10-25-2017

Reliability and Concurrent Validity of Select C3 Logix Test Components

Madeline Simon
New Hampshire Musculoskeletal Institute, Manchester, NH, mtsimon45@hotmail.com

Arthur C. Maerlender
University of Nebraska-Lincoln, amaerlender2@unl.edu

Katelyn Metzger
New Hampshire Musculoskeletal Institute, Manchester, NH

Laura Decoster
New Hampshire Musculoskeletal Institute, Manchester, NH

Amy Hollingworth
New Hampshire Musculoskeletal Institute, Manchester, NH

See next page for additional authors

Authors
Madeline Simon, Arthur C. Maerlender, Katelyn Metzger, Laura Decoster, Amy Hollingworth, and Tamara Valovich McLeod

Published in *Developmental Neuropsychology,* 2017. doi: 10.1080/87565641.2017.1383994

Published online 25 October 2017.

Reliability and Concurrent Validity of Select C3 Logix Test Components

Madeline Simon,[1] Arthur Maerlender,[2] Katelyn Metzger,[1] Laura Decoster,[1] Amy Hollingworth,[1] and Tamara Valovich McLeod[3]

1 New Hampshire Musculoskeletal Institute, Manchester, New Hampshire;
2 Center for Brain, Biology and Behavior, University of Nebraska - Lincoln, Lincoln, Nebraska;
3 Athletic Training Programs and School of Osteopathic Medicine, A.T. Still University, Mesa, Arizona

Corresponding author — Madeline Simon mtsimon45@hotmail.com; New Hampshire Musculoskeletal Institute, 35 Kosciuszko St, Manchester, NH 03101, USA

Abstract
We sought to investigate the one-week and within-session reliability of the instrumented balance error scoring system test and the concurrent validity/one-week reliability of two neurocognitive assessments available through C3 Logix. (n = 37) Participants completed two balance error scoring system tests separated by the Trails A, Trails B, and Symbol Digit Modality test available through C3 Logix, and with paper and pencil. We found that the instrumented balance error scoring system test demonstrated strong one-week reliability and that neuropsychological tests available through C3 Logix show acceptable concurrent validity with standard (comparable) paper and pencil measures.

Introduction

A concussion is a complex pathophysiological process affecting the brain, induced by biomechanical forces (McCrory et al., 2009; Nanda, 2012). It may be diagnosed in the presence of a variety of signs or symptoms such as headache, feeling in a fog, loss of consciousness, amnesia, neurological deficit, altered gait, irritability, slowed reaction time, and insomnia. After clinical assessment, such signs and symptoms can be categorized into one or more of the following areas: cognitive (e.g., confusion, loss of consciousness), physical (e.g., headache, nausea/vomiting, balance disturbance), emotional (e.g., depression, behavioral issues), and sleep disturbance (e.g., insomnia, sleeping too much) (Collins, Lovell, & McKeag, 1999; Kushner, 2001). Suspicion of a sports-related concussion should occur in the presence of any one or more of these signs or symptoms after the patient has sustained a possible concussive mechanism of injury. (McCrory et al., 2017). Appropriate management strategies should be implemented to facilitate a safe outcome for the athlete. Guidelines regarding the proper management of sport-related concussions are the topic of many consensus reports and position statements published over the past decade (Aubry et al. 2002; Broglio et al., 2014; Cantu et al., 2006; Herring et al., 2011; McCrory et al., 2013, 2017, 2009). This continues to be a rapidly-changing area with many new important studies being published regularly. Current recommendations suggest a concussion be assessed with a multi-modal paradigm that includes objective tests of neurocognitive performance, balance, and symptoms (Broglio et al., 2014; McCrory et al., 2013, 2017). As the science around the appropriate clinical management of concussions continues to evolve, it is important that clinicians stay updated on evidence-based recommendations and use appropriate tools and protocols to assess and manage concussions.

The protocols and technologies used to objectively assess balance and neurocognitive function following a concussion are rapidly developing. Although not clinically required, efforts are made to collect individual baseline data to be used to compare post-injury test performance, as opposed to using population norms (McCrory et al., 2017). The need to gather neurocognitive baseline information for many athletes in a short window of time has pushed the field away from traditional paper and pencil testing. These tests require considerable time and labor and can be costly (Coppel, 2011). Thus, the use of computerized assessment in the management of concussions has become increasingly common. C3 Logix (NeuroLogix Technologies, Cleveland, Ohio) is a tablet-based application available on the Apple iPad (Apple, Inc., Cupertino, CA) that offers promise as a customizable way to perform objective balance and neurocognitive assessments. However, it is important to understand the measurement properties of any testing (for both balance and neurocognitive assessment) before relying on them to inform clinical decision-making (Schatz & Zillmer, 2003). Specifically, it is crucial that clinicians understand the validity and reliability, as without this knowledge, use of these tests to return athletes to play is at best inappropriate (Valovich McLeod, Barr, McCrea, & Guskiewicz, 2006).

For years, the validated clinical test of choice used to assess balance has been the Balance Error Scoring System (BESS) (Guskiewicz, 2001). The BESS test has been shown be a valid assessment of balance deficits when large differences in error counts exist (Bell, Guskiewicz, Clark, & Padua, 2011). Additionally, the BESS test has been shown to have moderate to good test-retest reliability to assess static balance (Bell et al., 2011). Although it is important to be aware of large changes in balance, athletes often exhibit more-subtle post-injury changes that are undetectable with a subjective error scoring system alone. The BESS test is administered immediately post-injury, and administered again a few days post-injury per the clinician's discretion. Although it is sensitive 3–5 days post-injury, more sensitive measures and sophisticated gait analysis show longer deficits. Such analyses using force plate and optical motion systems have been shown to be reliable and valid measures of postural stability in the past (Brown et al., 2014). However, the use of these systems is not typically a realistic option for the majority of athletic trainers due to high cost and space restraints (Brown et al., 2014). An alternative to the use of force plate and optical motion systems is the use of accelerometers and gyroscope technology. This technology offers a less costly and more practical evaluation of postural sway. To date, their use in the clinical setting is not widespread (Alberts et al., 2015). C3 Logix utilizes the iPad's accelerometer and gyroscope to measure anterior-posterior, medial-lateral, and rotational postural sway. Assessing balance with gyroscope and accelerometer adjuncts appears to be a more sensitive and objective measure of postural stability than error scoring alone. The validity of this instrumented postural assessment has been demonstrated, however, test–retest reliability is not apparent in the literature (Alberts et al., 2015). Furthermore, although practice effects on the error score results have been examined, practice effects on postural sway as measured by the iPad have not been published.

Neuropsychological testing is an important component of concussion evaluation and management due to the ability to detect cognitive deficits that may not be reflected in the athlete's self-report of symptoms following injury (Van Kampen, Lovell, Pardini, Collins, & Fu, 2006). The use of C3 Logix allows examiners to objectively assess subtle changes in processing speed and reaction time that are not as easily obtained using the current gold-standard, paper and pencil tests (Barth et al., 1989). The C3 Logix tests include tablet-based applications of standard paper and pencil tests: Trail Making A, Trail Making B, and Symbol Digit Modality Test (Smith, 1982). C3 Logix also includes simple and choice reaction time tests. These allow the examiner to assess the athlete's speed of response to a single stimulus, and, in the case of the "choice reaction" test, to choose the appropriate response to two simultaneous stimuli. The original Trail Making and SDMT tests are recognized as valid assessments of psychomotor speed, and visual processing speed (Reitan, 1992). The tests available on the C3 Logix application are touch-screen adaptations of these traditional tests. However, there is a lack of data showing whether the analogous C3 Logix versions of these tests are valid or reliable.

Overall, the aims of this study were to assess stability and test–retest reliability of these tests in a college athlete sample, and to contribute to establishing concurrent validity of iBESS and neuropsychological tests. The primary goals regarding the balance assessment were to determine if the C3 Logix balance assessment is a reliable measure of postural stability and to characterize the practice effects associated with multiple administrations of this test. Our secondary goal was to determine the strength of the relationship between BESS errors and the instrumented postural stability measurements (iBESS volume). It was hypothesized that both BESS error counts and iBESS volume would demonstrate appropriate one-week test–retest reliability. Additionally, significant practice effects for both BESS error counts and postural stability were predicted to be present from BESS Trial 1 to all subsequent trials. In terms of concurrent validity, it was predicted that BESS errors and postural stability measurements in all single leg (SL) and tandem stance (TS) conditions, but not double leg (DL) conditions would be correlated.

Regarding the three C3Logix neuropsychological tests, we sought to assess one-week test–retest reliability and concurrent validity. In the realm of reliability, practice effects were predicted to follow specific patterns: Trail Making A and B times to completion would decrease (including the difference between A and B), and the number of correctly matched pairs during the SDMT would increase between session one and two. The concurrent validity goals were to determine whether the C3 Logix tests (i.e., Trail Making A, Trail Making B, and SDMT) are valid when compared to the paper-and-pencil forms of the tests. We hypothesized that the times to completion in Trails, and number of correctly matched pairs in SDMT would be highly correlated on both applications, thus demonstrating concurrent validity.

Finally, it was of interest to determine if participant sex or number of hours of sleep the night before testing had an impact on test scores at any session.

Methods

Research design

For the reliability studies, a repeated-measures design was employed with each participant serving as his or her own control. The independent variables for the BESS portion of the study were time (between sessions or within a single session), sex, and hours of sleep the night preceding the test. Dependent variables for the BESS portion of the study included the iBESS volume and number of BESS errors. The iBESS volume is the mathematical representation of combined accelerations in the anterior-posterior, medial-lateral and rotational planes. The iBESS value was computed by the application based on the motion that occurred within each plane (Alberts et al., 2015). We evaluated iBESS volume and errors for each stance, as well as for the BESS test overall. A single trained investigator administered all BESS tests. Inter-rater reliability (using a trained independent assessor viewing videotapes of the trials) was established to ensure quality results were obtained ($ICC > 0.90$).

For the concurrent validity studies, a correlational design of test results at the same time point (timepoint 1) was employed. Test order was counterbalanced to control for order effects.

For the neuropsychological tests, the independent variable was test session (one or two). The dependent variables were the time it took participants to complete the Trails assessments and the number of correctly matched symbols in the SDMT assessment. Validity was assessed by comparing the results participants obtained on the paper and pencil versions of the tests versus those they obtained using C3 Logix. To ensure quality results, a threshold of $r = 0.60$ was considered acceptable.

To assess the variables of sex and hours of sleep on test scores, a between-subjects ANOVA was completed for both test sessions (age, sex by test score).

Participants

Thirty-eight participants (47% males, 53% females, age 20.08 ± 1.44 years, height = 165.95 ± 18.49 cm, mass = 71.29 ± 12.55 kg) (Table 1) were recruited from all sports at two National Collegiate Athletic Association (NCAA) Division II colleges located in New England. We sought out-of-season varsity intercollegiate athletes between the ages of 18–25 for inclusion in this study. Seventy-nine percent of the participants were

Table 1. Participant demographics.

	Minimum	Maximum	Mean ± SD
Age	18	24	20.08 ± 1.44
Height (cm)	117	189	165.95 ± 18.48
Mass (kg)	48.7	100.6	71.29 ± 12.55
Sleep session 1 (hours)			7.45 ± 0.97
Sleep session 2 (hours)			7.36 ± 1.15

soccer athletes, 21% played various other sports. Potential participants were excluded if they had a diagnosed traumatic brain injury within the past year, acute lower extremity injury within the past 6 months, a history of ADD or ADHD, or used medication that could affect balance or neurocognitive performance. Once eligibility was determined, participants provided informed consent. The study received human subjects' approval from our institutional review board.

Instrumentation and measures

To ensure internal validity, two iPads with the C3 Logix application installed were designated for use throughout the research process. We created a test workflow within the application that featured a balance assessment, followed by the two neurocognitive assessments, and ending with the final balance assessment. The iPads were charged the night prior to testing and powered off after each balance assessment or neurocognitive test session. During testing, the iPads remained in airplane mode with the Wi-Fi off and all background applications closed. The iPads were charged between participants if the charge went below 80%. The same stylus was used for neurocognitive testing for all test sessions. The same foam Airex pad (Airex Balance Pad, Airex AG, Switzerland) was used for all balance assessments. The Trail Making A test involved the participants connecting the numbers 1–25 as quickly as they could (either on paper or the iPad screen), without making any mistakes. Trail Making B test was completed in a similar fashion with participants being instructed to connect both numbers and letters in alternating numerical and alphabetic order (Reitan, 1992). The SDMT test involved the participant matching by writing the corresponding number that each symbol corresponded to. A minute and a half was allowed on the paper-and-pencil test, while 120 seconds was allowed on C3 Logix (Smith, 1982). When a mistake was made during the C3 Logix SDMT, the instructions require that the test-taker move on and not fix it; participants were instructed to fix a self-identified mistake made on the paper and pencil form of the test. The test administrator was present to alert the participant of any mistakes made in Trail Making A and B; participants were not alerted of a mistake made during the SDMT by the test administrator.

Procedures

Participants completed 2 test sessions 1 week apart in the same environment and at the same time of day. The first test session began with participants reading and signing an informed consent form, completing a demographic and health history questionnaire and being measured for height and weight. All questions on these forms can be found in Figure 1.

Next, participants began their first balance assessment. All instructions to participants were given using a common script. The same examiner administered all the balance assessments. The BESS test was administered through the C3 Logix application, with the use of a special belt, according to C3 Logix instructions. Participants were given an overview of the BESS test protocol. The test administrator demonstrated each stance to the participant prior to testing and asked participants to demonstrate each stance to ensure comprehension before the trial began. Participants completed the double leg (DL), single leg (SL), and tandem leg (TL) stances first on the floor and then on a foam pad.

Session 1 participant questionnaire

1. How many hours of sleep did you get last night?
2. Have you ever had an injury to your face, head, skull, or brain that resulted in confusion, memory loss, or headache from a hit to your head, having your "bell rung" or getting "dinged" while participating in sports or recreational activities?
   If yes:
   Number of concussions:
   What sport or activity did it occur?
   Year of injury?
3. Have you been diagnosed with a concussion within the past 12 months?
4. Do you wear corrective lenses or glasses?
5. Have you been diagnosed with a hearing impairment?
6. Have you been diagnosed with a lower extremity musculoskeletal injury within the past 6 months?
7. Have you been diagnosed with a learning disability or ADD/ADHD?
8. Have you been diagnosed with a vestibular or neurological disorder that affects your balance?
   If answered yes to question 7 or 8, please name the disorder(s) below:
9. Are you currently taking any medications daily?

Figure 1. Questionnaire administered to participants prior to Session 1.

Following the balance assessment, a single investigator administered both sessions of the neuropsychological portion of testing with a common script. Test administrators were trained by a board-certified neuropsychologist in appropriate test administration. This portion of testing involved the participants completing both the adapted Trail Making A, Trail Making B, and SDMT using C3 Logix, and the analogous pencil and paper tests. An online random number generator (randomizer. org) to determine the order between iBESS and neuropsychological tests, A second randomization determined whether Trail Making or SDMT would be conducted first.

The Trails A, Trails B, and SDMT were administered according to C3 Logix written procedures. Images of these tests can be found in Figures 2, 3, and 4.

Between administration of the C3 Logix and paper-and-pencil neurocognitive tests, there was a five-minute rest break.

For the second administration of the balance test, participants returned 1 week later to complete the second testing session. Prior to the start of the second session, participants were issued a second health history questionnaire (Figure 5) to ensure there were no changes to their medical history since the first session (e.g., that they had not suffered a concussion, injury, or started any new drug therapies).

Statistical analyses

All data were analyzed using SPSS (Version 23). The alpha level was set a priori at .05, for all analyses. Test scores were checked for outliers using a 4 standard deviation from mean threshold (Grubbs, 1969; Stefansky, 1972). To determine within-session and 1-week test-retest reliability of balance assessments, a two-way random effects model using the average of ratings was used to calculate intraclass correlation coefficients (ICC) for BESS errors and iBESS volume. In order to compare differences in both BESS errors and iBESS volume vs. time, separate Repeated-measures analysis of variance (ANOVA) with Greenhouse Geisser Correction and paired samples *t*-tests posthoc analyses for both variables were used. Pearson correlation was used to demonstrate the relationship between BESS errors and iBESS volume.

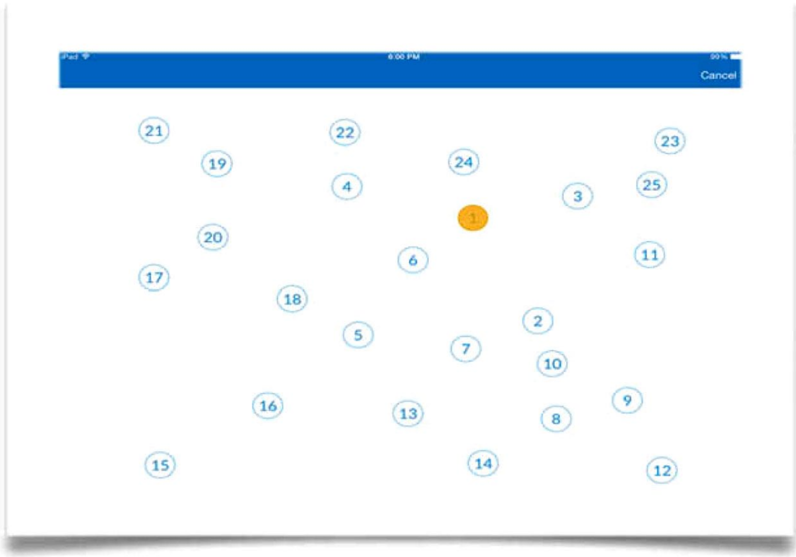

Figure 2. Picture of C3 Logix trail making A test. Screen shot of C3 Logix Trail Making A test administered on Apple iPad. © NHMI reproduced by permission of C3 Logix.

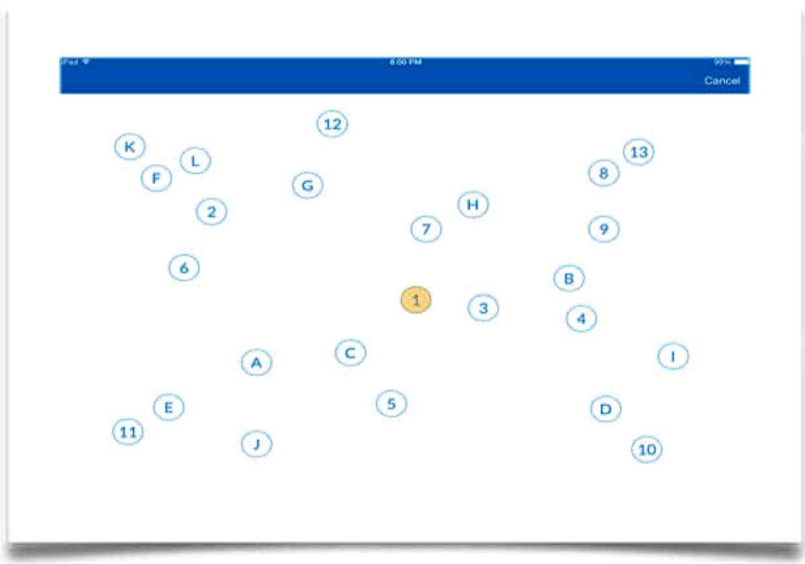

Figure 3. Picture of C3 Logix trail making B test. Screen shot of C3 Logix Trail Making B Test administered on Apple iPad. © NHMI reproduced by permission of C3 Logix.

To assess test–retest reliability of the neuropsychological tests, intraclass correlations were calculated for each pair of tests. A two-way random effects model using the average of ratings, with an absolute definition was used to calculate ICC. To assess for practice effects, a series of paired sample *t*-tests were calculated between time 1 and time 2 for each test to determine the magnitude of change. To assess concurrent validity of the neuropsychological tests, a Pearson correlation was calculated.

The demographic variables of sex and amount of sleep were analyzed for their effects on neuropsychological and balance scores at each test session. For sex as the independent variable, *t*-test was calculated. For hours of sleep, Pearson correlations were calculated.

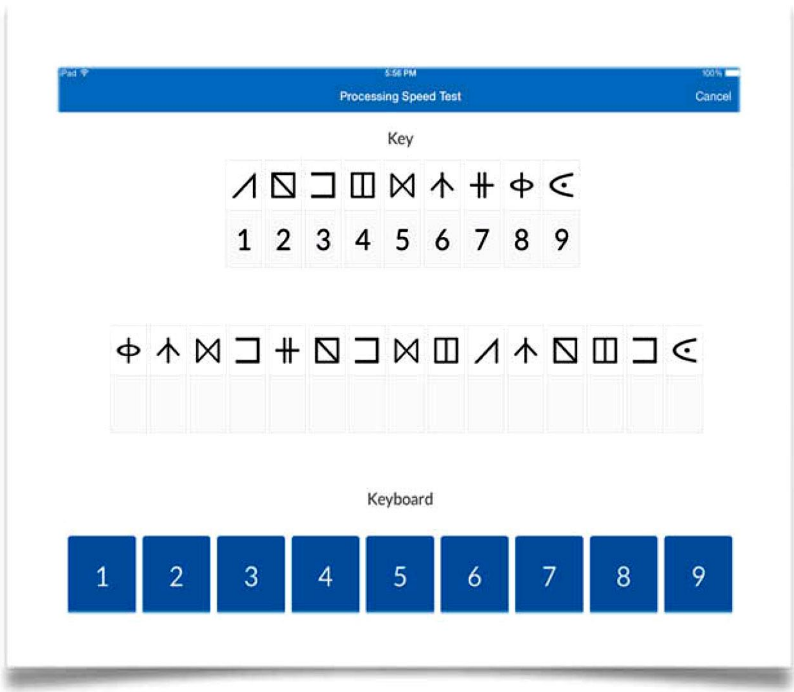

Figure 4. Picture of C3 Logix SDMT. Screen shot of C3 Logix SDMT administered on Apple iPad. © NHMI reproduced by permission of C3 Logix.

Session 2 Participant Questionnaire

1. How many hours of sleep did you get last night?
2. Have you ever had an injury to your face, head, skull, or brain that resulted in confusion, memory loss, or headache from a hit to your head, having your "bell rung" or getting "dinged" while participating in sports or recreational activities since your last testing session?
3. Have you been diagnosed with a lower extremity musculoskeletal injury since your last testing session?
4. Has there been any change in your medical history or medication use since your last testing session?

If you answered yes, please list any changes below:

Figure 5. Questionnaire administered to participants prior to Session 2.

Results

Balance assessment

The descriptive data of errors among all stances and trials can be found in Table 2. Across test conditions, within-session reliability ranged from acceptable to excellent for BESS errors ($ICC_{(2,1)} = 0.76–0.93$, $p < 0.001$) and was excellent for iBESS volume ($ICC_{(2,1)} = 0.76–0.93$, $p < 0.001$). One-week test-retest reliability was acceptable for BESS errors: $ICC_{(2,1)} = 0.80$ (95% CI:0.62–0.90, $p < 0.001$, and excellent for iBESS volume ($ICC_{(2,1)} = 0.91$(95% CI:0.82–0.95, $p < 0.001$). Reliability data is provided in Table 3. Repeated measures ANOVAs revealed significant differences between trials for errors and iBESS volume. Post-hoc analysis revealed BESS errors were significantly higher in trial 1 (11.5 ± 5.4) compared to trials 2 (9.2 ± 4.4), 3 (8.8 ±

Table 2. BESS results.

Mean errors ± SD				
Condition	Trial 1	Trial 2	Trial 3	Trial 4
DL floor[a]	0	0	0	0
SL floor	2.500 ± 2.166	1.579 ± 2.176	1.553 ± 1.927	1.395 ± 1.285
TS floor	0.737 ± 1.155	0.184 ± 0.457	0.369 ± 0.633	0.211 ± 0.413
DL foam[a]	0	0	0	0
SL foam	5.605 ± 2.034	4.895 ± 1.984	4.974 ± 1.966	4.447 ± 2.101
TS foam	2.684 ± 1.338	2.500 ± 1.371	1.895 ± 1.467	2.000 ± 1.186
Total error	11.526 ± 5.446	9.158 ± 4.421	8.789 ± 4.001	8.053 ± 3.296
Total iBESS volume	−5.863 ± 5.580	−7.562 ± 5.503	−7.260 ± 5.751	−7.495 ± 5.160
Errors vs. sleep obtained	$r = 0.07$, $p = 0.68$	$r = 0.054$, $p = 0.75$	$r = 0.23$, p = 0.16	$r = 0.18$, $p =$ 0.27
Errors vs. sex of participant	$t(36) = 0.031$, $p = 0.975$	$t(36) = -0.302$, p = 0.765	$t(36) = 0.58$, $p = 0.565$	$t(36) = 1.082$, $p = 0.287$

DL = double leg; SL = single leg; TS = tandem stance. T1 = trial 1; T2 = trial 2; T3 = trial 3; T4 = trial 4.
a. zero variance in DL floor and foam conditions among all 4 trails.

4.0), and 4 (8.1 ± 3.3). A significantly higher error count was also found in trial 2 (9.2 ± 4.4) compared to trial 4 (8.1 ± 3.3). Post-hoc analysis revealed iBESS volume was significantly greater in trial 1 (−5.9 ± 5.6) compared to trials 2 (−7.6 ± 5.5), 3 (−7.3 ± 5.8), and 4 (−7.5 ± 5.2) (Table 4). Pearson correlations revealed a significant moderate-strong, positive relationship between errors and iBESS volume, in the SL firm ($r = 0.44$, $p = 0.002$), TS firm ($r = 0.63$, $p < 0.001$), TS foam ($r = 0.61$, $p < 0.001$) and total stance (r = 0.41, $p = 0.011$) conditions (Table 5). Table 4 presents the data reporting practice effects among all trials.

There was no correlation between sleep and BESS errors in any of the four trials. There was also no significant difference in BESS errors between sexes in any of the trials (Table 2).

Neurocognitive assessment

Means and standard deviations for all neuropsychological tests appear in Table 6. One subject had one neuropsychological test score that was more than 4 standard deviations from the group mean (Grubbs' test, $p < .05$). That score was removed from the data.

Test-retest analyses found significant practice effects for most test-pairs. Student *t*-tests showed significant differences between test session one and two for all C3 Logix and paper-and-pencil tests (Table 7). These results can be found in Table 7. Intraclass correlations were acceptable for all but paper-and-pencil Trails B, the Trails B minus Trails A difference score, and the C3Logix Trails A. Paper-and-pencil Trails A was marginal for reliability.

One-week test stability demonstrated significant improvement across all tests, with the exception of paper-and-pencil Trails B-Trails A difference score. Some tests varied more than others (Table 7).

There were moderate to strong correlations between paired paper-and-pencil and C3Logix tests for Trails B, SDMT, and the Trails B minus A score. Trails A did not show a significant relationship between formats. The intercorrelation matrix of neuropsychological tests at Time 1 appears in Table 8. The intercorrelations among C3 Logix tests were almost all significant: C3Logix Trails A and Trails B correlating with each other, Trails B correlating with, and the Trails B minus A time correlated with Trails B and SDMT, but not Trails A. A different pattern emerged for the paper-and-pencil tests with both Trails A and B correlating with SDMT but not each other. The paper-and-pencil Trails B minus Trails A had the same pattern as the C3Logix analog.

Similar to the findings about the balance assessments reported above, participants' hours of sleep the night prior to testing had no effect on test performance (C31 Trails A: $r = 0.007$, $p = 0.967$, C31 Trails B: $r = -0.088$, $p = 0.599$, C31 SDMT: $r = 0.069$, $p = 0.683$, PP1 Trails A: $r = 0.161$, $p = 0.334$, PP1 Trails B: $r = -0.065$, $p = 0.701$, PP1 SDMT: $r = 0.107$, $p = 0.521$) while participant sex was only related to C3Logix SDMT scores,

Table 3. Test–retest reliability.

BESS ERRORS		ICC $_{(2,1)}$	95% CI	*p*-value
Within Session (1 vs. 2)	DL floor	0	0	0
	DL foam	0	0 0	
	SL floor	0.811	0.637,0.902	0.000
	SL foam	0.804	0.623,0.898	0.000
	TS floor	0.473	−0.014,0.725	0.027
	TS foam	0.706	0.434,0.847	0.000
	All stances	0.934	0.872,0.965	0.000
Within Session (3 vs. 4)	DL floor	0	0	0
	DL foam	0	0	0
	SL floor	0.602	0.234,0.793	0.003
	SL foam	0.586	0.203,0.785	0.004
	TS floor	0.325	−0.299,0.649	0.118
	TS foam	0.501	0.040,0.741	0.019
	All stances	0.760	0.537, −0.875	0.000
1 Week (2 vs. 3)	DL floor	0	0	0
	DL foam	0	0	0
	SL floor	0.714	0.450,0.851	0.000
	SL foam	0.627	0.282,0.806	0.002
	TS floor	0.466	−0.28,0.722	0.030
	TS foam	0.517	0.071,0.749	0.015
	All stances	0.803	0.621,0.898	0.000
iBESS VOLUME		**ICC $_{(2,1)}$**	**95% CI**	***p*-Value**
Within Session (1 vs. 2)	DL floor	0.900	0.807,0.948	0.000
	DL foam	0.909	0.825,0.953	0.000
	SL floor	0.799	0.613,0.895	0.000
	SL foam	0.830	0.672,0.911	0.000
	TS floor	0.855	0.741,0.930	0.000
	TS foam	0.660	0.346,0.823	0.001
	All stances	0.966	0.935,0.982	0.000
Within Session (3 vs. 4)	DL floor	0.893	0.795,0.945	0.000
	DL foam	0.902	0.831,0.954	0.000
	SL floor	0.388	−0.178,0.682	0.070
	SL foam	0.778	0.572,0.884	0.000
	TS floor	0.686	0.396,0.837	0.000
	TS foam	0.830	0.673,0.912	0.000
	All stances	0.923	0.852,0.960	0.000
1 Week (2 vs. 3)	DL floor	0.833	0.679,0.913	0.000
	DL foam	0.909	0.826,0.953	0.000
	SL floor	0.719	0.460,0.854	0.000
	SL foam	0.839	0.690,0.916	0.000
	TS floor	0.792	0.600,0.892	0.000
	TS foam	0.737	0.494,0.863	0.000
	All stances	0.906	0.820,0.951	0.000

ICC = intraclass correlations; CI = confidence interval.

with females getting significantly more correct on average than males: Time 1 female mean = 73.00, *SD* = 10.36, male mean = 66.56, *SD* = 7.83; $F(1,36) = 4.595$, $p = .039$). Time 2 findings were similar as none of the tests were significantly correlated with sex (C32 Trails A: $F(1,36) = 0.299$, $p = .588$, C32 Trails B: $F(1,36) = 0.488$, $p = .489$, PP2 Trails A: $F(1,36) = 1.953$, $p = .171$, PP2 Trails B: $F(1,36) = 0.308$, $p = .582$, PP2 SDMT: $F(1,36) = 2.745$, $p = .106$) except for C32 SDMT: ($F(1,36) = 11.138$, $p = .002$). Sleep the night before had no effect on any test score.

Table 4. Practice effects (Post-Hoc): BESS errors and iBESS volume.

Practice Effects: Comparison of Differences Between Trials
BESS Errors iBESS Volume

	Mean Trial 1-Trial 2 ± SD	*t*	*p*-value	Mean Trial 1-Trial 2 ± SD	*t*	*p*-value
Session 1 vs. session 2	2.37 ± **2.48**	5.90	<0.001*	1.70 ± **2.01**	5.22	0.001*
Session 1 vs. session 3	2.74 ± **4.25**	3.97	<0.001*	1.40 ± **3.15**	2.74	0.009*
Session 1 vs. session 4	3.47 ± **3.64**	5.88	<0.001*	1.63 ± **2.81**	3.58	0.001*
Session 2 vs. session 3	0.37 ± **3.42**	0.66	0.511	−0.30 ± **3.29**	−0.57	0.575
Session 2 vs. session 4	1.11 ± **2.86**	2.39	0.022*	−0.07 ± **2.66**	−0.16	0.878
Session 3 vs. session 4	0.74 ± **3.23**	1.41	0.168	0.24 ± **2.92**	0.50	0.623

df value among all trials was 37.
* $p < 0.05$.

Table 5. Correlation between errors and iBESS volume.

	Error Mean ± SD	iBESS mean ± SD	Pearson *r*	*p*-value
SL Floor	1.40 ± 1.28	−0.35 ± 1.64	0.444	0.002
TS Floor	0.21 ± 0.41	−2.21 ± 1.39	0.632	<0.001
SL Foam	4.45 ± 2.10	1.40 ± 1.10	0.006	0.973
TS Foam	2.00 ± 1.19	1.44 ± 1.46	0.608	<0.001
Total Stances	8.05 ± 3.30	−7.50 ± 5.16	0.409	0.011

Table 6. Means and standard deviations of neuropsychological tests.

	Time 1		Time 2	
Test	Paper Pencil 1	C3 Logix 1	Paper Pencil 2	C3 Logix 2
TM A[a]	17.214 ± 3.865	17.575 ± 3.162	15.631 ± 3.577	15.910 ± 3.388
TM B[a]	40.803 ± 10.090	35.412 ± 6.770	34.576 ± 8.015	30.463 ± 7.107
SDMT[a]	67.789 ± 8.826	69.947 ± 9.692	73.842 ± 10.466	73.658 ± 10.861
TrB-TrA	23.754 ± 9.738	17.903 ± 6.119	18.945 ± 7.776	14.553 ± 6.852

SDMT = Symbol Digit Modality Test; TrB-TrA = Time for Trails B minus Time for Trails A (seconds); All test sessions had 37 participations except for session 2 trail Making B, which collected data on 36 participants.
a. Mean test scores were time to complete in seconds for Trails A/B, and number of correctly matched symbols for SDMT.

Table 7. Test–retest statistics for neuropsychological tests.

Variable	*t*	*r*	ICC
PPTM A	2.602*	.494**	.628***
PPTM B	3.807***	.420**	.504**
PP SDMT	−5.300***	.746**	.765***
PP TrB-TrA	2.603*	0.252	0.361
C3 TM A	2.823**	.388*	.516**
C3 TM B	7.356***	.759***	.719***
C3 SDMT	−3.917***	−.424**	.883***
C3TrB-TrA	4.296***	.625***	.694***

t = *t*-value for paired subject's Student's *t*-test; *r* = Pearson correlation coefficient; ICC = intraclass correlation; PPTM A = paper and pencil Trails A; PPTM B = paper and pencil Trails B;PP SDMT = paper and pencil Symbol Digit Modalities Test; PP TrB-TrA = paper and pencil Trails B time minus Trails A time; C3 TM A = C3 Logix Trails A; C3 TM B = C3 Logix Trails B; C3 SDMT – C3 Logix Symbol Digit Modalities Test; C3TrBTrA = C3 Logix Trails B time minus Trails A time.
* $p < .05$; ** $p < .01$; *** $p < .001$

Table 8. Intercorrelations of neuropsychological tests: Session 1.

	PP1_Trails A_time	PP1_Trail B_time	PP1_SDMT_tot. corr	Tr1B_A	C31_Trails A_time	C31_Trails B_time	C31_Speed_Corr	C3Tr 1_B_A
PP1_TrailsA_time	1.000	0.279	−0.350*	−0.099	0.124	0.334*	−0.099	0.420**
PP1_TrailB_time	0.279	1.000	−0.649**	0.928**	0.224	0.519**	−0.621**	0.473**
PP1_SDMT_tot.corr	−0.350*	−0.649**	1.000	−0.544**	−0.323*	−0.649***	0.757**	−0.563***
Tr1B_A	−0.099	0.928**	−0.544**	1.000	0.179	0.538**	−0.621**	0.299
C31_TrailsA_time	0.124	0.224	−0.323*	0.179	1.000	0.430**	−0.308	−0.044
C31_TrailsB_time	0.437**	0.519**	−0.649***	0.334*	0.430**	1.000	−0.445**	0.883***
C31_Speed_Corr	−0.099	−0.621**	0.757**	−0.621**	−0.308	−0.445**	1.000	−0.345*
C3Tr1_B_A	0.420**	0.473**	−0.563***	0.299	−0.044	0.883***	−0.345*	1.000

Discussion

Our results suggest that in general, the C3 Logix balance assessment demonstrates moderate-strong one-week test-retest reliability when measuring postural stability. Additionally, we found that most of the neurocognitive tests assessed demonstrated concurrent validity compared to their analogous paper and pencil forms. This is the first known independent concurrent validity study of C3 Logix tests.

These results support the hypotheses that the balance measurement available through C3 Logix demonstrates moderate-strong within-session test–retest reliability for both errors and iBESS volume. This finding is similar to other studies in which ICC values for BESS test-retest reliability ranged from 0.70–0.90 (Amin, Coleman, & Herrington, 2014; Hansen et al., 2016; Valovich McLeod et al., 2006). Our ICC values obtained for within-session reliability between the first and second trials are above these previously-reported ranges. This finding may be related to the shorter intervals between our BESS trials (30 minutes compared to days, weeks, or even months). Our one-week test–retest reliability results are similar to those found in a study performed by Amin et al.

In each analysis performed, the iBESS volume demonstrated stronger test–retest reliability than that for analyses of error count. This finding suggests that studying postural sway is a more reliable variable to measure over time than is error counting. This finding is not surprising as the clinical practice of counting BESS errors emerged as a low-tech, low-price means to grossly assess balance. Before the gyroscope and accelerometer became available through the Apple iPad, advanced motion tracking and force plate systems allowed clinicians to gather precise measurements about postural sway and dynamic stability (Brown et al., 2014). However, cost, space limitations, and detailed training/set-up process (Brown et al., 2014) make it unlikely the average clinician will have these options. The iPad-based technology provides considerably more objective information and consistency for the assessment of balance. This is useful when performing multiple balance assessments over time to manage an athlete's return to play status.

The hypothesis regarding practice effects was also accepted: error count and iBESS volume decreased for all trials after Trial 1. There were significantly more errors in Trial 1 as compared to Trials 2, 3, and 4. These findings are similar to those observed by Valovich McLeod et al. (2004). In addition, when directly comparing Trial 2 and Trial 4, there was a significant error difference between the two trials. Our iBESS volume findings were similar in that there was significantly more postural sway in Trial 1 as compared to Trials 2, 3, and 4. These findings suggest that when performing the non-instrumented BESS test (i.e., without measuring postural sway), three practice sessions should be allowed to eliminate the effects of practice during testing. However, when considering iBESS volume, two practice sessions may adequately reduce the effects of practice. Eliminating the need for one full practice session by using C3 Logix may help clinicians perform balance assessments more efficiently and effectively.

One of the secondary goals was to determine the strength of the relationship between BESS errors and iBESS volume. There was a strong relationship between error count and postural stability measurements (iBESS volume) in 3 of 8 conditions and overall: SL floor, TS floor, TS foam. There was not a strong correlation between the error count and iBESS volume in the DL stance positions. Although none of the participants committed any errors during either DL stance (floor or foam), the accelerometer and gyroscope could detect differences in postural sway. This supports conclusion that the iBESS volume provides more sensitive measures of postural stability in both DL stance conditions (Alberts et al., 2015). Together with the improved reliability of iBESS, these findings suggest that the balance tests on the C3 Logix iPad application may allow clinicians to better detect lingering balance deficits than BESS error counts when assessing concussed athletes.

Another secondary goal was to identify relationships between BESS performance and sex or hours of sleep the night preceding the testing. No relationship between sleep or sex and BESS performance was found. It is important to note that all participants had a similar amount of sleep for both sessions (session-1 = 7.45 hours ± 0.97, session-2 = 7.36 ± 1.15). It is not possible to generalize these results to athletes who were not as well-rested. Previous research has shown a strong negative correlation between amount of sleep and balance stability (Siu, Huang, Beacom, Bista, & Rautiainen, 2015).

In terms of the neuropsychological tests the intraclass correlations for the 1-week retest period were generally good and consistent with previous studies of the paper and pencil if not better. However, both paper and pencil Trail Making tests (and the difference score) had ICC's that were below or marginal for acceptability, while the C3 Logix Trails A was below acceptability. Thus, reliability of Trail Making A is called into question across formats.

A practice effect was evident on all tests across formats. In the future, stability and reliability beyond 1 week should be assessed to allow a more robust description of practice effects. Clinicians should remain aware of the practice effect between session one and two. Additionally, it should be noted that C3 Logix generates new tests at random and participants may not have taken the same version of the test in their two test sessions. This is done purposely to reduce the chances of memorizing stimuli. The strong test-retest reliability should allay fears that changing stimuli impacts score outcomes.

The results generally support the hypotheses that the neuropsychological tests available through C3 Logix show acceptable concurrent validity with standard (comparable) paper and pencil measures. The performance of the paper and pencil Trails A was somewhat unusual as it did not correlate with either paper and pencil Trails B or C3Logix Trails A as expected. Somewhat similar findings were observed for C3Logix Trails A, as it only correlated with C3Logix Trails B (and paper and pencil SDMT). Because performance on Trails B depends on one's ability to do the simpler Trails A task, these findings raise questions about the validity of Trails A, and by extension, the Trails B minus A calculation. Interestingly, that score seemed to hold up across analyses for both formats.

We reiterate the caution previously mentioned: most of the participants in this study obtained the recommended 7–9 hours of sleep their age range (sleepfoundation.org). This may not be the case for in-season or concussed athletes. There is mixed evidence for the effect of sleep the night before testing with a recent study showing that less than 7-hours of sleep the night before ImPACT testing has a detrimental effect of test scores (McClure, Zuckerman, Kutscher, Gregory, & Solomon, 2014; Silverberg, Berkner, Atkins, Zafonte, & Iverson, 2016). Other recent studies have found that sleep differences in sleep amount and quality did not result in any differences in neurocognitive or balance performance (Mihalik et al., 2013; Silverberg et al., 2016).

Limitations

Several limitations were present in this study. First, all participants in the study were NCAA intercollegiate athletes and between the ages of 18–25 compromising generalizability to younger or older populations of varying athletic ability. Most participants were soccer players. Findings among athletes from other sports may be different. In addition, information about race/ethnicity or the presence of psychological conditions was not obtained.

The study design was not intended to validate these tests relative to diagnosis or monitoring of recovery. It was also only possible to assess three of the C3Logix five tests as no stand-alone test of reaction time was available. The sample size was also somewhat small so the ability to detect sex effects was limited. Obtaining comparable data across institutions is needed.

Conclusions

This independent study of the concurrent validity and reliability of several of the C3Logix tests showed that the instrumented balance assessment in the C3 Logix concussion testing suite demonstrated strong concurrent validity with the BESS error system and indicated improved metrics from the BESS. Moderate-strong one-week test–retest reliability was also demonstrated. Additionally, Trails B, SDMT, and the Trails difference score all demonstrated acceptable concurrent validity.

Our findings also suggest that measuring the iBESS volume measured through the highly-portable iPad offers clinicians a potentially stronger means to detect post-injury balance deficits than using BESS errors alone.

References

Alberts, J. L., Hirsch, J. R., Koop, M. M., Schindler, D. D., Kana, D. E., Linder, S. M., … Thota, A. K. (2015). Using accelerometer and gyroscopic measures to quantify postural stability. *Journal of Athletic Training*, *50*(6), 578–588.

Amin, D. J., Coleman, J., & Herrington, L. C. (2014). The test–retest reliability and minimal detectable change of the balance error scoring system. *Journal of Sports Sciences*, *2*, 200–207.

Aubry, M., Cantu, R., Dvorak, J., Graf-Baumann, T., Johnston, K., Kelly, J., … Schamasch, P. and G. Concussion in Sport. (2002). Summary and agreement statement of the First International Conference on Concussion in Sport, Vienna 2001. Recommendations for the improvement of safety and health of athletes who may suffer concussive injuries. *British Journal of Sports Medicine*, *36*(1), 6–10.

Barth, J. T., Alves, W. M., Ryan, T. V., Macciocchi, S. N., Rimel, R. W., Jane, J. A., Nelson, W. E. (1989). *Mild head injury in sports: Neuropsychological sequelae and recovery of function. Mild head injury* (Vol. 1, pp. 257–275). Levin, H. S., Eisenberg, M.D., & Benton, A. L. (Eds.). New York, NY: Oxford University Press.

Bell, D. R., Guskiewicz, K. M., Clark, M. A., & Padua, D. A. (2011). Systematic review of the balance error scoring system. *Sports Health*, *3*(3), 287–295.

Broglio, S. P., Cantu, R. C., Gioia, G. A., Guskiewicz, K. M., Kutcher, J., Palm, M., & McLeod, T. C. (2014). National athletic trainers' association position statement: Management of sport concussion. *Journal of Athletic Training*, *49*(2), 245–265.

Brown, H. J., Siegmund, G. P., Guskiewicz, K. M., Van Den Doel, K., Cretu, E., & Blouin, J. S. (2014). Development and validation of an objective balance error scoring system. *Medicine and Science in Sports and Exercise*, *46*(8), 1610–1616.

Cantu, R. C., Aubry, M., Dvorak, J., Graf-Baumann, T., Johnston, K., Kelly, J., … McLeod, T. C. (2006). Overview of concussion consensus statements since 2000. *Neurosurgical Focus*, *21*(4), E3.

Collins, M. W., Lovell, M. R., & McKeag, D. B. (1999). Current issues in managing sports-related concussion. *JAMA*, *282*(24), 2283–2285.

Coppel, D. B. (2011). Use of neuropsychological evaluations. *Physical Medicine and Rehabilitation Clinics of North America*, *22*(4), 653–664, viii.

Grubbs, F. (1969). Procedures for detecting outlying observations in samples. *Technometrics*, *11*(1), 1–21.

Guskiewicz, K. M. (2001). Postural stability assessment following concussion: One piece of the puzzle. *Clinical Journal of Sport Medicine*, *11*(3), 182–189.

Hansen, C., Cushman, D., Anderson, N., Chen, W., Cheng, C., Hon, S. D., & Hung, M. (2016). A normative dataset of the balance error scoring system in children aged between 5 and 14. *Clinical Journal of Sport Medicine*, *26*(6), 497–501.

Kushner, D. S. (2001). Concussion in sports: Minimizing the risk for complications. *American Family Physician*, *64*(6), 1007–1014.

McClure, D. J., Zuckerman, S. L., Kutscher, S. J., Gregory, A. J., & Solomon, G. S. (2014). Baseline neurocognitive testing in sports-related concussions: The importance of a prior night's sleep. *The American Journal of Sports Medicine*, *42*(2), 472–478.

McCrory, P., Meeuwisse, W., Aubry, M., Cantu, B., Dvorak, J., Echemendia, R., … Turner, M. (2013). Consensus statement on concussion in Sport-The 4th International Conference on Concussion in Sport held in Zurich, November 2012. *Journal of Science and Medicine in Sport*, *16*(3), 178–189.

McCrory, P., Meeuwisse, W., Dvorak, J., Aubry, M., Bailes, J., Broglio, S., ... Vos, P. E. (2017). Consensus statement on concussion in sport-the 5th international conference on concussion in sport held in Berlin, October 2016. *British Journal of Sports Medicine*. Br J Sports Med Published Online First: 26 April 2017. doi: 10.1136/bjsports-2017-097699

McCrory, P., Meeuwisse, W., Johnston, K., Dvorak, J., Aubry, M., Molloy, M., & Cantu, R. (2009). Consensus statement on Concussion in Sport-The 3rd International Conference on Concussion in Sport held in Zurich, November 2008. *Journal of Science and Medicine in Sport*, *16*(2), 83–92.

Mihalik, J. P., Lengas, E., Register-Mihalik, J. K., Oyama, S., Begalle, R. L., & Guskiewicz, K. M. (2013). The effects of sleep quality and sleep quantity on concussion baseline assessment. *Clinical Journal of Sport Medicine*, *23*(5), 343–348.

Nanda, A., Khan, I. S., Goldman, R., & Testa, M. (2012). Sports-related concussions and the Louisiana youth concussion act. *The Journal of the Louisiana State Medical Society: Official Organ of the Louisiana State Medical Society*, *164*(5), 246–250.

Reitan, R. M. (1992). *Trail making test*. Manual for administration and scoring, Reitan Neuropsychology Laboratory. Tucson, AZ.

Schatz, P., & Zillmer, E. A. (2003). Computer-based assessment of sports-related concussion. *Applied Neuropsychology*, *10*(1), 42–47.

Silverberg, N. D., Berkner, P. D., Atkins, J. E., Zafonte, R., & Iverson, G. L. (2016). Relationship between short sleep duration and preseason concussion testing. *Clinical Journal of Sport Medicine*, *26*(3), 226–231.

Siu, K. C., Huang, C. K., Beacom, M., Bista, S., & Rautiainen, R. (2015). The association of sleep loss and balance stability in farmers. *Journal of Agromedicine*, *20*(3), 327–331.

Smith, A. (1982). *Symbol digit modalities test (SDMT) manual (revised) Western psychological services*. Los Angeles, CA:.

Stefansky, W. (1972). Rejecting outliers in factorial designs, technometrics. Journal of Technometrics. *14*, 469–479. Valovich McLeod, T. C., Barr, W. B., McCrea, M., & Guskiewicz, K. M. (2006). Psychometric and measurement properties of concussion assessment tools in youth sports. *Journal of Athletic Training*, *41*(4), 399–408.

Valovich McLeod, T. C., Perrin, D. H., Guskiewicz, K. M., Shultz, S. J., Diamond, R., & Gansneder, B. M. (2004). Serial administration of clinical concussion assessments and learning effects in healthy young athletes. *Clinical Journal of Sport Medicine*, *14*(5), 287–295.

Van Kampen, D. A., Lovell, M. R., Pardini, J. E., Collins, M. W., & Fu, F. H. (2006). The “value added” of neurocognitive testing after sports-related concussion. *The American Journal of Sports Medicine*, *34*(10), 1630–1635.