Test−Retest Reliability of Computerized Neurocognitive Testing in Youth Ice Hockey Players

Melissa N. Womble[1,*], Erin Reynolds[1], Philip Schatz[2], Kishan M. Shah[1,3], Anthony P. Kontos[1]

[1]*Department of Orthopaedic Surgery, University of Pittsburgh Medical Center Sports Medicine Concussion Program, Pittsburgh, PA, USA*
[2]*Department of Psychology, Saint Joseph's University, Pittsburgh, PA, USA*
[3]*Department of Orthopedic Surgery, Indiana University School of Medicine, Indianapolis, IN, USA*

*Corresponding author at: Department of Orthopaedic Surgery, University of Pittsburgh Medical Center Sports Medicine Concussion Program, 3200 South Water Street, Pittsburgh, PA 15203, USA. Tel.: +1-412-432-3681; fax: +1-412-432-3644.
E-mail address: womblemn@upmc.edu; meliss0625@hotmail.com (M.N. Womble).

Abstract

Computerized neurocognitive tests are frequently used to assess pediatric sport-related concussions; however, only 1 study has focused on the test−retest reliability of the Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT) in high school athletes and age influences have largely been ignored. Therefore, the purpose was to investigate the test−retest reliability of ImPACT and underlying age influences in a pediatric population. Two hundred (169 men and 31 women) youth ice hockey players completed ImPACT before/after a 6-month season. Reliability was assessed using Pearson correlation coefficients, intraclass correlation coefficients (ICCs), and regression-based methods (RBz). ICCs for the sample ranged from .48 to .75 (single)/.65 to .86 (average). In general, the older athletes (15–18: Single/Average ICCs = .35−.75/.52−.86) demonstrated greater reliability across composites than the younger athletes (11–14: Single/Average ICCs = .54−.63/.70−.77). Although there was variation in athletes' performance across two test administrations, RBz revealed that only a small percentage of athletes performed beyond 80%, 90%, and 95% confidence intervals. Statistical metrics demonstrated reliability coefficients for ImPACT composites in a pediatric sample similar to previous studies, and also revealed important age-related influences.

Keywords: Head injury; Traumatic brain injury; Practice effects/reliable change; Statistical methods

Introduction

Researchers and clinicians have recommended a multifaceted approach to the assessment of sport-related concussion (SRC). One integral component of this approach is computerized neurocognitive testing (CNT) (Collins, Kontos, Reynolds, Murawski, & Fu, 2014; Elbin, Schatz, Lowder, & Kontos, 2014; Guskiewicz et al., 2004; Reynolds, Collins, Mucha, & Troutman-Ensecki, 2014). The use of CNTs, when compared with typical paper and pencil neuropsychological tests, allows for rapid administration, comparison of data across testing sessions, and minimization of practice effects (Schatz & Browndyke, 2002). These tests typically assess domains including verbal memory, visual memory, processing speed, and reaction time, as well as symptom reports. Despite the benefits of utilizing CNTs, some researchers have questioned their psychometric properties and clinical utility. Specifically, researchers have reviewed the psychometric properties of CNTs, from selected published studies, and suggested that there is limited evidence of their reliability and validity (Resch et al., 2013). However, other researchers have called into question these selected reviews as being biased by the results of specific studies demonstrating low test−retest values (Nakayama, Covassin, Schatz, Nogle, & Kovan, 2014; Schatz & Ferris, 2013) as there have been many published studies reporting on test−retest values for CNTs. Test−retest reliability is particularly important in SRC management, as CNTs are typically administered pre-injury and are then used to track cognitive changes post-injury. Poor test−retest reliability negatively influences reliable change indexes, which may then affect the overall sensitivity of the CNT. To further support the use of CNTs, additional test−retest reliability studies using prospective research designs and evidence-based indicators of neuropsychological change (e.g., reliable change index, regression-based methods [RBz]) are needed (Duff, 2012; Randolph, 2011).

During the past decade, much of the research on test–retest reliability of CNTs has focused on the Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT) test, one of the most commonly used CNTs (Kinnaman, Mannix, Comstock, & Meehan, 2013). Results from these studies have been inconsistent, with low (e.g., intraclass correlation coefficients [ICCs] = .23 for verbal memory) to high (e.g., ICC = .85 for visual motor speed) test–retest values (Bruce, Echemendia, Meeuwisse, Comper, & Sisco, 2014). Thus, researchers have suggested that ImPACT may have limited clinical utility due to these inconsistent psychometric properties. However, the influence of other factors (i.e., time interval, age, sex, and effort) on test–retest reliability has largely been ignored. In fact, previous studies have examined CNTs using varied test–retest intervals (i.e., 7 days [Resch et al., 2013] to 2 years [Schatz, 2010]) and populations including student volunteers (Broglio, Ferrara, Macciocchi, Baumgartner, & Elliott, 2007; Nakayama et al., 2014; Resch et al., 2013; Schatz & Ferris, 2013), military service members (Cole et al., 2013), high school athletes (Elbin, Schatz, & Covassin, 2011), college athletes (Schatz, 2010), and professional hockey players (Bruce et al., 2014). Moreover, despite the widespread use of CNTs in younger children (i.e., elementary to high school aged athletes/students), little is known about the reliability of these tests in this age group. Finally, the criteria used to assess reliability (i.e., Single/Average ICC measures) have varied.

Therefore, the primary purpose of the current study was to explore the test–retest reliability of a commonly used CNT (i.e., ImPACT) in a sample of youth (11–18 years) ice hockey players over a 6-month time interval. We hypothesized that the ImPACT composite measures (verbal and visual memory, visual motor speed, and reaction time) and symptom scores would reflect ICC values similar to those from studies utilizing a 45-day to 1-year test–retest interval (Bruce et al., 2014; Elbin et al., 2011; Nakayama et al., 2014; Resch et al., 2013). Additionally, per previous research (Resch et al., 2013; Schatz, 2010; Schatz & Ferris, 2013), we hypothesized that the visual motor speed and reaction time composite measures would demonstrate greater test–retest values than the verbal and visual memory composite measures. The secondary purpose of this study was to explore the influence of age on the resultant test–retest values. We hypothesized that the 11- to 14-year olds would demonstrate lower reliability values (i.e., ICC, Pearson correlations) and greater variability (i.e., neuropsychological change as determined by RBz) than the 15- to 18-year olds for the composite measures.

Methods

Participants

Participants were recruited from a multisite study of concussions in youth ice hockey in the mid-Atlantic and southeastern regions of the United States during the 2012–2013 ($N = 165$) and 2013–2014 ($N = 236$) seasons. A total of 204/401 (51%) players (173 men and 31 women) with a mean age of 14.64 ± 1.89 years completed baseline and post-season testing. Participants were included if they met the following criteria: (1) completed baseline and post-season testing, (2) no concussion during the study period, and (3) 11–18 years old. Exclusion criteria included a history of one or more of the following: moderate and/or severe traumatic brain injury, brain surgery, neurologic disorder, or substance abuse. Athletes with a history of prior concussion (i.e., less than three diagnosed concussions), attention-deficit/hyperactivity disorder (ADHD), learning disorder (LD), and treatment for migraines were included, consistent with population prevalence estimates. Four athletes with invalid scores per the test manufacturer's recommendations (i.e., excluded solely based on their composite Impulse Control ≥ 30) during baseline or post-season testing were also excluded. The resulting sample included 200 (50% of original sample) players (169 men and 31 women).

Procedure

This study was approved under an expedited protocol by the university's institutional review board. All participants and their parents signed written assent (child) and consent (parent) forms prior to the study. Testing was conducted using the online version of ImPACT on desktop computers with a mouse, in groups of 10 or fewer separated by one or more computer stations. One or more research assistants supervised each testing session. The testing protocol was conducted in the same location and under the same conditions prior to and after the 6-month season. Total test time ranged from 25 to 40 min per session. Alternate forms of the ImPACT test were used at baseline and after the 6-month season to minimize practice effects.

Measures

The ImPACT is a computer-based neurocognitive test battery composed of six subtests including: (1) Word Discrimination (verbal recognition memory immediately/after a delayed period), (2) Design Memory (visual recognition memory immediately/after a delayed period), (3) X's and O's (visual working memory, visual processing speed, and reaction time), (4) Symbol Matching (visual processing speed and memory), (5) Color Match (reaction time), and (6) Three Letter Memory (working memory and visual processing speed).

The ImPACT subtests yield four composite scores including verbal and visual memory (% correct), visual motor speed (measure of processing speed; #; higher = better performance), and reaction time (response time to stimuli, measured in 1/100th seconds; lower = better performance). All composites are expressed in percentiles and are normed according to age groups. The Post-Concussion Symptom Scale is a computerized self-report inventory embedded into the ImPACT test that includes 22 items representing somatic (e.g., headache and dizziness), cognitive (e.g., difficulty concentrating and memory problems), affective (e.g., depression and anxiety), and sleep-related symptoms. Participants rate each symptom on a 7-point Likert scale from 0 (none) to 6 (severe). The individual symptom ratings are totaled to provide a total symptom score.

Data Analysis

To evaluate mean differences for the ImPACT composites across the baseline and post-season assessments, paired-samples *t*-tests were conducted. When completing analyses for the age cohorts, the data were split by the younger (11–14 year olds) and older (15–18 year olds) age cohorts prior to conducting the *t*-tests. Pearson correlation coefficients (*r*) were calculated as a measure of test–retest reliability. To assess the significance of the difference between Pearson correlation coefficients, Fisher *r*-to-*z* transformations were conducted. ICCs using a two-way mixed model with the consistency type were also calculated as another indicator of test–retest reliability (Schatz, 2010). Both single and average measure ICCs were reported to provide comparable outcomes in that results from the current study can be compared to multiple papers and not just papers that reported single or average measure ICCs. Consistent with previous studies, a 0.70 reliability value was used to denote adequate reliability (Baumgartner & Chung, 2001; Bruce et al., 2014; Slick, 2006). However, for further clarification of resulting reliability values, the Slick (2006) classification system was utilized: ≥ 0.90 = very high, 0.80–0.89 = high, 0.70–0.79 = adequate, 0.60–0.69 = marginal, and <0.60 = low. Finally, for the RBz, ImPACT scores from the pre-season assessment were placed into a regression analysis, using post-season scores as the dependent variable, with the resulting equation providing an adjustment for the effect of initial performance level, as well as controlling for any regression to the mean (McCrea et al., 2005). Using this technique, regression equations were built to predict each participant's post-season level of performance based on initial testing performance (Crawford & Garthwaite, 2006). Deviation from the expected (predicted) scores was documented in terms of falling within the 80%, 90%, and 95% confidence intervals (CIs), which correspond to *z*-scores of 1.28, 1.64, and 1.96. RBz for multiple CIs were provided to facilitate comparison with previous studies (Elbin et al., 2011; Nakayama et al., 2014; Schatz, 2010). All analyses were conducted on SPSS version 22.

Results

Demographics

Two hundred youth ice hockey players completed baseline and post-season testing (Table 1). The 11- to 14- and 15- to 18-year-old samples included 91 athletes (19 women and 72 men) and 109 athletes (12 women and 97 men), respectively. Forty-three (21.5%) athletes reported at least one previous concussion, which included 36 men (21.3%) and 7 women (22.6%). χ^2 analyses did not support any differences on reported medical history (i.e., history of prior concussion, ADHD/LD, and treatment for migraine) for the age cohorts.

Results for the Total Sample

Results from paired *t*-tests supported small but significant improvement for the total sample across baseline and post-season testing for the visual motor speed, reaction time, and total symptom scores (Table 2). Pearson correlations between pre- and post-

Table 1. Demographic information for the entire sample, age samples, and sex samples

		Total sample	Age sample	
			11–14	15–18
Number	*N*	200	91	109
Age	*M (SD)*	14.62 (1.9)	12.87 (1.0)	16.07 (1.0)
Sex	Men, *n*	169 (84.5%)	72 (79.1%)	97 (89.0%)
	Women, *n*	31 (15.5%)	19 (20.9%)	12 (11.0%)
Conc. Hx	*n* (%)	43 (21.5%)	15 (16.5%)	28 (25.7%)

Note: Conc. Hx = concussion history.

season scores ranged from .49 (verbal memory; low) to .75 (visual motor speed; adequate; Table 2). Using a Fisher r-to-z transformation, significant differences were supported for the test–retest values of the total sample between (1) verbal memory and visual motor speed ($z = -4.34$, $p < .001$) and (2) visual memory and visual motor speed ($z = -2.93$, $p = .002$). No other significant differences were supported between composite measures for the total sample. Results from the ICCs indicated that visual motor speed (Single/Average ICC = .75/.86; adequate/high) and verbal memory (Single/Average ICC = .48/.65; low/marginal) demonstrated the largest and smallest ICCs, respectively (Table 2). Results from RBz indicated that 82.5%–88.0% of post-season composite scores and 85% of post-season symptom scores fell within an 80% CI. For the 90% CI, 87.0%–92.5% of post-season composite scores and 88.0% of post-season symptom scores fell within a 90% CI. For the 95% CI, 92.5%–96.5% of post-season composite scores and 92.5% of post-season symptom scores fell within a 95% CI (Table 3).

Comparison of Age Cohorts

Results of paired t-tests supported significant differences across baseline and post-season testing for the younger and older samples on visual motor speed, younger sample for reaction time, and older sample for the total symptom score (Table 4). Pearson correlations for the 11- to 14- and 15- to 18-year-old samples ranged from .54 (visual memory/reaction time; low) to .69 (total symptom score; marginal) and .35 (verbal memory; low) to .75 (visual motor speed; adequate), respectively (Table 4). Using a Fisher r-to-z transformation, significant differences were supported for the test–retest values of the 15- to 18-year-old sample between verbal memory and visual motor speed ($z = -4.4$, $p < .001$). No other significant differences were supported between composite measures for the age cohorts, including those comparing the two age cohorts. Results from the ICCs indicated that visual motor speed (11–14 Single/Average ICC = .63/.77; marginal/adequate; 15–18 Single/Average ICC = .75/.86; adequate/high) demonstrated the largest ICC value (Table 4). Between age cohorts, the older sample demonstrated greater reliability values across the composite measures with the exception of verbal memory and total symptom score. Results from RBz indicated that 82.4%–86.8% (11–14) and 80.7%–87.2% (15–18) of post-season composite scores and 84.6 (11–14 and 15–18) of post-season symptom scores fell within an 80% CI for the age cohorts, respectively. For the 90% CI, 85.7%–95.6% (11–14) and 88.1%–92.7% (15–18) of post-season composite scores and 89.0% (11–14) and 88.1% (15–18) of post-season symptom scores fell within a 90% CI for the age cohorts, respectively. For the 95% CI, 93.4%–96.7% (11–14) and 90.8%–95.4% (15–18) of post-season composite scores and 91.2% (11–14) and 91.7% (15–18) of post-season symptom scores fell within a 95% CI for the age cohorts, respectively (Table 5).

Discussion

The current study was the first to examine the test–retest reliability of a commonly used assessment for SRC (i.e., ImPACT) in a sample of youth ice hockey players over a 6-month season. As hypothesized, results supported ICC values similar to those previously

Table 2. ImPACT test–retest reliability values for the entire sample

Composite measure	Time 1 testing		Time 2 testing		t-Value	p	Single ICC	Average ICC	r
	Mean$_1$	SD$_1$	Mean$_2$	SD$_2$					
Verbal memory	85.18	9.35	85.54	11.3	−0.47	.636	.48	.65	.49
Visual memory	75.93	12.29	77.20	13.32	−1.55	.124	.59	.74	.59
Visual motor	36.04	7.21	38.45	7.41	−6.55	.001	.75	.86	.75
Reaction time	0.61	0.09	0.60	0.09	2.54	.012	.62	.76	.62
Total symptoms	1.52	2.86	2.24	4.37	−3.21	.002	.63	.77	.69

Note: p = significance value for the t-test; ICC = intraclass correlation coefficient ; r = Pearson product moment correlation coefficient.

Table 3. Rates of improvement and decline using the RBz for the entire sample

	80% CI			90% CI			95% CI		
	Improvement	Decline	Total	Improvement	Decline	Total	Improvement	Decline	Total
Verbal memory	6.5	11.0	17.5	1.0	6.5	7.5	1.0	2.5	3.5
Visual memory	6.5	11.0	17.5	4.0	5.5	9.5	2.0	3.5	5.5
Visual motor speed	5.5	6.5	12.0	2.5	6.0	8.5	1.0	4.0	5.0
Reaction time	6.0	11.5	17.5	4.0	9.0	13.0	2.0	5.5	7.5
Symptom score	5.0	10.0	15.0	4.0	8.0	12.0	2.0	5.5	7.5

Note: Values represent the percentage scoring beyond cut-off values set at the 80% (1.28), 90% (1.64), and 95% (1.96) CIs.

Table 4. ImPACT test–retest reliability values for 11- to 14-year-old ($N = 91$) and 15- to 18-year-old ($N = 109$) youth hockey players

Composite measure	Age range	Time 1 testing		Time 2 testing		t-Value	p	Single ICC	Average ICC	r
		Mean$_1$	SD$_1$	Mean$_2$	SD$_2$					
Verbal memory	11–14	83.26	9.61	82.91	12.20	0.32	.747	.55	.71	.57
	15–18	86.78	8.85	87.72	10.03	−0.91	.363	.35	.52	.35
Visual memory	11–14	73.60	11.35	75.34	12.15	−1.47	.146	.54	.70	.54
	15–18	77.86	12.75	78.74	14.09	−0.77	.443	.61	.75	.61
Visual motor	11–14	32.33	6.44	35.38	6.83	−5.10	.001	.63	.77	.63
	15–18	39.15	6.31	41.01	6.91	−4.16	.001	.75	.86	.75
Reaction time	11–14	0.65	0.10	0.63	0.10	2.03	.045	.54	.70	.54
	15–18	0.58	0.07	0.57	0.08	1.52	.130	.61	.76	.61
Total symptoms	11–14	1.30	2.42	1.82	3.74	−1.86	.067	.63	.77	.69
	15–18	1.70	3.17	2.58	4.82	−2.61	.010	.63	.77	.68

Note: p = significance value for the t-test; ICC = intraclass correlation coefficient; r = Pearson product moment correlation coefficient.

Table 5. Rates of improvement and decline using the RBz for 11- to 14-year-old ($N = 91$) and 15- to 18-year-old ($N = 109$) youth hockey players

	Age	80% CI			90% CI			95% CI		
		Improvement	Decline	Total	Improvement	Decline	Total	Improvement	Decline	Total
Verbal memory	11–14	5.5	8.8	14.3	2.2	2.2	4.4	1.1	2.2	3.3
	15–18	1.8	11.0	12.8	0.9	8.3	9.2	0.0	4.6	4.6
Visual memory	11–14	4.4	9.9	14.3	1.1	5.5	6.6	1.1	3.3	4.4
	15–18	7.3	11.9	19.3	3.7	5.5	9.2	1.8	2.8	4.6
Visual motor speed	11–14	6.6	6.6	13.2	3.3	6.6	9.9	2.2	4.4	6.6
	15–18	5.5	10.1	15.6	1.8	5.5	7.3	1.8	4.6	6.4
Reaction time	11–14	5.5	12.1	17.6	4.4	9.9	14.3	0.0	4.4	4.4
	15–18	6.4	9.2	15.6	3.7	8.3	11.9	0.9	8.3	9.2
Symptom score	11–14	6.6	8.8	15.4	4.4	6.6	11.0	3.3	5.5	8.8
	15–18	5.5	10.1	15.6	3.7	8.3	11.9	2.8	5.5	8.3

Note: Values represent the percentage scoring beyond cut-off values set at the 80% (1.28), 90% (1.64), and 95% (1.96) CIs.

reported in studies utilizing 45-day to 1-year test–retest intervals (Bruce et al., 2014; Elbin et al., 2011; Nakayama et al., 2014; Resch et al., 2013). Additionally, Pearson correlation coefficients were similar to those previously reported for a high school sample (Elbin et al., 2011) and English-speaking professional hockey players (Bruce et al., 2014). The average measure ICCs, with the exception of verbal memory, met or exceeded the 0.70 level that has been considered an adequate level of test–retest reliability (Baumgartner & Chung, 2001; Bruce et al., 2014; Slick, 2006). Consistent with previously published studies, the visual motor speed and reaction time composite measures demonstrated greater reliability coefficients than the verbal and visual memory composite measures. Also, single measure ICCs revealed lower reliability coefficients (low to adequate range) than the average measure ICCs (marginal to high) (Slick, 2006). With regard to age, as expected, the 15- to 18-year olds demonstrated greater reliability coefficients for the composite measures across the season than the 11- to 14-year olds with the exception of verbal memory. Similarly, regression-based findings indicated that the younger athletes aged 11–14 demonstrated greater variability for the timed tasks (i.e., reaction time and processing speed measures) across the two test administrations than the older 15- to 18-year olds.

Overall, findings from the current study revealed CNT performance across a season for youth ice hockey players, which were consistent with previously published studies (Broglio et al., 2007; Cole et al., 2013; Elbin et al., 2011; Resch et al., 2013; Schatz, 2010; Schatz & Ferris, 2013). Researchers have commonly discussed inconsistencies among reported test–retest values for the ImPACT test without addressing the current interpretative limitations (i.e., limited age of studied populations) and the influence of other factors on these findings (Randolph, 2011; Randolph, McCrea, & Barr, 2005). We compared our findings with previously published test–retest findings for ImPACT to provide context for the differences (i.e., test–retest intervals, sample populations, sexes, effort criteria, and ICC measures) across studies (Table 6) (Bruce et al., 2014). As is evident from this comparison, a key methodological difference in the studies is that researchers have alternated between reporting single and average measure ICCs (Bruce et al., 2014). This discrepancy is likely one contributing factor to inconsistencies in reported reliability values, as single measure ICCs for our data were .11 to .17 lower than average measure ICCs. Questions have arisen regarding the appropriateness of using average measure ICCs for these types of analyses, though formal answers regarding their use have not been well

Table 6. Intraclass correlation coefficients for ImPACT reported in the literature and the current study[a]

	Resch and colleagues (2013)	Cole and colleagues (2013)	Schatz and Ferris (2013)	Resch and colleagues (2013)	Broglio and colleagues (2007)	Nakayama and colleagues (2014)	Present study	Elbin and colleagues (2011)	Bruce and colleagues (2014)	Schatz (2010)
Time frame	7 days	30 days	30 days	45 days	45 days	45 days	6 months	1 year	1 year	2 years
Population description	Student volunteer	Service member	Student volunteer	Student volunteer	Student volunteer	Student volunteer	Youth hockey athlete	HS athlete	Pro hockey player	College athlete
Sample size, n	46	44	25	45	73	85	200	369	119	95
Mean age	22.4 (1.9)	N/A	N/A	20.9 (1.7)	21.4 (2.8)	N/A	14.6 (1.9)	14.8 (0.9)	25.5 (4.6)	18.8 (0.6)
Men/women	25/21	N/A	6/19	17/28	N/A	51/34	169/31	168/201	119	51/44
PVT	ImPACT	ImPACT	ImPACT	WMT	ImPACT	ImPACT	ImPACT	ImPACT	ImPACT	ImPACT
ICC type	N/A	N/A	Avg ICC	N/A	Single ICC	Avg ICC	Single/Avg ICC	Avg ICC	Single ICC	Avg ICC
Verbal memory	0.56	0.60	0.79	0.45	0.23	0.76	0.48/0.65	0.62	0.38	0.46
Visual memory	0.26	0.50	0.60	0.52	0.32	0.72	0.59/0.74	0.70	0.54	0.65
Motor speed	0.78	0.83	0.88	0.76	0.38	0.87	0.75/0.86	0.85	0.52	0.74
Reaction time	0.84	0.53	0.77	0.57	0.39	0.67	0.62/0.76	0.76	0.74	0.68
Symptom scale	–	–	0.81	–	–	–	0.63/0.77	0.57	0.46	0.43

Note: [a]Table adapted from Bruce and colleagues (2014). N/A = not available; PVT = performance validity test; Verbal memory = ImPACT memory composite Verbal; Visual memory = ImPACT memory composite visual; Motor speed = ImPACT visual motor speed composite; Reaction time = ImPACT reaction time composite; Symptom scale = Impact total symptom score.

established. As such, until further research is conducted, an argument can be made for both single and average measure ICCs in the current context of CNT reliability. However, because Pearson values are almost always used for test–retest studies, single ICCs may better reflect test sensitivity in practice. With that said, we suggest that researchers employ multiple statistical approaches when examining the reliability of CNTs to provide a less biased and more comprehensive analysis to help further research and statistical practices in this regard. To that end, we utilized paired *t*-tests, Pearson correlation coefficients, ICCs, and RBz to further understand differences. We believe that RBz were particularly helpful in understanding the influence of age on resultant baseline scores while accounting for other variables (i.e., practice effects, Time 1 scores, variability in Time 2 scores) (Duff, 2012). Specifically, RBz were helpful in identifying trends (i.e., % decline, % improve) that can be helpful in understanding the performances of specific age cohorts.

Given the widespread use of CNTs in pediatric populations and the current findings, clinicians and researchers should consider the effects of age and developmental changes on test–retest reliability. Findings from the current study indicate that the 11- to 14-year-old sample demonstrated lower reliability values (Single/Average ICCs = Low to Marginal/Adequate) than the 15- to 18-year-old sample (Single/Average ICCs = Low to Adequate/ Low to High) on all composite measures except for verbal memory. Likewise, the 11- to 14-year-old sample demonstrated greater variability for the timed tasks (i.e., reaction time and processing speed measures) across the hockey season than the 15- to 18-year-old sample. These findings are consistent with previous research demonstrating neurodevelopmental changes that transpire secondary to neuronal maturation during late childhood and early adolescence. These neurodevelopmental changes typically result in significant changes in reaction time, processing speed, and executive functioning (Casey, Giedd, & Thomas, 2000; Fry & Hale, 2000; Nagy, Westerberg, & Klingberg, 2004; Stevens, Skudlarski, Pearlson, & Calhoun, 2009), which further support our findings.

Strengths and Limitations

Strengths of the current research include utilization of a large sample size, focus on an at-risk and overlooked pediatric population, inclusion of men and women, prospective design, and use of multiple statistical analyses to assess test–retest reliability. Despite these strengths, this study is not without limitations. First, the current sample was over representative of men, including only a small sample of female athletes. As such, results for the entire sample may not be generalizable to women. Second, we did not employ stand-alone performance validity tests to assess effort among the sample. As such, the current findings may have included some invalid CNTs due to low effort. Finally, although this study examined ice hockey players—a group that had been absent from previous studies—the narrow focus of this sample may not be generalizable to children from other sports.

Clinical Implications and Future Research

Results from the current study are important in terms of better understanding of the reliability of a CNT that is often employed in pediatric populations to help manage SRC. Clinically, results from this study highlighted the relevance of age to CNT performance. Specifically, clinicians need to consider the greater variability in CNT performance among younger children when interpreting changes in these SRC-related parameters. To further understand the differences reported in the current study, researchers should focus on replicating this study in a sample that includes a larger number of women and non-athletic controls, as well as with longer test–retest time frames. In addition, researchers should assess developmental and maturational changes directly rather than use an age proxy per the current study. Finally, the influence of motivation on test–retest reliability should be examined through the use of external performance validity measures.

Conclusion

In summary, youth ice hockey players assessed with the ImPACT test prior to and after a 6-month season demonstrated test–retest values, utilizing multiple statistical approaches, similar to those previously reported in studies utilizing 45-day to 1-year test–retest intervals and different aged samples. Additional analyses suggested that age could be one factor contributing to discrepancies in the reliability of CNTs previously reported by researchers. As such, clinicians should account for age when employing CNT as part of a comprehensive assessment approach to SRC in a pediatric population. This study also demonstrated the importance of using multiple statistical methods to further understand test–retest values and potential confounding factors.

Funding

Conflict of Interest

PS and ER have served as independent consultants to ImPACT applications, Inc. PS's role has included analyzing normative data and establishing age and gender-based norms. ER's role has included teaching and training on ImPACT technology and best clinical practices. However, ImPACT Applications, Inc. had no role in the conceptualization of the study, the collection or analysis of data, the writing of the article, or the decision to submit it for publication. None of the other authors have financial disclosures or conflicts of interest to declare.

References

Baumgartner, T., & Chung, H. (2001). Confidence limits for intraclass reliability coefficients. *Measurement in Physical Education and Exercise Science*, *5*, 179–188.

Broglio, S. P., Ferrara, M. S., Macciocchi, S. N., Baumgartner, T. A., & Elliott, R. (2007). Test-retest reliability of computerized concussion assessment programs. *Journal of Athletic Training*, *42*, 509–514.

Bruce, J., Echemendia, R., Meeuwisse, W., Comper, P., & Sisco, A. (2014). 1 year test-retest reliability of ImPACT in professional ice hockey players. *Clinical Neuropsychology*, *28*, 14–25.

Casey, B. J., Giedd, J. N., & Thomas, K. M. (2000). Structural and functional brain development and its relation to cognitive development. *Biological Psychology*, *54*, 241–257.

Cole, W. R., Arrieux, J. P., Schwab, K., Ivins, B. J., Qashu, F. M., & Lewis, S. C. (2013). Test-retest reliability of four computerized neurocognitive assessment tools in an active duty military population. *Archives of Clinical Neuropsychology*, *28*, 732–742.

Collins, M. W., Kontos, A. P., Reynolds, E., Murawski, C. D., & Fu, F. H. (2014). A comprehensive, targeted approach to the clinical care of athletes following sport-related concussion. *Knee Surgery, Sports Traumatology, Arthroscopy*, *22*, 235–246.

Crawford, J. R., & Garthwaite, P. H. (2006). Comparing patients' predicted test scores from a regression equation with their obtained scores: A significance test and point estimate of abnormality with accompanying confidence limits. *Neuropsychology*, *20*, 259–271.

Duff, K. (2012). Evidence-based indicators of neuropsychological change in the individual patient: Relevant concepts and methods. *Archives of Clinical Neuropsychology*, *27*, 248–261.

Elbin, R. J., Schatz, P., & Covassin, T. (2011). One-year test-retest reliability of the online version of ImPACT in high school athletes. *The American Journal of Sports Medicine*, *39*, 2319–2324.

Elbin, R. J., Schatz, P., Lowder, H. B., & Kontos, A. P. (2014). An empirical review of treatment and rehabilitation approaches used in the acute, sub-acute, and chronic phases of recovery following sports-related concussion. *Current Treatment Options in Neurology*, *16*, 320.

Fry, A. F., & Hale, S. (2000). Relationships among processing speed, working memory, and fluid intelligence in children. *Biological Psychology*, *54*, 1–34.

Guskiewicz, K. M., Bruce, S. L., Cantu, R. C., Ferrara, M. S., Kelly, J. P., McCrea, M., et al. (2004). National Athletic Trainers' Association position statement: Management of sport-related concussion. *Journal of Athletic Training*, *39*, 280–297.

Kinnaman, K. A., Mannix, R. C., Comstock, R. D., & Meehan, W. P., III (2013). Management strategies and medication use for treating paediatric patients with concussions. *Acta Paediatrica*, *102*, e424–e428.

McCrea, M., Barr, W. B., Guskiewicz, K., Randolph, C., Marshall, S. W., Cantu, R., et al. (2005). Standard regression-based methods for measuring recovery after sport-related concussion. *Journal of the International Neuropsychological Society*, *11*, 58–69.

Nagy, Z., Westerberg, H., & Klingberg, T. (2004). Maturation of white matter is associated with the development of cognitive functions during childhood. *Journal of Cognitive Neuroscience*, *16*, 1227–1233.

Nakayama, Y., Covassin, T., Schatz, P., Nogle, S., & Kovan, J. (2014). Examination of the test-retest reliability of a computerized neurocognitive test battery. *The American Journal of Sports Medicine*, *42*, 2000–2005.

Randolph, C. (2011). Baseline neuropsychological testing in managing sport-related concussion: Does it modify risk? *Current Sports Medicine Reports*, *10*, 21–26.

Randolph, C., McCrea, M., & Barr, W. B. (2005). Is neuropsychological testing useful in the management of sport-related concussion? *Journal of Athletic Training*, *40*, 139–152.

Resch, J., Driscoll, A., McCaffrey, N., Brown, C., Ferrara, M. S., Macciocchi, S., et al. (2013). ImPact test-retest reliability: Reliably unreliable? *Journal of Athletic Training*, *48*, 506–511.

Reynolds, E., Collins, M. W., Mucha, A., & Troutman-Ensecki, C. (2014). Establishing a clinical service for the management of sports-related concussions. *Neurosurgery*, *75*, S71–S81.

Schatz, P. (2010). Long-term test-retest reliability of baseline cognitive assessments using ImPACT. *The American Journal of Sports Medicine*, *38*, 47–53.

Schatz, P., & Browndyke, J. (2002). Applications of computer-based neuropsychological assessment. *Journal of Head Trauma Rehabilitation*, *17*, 395–410.

Schatz, P., & Ferris, C. S. (2013). One-month test-retest reliability of the ImPACT test battery. *Archives of Clinical Neuropsychology*, *28*, 499–504.

Slick, D. J. (2006). Psychometrics in neuropsychological assessment. In E. Strauss, & E. Sherman (Eds.), *A compendium of neuropsychological tests: Administration, norms, and commentary* (pp. 3–43). New York, NY: Oxford University Press.

Stevens, M. C., Skudlarski, P., Pearlson, G. D., & Calhoun, V. D. (2009). Age-related cognitive gains are mediated by the effects of white matter development on brain network integration. *Neuroimage*, *48*, 738–746.