Pilot study of a new concussion assessment tool using computerized cognitive testing and instrumented balance: Test-retest reliability

Neville C[1]*, Rieger B[2], Viterise T[1] and Brindle A[3]

[1]Department of Physical Therapy Education, Upstate Medical University, Syracuse, NY, USA

[2]Upstate Concussion Center, Upstate Medical University, Syracuse, NY, USA

[3]Quadrant Biosciences Inc, Syracuse, NY, USA

Abstract

Objective: Computerized cognitive tests and instrumented balance have become common in assessment systems for concussion injuries but reliability of novel systems and protocols have not been fully investigated. Therefore, the purpose of this study is to examine the reliability of a computerized tablet-delivered trail-making cognitive test and instrumented balance test in a sample of healthy participants.

Design: Test-retest reliability study

Setting: Laboratory testing

Participants: 20 healthy participants were included (5 males, 15 females).

Interventions: Not Applicable

Main outcome measures: Test-retest reliability was evaluated for a concussion screening assessment that included a tablet-delivered trail making cognitive test, an instrumented balance assessment, and a dual-task assessment that included both balance and cognitive testing simultaneously.

Results: Reliability coefficients for cognitive, balance, and dual-task testing ranged across tests and data processing options but the use of average speed for the digital trail making test and the power integral calculated from the balance tests yielded ICC's greater than 0.75. Further the dual-task tests also remained consistent across testing sessions with ICC's that were greater than 0.84.

Conclusions: The components of an objective assessment system (digital trail making and instrumented balance tests) when used in a concussion assessment show good reliability. Test-retest reliability for a computerized trail making test and instrumented balance is influenced by data processing strategies. Metrics used in instrumented balance assessment and computerized neurocognitive testing can show high retest reliability suggesting they may be helpful for clinical decision making. The ongoing development of concussion assessment tools requires clinicians to be aware of novel tools and their tested test-retest reliability.

Abbreviations: mTBI - mild traumatic brain injury; BESS – balance error scoring system; TMT – trail making test; RMS – root mean square; PSD – power spectral density; TLEO – two-leg eyes open; TLEC – two leg eyes closed; TSEO – tandem stance eyes open; TSEC - tandem stance eyes closed; TLEOfp – two-leg eyes open foam pad; TLECfp – two leg eyes closed foam pad; TSEOfp – tandem stance eyes open foam pad; TSECfp - tandem stance eyes closed foam pad; ICC – intraclass correlation coefficient;

Introduction

The incidence of mild traumatic brain injury (mTBI), also referred to as concussion injury, ranges between a low 1% to a high 24% depending on the sample [1,2]. There has been considerable interest in examining the sensitivity of new tests to identify impairments following concussion injuries [3,4]. However, test-retest reliability is under-evaluated and makes clinical decision making about injury difficult [5]. Numerous balance and neurocognitive assessments have been used as the availability of computer aided and portable tests have expanded with technology [6-12]. Further, the increased use of computerized

neurocognitive testing and instrumented balance assessments has allowed for greater opportunity to develop novel metrics. These new metrics and novel testing methods need further study to establish test-retest reliability for use in concussion screening.

The Balance Error Scoring System (BESS) is a widely used, and convenient, balance-screening tool with a range of reliability values reported across studies [13-16]. The test is visually scored by a trained rater who observes the participant complete a series of balance tests on a firm surface and also on a foam pad. Intra-rater reliability ranges from 0.60 to 0.92 for total scores and 0.50 to 0.98 for individual stances [13-15]. Training the rater on scoring (what counts as an error) has

been shown to impact reliability across studies [13]. Additionally, averaging multiple test attempts, or using a video review of performance improves reliability [17]. Removing the subjective nature of error counting has also been considered by using inertial sensors to divide up the test and identify or "count" balance errors [18]. Instrumenting the BESS does mean that data analysis options may impact retest reliability and is under-studied. Use of inertial measurement units to objectively capture performance on the BESS has the advantage of cost, portability, and ease of use, which may be factors explaining their increased use [9,19,20]. It remains an interest to find objective tests that can be used outside the confines of the research laboratory while still being easy to administer and demonstrate test-retest reliability similar to the widely used BESS test.

Concussion is known to affect several domains of cognitive functioning including attention, processing speed, memory and executive functioning [21]. Neurocognitive tests are common in a comprehensive concussion battery to capture these decrements. The Trail Making Test (TMT) is a neurocognitive test with two separate parts, Part A and B, and has a long history of use in the management of traumatic brain-injury [22]. Further, efforts to develop computerized versions of the Trails tests show promise to improve objective measurement [6,8]. Part A of the test involves connecting ascending encircled numbers in order (1-2-3) while Part B requires alternating between numbers and letters (1-A-2-B). Part B has been shown to be more difficult due to the more complex cognitive task that involves alternating between a sequence of letters and numbers to the complete the "trail" [23]. Motor speed and visual scanning are important to both A and B. Test-retest reliability for Part A has been reported to be 0.55, while part B is higher at 0.75 [24]. Across studies the reliability coefficients range, perhaps based on retest time frames or sample differences, but usually remain above a 0.70 cut-off some use for clinical acceptability [2,25-27]. However limited data on reliability for digital versions of the Trails tests and efforts to combine these in dual-task paradigms are even more limited.

A dual-task testing paradigm that combines cognitive and motor tests may increase sensitivity to the subtle deficit seen with concussion injury [28-30]. Interestingly, high reliability values for dual-task tests have also been reported, in some cases more consistent than the same tests done in isolation [31-36]. In a systematic review of dual task paradigms related to concussion there was evidence of high reliability for dual-task tests but an overall limitation of too few studies that report reliability psychometrics [37]. Further, the authors identify a lack of time and cost effective options for use of dual-task testing [33].

The need to evaluate test-retest reliability of assessment tools used for concussion injury is required before interpretation and application in clinical care is possible. There are tools available for the assessment of balance, cognitive function, and dual-task testing, but new methods of computerized testing and instrumented balance require examination of test-retest reliability. Therefore the purpose of this study is to pilot the test-retest reliability of balance, cognitive performance, and dual-task performance in a small sample to determine if expanded testing across more normative samples and injury patients is warranted. A secondary purpose was to examine the various data processing options inherently possible with computerized and instrumented testing to inform further work with these tools. The pilot nature of this study in a small homogenous sample was thought to be ideal for documenting the various data processing and analysis strategies to inform future testing with similar novel digital and instrumented measures.

Methods

Subjects

Twenty healthy college students between the ages of 20-30 volunteered to participate in the study. The Institutional Review Board at Upstate Medical University, Syracuse, NY approved the study and all subjects gave written informed consent before participation. The subjects included 5 males and 15 females. All subjects were graduate students who averaged 6.0±1.2 years of postsecondary (college level) education at the time of testing. All subjects were considered healthy and free from any current orthopedic or neurologic injuries that might affect their balance or cognition. Five subjects reported a prior concussion injury, one injury occurring 12 months prior; the other 4 injuries were all greater than 24 months prior to testing. All reported having fully recovered from their last concussion with no lingering symptoms at the time of testing. All 20 subjects reported engagement in sports activities that included yoga, cycling, and organized team sports such as basketball, soccer, and football. None currently played competitive sports but most continued to engage in recreational activity including running and a range of intramural type sports organized by the university. All subjects were currently enrolled graduate allied health students in good academic standing.

Procedure

All subjects participated in two sessions averaging 48 hours apart to establish short-term test-retest reliability. The assessments consisted of cognitive testing, balance testing, and a dual-task measure. Using a screening type format all participants completed the testing in the same order (cognitive tests, balance tests, dual-task test) using a tablet computer for all instructions and data collection. Details related to the specific cognitive and balance testing is included below.

Neurocognitive testing

All subjects completed a digital version of the Trail Making Test (Trails A and B) on a tablet computer (Toshiba, Model: WTB-B) using the touch interface and an accompanying stylus. The digital version used was modeled after the traditional paper-and-pencil version of the Trail Making Test in appearance and instructions [22]. To complete the test subjects are asked to connect, as quickly as possible, an ascending series of encircled numbers in Trails A (1-2-3 etc.) and ascending series of both encircled numbers and letters in Trails B, alternating between them (1-A-2-B-3-C etc.). Unlike the paper and pencil version, where the subject draws a "trail" to successively connect each number or letter to the next, the digital version only shows the last segment of the trail. Hence the board does not "fill" as you progress through the test, making it harder to identify encircled numbers or letters that haven't already been selected. One additional characteristic of the digital tablet version was that errors were not possible, as the trail segment would not be drawn if the numbers or letters were not touched in the required order. The digital administration of the test allowed for time and speed metrics to be calculated for the test total, consistent with a standard paper and pencil administration, and for each of the segments, which is only possible with the digital administration. Before each test, subjects were asked to complete a brief sample test to familiarize them with the process and the digital interface. The primary metric of interest was total time taken to complete the test consistent with the historic use of the paper-and-pencil version of the test while a segment-by-segment speed served as a secondary metric. Taking advantage of the digital administration of this test, reliability of the total time was examined as well as total time after dropping the slowest segment. The rationale

for investigating this additional time variable was an observation that subjects would at times get stuck during the test suggesting reliability of the underlying construct of speed would be improved with the removal of one "missed" item. Speed was calculated as the time taken to span the distance (in centimeters) between each of the circles (distance/time). The reliability for the average speed and average speed with the slowest speed dropped was evaluated to determine test, retest reliability.

Balance testing

Participants were asked to complete balance tests without shoes and instructed to maintain an upright standing position while being as still as possible. Measurement of balance was done with an inertial sensor (Quandrant Biosciences Inc, Syracuse, NY) affixed to a belt worn around the waist and placed at the approximate L5 level. The balance tests have been previously described, and validated, with use of an inertial sensor compared to other force plate and motion capture derived measures of balance but are summarized here as pertains to the current study [38]. Data from the inertial sensor was output at 250 Hz for 3-D linear and angular acceleration ($\pm$ 2.0 g range). The inertial sensor data in each axis was converted from g's to m/s^2. The acceleration due to gravity (estimated a 9.81 m/s^2) and any sensor bias was then subtracted before the signal was transformed into the frequency domain using a fast-fourier transform (FFT) algorithm. The two primary metrics of interest from the acceleration data included the power integral in the frequency domain, and the root mean square (RMS) calculated in the time domain. The power integral was calculated by taking the area under the power-spectral density (PSD) curve. These metrics were chosen because they have been previously used in the literature [9,11,39-41], allowing comparison to previous data, and represent both a time-domain and frequency-domain measure thought to be representative of postural stability. Subjects were asked to complete a series of 8 balance tasks chosen to challenge their balance while wearing an inertial sensor. The balance tests were done under controlled laboratory conditions that limited distractions and lasted 30 seconds each. The 8 balance tasks included:

- two-leg stance feet together, hands on hips, with eyes open (two-leg, eyes open: TLEO)

- two-leg stance feet together, hands on hips, with eyes closed (TLEC)

- tandem stance (dominant foot forward), hands on hips, with eyes open (tandem stance, eyes open: TSEO)

- tandem stance (dominant foot forward), hands on hips, with eyes closed (TSEC)

- two-leg stance feet together, hands on hips, with eyes open, on a foam-pad (TLEOfp)

- two-leg stance feet together, hands on hips, with eyes closed, on a foam-pad (TLECfp)

- tandem stance (dominant foot forward), hands on hips, with eyes open, on a foam-pad (TSEOfp)

- tandem stance (dominant foot forward), hands on hips, with eyes open, on a foam pad (TSECfp).

To examine effects on reliability each of the metrics (RMS and Power integral) were output for the entire 30-second test and then output following a method to divide the test into smaller windows of time (5-10 seconds) and evaluating each window, similar to collecting multiple trials. This windowing approach was utilized to consider the impact of removing excursions or losses-of-balance that may occur

across the test leaving the most stable, at-rest, and representation of balance to output. This is in contrast to another study that utilized a window approach to best capture the errors, or looses of balance, during the BESS test [18]. The windows were examined and the 3 windows with the greatest values were removed to allow investigation of "at-rest" postural stability. Windowing approaches have been frequently used but effects on retest reliability have been rarely described [9,11,18]. This windowing approach may be similar to other options such as repeating trials or filtering out specific frequencies described in previous studies [9-11,19]. It was hypothesized that windowing would improve reliability by removing inconsistent loses in balance between trials.

Dual-task testing

For the dual-task testing, subjects were asked to assume the TSEO position as completed during the balance portion of the screening but while also holding the tablet and completing the Trails B test. Subjects were instructed to correct any loss of balance as quickly as possible and continue to maintain the tandem-stance position until completing the Trails B test on the tablet computer. The data processing and outputs for the dual-task testing followed the same procedures described above for balance and neurocognitive testing done separately.

Data analysis

Descriptive statistics were used for all tests to examine means, standard deviations, and distributions. SPSS version 23 was used to obtain ICC's for each output metric of interest. Specifically, a 2-way random-effects analysis of variance ICC for absolute agreement (model 3,1) was calculated to estimate the test-retest reliability of the screening assessment for the short-term retest reliability. There are a variety of recommendations for ICC interpretation with several suggestions that an ICC of 0.70 to be minimally acceptable, however the intent of this analysis was to report the ICC's as a measure of reliability while also reporting the variability of differences (means and standard deviations) across time points to provide insight into any systematic changes in scores that might be a result of "learning or acclimation" effects. For this pilot study targeting the description of various analysis options the interpretation of acceptable reliability is largely left to the reader unless suggestions can be made when comparing previously published findings.

Results

Balance measures

The ICC's across the 8 balance tests for the RMS output metric range from 0.28 to 0.92 (Table 1). For the power integral metric, the ICC ranged from 0.23 to 0.93. Using a windowing approach improved the ICC's to a minimum of 0.76, presumably by limiting variability caused by inconsistent losses in balance. With a windowing approach, the TSECfp test was the least reliable test while the TSEC test was the most reliable (Table 2).

Neurocognitive measures

For the measure of total time taken to complete Trails A the average for trial 1 was 26.6 seconds while the average for trial 2 was 23.5 seconds (ICC 0.80) (Table 3). For Trails B the average for trial 1 was 51.1 seconds and 38.3 seconds for trial 2 (ICC 0.54) (Table 3). For the measure of average speed in Trails A for trial 1 was 4.64cm/second while the average for trial 2 was 5.03 cm/second (ICC 0.81). For Trails B the average for trial 1 was 2.66 cm/second and 3.42 cm/second for trial 2 (ICC 0.81) (Table 3). The reliability is improved (higher ICC's)

Table 1. Calculated variables of interest in the balance testing (RMS and Power Integral) with all 30 seconds of the data included (No Windowing).

	RMS				Power Integral			
	Trial 1 mean (SD)	Trial 2 mean (SD)	Difference	ICC (3,1)	Trial 1 mean (SD)	Trial 2 mean (SD)	Difference	ICC (3,1)
TLEO	0.045 (0.015)	0.044 (0.012)	<0.001	0.83	0.019 (0.007)	0.019 (0.006)	<0.001	0.76
TLEC	0.054 (0.016)	0.055 (0.018)	0.002	0.89	0.023 (0.007)	0.023 (0.008)	<0.001	0.91
TSEO	0.081 (0.038)	0.078 (0.027)	0.003	0.92	0.034 (0.016)	0.034 (0.012)	<0.001	0.93
TSEC	0.241 (0.134)	0.163 (0.072)	0.078	0.28	0.117 (0.034)	0.074 (0.036)	0.043	0.23
TLEOfp	0.079 (0.021)	0.078 (0.019)	0.001	0.87	0.034 (0.086)	0.034 (0.009)	<0.001	0.88
TLECfp	0.185 (0.052)	0.168 (0.055)	0.0108	0.89	0.078 (0.023)	0.070 (0.023)	0.008	0.92
TSEOfp	0.182 (0.073)	0.192 (0.077)	0.009	0.90	0.083 (0.034)	0.086 (0.036)	0.003	0.89
TSECfp	0.824 (0.340)	0.817 (0.328)	0.007	0.59	0.405 (0.184)	0.387 (0.162)	0.018	0.52

Table 2. Calculated variables of interest in the balance testing (RMS and Power Integral) using a windowing approach where equal windows of 5-10 seconds of the 30 second test were compared and the highest removed.

	RMS				Power Integral			
	Trial 1 mean (SD)	Trial 2 mean (SD)	Difference	ICC (3,1)	Trial 1 mean (SD)	Trial 2 mean (SD)	Difference	ICC (3,1)
TLEO	0.041 (0.013)	0.040 (0.011)	<0.001	0.93	0.036 (0.012)	0.036 (0.011)	<0.001	0.94
TLEC	0.047 (0.013)	0.050 (0.017)	0.003	0.81	0.042 (0.013)	0.045 (0.015)	0.003	0.84
TSEO	0.072 (0.034)	0.068 (0.024)	0.003	0.92	0.065 (0.029)	0.064 (0.022)	0<.001	0.93
TSEC	0.143 (0.057)	0.119 (0.046)	0.024	0.96	0.137 (0.057)	0.112 (0.050)	0.026	0.97
TLEOfp	0.069 (0.017)	0.068 (0.018)	0.001	0.86	0.063 (0.015)	0.062 (0.017)	0<.001	0.89
TLECfp	0.152 (0.044)	0.145 (0.052)	0.007	0.81	0.138 (0.041)	0.133 (0.047)	0.006	0.85
TSEOfp	0.153 (0.059)	0.160 (0.064)	0.007	0.87	0.152 (0.060)	0.152 (0.063)	<0.001	0.86
TSECfp	0.481 (0.216)	0.547 (0.235)	0.065	0.76	0.459 (0.209)	0.509 (0.206)	0.051	0.78

Table 3. Metrics of interest from the digital administration of the Trail Making A and B tests.
Neurocognitive Testing

	Time (seconds) (Total Time)				Speed (cm per second) (Average Speed)			
	Trial 1 mean (SD)	Trial 2 mean (SD)	Difference	ICC (3,1)	Trial 1 mean (SD)	Trial 2 mean (SD)	Difference	ICC (3,1)
Trails A	26.6 (6.12)	23.5 (8.8)	2.5	0.80	4.64 (0.68)	5.03 (0.98)	0.39	0.81
Trails B	51.1 (14.5)	38.3(8.7)	12.9	0.54	2.66 (0.50)	3.42 (0.61)	0.76	0.81
	Time (seconds) (Total Time - Slowest segment)				Speed (cm per second) (Average Speed – Slowest)			
	Trial 1 mean (SD)	Trial 2 mean (SD)	Difference	ICC (3,1)	Trial 1 mean (SD)	Trial 2 mean (SD)	Difference	ICC (3,1)
Trails A	21.9 (4.7)	20.1 (5.8)	1.8	0.84	4.69 (0.70)	5.16 (1.11)	0.47	0.82
Trails B	42.9 (11.3)	33.3(7.5)	9.6	0.68	2.74 (0.56)	3.49 (0.62)	0.75	0.82

only minimally by dropping the slowest segment in either metric however total time did improve with an increase in the ICC from 0.54 to 0.68 although this may still not be high enough to allow use for injury assessment.

Dual-task measures

For the balance metrics the RMS had an ICC of 0.92 while the power integral had an ICC of 0.94. For the Trails B tested while balancing the ICC was 0.71 for the total time taken and 0.84 for speed completing the task (Table 4).

Discussion

Overall, the purpose of this study was to evaluate the test-retest reliability of cognitive, balance, and dual-task components of a concussion screening assessment. Although reliability of cognitive and balance components have been evaluated previously, new to this study is the reliability of novel digital cognitive measures, instrumented balance measures, and dual-task measures collected as part of a comprehensive screening.

The reliability of the Trail Making Test has been well studied and data from this study using a digital tablet-based administration of the test shows consistent results to those previously gathered. In a large sample of 155 university students of a similar age to those tested in this study the 50th percentile time taken to complete Trails A and

B was 22 and 47 seconds, respectively [42]. The 30th percentile was 25 and 54 seconds for Trails A and B, making it a closer match to the 26.6 and 51.1 seconds in this study using a tablet-based digital administration. This may suggest the tablet-based version was more difficult, taking on-average, more time than the paper-and-pencil version and administration. Upon retest our sample did demonstrate improvement with, on average, faster times completing both the A and B tests in trial 2. Although various time spans separate the retests in prior studies, coefficients from 0.46 to 0.79 have been reported for the Trail making test [6,25,27]. The metric of total time collected in this study had an ICC of 0.80 and 0.54 for the Trails A and B respectively, suggesting similar reliability in this homogenous group of healthy controls taking a tablet based version to those samples completing the paper-and-pencil versions. Lower reliability for the Trails B test may suggest a more pronounced learning effect in the short retest time frame consistent with the average improvement of 12.8 seconds between trials. Although addressing only one part of the test properties needed for clinical use the test-retest reliability of the digital cognitive measures suggests stable results across testing when using the average speed. Further study using digital administration of the trails test is needed based on this promising pilot study.

Across the instrumented balance tests used in this study reliability was high for many of the tests while also influenced by data analysis choices that remove inter-trial variability. The original BESS included

Table 4. Balance and Cognitive metrics from a dual-task testing including tandem stance and completing the Trails B test.
Dual-Task Testing

	Trial 1 mean (SD)	Trial 2 mean (SD)	Difference	ICC (3,1)	Trial 1 mean (SD)	Trial 2 mean (SD)	Difference	ICC (3,1)
	RMS (with windowing)				Power Integral (with windowing)			
TSEOTrailsB	0.116 (0.052)	0.119 (0.059)	0.003	0.92	0.110 (0.052)	0.111 (0.053)	0.003	0.94
	Time (seconds) (Total Time - Slowest segment)				Speed (cm per second)(Average Speed)			
Trails B	37.7 (9.8)	31.0(8.0)	6.7	0.71	3.20 (0.71)	3.67 (0.77)	0.47	0.84

6 tests, each lasting 20 seconds, and is scored visually based on errors. Reliability for the test has been reported to be 0.60 when examining all six trials collected [15]. However, after removing the double-leg stance trials that have limited variance the reliability improved to 0.71. If scores are averaged across multiple trials for each stance the reliability is reported to be as high as 0.88 [15]. Comparing the same stance and similar conditions (eyes closed) the reliability of the instrumented 30 second tests used in this study are similar with ICCs that range from 0.23 to 0.92 but improve to 0.78 to 0.97 after windowing which may have a similar effect to averaging across multiple trials. Comparing an instrumented balance test using an inertial sensor Mancini, *et al.* reports reliability for 3 trials of quiet standing on a firm surface of between 0.25 and 0.89 depending on the metric used. For the RMS and Power Integral the ICC's were 0.61 and 0.71, respectively. This compares to the TLEO ICC of 0.83 and 0.76 found in this study when using the entire 30-second trial. However, using a windowing method to remove potential losses of balance, similar to that reported by Manicini, *et al.* [11] the ICC's are 0.93 and 0.94 for the RMS and Power integral. Across these studies of balance, reliability seems to increase (>0.70) with efforts to average across more trials or across more windows in the data. This likely improves reliability by limiting the impact of single outliers and makes the variability between retest trials less.

Particularly interesting from the dual-task testing is the high reliability observed in the cognitive and balance components with a concomitant increase in postural sway. There was a 69% increase in the power integral for trial 1 between the single task balance (TSEO) and dual task (Table 1, Figure 1). This increase in postural sway is in contrast to other studies that have shown decreases in postural sway under dual-task testing consistent with a posture-first or constrained-action hypothesis [34,43]. Although the hypothesis would suggest that a shift to an external focus on the cognitive test would increase motor performance the current cognitive test require the user to hold a tablet and input answers [44]. The use of the upper extremities to complete the task may make this cognitive dual-task unique compared to auditory response tests or the difference may be due to the difficulty of the cognitive task as this has also influenced the response [44]. Interesting is the increase in sway and variability (larger standard deviation) between people in the dual-task test compared to the balance alone TSEO test while maintaining the consistency of the response between trials resulting in the high reliability. Further testing to establish sensitivity to injury will provide insight into if these observations from reliability data are helpful for use in patient populations but high variability between people but consistent responses between trials may be good for clinical testing.

The findings from this study reflect high reliability for novel tests that can be used as part of concussion screening tools. This was a pilot study to test and develop these novel digital and instrumented tests as well as explore possible data analysis options. The findings are influenced by the small homogenous sample that was ideal for development of output metrics and to test analysis options such as the windowing used in the balance testing. The sample tested were students from an enrolled graduate health sciences program at the local university. They all willingly volunteered to participate and were
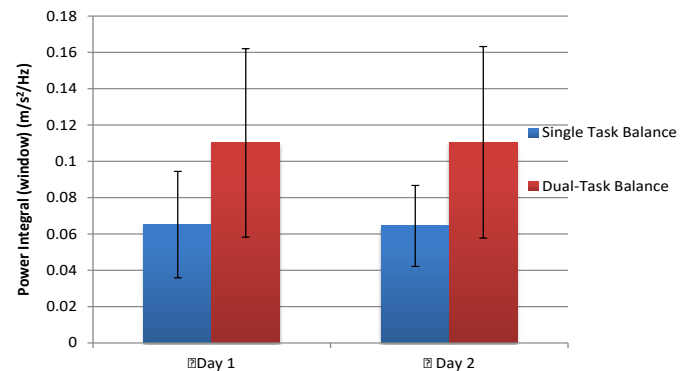

Figure 1. Balance performance across single and dual-task conditions across the two testing sessions. (48 hours apart).

largely all health conscience individuals who regularly exercised and were, at the time of testing, actively engaged in graduate science courses. These factors may influence the average balance and cognitive scores described for comparison to baseline tests of other samples. Baseline testing using balance and cognitive measures has been criticized if effort is not monitored, as this may impact results [45]. The current data was collected with individual proctoring, which may influence the reliability findings. The short time frame for retesting (average 48 hours) was important in this study to limit changes that may occur over time. While many balance reliability studies have included shorter retest intervals (minutes or hours) [11], cognitive retest intervals vary and may be important when considering any learning or acclimation effects that may occur from exposure to the test. Further study of the digital cognitive tests used in this study should include investigation of the effects of multiple exposures across varied time points. Additionally, the balance metrics used in this study were shown to be reliable but are impacted by the processing and data management choices described. This study has described the impact windowing has on reliability while highlighting how the choices for processing data from inertial measurement units impacts reliability. Future studies should continue to explore how various signal-processing choices impact balance scores using inertial measurement units in expanded samples and those with injuries.

Conclusions

The components of a novel objective assessment system for use in a concussion screening show good reliability. A small improvement in cognitive performance was observed at retest, while balance scores remained consistent across tests even under dual-task testing conditions. Greater postural sway was observed under the dual-task that consisted of a tablet-based digital trail-making test while maintaining tandem stance. Reliability coefficients from this study could be used in future studies to establish sensitivity to injury cut-offs

Acknowledgements

Drs. Neville and Rieger have significant financial interests in Quadrant Biosciences Inc, a company that has a commercial interest in the results of this research. This potential conflict of interest has been reviewed and managed by the Upstate Medical University Conflict of Interest Committee. Mr. Brindle is an employee of Quadrant Biosciences Inc.

All authors were fully involved in the study including the formulation, data collection and interpretation, and manuscript preparation. The material in the manuscript has not been submitted for publication elsewhere.

References

1. Hecimovich MD, King D (2016) Prevalence of head injury and medically diagnosed concussion in junior-level community-based Australian Rules Football. *J Paediatr Child Health* 53: 246-251. [Crossref]

2. Marshall SW, Guskiewicz KM, Shankar V, McCrea M, Cantu RC (2015) Epidemiology of sports-related concussion in seven US high school and collegiate sports. *Inj Epidemiol* 2: 13. [Crossref]

3. Broglio SP, Macciocchi SN, Ferrara MS (2007) Sensitivity of the concussion assessment battery. *Neurosurgery* 60: 1050-1057. [Crossref]

4. Pardasaney PK, Latham NK, Jette AM, Wagenaar RC, Ni P, *et al.* (2012) Sensitivity to change and responsiveness of four balance measures for community-dwelling older adults. *Phys Ther* 92: 388-97. [Crossref]

5. Goncalves MM, Pinho MS, Simoes MR (2016) Test-retest reliability analysis of the Cambridge Neuropsychological Automated Tests for the assessment of dementia in older people living in retirement homes. *Appl Neuropsychol Adult* 23: 251-263. [Crossref]

6. Dahmen J, Cook D, Fellows R, Schmitter-Edgecombe M (2016) An analysis of a digital variant of the Trail Making Test using machine learning techniques. *Technol Health Care* 25: 251-264. [Crossref]

7. Fratti S, Bowden SC, Cook MJ (2016) Reliability and validity of the CogState computerized battery in patients with seizure disorders and healthy young adults: comparison with standard neuropsychological tests. *Clin Neuropsychol* 31: 569-586. [Crossref]

8. Woods DL, Wyma JM, Herron TJ, Yund EW (2015) The Effects of Aging, Malingering, and Traumatic Brain Injury on Computerized Trail-Making Test Performance. *PLoS One* 10: e0124345. [Crossref]

9. Furman GR, Lin CC, Bellanca JL, Marchetti GF, Collins MW, et al. (2013) Comparison of the balance accelerometer measure and balance error scoring system in adolescent concussions in sports. *Am J Sports Med* 41: 1404-1410. [Crossref]

10. Howell D, Osternig L, Chou LS (2015) Monitoring recovery of gait balance control following concussion using an accelerometer. *J Biomech* 48: 3364-3368. [Crossref]

11. Mancini M, Salarian A, Carlson-Kuhta P, Zampieri C, King L, et al. (2012) ISway: a sensitive, valid and reliable measure of postural control. *J Neuroeng Rehabil* 9: 59. [Crossref]

12. Seimetz C, Tan D, Katayama R, Lockhart T (2012) A comparison between methods of measuring postrual stability: force plates versus accelerometers. *Biomed Sci Instrum* 48: 386-392. [Crossref]

13. Bell DR, Guskiewicz KM, Clark MA, Padua DA (2011) Systematic review of the balance error scoring system. *Sports Health* 3: 287-295. [Crossref]

14. Finnoff JT, Peterson VJ, Hollman JH, Smith J (2009) Intrarater and interrater reliability of the Balance Error Scoring System (BESS). *PM R* 1: 50-54. [Crossref]

15. Hunt TN, Ferrara MS, Bornstein RA, Baumgartner TA (2009) The reliability of the modified Balance Error Scoring System. *Clin J Sport Med* 19: 471-475. [Crossref]

16. Valovich McLeod TC, Barr WB, McCrea M, Guskiewicz KM (2006) Psychometric and measurement properties of concussion assessment tools in youth sports. *J Athl Train* 41: 399-408. [Crossref]

17. Broglio SP, Sosnoff JJ, Ferrara MS (2009) The relationship of athlete-reported concussion symptoms and objective measures of neurocognitive function and postural control. *Clin J Sport Med* 19: 377-382. [Crossref]

18. Brown HJ, Siegmund GP, Guskiewicz KM, Van Den Doel K, Cretu E, et al. (2014) Development and validation of an objective balance error scoring system. *Med Sci Sports Exerc* 46: 1610-1616. [Crossref]

19. Alberts JL, Thota A, Hirsch J, Ozinga S, Dey T, et al. (2015) Quantification of the Balance Error Scoring System with Mobile Technology. *Med Sci Sports Exerc* 47: 2233-2240. [Crossref]

20. Alsalaheen BA, Haines J, Yorke A, Stockdale K, Broglio SP (2015) Reliability and concurrent validity of instrumented balance error scoring system using a portable force plate system. *Phys Sportsmed* 43: 221-226. [Crossref]

21. Covassin T, Elbin RJ (2010) The cognitive effects and decrements following concussion. *Open Access J Sports Med* 1: 55-61. [Crossref]

22. Reitan RM (1971) Trail making test results for normal and brain-damaged children. *Percept Mot Skills* 33: 575-581. [Crossref]

23. Gaudino EA, Geisler MW, Squires NK (1995) Construct validity in the Trail Making Test: what makes Part B harder? *J Clin Exp Neuropsychol* 17: 529-535. [Crossref]

24. Bornstein RA, Baker GB, Douglass AB (1987) Short-term retest reliability of the Halstead-Reitan Battery in a normal sample. *J Nerv Ment Dis* 175: 229-232. [Crossref]

25. Dikmen SS, Heaton RK, Grant I, Temkin NR (1999) Test-retest reliability and practice effects of expanded Halstead-Reitan Neuropsychological Test Battery. *J Int Neuropsychol Soc* 5: 346-356. [Crossref]

26. Levine AJ, Miller EN, Becker JT, Selnes OA, Cohen BA (2004) Normative data for determining significance of test-retest differences on eight common neuropsychological instruments. *Clin Neuropsychol* 18: 373-384. [Crossref]

27. Matarazzo JD, Wiens AN, Matarazzo RG, Goldstein SG (1974) Psychometric and clinical test-retest reliability of the Halstead impairment index in a sample of healthy, young, normal men. *J Nerv Ment Dis* 158: 37-49. [Crossref]

28. Catena RD, van Donkelaar P, Chou LS (2007) Altered balance control following concussion is better detected with an attention test during gait. Gait Posture 25: 406-411. [Crossref]

29. Fait P, Swaine B, Cantin JF, Leblond J, McFadyen BJ (2013) Altered integrated locomotor and cognitive function in elite athletes 30 days postconcussion: a preliminary study. *J Head Trauma Rehabil* 28: 293-301. [Crossref]

30. Howell DR, Osternig LR, Chou LS (2015) Return to Activity after Concussion Affects Dual-Task Gait Balance Control Recovery. *Med Sci Sport Exerc* 47: 673-680.

31. Ross LM, Register-Mihalik JK, Mihalik JP, McCulloch KL, Prentice WE, et al. (2011) Effects of a single-task versus a dual-task paradigm on cognition and balance in healthy subjects. *J Sport Rehabil* 20: 296-310. [Crossref]

32. Teel EF, Register-Mihalik JK, Troy Blackburn J, Guskiewicz KM (2013) Balance and cognitive performance during a dual-task: preliminary implications for use in concussion assessment. *J Sci Med Sport* 16: 190-194. [Crossref]

33. Yang L, Liao LR, Lam FM, He CQ, Pang MY (2015) Psychometric properties of dual-task balance assessments for older adults: a systematic review. *Maturitas* 80: 359-369. [Crossref]

34. Siu KC, Woollacott MH (2007) Attentional demands of postural control: the ability to selectively allocate information-processing resources. *Gait Posture* 25: 121-126.

35. Sasai S, Boly M, Mensen A, Tononi G (2016) Functional split brain in a driving/listening paradigm. *Proc Nat Acad Sci USA* 113: 14444-14449.

36. Yang L, Lam FM, Liao LR, Huang MZ, He CQ, et al. (2016) Psychometric properties of dual-task balance and walking assessments for individuals with neurological conditions: A systematic review. *Gait Posture* 52: 110-123. [Crossref]

37. Register-Mihalik JK, Littleton AC, Guskiewicz KM (2013) Are divided attention tasks useful in the assessment and management of sport-related concussion? *Neuropsychol Rev* 23: 300-313. [Crossref]

38. Neville C, Ludlow C, Rieger B (2015) Measuring postural stability with an inertial sensor: validity and sensitivity. *Med Devices (Auckl)* 8: 447-455. [Crossref]

39. King LA, Horak FB, Mancini M, Pierce D, Priest KC, et al. (2014) Instrumenting the balance error scoring system for use with patients reporting persistent balance problems after mild traumatic brain injury. *Arch Phys Med Rehabil* 95: 353-359. [Crossref]

40. Moe-Nilssen R (1998) Test-retest reliability of trunk accelerometry during standing and walking. *Arch Phys Med Rehabil* 79: 1377-1385. [Crossref]

41. Whitney SL, Roche JL, Marchetti GF, Lin CC, Steed DP, et al. (2011) A comparison of accelerometry and center of pressure measures during computerized dynamic posturography: a measure of balance. *Gait Posture* 33: 594-599. [Crossref]

42. Tombaugh TN (2004) Trail Making Test A and B: normative data stratified by age and education. *Arch Clin Neuropsychol* 19: 203-214. [Crossref]

43. Broglio SP, Tomporowski PD, Ferrara MS (2005) Balance performance with a cognitive task: a dual-task testing paradigm. *Med Sci Sports Exerc* 37: 689-695. [Crossref]

44. Swan L, Otani H, Loubert PV, Sheffert SM, Dunbar GL (2004) Improving balance by performing a secondary cognitive task. *Br J Psychol* 95: 31-40.

45. Nelson LD, Pfaller AY, Rein LE, McCrea MA (2015) Rates and Predictors of Invalid Baseline Test Performance in High School and Collegiate Athletes for 3 Computerized Neurocognitive Tests: ANAM, Axon Sports, and ImPACT. *Am J Sports Med* 43: 2018-2026. [Crossref]