Published in final edited form as:

Percept Mot Skills. 2015 June ; 120(3): 841–859. doi:10.2466/25.15.PMS.120v19x6.

Reliability and Criterion Validity of a Novel Clinical Test of Simple and Complex Reaction Time in Athletes¹

James T. Eckner,

Department of Physical Medicine and Rehabilitation and Michigan NeuroSport, University of Michigan, Ann Arbor, Michigan

James K. Richardson,

Department of Physical Medicine and Rehabilitation, University of Michigan, Ann Arbor, Michigan

Hogene Kim,

Department of Mechanical Engineering, University of Michigan, Ann Arbor, Michigan

Monica S. Joshi,

Department of Mechanical Engineering, University of Michigan, Ann Arbor, Michigan

Youkeun K. Oh, and

Department of Mechanical Engineering, University of Michigan, Ann Arbor, Michigan

James A. Ashton-Miller

Department of Mechanical Engineering and Department of Biomedical Engineering, University of Michigan, Ann Arbor, Michigan

Summary

Slowed reaction time (RT) represents both a risk factor for and a consequence of sport concussion. The purpose of this study was to determine the reliability and criterion validity of a novel clinical test of simple and complex RT, called RTclin, in contact sport athletes. Both tasks were adapted from the well-known ruler drop test of RT and involve manually grasping a falling vertical shaft upon its release, with the complex task employing a go/no-go paradigm based on a slight cue. In 46 healthy contact sport athletes (24 males; M = 16.3 yr., SD = 5.0; 22 women: M age = 15.0 yr., SD = 4.0) whose sports included soccer, ice hockey, American football, martial arts, wrestling, and lacrosse, the latency and accuracy of simple and complex RTclin had acceptable test-retest and inter-rater reliabilities and correlated with a computerized criterion standard, the Axon Computerized Cognitive Assessment Tool. Medium to large effect sizes were found. The novel RTclin tests have acceptable reliability and criterion validity for clinical use and hold promise as concussion assessment tools.

¹This project was supported by the Rehabilitation Medicine Scientist Training Program (5K12HD001097). The authors wish to thank Ms. Lea Franco for her editorial assistance in reviewing and submitting this manuscript. We would also like to thank Dr. Jason Cromer for his assistance in addressing our questions about the Axon CCAT.

¹Address correspondence James T. Eckner, M.D., M.S., Department of Physical Medicine & Rehabilitation, University of Michigan, 325 E. Eisenhower Pkwy., Ann Arbor MI 48108 or (jeckner@med.umich.edu).

²Hogene Kim is currently at Department of Rehabilitation & Assistive Technology, National Rehabilitation Center Research Institute, Ministry of Health and Welfare, Seoul, South Korea. Monica S. Joshi is currently at The Boeing Company, Boeing Research and Technology, Human-System Integrated Technologies, North Charleston, SC. Youkeun K. Oh is currently at Department of Mechanical Engineering, Hongik University, Seoul, South Korea.

Keywords

athlete; assessment; reaction time

Reaction time (RT) is a clinically relevant measure in terms of assessing both function and health. Until recently, RT assessment required use of a computer running specialized software. Despite the ubiquitous nature of computers in the modern health care environment, this dependence makes it difficult to do routine assessment of RT by most health care providers during the delivery of clinical care in both in- and out-patient settings. To allow RT assessment in a wide variety of circumstances, we developed a simple, computer-independent clinical measure of RT by standardizing the simple “ruler drop test” that has long been used in physics classrooms to teach students the relationships between distance, acceleration, and time for a free-falling body (Chudler, 2009). While based on the same general “ruler drop” paradigm, our methodology differs from that employed by Montare (2009). The original device and method for measuring simple clinical reaction time, or simple RT_{clin}, has been previously described (Eckner, Whitacre, Kirsch, & Richardson, 2009).

Throughout this report, the abbreviation “RT” is used in general discussion of reaction time without implying a specific measurement tool or method. When applicable, the abbreviation “RT_{clin}” is used to describe “clinical” reaction times measured using the clinical reaction time device and method that is the subject of this report or our prior related work. Alternatively, the term “RT_{comp}” is used to describe “computerized” reaction times measured using the Axon Computerized Cognitive Assessment Tool (CCAT), which we used as the criterion standard for comparison in this study. The reader should note that since the time of data collection, the name of the Axon CCAT has been changed to the CogState CCAT. Additionally, the modifier, “simple” is used to describe any RT task during which the participant performs the same response to a single stimulus during every trial, while the modifier, “complex” is used to describe any RT task involving more than one stimulus and/or response across trials. Simple RT tasks are associated with a single outcome measure, namely RT latency (in ms), while complex RT tasks are associated with measures of both RT latency (in ms) and accuracy (% of correctly performed trials). In the case of complex RT_{clin}, a go/no-go testing paradigm is employed, so complex RT_{clin} accuracy refers to the percentage of complex RT_{clin} trials during which the participant performed the correct go/no-go response.

Reaction time has clinical and functional relevance in many contexts, including sport-related concussion, where it represents both a risk factor for, and a measurable consequence of brain injury. A slower RT is common after concussion, and is, in fact, one of the most sensitive indices of neurocognitive change following brain injury (Erlanger, Saliba, Barth, Almquist, Webright, & Freeman, 2001; Collie, Maruff, Makdissi, McCrory, McStephen, & Darby, 2003). Reaction time also has prognostic value following concussion (Lau, Lovell, Collins, & Pardini, 2009), with incremental value over post-concussion symptom monitoring alone, given that slower RT may indicate persisting injury beyond the time of symptom resolution (Warden, Bleiberg, Cameron, Ecklund, Walter, Sparling, et al., 2001; Broglio, Macciocchi,

& Ferrara, 2007a; Makdissi, Darby, Maruff, Ugoni, Brukner, & McCrory, 2010). Simple RTclin is a valid tool to assess concussions in athletes, with 75% sensitivity and 68% specificity for the effects of sport-related concussion using a cutoff score of 0 msec., i.e., interpreting any slowing of simple RTclin compared to an athlete's own pre-season baseline as abnormal (Eckner, Kutcher, & Richardson, 2010, 2011a, 2011b; Eckner, Kutcher, Broglio, & Richardson, 2014). In addition, simple RTclin predicts an athlete's ability to rapidly protect their head in a simulated sport environment (Eckner, Lipps, Kim, Richardson, & Ashton-Miller, 2011).

While simple RTclin has demonstrated promise as a concussion assessment tool in athletes, it is possible that additional modifications to the device and technique might improve its sensitivity and specificity for detecting concussion. Therefore the original simple RTclin device and technique were modified to permit assessment of a complex form of RTclin (Eckner, Richardson, Kim, Lipps, & Ashton-Miller, 2012). Briefly, the simple RTclin “ruler drop” paradigm was modified so that a light affixed to the device, illuminating (or not) at the instant the device begins to fall, signals whether to catch the device or allow it to fall. This complex RTclin protocol therefore represents a go/no-go testing paradigm. In the initial investigation using complex RTclin in healthy adults ages 19 to 83 years, the latency of complex RTclin was significantly greater than simple RTclin latency, with 71% of the slowing attributable to pre-movement time (Eckner, et al., 2012). This suggests that the complex RTclin task predominantly reflects cognitive processing, and therefore may be more affected by concussion than simple RTclin. Furthermore, there was a strong negative relationship between age and complex RTclin accuracy ($r = -.603$), with participants having greater difficulty in withholding the catching response during “light-off” trials than in catching the device during “light-on” trials (Eckner, et al., 2012). This further confirms that the complex RTclin task is measuring inhibitory function.

Before conducting additional field research to compare the sensitivities and specificities for assessing concussions based on simple and complex RT_{clin} measurements obtained using the new device, we assessed their test-retest and inter-rater reliabilities in a population of contact sport athletes. We also sought to determine their criterion validities with respect to existing validated computer-based measures of simple and complex reaction time (RT_{comp}). We elected to recruit contact sport athletes for this study because they represent the population of athletes at greatest risk for concussion, and therefore with the greatest potential to benefit from the new test. Athletes of both sexes, representing a wide age range from 8 to 30 years, and participating in multiple contact sports over a range of competitive levels were recruited to ensure broad representation of contact sport athletes and increase the external validity of the results. We selected the simple and choice modules of the Axon CCAT as the criterion standard RT_{comp} measures against which to compare our simple and complex RT_{clin} results given that the Axon CCAT is an established cognitive assessment tool within the sports medicine community that directly reports both latency and accuracy values for these measures.

A secondary purpose of this study was to determine the stability of simple and complex RT_{clin} latency and complex RT_{clin} accuracy scores when reduced numbers of trials were included in their calculations. This secondary analysis evaluated the fewest number of

simple and complex RT_{clin} trials that would yield stable results for RT_{clin} , i.e., the most time-efficient testing protocol.

Method

Participants

Forty-six healthy contact sport athletes ages 8 to 30 years participated in this study. The athletes included 24 males (M age = 16.3 yr., SD = 5.0; M height = 167.2 cm, SD = 12.4; M weight = 70.3 kg, SD = 28.8) and 22 females (M age = 15.0 yr., SD = 4.0; M height = 155.9 cm, SD = 12.8; M weight = 53.3 kg, SD = 14.4) representing six sports: soccer (n = 15), ice hockey (n = 14), American football (n = 9), martial arts (n = 5), wrestling (n = 2), and lacrosse (n = 1). The athletes participated at a wide range of competitive levels ranging from local youth leagues to National Collegiate Athletics Association Division I sports. Written informed consent, or assent with parental consent for minors, was obtained using documents approved by the institutional review board at the authors' institution.

Procedure

Participants completed two testing sessions in a university-based biomechanics laboratory on two different days, separated by approximately 1 wk. (M = 6.3 days, SD = 2.3). During the first testing session, participants completed initial assessments of simple and complex RT_{clin} , as well as assessments of simple and complex RT_{comp} using the Axon CCAT. During the second testing session, participants completed a second set of simple and complex RT_{clin} assessments administered by the same member of the study team (Examiner A), as well as a third set of assessments of simple and complex RT_{clin} administered by a different member of the study team (Examiner B). One participant did not attend the second testing session and therefore was not included in the reliability analyses. All testing was conducted in a quiet, well-lit room.

Clinical Reaction Time

Simple and complex RT_{clin} were measured using a custom-built device that was functionally identical to the complex RT_{clin} device used in a prior study (Eckner, et al., 2012). The apparatus used in the present study was slightly modified to improve durability and portability (Fig. 1). The device consisted of a collapsible, rigid, lightweight 107 cm shaft having a dumbbell-shaped cross-section with a concavity on each long side; affixed to the bottom was an $11 \times 6 \times 2.5$ cm box housing a linear accelerometer, timing circuit, microprocessor, battery, liquid crystal display (LCD), and light-emitting diode (LED). The box also acted to standardize starting thumb-finger spacing by providing a visual reference against which to position participants' hands prior to each trial. Shock-absorbing material on the bottom of the spacer box dissipated energy upon contact with the ground. The accelerometer sensed the onset of device acceleration (defined at a downward acceleration threshold of 0.5 gravitational units, g) and deceleration (also defined at a downward acceleration threshold of 0.5 g). The microprocessor sampled at a rate of 2,000 Hz, and the elapsed time for each trial was displayed on the LCD to the nearest 1 msec. The device was programmed so that the LED did not illuminate when in simple mode; but when in complex

mode, it randomly illuminated a green light during 50% of trials upon onset of device movement when detected by the linear accelerometer.

Simple and complex RTclin were assessed using a modified version of the previously described testing protocol (Eckner, et al., 2012). Briefly, participants stood with their dominant forearm resting on an adjustable table so their dominant hand was positioned at the edge of the surface. The examiner suspended the device vertically so that the spacer box rested between the participant's open digits, without making contact with any part of their hand, and so that the top of the spacer box was aligned with the superior-most aspect of the first and second digits and the vertical walls of the spacer box defined the standardized distance between the thumb and the second through fifth digits (Fig. 1). The examiner released the device after a randomly assigned delay time between 2 and 5 sec. During simple RTclin trials, the participant was instructed to catch the falling device using a pincer grasp as quickly as possible during every trial. During complex RTclin trials, participants were instructed to catch the device only during trials when the green LED came on, and to allow it to fall to the ground during trials when the LED did not come on. Participants were instructed to perform the complex RTclin task "as accurately as possible." They were further instructed that, "it is still important to be fast, but your primary goal should be to catch the device only when the light turns on." For simple RTclin, participants completed four practice trials followed by 12 data acquisition trials. For complex RTclin, participants first completed two practice trials during which they were notified that the LED would not illuminate, followed by four practice trials during which the LED randomly illuminated on 50% of the trials, and finally 30 data acquisition trials, again during which the LED randomly came on for 50% of trials. The purpose of the initial "light-off" complex RTclin practice trials was to familiarize participants with allowing the device to fall when the LED did not light up. Simple RTclin latencies were recorded for each trial as the elapsed time from the onset of device acceleration to the onset of deceleration as displayed on the device LCD. During measurement of complex RTclin, the accuracy of the participant's response was recorded for each trial (correct vs incorrect), and the response latency was recorded as the elapsed time from the onset of acceleration to the onset of deceleration during those trials when the participant correctly caught the device when the LED did illuminate. Due to the cognitive decision with inhibitory processing involved, this was slower than the simple measure. For each participant, overall mean latency was calculated for simple RTclin and mean latency and accuracy were calculated for complex RTclin.

Computerized Reaction Time

Simple and complex RTcomp were measured using a laptop personal computer running the Axon CCAT (now CogState; formerly CogSport, or CogState-Sport. Axon Sports, Wausau, WI, USA; Collie, Maruff, Darby, Makdissi, McCrory, & McStephen, 2006). The Axon CCAT is an approximately 15 min. computerized test including four separate modules, each based on the participant's response to playing cards presented on the computer monitor. The first two modules, Processing Speed and Attention, were utilized during this study. The Processing Speed task requires the participant to respond to the question, "Has the card turned over?" by pressing the "K" key, for "yes," as quickly as possible with their right index or middle finger when the initially face-down playing card presented in the middle of

the screen instantaneously turns face up. The same joker card is presented for a minimum of 35 trials. The Attention task requires the participant to respond to the question, “Is the card red?” by pressing the “K” key, for “yes,” as quickly as possible with the right index or middle finger when a red joker card is displayed and the “D” key, for “no,” as quickly as possible with the left index or middle finger when a black joker card is displayed. A single, initially face-down card is instantaneously turned over to reveal either a red or a black joker card for a minimum of 30 trials. Additional trials are added to both tasks following incorrect responses or anticipatory button presses prior to the card turning over. The instructions provided by the Axon CCAT for both tasks are, “Go as fast as you can and try not to make any mistakes.” The system-reported reaction time (msec.) on the Processing Speed task was recorded as the participant's simple RT_{comp} latency score, while the system-reported reaction time (msec.) and accuracy (%) scores on the Attention task were recorded as the participant's complex RT_{comp} latency and accuracy scores, respectively. If a participant's performance on either the Processing Speed or Attention tasks did not pass the Axon program's internal integrity checks then the entire Axon CCAT was repeated until the internal integrity checks were passed. Reasons for integrity check failure included non-anticipatory response rate on the simple RT_{comp} task 90%, complex RT_{comp} accuracy 80%, and simple RT_{comp} latency slower than complex RT_{comp} latency.

Analysis

Statistical analyses were conducted using SAS (Version 9.3, SAS Institute Inc., Cary NC, USA) and SPSS (Version 22, IBM Corporation, Armonk, NY, USA). To assess the test-retest and inter-rater reliabilities, the intraclass correlation coefficients (ICCs) were calculated for the overall simple RT_{clin} latency, and for the overall complex RT_{clin} latency and accuracy scores. The test-retest reliability scores were calculated using the results of the test sessions administered by Examiner A, one week apart. The inter-rater reliability scores were calculated using the results of the test sessions administered by Examiners A and B on the same day. To assess criterion validity, overall simple RT_{clin} latency, complex RT_{clin} latency, and complex RT_{clin} accuracy were compared to the corresponding criterion standard RT_{comp} measures using Pearson's correlation coefficient (r).

Secondary analyses were then carried out to determine the effect of reducing the number of trials on simple RT_{clin} latency and complex RT_{clin} latency and accuracy scores. Descriptive statistics were re-calculated for each measure of the participants' first testing session while successively removing the results of the last trial, to a minimum number of 8 trials. For every new mean value, the difference and absolute differences between the new mean value and the original overall mean value were calculated, as well as the percent of the original overall mean value represented by each difference. In addition the correlations between the new mean values and the original overall mean values were calculated for each measure using Pearson's correlation coefficient (r).

Results

Across all testing sessions, mean simple RT_{clin} latency range was 135–258 msec. ($M = 176$, $SD = 23$), mean complex RT_{clin} latency range was 178–302 msec. ($M = 236$, $SD = 24$), and

complex RTclin accuracy range was 46.7–100% ($M = 86.4\%$, $SD = 11.1\%$). During the complex RTclin task, participants were more successful in catching the falling device during light-on trials ($M = 94.7\%$, $SD = 8.3\%$) than they were in allowing the device to fall during light-off trials ($M = 70.3\%$, $SD = 21.0\%$). Test-retest reliabilities were $ICC = .76$ for simple RTclin latency, $ICC = .79$ for complex RTclin latency, and $ICC = .70$ for complex RTclin accuracy. Inter-rater reliabilities were $ICC = .74$ for simple RTclin latency, $ICC = .87$ for complex RTclin latency, and $ICC = .78$ for complex RTclin accuracy. Across all testing sessions, mean simple RTcomp latency range was 229–448 msec. ($M = 308$, $SD = 53$), mean complex RTcomp latency range was 320–680 msec. ($M = 437$, $SD = 79$), and complex RTcomp accuracy range was 81.1–100% ($M = 95.4\%$, $SD = 4.3\%$). The correlations between simple RTclin latency, complex RTclin latency, complex RTclin accuracy, and their corresponding criterion standard RTcomp values were $r = .54$, $r = .44$, and $r = .33$, respectively.

Secondary analyses investigating the effect of trial number on each RTclin measure are presented in Tables 1, 2, and 3. Reducing the number of simple and complex RTclin trials to as few as eight had relatively small effects on the overall mean values of simple RTclin latency and complex RTclin latency and accuracy at the group level. However, inclusion of fewer trials resulted in greater absolute differences in each measure for individual participants. RTclin values calculated from reduced numbers of trials had correlations of $r = .9$ with the corresponding RTclin values calculated from all trials when at least eight trials were included for simple RTclin latency, at least 19 trials were included for complex RTclin latency, and at least 13 trials were included for complex RTclin accuracy.

Discussion

Reliability

This study investigated the reliability and criterion validity of a clinical measure of simple and complex RT in a population of healthy contact sport athletes ages 8 to 30 years. The 1 wk. test-retest and inter-rater reliabilities were high for simple and complex RTclin latencies (.76, .79) as well as complex RTclin accuracy (.70). The ICC values fell broadly within the range of values previously reported for simple RTclin measured using the original clinical reaction time device. These ICC's are lower than the 1 wk. test-retest and inter-rater reliabilities ($ICC = .86$ and $.92$, respectively) originally reported in a general healthy adult population (Eckner, et al., 2009), but they are higher than the 1 yr. value ($ICC = .65$) reported in collegiate athletes (Eckner, et al., 2011a), as well as the 1 yr. value reported by MacDonald, Wilson, Young, Duerson, Swisher, Collins, et al. (in press) in high school athletes ($ICC = .61$).

The simple and complex RTclin reliability indices obtained also compare favorably with reliability indices previously reported for various computerized measures of RT in existing computerized cognitive assessment tools (CCAT) currently in widespread use (Table 4). Four studies have reported on the reliability of the ANAM, which includes measures of simple and choice RT, with test-retest reliabilities ranging from .14 to .60 (Cernich, Reeves, Sun, & Bleiberg, 2007; Segalowitz, Mahaney, Santesso, MacGregor, Dywan, & Willer, 2007; Cole, Arrieux, Schwab, Ivins, Qashu, & Lewis, 2013; Register-Mihalik Guskiewicz,

Mihalik, Schmidt, Kerr, & McCrea, 2013). Also, four studies have reported on the reliability of Axon (formerly CogState-Sport, CogSport, or Concussion Sentinel), which includes measures of simple and choice RT, with test-retest reliabilities ranging from .55 to .90 (Collie, et al., 2003; Broglio, Ferrara, Macciocchi, Baumgartner, & Elliott, 2007; Cole, et al., 2013; MacDonald & Duerson, in press). Two studies have reported on the reliability of CNS Vital Signs, which includes a complex RT index score, with test-retest reliabilities ranging from .75 to .79 (Gualtieri & Johnson, 2006; Cole, et al., 2013). Two studies have reported on the test-retest reliability of the Concussion Resolution Index, which includes simple and complex measures of RT, with reliabilities ranging from .36 to .70 (Erlanger, Feldman, Kutner, Kaushik, Kroger, Festa, et al., 2003; Broglio, et al., 2007). Finally, six studies have reported on the reliability of ImPACT, which includes a single RT composite score that incorporates the results of both simple and complex RT tasks, with reliabilities ranging from .39 to .79 (Iverson, Lovell, & Collins, 2003; Broglio, et al., 2007; Schatz, 2010; Elbin, Schatz, & Covassin, 2011; Register-Mihalik, Kontos, Guskiewicz, Mihalik, Conder, & Shields, 2012; Cole, et al., 2013).

The minimum reliability threshold deemed acceptable for clinical use varies among authors, with suggested cutoffs ranging from .6 to .9 (Anastasi, 1998; Randolph, McCrea, & Barr, 2005). Cole, et al. (2013) have argued that tests capable of precisely measuring a construct like reaction time to the hundredth of a second may be more sensitive to changes in performance due to state variables like fatigue, pain, and testing environment, than traditional neuropsychological tests and therefore may need to be held to less rigorous psychometric standards. Practically speaking, greater reliability leads to greater sensitivity to detect meaningful change, so tests with higher reliability are more desirable for clinical decision making. Based on a classification of test-retest reliability coefficients .90 as very high, .80–.89 as high, .70–.79 as adequate, .60–.69 as marginal, and < .60 as low (Lezak, Howleson, Bigler, & Tranel, 2012; Cole, et al., 2013), all of the ICC indices reported in this study for simple and complex RTclin fall within the range that is considered acceptable for clinical use.

In addition to quantifying test-retest reliability in terms of ICC values, it may also be useful to consider the magnitude of RTclin change during subsequent test administrations to the change anticipated to occur following concussion. In prior work involving 28 concussed and control athletes who completed simple RTclin testing sessions in the preseason and again within 48 hr. of sustaining a concussion, simple RTclin latency was 17 msec. slower on average in the concussed athletes, while it was 9 msec. faster on average in the controls. This average preseason-to-after-injury simple RTclin latency difference of 26 msec. between concussed and control athletes is more than three times greater than the average observed magnitude of test-retest simple RTclin latency change in the present study ($M = 8$ msec, $SD = 19$). The effect of concussion on complex RTclin latency and accuracy is unknown, so a similar comparison cannot be made at the present time for complex RTclin.

Criterion Validity

When comparing simple RTclin latency to its corresponding criterion standard, simple RTcomp latency, the strength of association was slightly greater than the previously reported

value ($r = .45$) obtained in a population of collegiate football players (Eckner, et al., 2010) and much greater than was reported by MacDonald, et al. (in press) in a population of high school athletes. In contrast, the strength of association between the clinical and computerized measures was lower for both complex RTclin latency and accuracy. The lower correlations between the complex RTclin measures and their corresponding RTcomp criterion standards for comparison are likely attributable to several differences in the two complex RT tasks. First, complex RTcomp involves a choice paradigm in which the participant is instructed to perform one response (striking the “K” key) when presented with one stimulus (a red joker), and a distinct, second response (striking the “D” key) when presented with a second stimulus (a black joker). In contrast, complex RTclin uses a recognition, or go/no-go, paradigm in which the participant is instructed to perform a single response (catching the falling device) when presented with one stimulus (illumination of the LEDs), but to withhold that response (i.e., allow the device to fall) when presented with a second stimulus (non-illumination of the LEDs). Furthermore, the complex RTclin task has a natural floor in how slowly the task can be completed before the falling device strikes the ground, whereas complex RTcomp has no such time restriction. As a result of this floor effect, the difficulty associated with the complex RTclin task is much greater than that associated with complex RTcomp. While this assessment is based primarily on the subjective impression of the investigators and participants, it is supported by the lower accuracy values observed on the complex RTclin task as compared to complex RTcomp. Finally, in addition to being more challenging, the complex RTclin task is also likely more engaging than the complex RTcomp task, and this may have led to greater motivation and effort during the complex RTclin task. While motivation was not directly assessed in this study, prior work has demonstrated simple RTclin to be more motivating than a computerized simple RT task that was different from the Axon CCAT used in this study (Eckner, Chandran, & Richardson, 2011).

Given the fundamental differences between complex RTclin and complex RTcomp, the Axon CCAT choice RT task may not be an optimal criterion standard against which to compare complex RTclin. The rationale for using the Axon CCAT in this study was based on its status as an established, validated, and generally accepted concussion assessment tool within the sports medicine community, containing both a simple and complex RT task with direct speed and accuracy values reported for each. The differential floor effects between complex RTclin and complex RTcomp may be the greatest contributor to the low correlation between their accuracy indices, as it likely resulted in a differential ability to trade off between speed and accuracy between the tasks. It has been long established that one can trade off accuracy for speed in any task (Wickelgren, 1977). In the case of complex RTclin, very little time could be traded off to improve accuracy, whereas an unlimited amount of time could be traded off for accuracy during complex RTcomp. Furthermore, with respect to the potential use of complex RTclin in the context of concussion assessment in athletes, criterion validity is a less important measure of test validity than diagnostic validity, which was not addressed by this study.

Number of Trials

The rationale for the secondary analysis determining the effect of reduced number of trials on each measure was to assess whether using fewer trials was justifiable. In the context of concussion assessment in athletes, there are several advantages to minimizing testing time. One is efficiency, given that baseline testing of an entire team may involve more than 100 athletes. Given practical limitations both in athletes' time and in resources available to athletics programs, especially at the high school and youth levels, a more efficient test battery is highly desirable. Additionally, there is a general consensus that a multi-faceted concussion assessment battery should be used when assessing concussion (Broglia, Macciocchi, & Ferrara, 2007b; Harmon, Drezner, Gammons, Guskiewicz, Halstead, Herring, et al., 2013; McCrory, Meeuwisse, Aubry, Cantu, Dvorak, Echemendia, et al., 2013). Shorter, more time-efficient measures permit inclusion of more tools for assessment. Finally, athletes' effort and motivation decline over the course of a longer baseline testing session. Poor effort and motivation during baseline testing can jeopardize comparisons of baseline to post-injury results (Green, Rohling, Lees-Haley, & Allen, 2001; Bailey, Echemendia, & Arnett, 2006). Therefore, the shortest RTclin testing protocol that would yield stable results was sought. While there was relatively little change in the overall mean simple and complex RTclin latency and complex RTclin accuracy values at the group level when the number of trials analyzed was decreased to as few as eight, the effect of trial reduction on RTclin results at the individual level was of greater interest.

To quantify the effects of numbers of trials chosen, review of Table 1 shows that inclusion of 10 simple RTclin trials yields results that differ from the original values by no more than 4.1% and a review of Tables 2–3 shows that including 20 complex RTclin trials yields results that differ no more than 9.9% from the original latency values and 11.7% from the original accuracy values. If the RTclin test-retest and inter-tester reliabilities are recalculated using 10 simple and 20 complex RTclin trials, the ICC values decrease slightly, but remain above the minimum .6 threshold of acceptability (Anastasi, 1998; Lezak et al., 2012; Cole, et al., 2013). Test-retest reliabilities become ICC = .73 for simple RTclin latency, ICC = .68 for complex RTclin latency, and ICC = .66 for complex RTclin accuracy. Inter-rater reliabilities become ICC = .73 for simple RTclin latency, ICC = .80 for complex RTclin latency, and ICC = .63 for complex RTclin accuracy. Therefore, future work using the complex RTclin device may employ a minimum protocol of 10 simple and 20 complex RTclin trials.

In addition to the dissimilarities between complex RTclin and its criterion standard RTcomp described above, another limitation of this study is that the one-week test-retest timeframe is shorter than the test-retest interval typically used in clinical practice during concussion assessment and return-to-play decision making, which has been reported to be approximately 6–7 wk. (Lovell, Collins, Iverson, Field, Maroon, Cantu, et al., 2003; Broglia, et al., 2007). A third study limitation related to the speed-accuracy trade-off discussion above is that we did not attempt to obtain the entire speed-accuracy trade-off function for either complex RT task. Participants were simply instructed to perform as quickly and accurately as possible during both. Because the intrinsic floor effect associated with the complex RTclin task already forces a fast response, accuracy was emphasized

during this test. A fourth limitation is that participants' effort was not measured; effort can have a greater effect on neurocognitive test performance than severe brain injury (Green, et al., 2001). While we cannot quantitatively comment on participant effort, prior work has demonstrated that the simple RTclin task is intrinsically motivating (Eckner, et al., 2011), and our subjective impression was that the athletes participating in this study were motivated to perform their best on the RTclin tasks. The most important limitation of this study is that it included only healthy participants and addressed only one form of validity, namely criterion validity. Additional prospective research including concussed athletes is necessary to investigate the diagnostic validity of complex RTclin for concussion.

In conclusion, the test-retest and inter-rater reliabilities of simple and complex RTclin were found to be acceptable for clinical use. Furthermore, RTclin results correlated with the criterion standard RTcomp. Inclusion of 10 simple RTclin trials and 20 complex RTclin trials optimized the balance between test stability and efficiency. Future research addressing the diagnostic validity of complex RTclin is necessary to establish its sensitivity and specificity, both as a concussion assessment tool and a tool for assessing other medical conditions known to influence RT.

References

- Anastasi, A. *Psychological Testing*. 6th. New York, NY: Macmillan; 1998.
- Bailey CM, Echemendia RJ, Arnett PA. The impact of motivation on neuropsychological performance in sports-related mild traumatic brain injury. *J Int Neuropsychol Soc*. 2006; 12(4):475–484. [PubMed: 16981599]
- 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. [PubMed: 18174939]
- Broglio SP, Macciocchi SN, Ferrara MS. Neurocognitive performance of concussed athletes when symptom free. *J Athl Train*. 2007a; 42(4):504–508. [PubMed: 18174938]
- Broglio SP, Macciocchi SN, Ferrara MS. Sensitivity of the concussion assessment battery. *Neurosurgery*. 2007b; 60(6):1050–1057. discussion 1057-1058. [PubMed: 17538379]
- Cernich A, Reeves D, Sun W, Bleiberg J. Automated Neuropsychological Assessment Metrics sports medicine battery. *Arch Clin Neuropsychol*. 2007; 22 Suppl 1:S101–114. [PubMed: 17118625]
- Chudler, EH. Neuroscience for Kids. 2009. Retrieved August 31, 2009, *University of Washington in Seattle WA.*, <http://faculty.washington.edu/chudler/chreflex.html>
- Cole WR, Arrieux JP, Schwab K, Ivins BJ, Qashu FM, Lewis SC. Test-retest reliability of four computerized neurocognitive assessment tools in an active duty military population. *Arch Clin Neuropsychol*. 2013; 28(7):732–742. [PubMed: 23819991]
- Collie, A.; Maruff, P.; Darby, D.; Makdissi, M.; McCrory, P.; McStephen, M. CogSport. In: Echemendia, RJ., editor. *Sports Neuropsychology: Assessment and Management of Traumatic Brain Injury*. New York: Guildford Publications; 2006. p. 240-262.
- 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. [PubMed: 12544161]
- Eckner JT, Chandran S, Richardson JK. Investigating the role of feedback and motivation in clinical reaction time assessment. *PM&R*. 2011; 3(12):1092–1097. [PubMed: 21852221]
- Eckner JT, Kutcher JS, Broglio SP, Richardson JK. Effect of sport-related concussion on clinically measured simple reaction time. *Br J Sports Med*. 2014; 48(2):112–118. [PubMed: 23314889]
- 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. [PubMed: 20617905]

- Eckner JT, Kutcher JS, Richardson JK. Between-seasons test-retest reliability of clinically measured reaction time in National Collegiate Athletic Association Division I athletes. *J Athl Train*. 2011a; 46(4):409–414. [PubMed: 21944073]
- Eckner JT, Kutcher JS, Richardson JK. Effect of concussion on clinically measured reaction time in 9 NCAA division I collegiate athletes: a preliminary study. *Pm R*. 2011b; 3(3):212–218. [PubMed: 21402367]
- Eckner JT, Lipps DB, Kim H, Richardson JK, Ashton-Miller JA. Can a clinical test of reaction time predict a functional head-protective response? *Med Sci Sports Exerc*. 2011; 43(3):382–387. [PubMed: 20689458]
- Eckner JT, Richardson JK, Kim H, Lipps DB, Ashton-Miller JA. A novel clinical test of recognition reaction time in healthy adults. *Psychol Assess*. 2012; 24(1):249–254. [PubMed: 21859222]
- 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. [PubMed: 19725308]
- Elbin RJ, Schatz P, Covassin T. One-year test-retest reliability of the online version of ImPACT in high school athletes. *Am J Sports Med*. 2011; 39(11):2319–2324. [PubMed: 21841066]
- Erlanger D, Feldman D, Kutner K, Kaushik T, Kroger H, Festa J, Barth J, Freeman J, Broshek D. Development and validation of a web-based neuropsychological test protocol for sports-related return-to-play decision-making. *Arch Clin Neuropsychol*. 2003; 18(3):293–316. [PubMed: 14591461]
- Erlanger D, Saliba E, Barth J, Almquist J, Webright W, Freeman J. Monitoring Resolution of Postconcussion Symptoms in Athletes: Preliminary Results of a Web-Based Neuropsychological Test Protocol. *J Athl Train*. 2001; 36(3):280–287. [PubMed: 12937497]
- Green P, Rohling ML, Lees-Haley PR, Allen LM III. Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Inj*. 2001; 15(12):1045–1060. [PubMed: 11712951]
- Gualtieri CT, Johnson LG. Reliability and validity of a computerized neurocognitive test battery, CNS Vital Signs. *Arch Clin Neuropsychol*. 2006; 21(7):623–643. [PubMed: 17014981]
- Harmon KG, Drezner JA, Gammons M, Guskiewicz KM, Halstead M, Herring SA, Kutcher JS, Pana A, Putukian M, Roberts WO. American Medical Society for Sports Medicine position statement: concussion in sport. *Br J Sports Med*. 2013; 47(1):15–26. [PubMed: 23243113]
- Iverson GL, Lovell MR, Collins MW. Interpreting change on ImPACT following sport concussion. *Clin Neuropsychol*. 2003; 17(4):460–467. [PubMed: 15168911]
- Lau B, Lovell MR, Collins MW, Pardini J. Neurocognitive and symptom predictors of recovery in high school athletes. *Clin J Sport Med*. 2009; 19(3):216–221. [PubMed: 19423974]
- Lovell MR, Collins MW, Iverson GL, Field M, Maroon JC, Cantu R, Podell K, Powell JW, Belza M, Fu FH. Recovery from mild concussion in high school athletes. *J Neurosurg*. 2003; 98(2):296–301. [PubMed: 12593614]
- MacDonald J, Duerson D. Reliability of a computerized neurocognitive test in baseline concussion testing of high school athletes. *Clin J Sport Med*. 2014 in press.
- Macdonald J, Wilson J, Young J, Duerson D, Swisher G, Collins CL, Meehan WP. Evaluation of a Simple Test of Reaction Time for Baseline Concussion Testing in a Population of High School Athletes. *Clin J Sport Med*. 2014 in press.
- Makdissi M, Darby D, Maruff P, Ugoni A, Brukner P, McCrory PR. Natural history of concussion in sport: markers of severity and implications for management. *Am J Sports Med*. 2010; 38(3):464–471. [PubMed: 20194953]
- McCrory P, Meeuwisse WH, Aubry M, Cantu B, Dvorak J, Echemendia RJ, Engebretsen L, Johnston K, Kutcher JS, Raftery M, Sills A, Benson BW, Davis GA, Ellenbogen RG, Guskiewicz K, Herring SA, Iverson GL, Jordan BD, Kissick J, McCrea M, McIntosh AS, Maddocks D, Makdissi M, Purcell L, Putukian M, Schneider K, Tator CH, Turner M. Consensus statement on concussion in sport: the 4th International Conference on Concussion in Sport held in Zurich, November 2012. *British Journal of Sports Medicine*. 2013; 47(5):250–258. [PubMed: 23479479]
- Montare A. The simplest chronoscope: group and interindividual differences in visual reaction time. *Perceptual & Motor Skills*. 2009; 108(1):161–172. [PubMed: 19425458]

- Lezak, MD.; Howleson, DB.; Bigler, ED.; Tranel, D. Neuropsychological Assessment. 5. New York, NY: Oxford University Press; 2012.
- Randolph C, McCrea M, Barr WB. Is neuropsychological testing useful in the management of sport-related concussion? J Athl Train. 2005; 40(3):139–152. [PubMed: 16284633]
- Register-Mihalik JK, Guskiewicz KM, Mihalik JP, Schmidt JD, Kerr ZY, McCrea MA. Reliable change, sensitivity, and specificity of a multidimensional concussion assessment battery: implications for caution in clinical practice. J Head Trauma Rehabil. 2013; 28(4):274–283. [PubMed: 22691965]
- Register-Mihalik JK, Kontos DL, Guskiewicz KM, Mihalik JP, Conder R, Shields EW. Age-related differences and reliability on computerized and paper-and-pencil neurocognitive assessment batteries. J Athl Train. 2012; 47(3):297–305. [PubMed: 22892411]
- Schatz P. Long-term test-retest reliability of baseline cognitive assessments using ImPACT. Am J Sports Med. 2010; 38(1):47–53. [PubMed: 19789333]
- Segalowitz SJ, Mahaney P, Santesso DL, MacGregor L, Dywan J, Willer B. Retest reliability in adolescents of a computerized neuropsychological battery used to assess recovery from concussion. NeuroRehabilitation. 2007; 22(3):243–251. [PubMed: 17917174]
- Warden DL, Bleiberg J, Cameron KL, Ecklund J, Walter J, Sparling MB, et al. Persistent prolongation of simple reaction time in sports concussion. Neurology. 2001; 57(3):524–526. [PubMed: 11502926]
- Wickelgren W. Speed-accuracy tradeoff and information processing dynamics. Acta Psychologica. 1977; 41:67–85.

Figure 1.

a. Novel device for measuring simple and complex RT_{clin} . b. Demonstration of testing protocol: Pre-drop. c. Post-drop.

Table 1

Effect of Trial Number Reduction In Simple RTclin Latency (msec.)

Trial	M	SD	Min.	Max.	M Diff.	M Abs. Diff.	% Diff.	% Abs. Diff.	Max % Abs. Diff.	R
All 12	181.9	24.9	142	258	N/A	N/A	N/A	N/A	N/A	N/A
First 11	182.3	25.0	142	256	-0.4	1.8	-0.2	1.0	4.2	1.00
First 10	183.0	25.4	144	260	-1.1	2.5	-0.6	1.4	4.1	.99
First 9	183.8	25.3	144	259	-1.9	3.5	-1.1	1.9	6.2	.99
First 8	184.1	25.5	145	263	-2.2	4.5	-1.3	2.4	7.7	.98

Table 2Effect of Trial Number Reduction on Complex RT_{clin} Latency (msec.)

Trial	<i>M</i>	<i>SD</i>	Min.	Max.	<i>M</i> Diff.	<i>M</i> Abs. Diff.	% Diff.	% Abs. Diff.	Max % Abs. Diff.	<i>R</i>
All 30	238.2	23.7	204	302	N/A	N/A	N/A	N/A	N/A	N/A
First 28	238.3	23.9	201	300	-0.1	2.8	-0.10	1.20	5.30	.99
First 26	237.1	24.0	200	309	1.1	4.8	0.40	2.00	6.90	.97
First 24	237.2	24.1	200	309	0.9	5.1	0.40	2.10	6.90	.97
First 22	237.3	24.8	200	317	0.9	5.4	0.40	2.20	8.80	.96
First 20	237.6	24.3	199	304	0.5	6.8	0.20	2.80	9.90	.93
First 18	236.9	24.3	199	303	1.3	8.9	0.40	3.70	11.50	.87
First 16	235.9	25.3	193	311	2.2	12.2	0.80	5.00	16.90	.80
First 14	236.9	26.2	192	311	1.3	14.1	0.40	5.90	17.50	.76
First 12	235.7	26.6	184	296	2.5	17.0	0.80	7.10	23.60	.66
First 10	235.7	26.5	185	296	2.4	17.3	0.80	7.30	23.60	.65
First 8	235.2	27.9	185	296	3.0	19.4	1.00	8.10	23.30	.57

Table 3
Effect of Trial Number Reduction on Complex RT_{clin} Accuracy (%)

Trial	<i>M</i>	<i>SD</i>	Min.	Max.	<i>M</i> Diff.	<i>M</i> Abs. Diff.	Max. Abs. Diff.	<i>R</i>
All 30	82.4	13.1	46.7	100.0	N/A	N/A	N/A	N/A
First 28	81.8	13.3	42.9	100.0	0.7	1.0	3.8	1.00
First 26	81.1	13.6	42.3	100.0	1.4	2.0	5.9	.99
First 24	80.5	14.0	41.7	100.0	2.0	2.4	6.7	.99
First 22	81.3	14.0	45.5	100.0	1.1	3.0	9.7	.97
First 20	81.7	13.6	45.0	100.0	0.8	2.8	11.7	.96
First 18	81.6	14.2	44.4	100.0	0.9	3.4	13.3	.95
First 16	81.2	14.2	43.8	100.0	1.3	3.9	11.3	.94
First 14	80.4	14.7	50.0	100.0	2.0	4.6	16.7	.92
First 12	80.1	14.1	41.7	100.0	1.5	5.1	21.7	.89
First 10	84.2	13.3	50.0	100.0	-1.8	5.1	13.3	.89
First 8	82.2	15.6	50.0	100.0	0.3	6.0	17.5	.89

Table 4
Reliability Indices for Measures of RT from Existing CCAT Programs and Current Study

CCAT/RT _{clin} Study	Population	RT Measure	Retest Interval	Reliability	
				R	ICC
CURRENT STUDY					
Eckner, <i>et al.</i> (2015)	46 contact sport athletes	Simple RT	1 wk.	.76	
		Complex RT _{clin}		.79	
		Complex RT _{clin} accuracy		.70	
ANAM					
Cernich, <i>et al.</i> (2007)	18 U.S. Military Academy cadets	Simple RT	166.5 days	.38	
	Segalowitz, <i>et al.</i> (2007)	Simple RT 1	1 wk.	.44	
Simple RT 2		.47			
Simple RT 1		46.7 days		.29	
Register-Mihalik, <i>et al.</i> (2013)	38 male collegiate football athletes	Simple RT 2		.14	
		Procedural RT (Choice RT)		.33	
		Simple RT 1	32 days	.60	
Cole, <i>et al.</i> (2013)	50 active duty military service members	Simple RT 2		.40	
		Procedural RT (Choice RT)		.51	
		Axon (CogState–Sport/CogSport/Concussion Sentinel)			
Collie, <i>et al.</i> (2003)	60 healthy young adults	Psychomotor Speed (Simple RT)	1 hr.	.90	
			1 wk.	.76	
		Decision Making Speed (Choice RT)	1 hr.	.69	
Broglia, <i>et al.</i> (2007)	118 college students		1 wk.	.69	
		Reaction Time (Simple RT)	45 days	.60	
			5 days	.55	
Cole, <i>et al.</i> (2013)	53 active duty military service members	Decision Making (Choice RT)	45 days	.56	
			5 days	.62	
		Detection Speed (Simple RT)	32 days	.78	
MacDonald, <i>et al.</i> (2014)	117 high school athletes	Identification Speed (Choice RT)		.77	
		Detection Speed (Simple RT)	1 yr.	.55	
		Identification Speed (Choice RT)		.67	

CCAT/RT _{clin} Study	Population	RT Measure	Retest Interval	Reliability		
				R	ICC	
CNS Vital Signs						
Gualtieri, <i>et al.</i> (2006)	99 healthy and neuropsychiatric patients	Reaction Time (Complex RT)	62 days	.80		
Cole <i>et al.</i> (2013)	50 Active duty military service members	Reaction Time (Complex RT)	32 days		.75	
Concussion Resolution Index						
Erlanger, <i>et al.</i> (2003)	175 healthy individuals ages 13–35 yr.	Simple RT Index Score	2 wk.	.70		
Brogljo, <i>et al.</i> (2007)	118 college students	Complex RT Index Score		.68		
		Simple RT Index Score	45 days		.65	
			5 days		.36	
		Complex RT Index Score	45 days		.43	
ImPACT			5 days		.46	
Iverson, <i>et al.</i> (2003)	56 healthy adolescents and young adults	RT Composite Score	5.8 days	.79		
Brogljo, <i>et al.</i> (2007)	118 college students	RT Composite Score	45 days		.39	
			5 days		.51	
Schatz, <i>et al.</i> (2010)	95 collegiate athletes	RT Composite Score	1.9 yr.		.68	
Elbin, <i>et al.</i> (2011)	484 high school athletes	RT Composite Score	1.2 yr.		.76	
Register-Mihalik, <i>et al.</i> (2012)	40 high school and collegiate athletes	RT Composite Score	24–72 hr.		.60	
Cole, <i>et al.</i> (2013)	44 active duty military service members	RT Composite Score	32 days		.53	