded if they had a lower extremity injury in the last 3 months resulting in physical activity time loss of ≥ 1 day; a history of lower extremity or low back orthopaedic surgery; a self-reported concussion within the past year; or any learning disability or attention-deficit/hyperactivity, psychiatric, balance, or mental health disorder. Participants received their honoraria only after completing testing sessions 1 and 2 separately to encourage maximal performance throughout and minimize potential attrition. This investigation was approved by the University of Georgia institutional review board, and all individuals provided written informed consent before the study.

Instrumentation and Procedures

Participants completed the assessments in a block-randomized (StART, computerized, and laboratory) order established *a priori*. The assessment order was identical between individual participant testing sessions to keep any learning or fatiguing effects consistent. All single-task conditions were completed before dual-task conditions, where applicable. The assessments consisted of (1) StART, (2) computerized reaction time,¹⁸ (3) laboratory jump landing,^{13,30} and (4) laboratory cutting.^{13,30} All tests were completed in a designated, isolated laboratory space in order to mimic recommended concussion assessment conditions.^{26,27} Jump landing and cutting were performed under both single-task (completing the motor task as quickly as possible) and dual-task (completing the cognitive task while simultaneously completing the motor task as quickly as possible) conditions. The cognitive task was standard across all assessments and was used in StART. Participants subtracted from a random number between 90 and 150 by either 6s or 7s throughout the motor task.^{13,30,35,36} Subtraction was initiated when they were instructed to “get set” and stopped once the trial was completed. The subtraction task was started from a different random number for each subsequent trial.

Standardized Assessment of Reaction Time. All trials were video and audio recorded on an iPad (model A1701; Apple Inc) using either the OnForm (version 1.95)³⁷ or Hudl (no longer available; sold and phased out to OnForm) application and recorded at 240 Hz (ie, Slo-Mo function on an iPhone [Apple Inc]) with 720-pixel resolution while fixed on a tripod approximately 3.05 m from participants to ensure that individuals of all heights would completely fit in the frame with space above and below their head and feet.²⁸

The video frame rate is equivalent to or faster than that of criterion standard motion-capture cameras.³⁶ A light-emitting diode penlight tip was placed in the camera recording frame and served as the time-synchronized visual stimulus to enable us to calculate reaction time. For all StART trials, participants were instructed to “get set,” and then the penlight was randomly illuminated 2 to 10 seconds later. Participants initiated and completed the specified condition as fast as possible after the penlight illumination. For visual depictions of each StART subtest, please see Lempke et al.²⁸

The StART consisted of 5 trials of 3 movements (standing, single-legged balance, and cutting) across 2 conditions (single and dual task, 30 trials total in each testing session) and took approximately 10 minutes to complete.²⁸ For the standing trials, individuals stood with their feet together and hands on hips. The standing trials were conducted as a relatively simple way of assessing reaction time that theoretically would display postconcussion deficits.¹⁰ For single-legged balance trials, participants stood on their *nondominant leg* (the leg they would not use to kick a ball) with hands on hips as they balanced throughout the trial. We chose single-legged balance because patients with concussion are known to have challenges integrating sensory information^{20,38,39} and may display a slower reaction time. For the cutting trials, individuals assumed a semisquatting athletic stance with hands on hips. The cutting trials were used to emulate sport-like functional movements that have been previously completed only in laboratory settings using motion-capture equipment.^{13,30}

Participants moved their hands off their hips until their arms were straight out in a “T” position (parallel to the ground) as soon as they saw the penlight illuminate for the standing and single-legged balance tasks. Cutting trials required participants to sprint from the starting position to the left- or right-side target positioned approximately 3.05 m away at 45° by performing an athletic cutting motion. The time (milliseconds) between penlight activation and the first frame of arm movement (eg, hands off hips, elbows bending, fingers raising) was deemed the reaction time for standing and single-legged trials, and the first body movement (eg, foot pivoting, hands coming off hips, torso or head laterally deviating) was identified as the reaction time for cutting trials. A secondary reaction time metric was calculated for cutting trials that included the time from light activation to the first frame of the torso moving, as indicated by white tape placed on the sternum to replicate a center-of-mass area starting to move, similar to prior motion-capture work.¹³ Each trial from the 6 StART reaction time movement and cognitive combinations was scored individually and then averaged for each movement (standing, single-legged balance, and cutting) and cognitive condition (single task and dual task) separately. Additionally, we derived composite single-task and dual-task StART outcomes by averaging their respective trials separately, as well as an omnibus StART outcome comprising all StART outcomes averaged as detailed in the Data Processing and Statistical Analysis section.

Computerized Reaction Time. The computerized reaction time was assessed via the Stroop Task test on a reliable and valid computerized neurocognitive platform (CNS Vital

Signs).^{18,40} The Stroop Task is widely used in computerized neurocognitive programs.^{18,41} Participants were presented with the embedded standard instructions for each of the 3 subtests along with a practice sample before testing.¹³ The test displayed color words (“red,” “yellow,” “blue,” or “green”) randomly in different font colors for the subtests. The 3 subtests required participants to press the spacebar as quickly as possible after (1) any word was presented, (2) the color word was presented in the same color font, and (3) the color word was not presented in the same color font. The computerized reaction time composite (milliseconds) was calculated as a weighted score from the subtests per the neurocognitive test’s standard method and was the main outcome measure.^{7,18}

Laboratory Jump Landing and Cutting Reaction Time. Laboratory-based jump landing and cutting occurred in an 8-camera 3-dimensional motion-capture space (model MIQUIS; Qualisys AB) recording at 240 Hz as previously described.^{13,30,36} Participants stood on a 30-cm box and placed at 50% of their body height behind the landing target with 3 reflective markers over their posterior-superior iliac spine and sacral body. They adopted an athletic stance after being told to “get set” by the research team, and then a green light was randomly triggered in 2 to 5 seconds. Individuals initiated the movement as quickly as possible after seeing the green light.⁴² The time (milliseconds) between visual stimulus and sacral marker movement of ≥ 3 cm in either the sagittal or frontal plane from its mean position 0.5 seconds before the visual stimulus was deemed the reaction time for each assessment.^{13,30,36,42} Participants were given at least 1 practice trial before data collection and took a 1-minute break between the jump landing and cutting tasks to minimize any physical fatigue.

They jumped forward off the box, landed on both legs, and performed a maximal-height countermovement jump for jump landing in 5 single-task and 5 dual-task trials.^{13,30,36} For cutting, participants jumped forward, landed on a single leg, and immediately executed a 45° athletic-cutting motion in the direction provided before the trial. They landed on their left foot and ran to the right side of the laboratory space for cuts to the right side and vice versa for left cuts. Individuals completed 5 trials in each direction under single- and dual-task conditions (10 trials total). The marker positional data for the laboratory jump landing and cutting sacral were processed and filtered with a fourth-order, low-pass Butterworth filter and a 10-Hz cutoff frequency. The laboratory jump landing and cutting data were imported, processed, and analyzed using Visual3D (version 2021.02.1; C-Motion Inc) to calculate reaction time across all trials as previously defined.^{13,30,36,42} Each laboratory jump landing and cutting trial was processed individually and averaged for each movement (jump landing and cutting) and cognitive condition (single task and dual task) separately.

Data Processing and Statistical Analysis

We averaged the 5 trials from all reaction time assessments and cognitive conditions to determine the 6 (single- and dual-task standing, single-legged balance, and cutting) main StART outcome scores separately. We used

Table 1. StART Reliability, Standard Error of Measurement, and Minimal Detectable Change

StART Outcome	Reliability, Intraclass Correlation Coefficient (95% CI)			Standard Error of Measurement, ms	Minimal Detectable Change, ms
	Interrater [2,k]	Intrarater [3,k]	Test-Retest [3,k]		
Single task					
Standing	0.93 (0.83, 0.97)	0.98 (0.96, 0.99)	0.79 (0.55, 0.90)	11	32
Single legged	0.97 (0.93, 0.98)	0.98 (0.95, 0.99)	0.78 (0.53, 0.90)	12	33
Cutting	0.92 (0.82, 0.96)	0.99 (0.97, 0.99)	0.88 (0.73, 0.94)	10	29
Dual task					
Standing	0.99 (0.98, 0.99)	0.99 (0.98, 1.00)	0.79 (0.54, 0.90)	23	64
Single legged	0.97 (0.93, 0.99)	0.98 (0.96, 0.99)	0.74 (0.44, 0.88)	31	85
Cutting	0.83 (0.63, 0.92)	0.91 (0.80, 0.96)	0.69 (0.33, 0.86)	25	69
Composite					
Single task	0.93 (0.86, 0.97)	0.99 (0.97, 0.99)	0.85 (0.68, 0.93)	8	21
Dual task	0.93 (0.86, 0.97)	0.95 (0.90, 0.98)	0.74 (0.43, 0.88)	20	57
StART	0.93 (0.84, 0.97)	0.97 (0.94, 0.99)	0.83 (0.64, 0.92)	12	34

Abbreviation: StART, Standardized Assessment of Reaction Time.

these 6 StART main outcomes to calculate the 3 following composite scores as the average score from the specified trials: single-task reaction time (all single-task StART trials), dual-task reaction time (all dual-task StART trials), and an omnibus StART composite reaction time (all StART trials combined) to comprehensively examine the measurement reliability and precision. The trials from laboratory-based jump landing and cutting were averaged separately to calculate mean single- and dual-task reaction time outcomes for each movement separately.

Descriptive statistics were computed for participant demographic variables and mean reaction time assessments. To determine interrater reliability, 2 raters (E.J.S. and T.A. P.) independently scored the first testing session StART reaction time for all trials and conditions while blinded to the other rater's scoring. Author E.J.S. scored all participants' first testing session StART trials on 2 occasions at a median (interquartile range) of 23 (21–24.5) days apart to assess intrarater reliability and then examined all participants' first and second testing sessions to assess test-retest reliability. Intrarater, interrater, and test-retest reliability were evaluated using ICCs^{31,43} with 95% CIs. Intrarater and interrater ICCs were generated from the first testing session. All ICCs were interpreted as *poor* (<0.50), *moderate* (0.50–0.74), *good* (0.75–0.89), or *strong* (≥ 0.90).³¹ The test-retest reliability ICC was used to calculate the standard error of measurement (SEM), which was used to calculate the minimal detectable change (MDC)⁴⁴ for each StART subtest and composite score using the equations described by Weir.⁴⁴ Test-retest reliability for center-of-mass cutting reaction time was poor (ICC [3,k] = 0.28), and therefore, we excluded this outcome from all subsequent analyses. Sensitivity analyses via separate repeated-measures analyses of variance (ANOVAs) were also conducted to determine the minimal number of trials needed for each StART subtest before the outcomes were significantly influenced. Specifically, we compared the average of all 5 trials versus the average of the first 4, the first 3, the first 2, and the first trial for each StART outcome separately with unadjusted post hoc *t* tests to detect any potential difference.

A 5 × 2 repeated-measures ANOVA was computed to compare the StART subtests (standing, single-legged balance, and cutting) and laboratory jump landing and

cutting reaction times under cognitive conditions (single task versus dual task) in the first testing session data. We conducted a 6 × 2 repeated-measures ANOVA to compare all reaction time measures across the 2 testing sessions in order to examine any time effect for the single-task condition and a 5 × 2 repeated-measures ANOVA (no computerized reaction time due to the absence of the dual-task) to examine any dual-task time effect. Any significant ANOVA interactions or main effects were followed up with post hoc Tukey *t* tests, mean differences, and 95% CIs.

To determine the relationship between StART and previously established reaction time measures at the first testing session, we calculated Pearson correlation coefficients between all reaction time measures. Correlations were interpreted as *negligible* (≤ 0.20), *low* (0.21–0.40), *moderate* (0.41–0.60), *high* (0.61–0.80), or *strong* (≥ 0.81). All data were assessed for general linear model assumptions before analysis and analyzed in the R Project for Statistical Programming (version 4.0.4; The R Foundation) with $\alpha = .05$ *a priori*. Specifically, all reaction time metrics were examined for the normality of their residuals visually via QQ plots and statistically via Shapiro-Wilk tests; no statistical or visual violations were present (*P* values $\geq .067$).

RESULTS

Participants

All 20 participants completed both testing sessions. Their mean \pm SD age was 20.3 \pm 1.8 years, height was 173.0 \pm 8.6 cm, and mass was 69.9 \pm 13.4 kg. They reported 7.0 \pm 1.2 and 7.0 \pm 1.7 hours of sleep the night before the first and second testing sessions, respectively. A total of 60% of participants were female, 75% were without a history of concussion (n = 4 with 1 prior concussion, n = 1 with 2 prior concussions), 85% were right-hand dominant (identified as the hand with which the individual would throw), and 100% were right leg dominant.

Reliability, SEM, and MDC of StART and Trial Sensitivity Analysis

The interrater reliability for all StART outcomes was good to strong, with the lowest value for dual-task cutting

(ICC [2,k] = 0.83; Table 1). Intrarater reliability was strong across all StART outcomes; the lowest value was for dual-task cutting (ICC [3,k] = 0.91; Table 1). Test-retest reliability was relatively lower and ranged from moderate to good across outcomes (ICC [3,k] range = 0.69–0.89). The associated StART SEM values ranged from 8 to 31 milliseconds, and MDC values ranged from 21 to 85 milliseconds, with the highest (worst) value for dual-task single-legged balance (Table 1).

The sensitivity analyses revealed that numerous StART outcomes were statistically influenced by the number of trials used (Table 2). Post hoc testing indicated that the StART outcomes differed when the 5-trial average was compared with only the single trial ($P \leq .005$) but not when compared with the first 2-, 3-, or 4-trial averages. Similarly, all 3 StART composite scores using the 5-trial average were different than the 1-trial or 2-trial averages but not different than the 3- or 4-trial averages.

Dual-Task and Time Effect Across all Reaction Time Measures

A significant reaction time outcome-by-cognitive condition interaction was observed ($F_{4,19} = 3.17$, $P = .018$, $\eta_p^2 = 0.14$) such that the dual-task increase in reaction time varied by each assessment outcome (Table 3). All measures resulted in increased reaction time during dual-task versus single-task ($F_{4,19} = 155.83$, $P < .001$, $\eta_p^2 = 0.67$) conditions, with StART single-legged balance having the largest mean difference and StART cutting having the smallest.

No significant reaction time outcome-by-testing session interactions were seen for either single-task ($F_{5,19} = 0.93$, $P = .466$, $\eta_p^2 = 0.05$) or dual-task ($F_{4,19} = 0.72$, $P = .582$, $\eta_p^2 = 0.04$) comparisons. No main time effect was identified for single-task outcomes ($F_{5,19} = 0.92$, $P = .351$, $\eta_p^2 = 0.01$); however, for dual-task outcomes ($F_{4,19} = 5.88$, $P = .026$, $\eta_p^2 = 0.08$; Table 3), reaction time overall at the second testing session was 25 milliseconds (95% CI = 5, 45 milliseconds) faster. Mean differences between testing sessions for all outcomes are presented in Table 3, with StART single-legged balance and cutting during the dual task being faster during the second testing session.

Correlations Between Clinical, Laboratory, and StART Reaction Time Measures

None of the StART outcomes were significantly correlated with computerized reaction time ($P \geq .05$; Figure). The StART single-task outcomes of standing ($r = 0.42$), single-legged balance ($r = 0.49$), and cutting ($r = 0.65$) were moderately to highly correlated with single-task cutting, whereas no StART outcomes were correlated with single-task jump landing. Under dual-task conditions, only the StART cutting condition ($r = 0.50$) and StART dual-task composite ($r = 0.47$) were moderately correlated with laboratory-based cutting. Among StART conditions, various correlation levels were noted (r range = -0.19 – 0.57). All single- and dual-task StART subtest scores were highly correlated with their respective composite scores (Figure).

DISCUSSION

Our findings provide foundational evidence for the StART measurement properties, which will be beneficial for future potential clinical use. Excellent overall intrarater, interrater, and test-retest reliability was coupled with a relatively small SEM and MDC for most StART outcomes. We also detected similar or stronger dual-task effects for StART than for the laboratory-based measures from which it was derived,¹³ correlations with laboratory-based reaction time, and no correlation with computerized reaction time. Significant time effects were present for dual-task single-legged balance and cutting StART conditions, reflecting performance improvements upon retesting, which were similar to those in other dual-task paradigms.^{30,45} Cumulatively, our results indicate that StART may serve as a quick, reliable, accurate, and clinically feasible measure that can be implemented using readily available tools. Yet research in participants postconcussion is warranted to understand its clinical utility and safety before it is used in clinical decision-making.

Measurement reliability is an important consideration for interpreting whether assessment outcome changes are due to error or injury. Our findings demonstrated similar reliability as those for laboratory-based, functional reaction time measures³⁰ and StART had overall excellent and

Table 2. Sensitivity Analysis for the Number of StART Trials Used

StART Outcome	Number of Trials Averaged							
	Mean Difference (95% CI), ^a P Value							
	5 vs 4		5 vs 3		5 vs 2		5 vs 1	
Single task								
Standing	2 (–3, 6)	.390	0 (–4, 5)	.909	1 (–4, 5)	.726	0 (–5, 4)	.856
Single legged ^b	2 (–5, 9)	.563	0 (–6, 7)	.919	–1 (–8, 6)	.721	–12 (–18, –5)	.001
Cutting ^a	–1 (–6, 5)	.766	–2 (–7, 3)	.447	–3 (–8, 3)	.298	–8 (–13, –2)	.005
Dual task								
Standing ^b	–5 (–18, 9)	.484	–10 (–24, 3)	.128	–13 (–26, 1)	.059	–33 (–47, –20)	<.001
Single legged	0 (–2, 2)	.979	–12 (–35, 10)	.289	–20 (–42, 3)	.086	1 (–21, 24)	.910
Cutting	0 (–13, 12)	.952	4 (–8, 17)	.511	–5 (–17, 8)	.475	6 (–7, 19)	.357
Composite								
Single task ^b	1 (–2, 3)	.694	–1 (–4, 2)	.546	–2 (–4, 1)	.299	–7 (–10, –4)	<.001
Dual task	–3 (–11, 6)	.549	–6 (–14, 3)	.172	–11 (–19, –3)	.011	–8 (–17, 1)	.054
StART ^b	–1 (–5, 3)	.647	–3 (–7, 1)	.127	–6 (–11, –2)	.003	–8 (–12, –4)	.001

Abbreviation: StART, Standardized Assessment of Reaction Time.

^a All mean differences and 95% CIs are presented in milliseconds. Bold values indicate significant post hoc *t* tests ($P \leq .05$).

^b Indicates significant omnibus analysis of variance ($P \leq .05$).

Table 3. Dual-Task and Time Effects Across Reaction Time Outcomes, ms

Assessment Category	Outcome	Session, Mean ± SD		Effect, Mean Difference (95% CI)	
		1	2	Dual Task (Dual vs Single)	Time (Session 2 vs 1)
StART	Standing, single task	187 ± 18	194 ± 35	118 (100, 136)	7 (−3, 17)
	Standing, dual task	305 ± 43	287 ± 58		−18 (−38, 2)
	Single legged, single task	191 ± 24	192 ± 28	139 (112, 166)	1 (−9, 11)
	Single legged, dual task	330 ± 67	298 ± 60		−32 (−57, −7)
	Cutting, single task	218 ± 30	216 ± 18	78 (60, 96)	−2 (−12, 8)
	Cutting, dual task	296 ± 37	269 ± 41		−27 (−51, −3)
	Single-task composite ^a	203 ± 22	205 ± 20		2 (−8, 12)
	Dual-task composite ^a	306 ± 37	281 ± 42		−25 (−48, −2)
	StART composite ^a	256 ± 24	243 ± 28		−13 (−29, 3)
Computer (Stroop)	Reaction time composite ^b	587 ± 48	592 ± 82		5 (−15, 25)
Laboratory (motion capture)	Jump landing, single task	462 ± 75	466 ± 55	121 (90, 152)	4 (−29, 37)
	Jump landing, dual task	584 ± 83	537 ± 70		−47 (−96, 2)
	Cutting, single task	424 ± 53	467 ± 58	121 (80, 162)	43 (8, 78)
	Cutting, dual task	545 ± 82	535 ± 45		−10 (−57, 37)

Abbreviation: StART, Standardized Assessment of Reaction Time.

^a Outcomes were not examined in analysis-of-variance models because they were derived from the other StART outcomes.

^b Outcome was not not examined in the dual-task analysis-of-variance model because no dual-task condition was completed. Bold effect values indicate a statistically significant dual-task or time effect ($P \leq .05$).

clinically suitable reliability. Specifically, intrarater and interrater reliability were mostly strong across all StART conditions (Table 1) and showed that the same examiner and different examiners can determine reaction time equally accurately. Test-retest reliability was lower than intrarater and interrater reliability but was still moderate to good (ICC [3,k] range = 0.69–0.89) with the lowest performing subtest being dual-task cutting. The relatively lower reliability for dual-task cutting may be attributed to it being the most cognitively and motor-demanding task in the StART, which may inherently introduce more variability as the cognitive-motor interference increases.⁴⁶ Although the StART test-

retest reliability was relatively lower, it was again similar to the laboratory-based, cutting reaction time measures previously established (ICC [3,k] range = 0.75–0.91).³⁰ We also observed a relatively small MDC range across StART subtests and composites (MDC range = 21–85 milliseconds) that was smaller than for laboratory-based, cutting reaction time³⁰ and may provide additional measurement diagnostic precision for future postconcussion use. Based on our sensitivity analyses, 5 trials are not necessary to measure stable StART performance. Two or three trials were needed for all StART outcomes to establish stable metrics (Table 2), and thus, we recommend 3 trials for each condition to ensure measurement stability and reduce StART administration time from approximately 10 minutes to 5 minutes.

Dual-task paradigms are growing in clinical use due to their functional applicability.^{13,30,47} We found that a dual-task effect was present in StART (mean difference range = 78–129 milliseconds; Table 3) that was similar to laboratory-based reaction time measures in our study and prior reports.¹³ Importantly, the StART random light stimulus window was 2 to 10 seconds, whereas the laboratory-based random light stimulus window was 2 to 5 seconds. The additional potential time provided to subtract during StART may have contributed to the optimal dual-task effects. Regardless, our results indicate that StART can sufficiently elicit the intended dual-task effect via simpler and clinically practical equipment.

Time effects during StART were assessed to understand whether improved performance occurred across repeat test administrations. Most outcomes did not show any time effects across the 7-day retest window (Table 3); however, single-legged balance (32 milliseconds) and cutting (27 milliseconds) under dual-task conditions did improve at the second testing session. Improved dual-task single-legged balance and cutting performance at the second testing session coupled with the test-retest reliability ICCs imply that learning or practice improvements took place. Better performance at the second testing session is commonly reported for dual-task^{30,45} and reaction time assessments

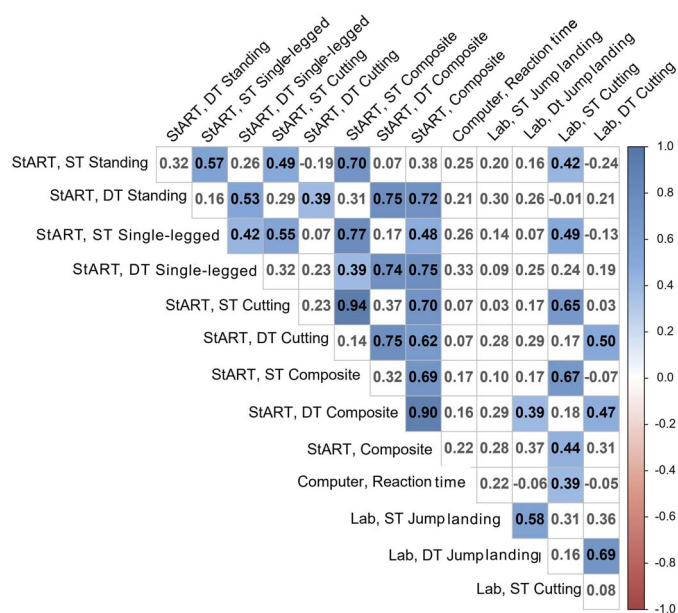


Figure. Pearson correlations across reaction time outcomes (n = 20). Only correlations with $P \leq .05$ are colored according to the correlation strength with black numeric values. White cells with grey numeric values indicate nonsignificant correlations (P values $> .05$). Abbreviations: StART, Standardized Assessment of Reaction Time; ST, single-task; DT, dual-task.

generally^{30,47,48} when individuals are retested within a relatively short time period. Improvements with repeat testing highlight the importance of determining test-retest reliability and using MDCs, which mathematically account for test-retest reliability,⁴⁴ in clinical practice to better isolate the source of the decreased performance.

Determination of the relationship between assessments is a key consideration to ensure that the constructs of interest are being measured accurately. We specifically examined the relationships among StART relative to laboratory-based reaction time and computerized reaction time to learn whether the appropriate factors were being assessed. The overall correlations supported our hypotheses (Figure): moderate to high correlations were evident between StART single-task subtests and the single-task laboratory-based cutting reaction time (r range = 0.42–0.65) and between StART dual-task cutting and dual-task laboratory-based cutting reaction time (r = 0.50). No StART outcomes correlated with single- or dual-task jump landing except the StART dual-task composite score (r = 0.39), which may indicate that only cutting, rather than general functional movement, was emulated in StART. Conversely, no to low correlations were present between any StART outcome and computerized reaction time (r range = 0.07–0.33). Thus, our cumulative findings support the notion that StART is a reliable and accurate assessment for measuring functional reaction time in clinical settings. Of note, moderate correlations between StART standing and single-legged balance outcomes (single task = 0.57, dual task = 0.53; Figure) may reflect redundancy. However, these possible redundancies may be attributed to the healthy cohort examined. Individuals experiencing a concussion often display postural stability impairments,²⁰ and therefore, evaluating the correlations in a sample of individuals experiencing a concussion may elicit different relationships. Thus, future researchers should address how the StART metrics and conditions relate to concussion in order to further optimize StART for maximal efficiency.

Limitations

We investigated college-aged, physically active individuals attending a university, but our results may not be generalizable to collegiate athletes or individuals younger or older than this cohort. Our research questions and study design required that we use numerous reaction time measures at each visit and that may have affected reaction time performance despite randomization. Hence, the outcome summary statistics may not represent expected performance in clinical practice and should not be used as healthy reference data. We also offered participant honoraria to support recruitment, the effort expended, and retention in the study, but this might have produced a biased sample or altered levels of effort than could be expected in a clinical setting. Therefore, future researchers should implement StART in a more naturalistic sports medicine setting, both pre- and postconcussion, to better understand StART performance and its potential diagnostic properties before it can be recommended for clinical decision-making.

CONCLUSIONS

The StART demonstrated good to strong intrarater, interrater, and test-retest reliability coupled with a relatively

small SEM and MDC. Dual-task effects that were similar or stronger than prior laboratory-based reaction time measures were identified,¹³ along with no correlation with computerized reaction time and correlations with laboratory-based reaction time. Excellent reliability combined with appropriate dual-task effects indicates that StART is an appropriate measure for examining functional reaction time by using relatively low-cost tools and may efficiently translate to clinical settings. However, future authors should study StART in patients after concussion to understand its clinical utility before it is used in clinical decision-making.

ACKNOWLEDGMENTS

We thank undergraduate students Abby Moss and Rachel Stupp for their assistance in data processing in this project, which was funded by the University of Georgia Louise E. Kindig Doctoral Grant.

REFERENCES

1. Wasserman EB, Kerr ZY, Zuckerman SL, Covassin T. Epidemiology of sports-related concussions in National Collegiate Athletic Association athletes from 2009–2010 to 2013–2014: symptom prevalence, symptom resolution time, and return-to-play time. *Am J Sports Med.* 2016;44(1):226–233. doi:10.1177/0363546515610537
2. O'Connor KL, Baker MM, Dalton SL, Dompier TP, Broglio SP, Kerr ZY. Epidemiology of sport-related concussions in high school athletes: National Athletic Treatment, Injury and Outcomes Network (NATION), 2011–2012 through 2013–2014. *J Athl Train.* 2017;52(3):175–185. doi:10.4085/1062-6050-52.1.15
3. Cancelliere C, Coronado VG, Taylor CA, Xu L. Epidemiology of isolated versus nonisolated mild traumatic brain injury treated in emergency departments in the United States, 2006–2012: sociodemographic characteristics. *J Head Trauma Rehabil.* 2017;32(4):E37–E46. doi:10.1097/HTR.0000000000000260
4. Harman DR, Hooper TL, Gackstetter GD. Aeromedical evacuations from Operation Iraqi Freedom: a descriptive study. *Mil Med.* 2005;170(6):521–527. doi:10.7205/MILMED.170.6.521
5. Barkhoudarian G, Hovda DA, Giza CC. The molecular pathophysiology of concussive brain injury—an update. *Phys Med Rehabil Clin N Am.* 2016;27(2):373–393. doi:10.1016/j.pmr.2016.01.003
6. Broglio SP, McAllister T, Katz BP, et al. The natural history of sport-related concussion in collegiate athletes: findings from the NCAA-DoD CARE Consortium. *Sports Med.* 2022;52(2):403–415. doi:10.1007/s40279-021-01541-7
7. Lempke LB, Schmidt JD, Lynall RC. Athletic trainers' concussion-assessment and concussion-management practices: an update. *J Athl Train.* 2020;51(1):17–26. doi:10.4085/1062-6050-322-18
8. McCrory P, Meeuwisse W, Dvorak J, et al. Consensus statement on concussion in sport—the 5th International Conference on Concussion in Sport held in Berlin, October 2016. *Br J Sports Med.* 2017;51(11):838–847. doi:10.1136/bjsports-2017-097699
9. Farnsworth JL II, Dargo L, Ragan BG, Kang M. Reliability of computerized neurocognitive tests for concussion assessment: a meta-analysis. *J Athl Train.* 2017;52(9):826–833. doi:10.4085/1062-6050-52.6.03
10. Lempke LB, Howell DR, Eckner JT, Lynall RC. Examination of reaction time deficits following concussion: a systematic review and meta-analysis. *Sports Med.* 2020;50(7):1341–1359. doi:10.1007/s40279-020-01281-0
11. Caccese JB, Eckner JT, Franco-MacKendrick L, et al. Interpreting clinical reaction time change and recovery after concussion: a baseline versus norm-based cutoff score comparison. *J Athl Train.* 2021;56(8):851–859. doi:10.4085/1062-6050-457-20

12. Covassin T, Stearne D, Elbin R. Concussion history and post-concussion neurocognitive performance and symptoms in collegiate athletes. *J Athl Train*. 2008;43(2):119–124. doi:10.4085/1062-6050-43.2.119
13. Lempke LB, Johnson RS, Schmidt JD, Lynall RC. Clinical versus functional reaction time: implications for postconcussion management. *Med Sci Sports Exerc*. 2020;52(8):1650–1657. doi:10.1249/MSS.0000000000002300
14. Eckner JT, Whitacre RD, Kirsch NL, Richardson JK. Evaluating a clinical measure of reaction time: an observational study. *Percept Mot Skills*. 2009;108(3):717–720. doi:10.2466/pms.108.3.717-720
15. Eckner JT, Kutcher JS, Richardson JK. Pilot evaluation of a novel clinical test of reaction time in National Collegiate Athletic Association Division I football players. *J Athl Train*. 2010;45(4):327–332. doi:10.4085/1062-6050-45.4.327
16. Catena RD, van Donkelaar P, Chou LS. Altered balance control following concussion is better detected with an attention test during gait. *Gait Posture*. 2007;25(3):406–411. doi:10.1016/j.gaitpost.2006.05.006
17. Heitger MH, Jones RD, Dalrymple-Alford JC, Frampton CM, Ardagh MW, Anderson TJ. Motor deficits and recovery during the first year following mild closed head injury. *Brain Inj*. 2006;20(8):807–824. doi:10.1080/02699050600676354
18. Gualtieri CT, Johnson LG. Reliability and validity of a computerized neurocognitive test battery, CNS Vital Signs. *Arch Clin Neuropsychol*. 2006;21(7):623–643. doi:10.1016/j.acn.2006.05.007
19. Avedesian JM, Covassin T, Baez S, Nash J, Nagelhout E, Dufek JS. Relationship between cognitive performance and lower extremity biomechanics: implications for sports-related concussion. *Orthop J Sports Med*. 2021;9(8):232596712111032246. doi:10.1177/232596712111032246
20. Büttner F, Howell DR, Ardern CL, et al. Concussed athletes walk slower than non-concussed athletes during cognitive-motor dual-task assessments but not during single-task assessments 2 months after sports concussion: a systematic review and meta-analysis using individual participant data. *Br J Sports Med*. 2020;54(2):94–101. doi:10.1136/bjsports-2018-100164
21. Lynall RC, Mauntel TC, Pohlign RT, et al. Lower extremity musculoskeletal injury risk after concussion recovery in high school athletes. *J Athl Train*. 2017;52(11):1028–1034. doi:10.4085/1062-6050-52.11.22
22. McPherson AL, Nagai T, Webster KE, Hewett TE. Musculoskeletal injury risk after sport-related concussion: a systematic review and meta-analysis. *Am J Sports Med*. 2019;47(7):1754–1762. doi:10.1177/0363546518785901
23. Buckley TA, Howard CM, Oldham JR, Lynall RC, Swanik CB, Getchell N. No clinical predictors of postconcussion musculoskeletal injury in college athletes. *Med Sci Sports Exerc*. 2020;52(6):1256–1262. doi:10.1249/MSS.0000000000002269
24. Oldham JR, Howell DR, Knight CA, Crenshaw JR, Buckley TA. Gait performance is associated with subsequent lower extremity injury following concussion. *Med Sci Sports Exerc*. 2020;52(11):2279–2285. doi:10.1249/MSS.0000000000002385
25. Howell DR, Buckley TA, Lynall RC, Meehan WP III. Worsening dual-task gait costs after concussion and their association with subsequent sport-related injury. *J Neurotrauma*. 2018;35(14):1630–1636. doi:10.1089/neu.2017.5570
26. Rahman-Filipiak AAM, Woodard JL. Administration and environment considerations in computer-based sports-concussion assessment. *Neuropsychol Rev*. 2013;23(4):314–334. doi:10.1007/s11065-013-9241-6
27. Rahn C, Munkasy BA, Barry Joyner A, Buckley TA. Sideline performance of the Balance Error Scoring System during a live sporting event. *Clin J Sport Med*. 2015;25(3):248–253. doi:10.1097/JSM.0000000000000141
28. Lempke LB, Oldham JR, Passalugo S, et al. Influential factors and preliminary reference data for a clinically feasible, functional reaction time assessment: the Standardized Assessment of Reaction Time. *J Athl Train*. 2023;58(2):112–119. doi:10.4085/1062-6050-0073.22
29. Lempke LB, Passalugo S, Baranker BT, et al. Relationship and latent factors between clinical concussion assessments and the functional, Standardized Assessment of Reaction Time (StART). *Clin J Sport Med*. 2022;32(6):e591–e597. doi:10.1097/JSM.0000000000001061
30. Lynall RC, Johnson RS, Lempke LB, Schmidt JD. Test–retest reliability of a functional reaction time assessment battery. *J Sport Rehabil*. 2021;30(8):1237–1241. doi:10.1123/jsr.2021-0021
31. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155–163. doi:10.1016/j.jcm.2016.02.012
32. Zou GY. Sample size formulas for estimating intraclass correlation coefficients with precision and assurance. *Stat Med*. 2012;31(29):3972–3981. doi:10.1002/sim.5466
33. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174. doi:10.2307/2529310
34. Haskell WL, Lee IM, Pate RR, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Med Sci Sports Exerc*. 2007;39(8):1423–1434. doi:10.1249/mss.0b013e3180616b27
35. Fino PC, Nussbaum MA, Broilinson PG. Locomotor deficits in recently concussed athletes and matched controls during single and dual-task turning gait: preliminary results. *J NeuroEng Rehabil*. 2016;13(1):65. doi:10.1186/s12984-016-0177-y
36. Lempke LB, Oh J, Johnson RS, Schmidt JD, Lynall RC. Single-versus dual-task functional movement paradigms: a biomechanical analysis. *J Sport Rehabil*. 2021;30(5):774–785. doi:10.1123/jsr.2020-0310
37. Video analysis app and coaching platform. OnForm. Accessed September 7, 2021. <https://www.getonform.com>
38. Lempke LB, Call JA, Hoch MC, Schmidt JD, Lynall RC. Lower extremity neuromuscular function following concussion: a preliminary examination. *J Sport Rehabil*. 2022;32(1):31–39. doi:10.1123/jsr.2022-0075
39. Manaseer TS, Gross DP, Mrazik M, Schneider K, Whittaker JL. Reconceptualizing postural control assessment in sport-related concussion: transitioning from the reflex/hierarchical model to the systems model. *Physiother Theory Pract*. 2021;37(7):763–774. doi:10.1080/09593985.2019.1648624
40. Broglio SP, Katz BP, Zhao S, McCrea M, McAllister T; CARE Consortium Investigators. Test-retest reliability and interpretation of common concussion assessment tools: findings from the NCAA-DoD CARE consortium. *Sports Med*. 2017;48(5):1255–1268. doi:10.1007/s40279-017-0813-0
41. Iverson GL, Lovell MR, Collins MW. Interpreting change on ImPACT following sport concussion. *Clin Neuropsychol*. 2003;17(4):460–467. doi:10.1076/clin.17.4.460.27934
42. Lynall RC, Blackburn JT, Guskiewicz KM, Marshall SW, Plummer P, Mihalik JP. Reaction time and joint kinematics during functional movement in recently concussed individuals. *Arch Phys Med Rehabil*. 2018;99(5):880–886. doi:10.1016/j.apmr.2017.12.011
43. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420–428. doi:10.1037/0033-2909.86.2.420
44. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005;19(1):231–240. doi:10.1519/15184.1
45. Howell DR, Brilliant AN, Meehan WP III. Tandem gait test-retest reliability among healthy child and adolescent athletes. *J Athl Train*. 2019;54(12):1254–1259. doi:10.4085/1062-6050-525-18

46. Plummer P, Eskes G. Measuring treatment effects on dual-task performance: a framework for research and clinical practice. *Front Hum Neurosci.* 2015;9:225. doi:10.3389/fnhum.2015.00225
47. Eckner JT, Richardson JK, Kim H, Joshi MS, Oh YK, Ashton-Miller JA. Reliability and criterion validity of a novel clinical test of simple and complex reaction time in athletes. *Percept Mot Skills.* 2015;120(3):841–859. doi:10.2466/25.15.PMS.120v19x6
48. Resch JE, Schneider MW, Munro Cullum C. The test-retest reliability of three computerized neurocognitive tests used in the assessment of sport concussion. *Int J Psychophysiol.* 2018;132(pt A):31–38. doi:10.1016/j.ijpsycho.2017.09.011

Address correspondence to Landon B. Lempke, PhD, ATC, Michigan Concussion Center, Exercise and Sport Science Initiative, School of Kinesiology, University of Michigan, 830 North University Avenue, Ann Arbor, MI 48109. Address email to llempke@umich.edu.