REVIEW

Psychometric Properties of Computerized Cognitive Tools and Standard Neuropsychological Tests Used to Assess Sport Concussion: A Systematic Review

Kristin Wilmoth[1] · Benjamin L. Brett[2] · Natalie A. Emmert[3] · Carolyn M. Cook[4] · Jeffrey Schaffert[5] · Todd Caze II[5] · Thomas Kotsonis[4] · Margaret Cusick[4] · Gary Solomon[6] · Jacob E. Resch[7] · C. Munro Cullum[8] · Lindsay D. Nelson[2] · Michael McCrea[2]

Abstract
Athletic programs are more frequently turning to computerized cognitive tools in order to increase efficiencies in concussion assessment. However, assessment using a traditional neuropsychological test battery may provide a more comprehensive and individualized evaluation. Our goal was to inform sport clinicians of the best practices for concussion assessment through a systematic literature review describing the psychometric properties of standard neuropsychological tests and computerized tools. We conducted our search in relevant databases including Ovid Medline, Web of Science, PsycINFO, and Scopus. Journal articles were included if they evaluated psychometric properties (e.g., reliability, sensitivity) of a cognitive assessment within pure athlete samples (up to 30 days post-injury). Searches yielded 4,758 unique results. Ultimately, 103 articles met inclusion criteria, all of which focused on adolescent or young adult participants. Test–retest reliability estimates ranged from .14 to .93 for computerized tools and .02 to .95 for standard neuropsychological tests, with strongest correlations on processing speed tasks for both modalities, although processing speed tasks were most susceptible to practice effects. Reliability was improved with a 2-factor model (processing speed and memory) and by aggregating multiple baseline exams, yet remained below acceptable limits for some studies. Sensitivity to decreased cognitive performance within 72 h of injury ranged from 45%–93% for computerized tools and 18%–80% for standard neuropsychological test batteries. The method for classifying cognitive decline (normative comparison, reliable change indices, regression-based methods) affected sensitivity estimates. Combining computerized tools and standard neuropsychological tests with the strongest psychometric performance provides the greatest value in clinical assessment. To this end, future studies should evaluate the efficacy of hybrid test batteries comprised of top-performing measures from both modalities.

Keywords Sport concussion · Computerized neurocognitive tests · Neuropsychological assessment · Reliability · Sensitivity · Validity

✉ Kristin Wilmoth
Kristin.Wilmoth@UTSouthwestern.edu

1 Departments of Psychiatry and Physical Medicine & Rehabilitation, UT Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75390-9055, USA

2 Departments of Neurosurgery and Neurology, Medical College of Wisconsin, Milwaukee, WI, USA

3 Department of Neurology, University at Buffalo, State University of New York, Buffalo, NY, USA

4 Department of Neurology, Medical College of Wisconsin, Milwaukee, WI, USA

5 Department of Psychiatry, UT Southwestern Medical Center, Dallas, TX, USA

6 Player Health and Safety Department, National Football League and Department of Neurosurgery, Vanderbilt University School of Medicine, Nashville, TN, USA

7 Department of Kinesiology, University of Virginia, Charlottesville, VA, USA

8 Departments of Psychiatry, Neurology, and Neurological Surgery, UT Southwestern Medical Center, Dallas, TX, USA

Sport concussion is a significant public health issue, with estimated rates ranging from 0.1 to 32.1 per 1,000 athletic exposures, depending on the sport, level of play, and injury surveillance methods employed (Clay et al., 2013; Kerr et al., 2018). Consensus guidelines recommend assessment of cognitive function following suspected concussion in athletes (McCrory et al., 2017). Two testing modalities currently exist for this purpose, including computerized cognitive tools and conventional tests administered as part of a neuropsychological evaluation.

While traditional "paper-and-pencil" neuropsychological tests are the gold standard for evaluating human cognitive function, the last two decades have seen the development of computerized neurocognitive tests for the evaluation of sport concussion (Meehan et al., 2012). These computerized tools have been widely adopted by athletic programs given practical advantages such as ease of administration and scoring, ability to test athletes in groups rather than individually, and portable test results (Collie et al., 2001). However, these tools are tailored to the areas of cognition most susceptible to concussion (e.g., processing speed) and as such do not evaluate other aspects of cognition including executive functioning, verbal fluency, and free recall memory (Belanger & Vanderploeg, 2005; Broglio & Puetz, 2008; Iverson & Schatz, 2015).

To inform best practices in sport concussion assessment, we sought to update and expand prior reviews evaluating the use of computerized and conventional cognitive testing in concussion assessment. Past literature reviews have identified highly variable psychometric estimates across tests (Farnsworth et al., 2017; Randolph et al., 2005; Resch et al., 2013a, b). Our goal was to utilize systematic review methodology to describe published psychometric data and to formally compare computerized cognitive tools with standard neuropsychological tests that may be used to assess athletes with a suspected concussion.

Methods

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were followed (Moher et al., 2009). We defined the population, intervention, comparison, and outcome (PICO), as recommended by the American Academy of Neurology guidelines for evidence-based medicine (Gronseth et al., 2011, pp. 3–4). Our clinical research question in PICO format was, "In English-speaking athletes, do computerized cognitive tools, compared to standard neuropsychological tests, provide sufficiently reliable and valid measures of cognitive functioning?" Upper thresholds for good reliability and sensitivity of clinical assessments, including subtest scores, were based on previously published

criteria and convention (Blake et al., 2002; Nunnally & Bernstein, 1994; Slick, 2006; Weissberger et al., 2017).

Search Strategy

Searches were developed with the assistance of a medical librarian at the Medical College of Wisconsin. Terms were combined for three main concepts, including concussion, cognitive assessment, and athletes (see Table S1 in supplementary materials). Databases queried included Ovid Medline, Web of Science (Core Collection), PsycINFO, Scopus, Cochrane (Database of Systematic Reviews, Register of Controlled Trials, and Protocols; searched via Wily), Cumulative Index of Nursing and Allied Health Literature (CINAHL; searched via Ebsco), and Education Resources Information Center (ERIC; searched via EBSCO). The initial query of databases was conducted in December 2018 (without restriction of date). Results were updated most recently in August 2021 with a follow-up search.

Study Selection

Studies were included if they evaluated the psychometric properties (e.g., sensitivity, specificity, reliability, convergent or discriminant validity) of one or more cognitive assessments. Strict inclusion criteria involved studies that solely examined athlete participants, although the cognitive assessments need not be designed only for athlete use. We included postconcussive evaluations conducted up to 30 days post injury, although evaluations within 72 h of injury were considered acute (Joseph et al., 2018). Results for non-English tests and studies that solely examined assessments of balance, vestibular function, or oculomotor function were excluded. As only journal articles were considered, conference abstracts, commentaries, letters, and editorials were omitted. References of relevant review articles yielded by the search were examined to identify additional studies for inclusion.

Initial screening was conducted based on titles and abstracts using Rayyan (http://rayyan.qcri.org/). Two authors made independent inclusion decisions, with a third providing an independent rating for tied articles to achieve consensus on inclusion. Articles that passed the initial screening underwent full-text review in a similar manner.

Data Extraction and Analysis

PROSPERO is an international prospective register of systematic reviews. Prior to data extraction, our review protocol was registered with PROSPERO (Record ID 180763; see supplementary materials). Sample characteristics and key findings were extracted from all studies meeting inclusion criteria and information was verified by a second reviewer.

Risk of bias was assessed via the Quality Assessment of Diagnostic Accuracy Studies—Version 2 (QUADAS-2; Whiting et al., 2011). The QUADAS-2 was modified to fit the aims of the review, as has previously been done to assess methodological quality of concussion studies (e.g., McCrea et al., 2017). Four domains assessed risk of bias, including patient selection, use and interpretation of the test, examination of psychometric properties, and patient flow (see supplementary material). All reviewers were authors trained on the tool using beta articles to ensure consistency with ratings and met regularly to refine shared understanding of rating criteria. An overall rating for each included study was generated based on the instrument, which was then confirmed by an independent second rater (with a third rater utilized as needed to break ties and reach consensus). Level of evidence was assessed using a Strength of Recommendation Taxonomy (SORT) scoring system (Ebell et al., 2004).

Results

A total of 10,590 combined search results were returned from Ovid, Web of Science, PsycINFO, Scopus, Cochrane, CINAHL, and ERIC. Deduplication revealed 5,409 unique results, which were screened by two authors (with screening by a third author needed for 355 results to achieve consensus). Figure 1 displays a standard PRISMA flowchart (Page et al., 2021) of inclusion and exclusion during screening and full-text review. Ultimately, 103 articles met inclusion criteria. Sample characteristics of included studies are summarized in Table 1.

The Immediate Post-Concussion Assessment and Cognitive Test (ImPACT; Riverside Publishing) was the most widely evaluated, included in 65 of the 103 studies (63%). Thirteen studies evaluated a hybrid battery, comprised of both computerized cognitive and traditional neuropsychological

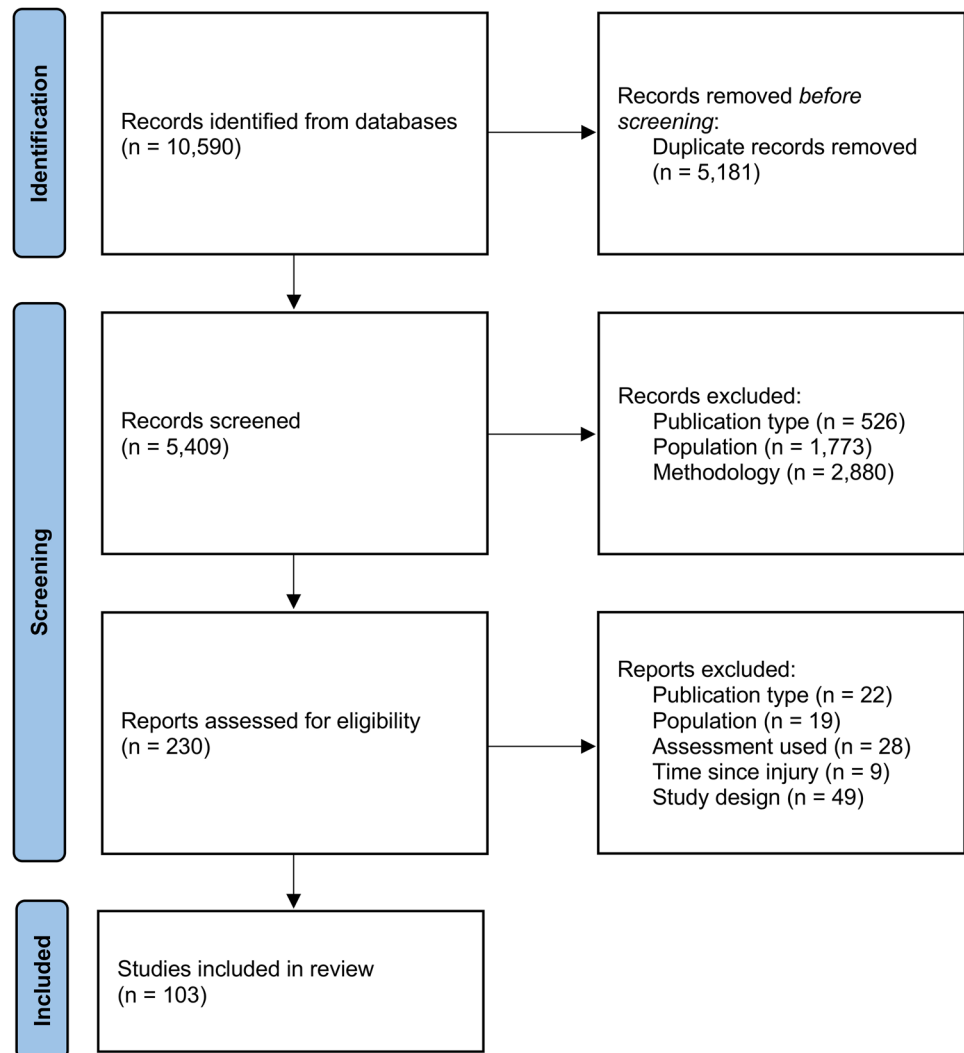**Fig. 1** PRISMA flowchart of inclusions and exclusions

Table 1 Sample characteristics of the 103 studies meeting systematic review inclusion

Study	Age of sample in years, mean (SD or range)	Total sample size	Size of concussed group
Abbassi and Sirmon-Taylor (2017)	16.2 (2.4)	30	30
Barr (2003)	15.9 (1.0)	48	NA
Bernstein et al. (2019)	20.0 (1.7)	54	54
Bialunska and Salvatore (2017)	Concussed: 18.2 (4.3), Control: 20.4 (3.2)	231	88
Brett and Solomon (2017)	15.1 (1.9)	1,146	NA
Brett et al. (2016)	15.1 (1.9)	1,510	NA
Brett et al. (2018)	15.0 (0.7)	978	NA
Broglio et al. (2007a)	19.8 (1.3)	21	21
Broglio et al. (2007b)	NR	75	75
Broglio et al. (2009)	19.7 (1.2)	36	36
Broglio et al. (2018)	19.2 (1.2)	4,874	NA
Bruce et al. (2014)	25.6 (4.9)	305	NA
Bruce et al. (2016)	25.6 (4.4)	115	NA
Bruce et al. (2017)	28.4 (4.2)	97	NA
Collie et al. (2006)	Control: 23.4 (3.6), Symptomatic: 22.3 (3.6), Asymptomatic: 23.3 (3.9)	615	61
Collins et al. (1999)	20.4 (1.7)	393	16
Collins et al. (2006)	NR	2,141	136
Covassin et al. (2007)	NR	79	79
Covassin et al. (2010)	15.8 (1.3)	72	72
Czerniak et al. (2021)	ImPACT Concussed: 19.2 (1.4), Control: 19.1 (1.2); CNS Vital Signs Concussed: 19.0 (1.2), Control: 19.3 (1.1); CogState Concussed: 19.3 (1.3), Control: 18.8 (1.0)	9,449	1,414
Echemendia et al. (2001)	NR	49	29
Echemendia et al. (2012)	19.2 (1.4)	223	223
Echemendia et al. (2016)	21.0 (3.1)	187	NA
Echlin et al. (2012)	NR	45	11
Elbin et al. (2011)	14.8 (0.9)	369	NA
Elbin et al. (2020)	14.3 (0.5)	47	NA
Erlanger et al. (2001)	15–21	26	26
Erlanger et al. (2003a, b)	Concussed: 18.7 (16–21)	823	26
Erlanger et al. (2003a, b)	17.6 (2.2)	1,603	47
Espinoza et al. (2021)	18 (16–23)	123	21
Fazio et al. (2007)	Symptomatic: 16.7 (12–27), Asymptomatic: 16.6 (14–22) Controls: 17.3 (14–22)	192	122
Gardner et al. (2012)	Concussed: 24.2 (4.2), Control: 24.3 (3.6)	87	46
Gerrard et al. (2017)	15.9 (1.3)	4,809	NA
Glutting et al. (2020)	17.0 (0.7)	218	NA
Goh et al. (2021)	Concussed Male: 25.6 (4.2), Female: 24.2 (4.1); Control Male: 23.9 (4.5), Female: 23.6 (4.4)	130	46
Guskiewicz et al. (1997)	18.6 (2.0)	22	11
Guskiewicz et al. (2001)	Concussed: 19.5 (1.3), Control: 20 (2.4)	72	36
Hang et al. (2015)	11–17	109	109
Higgins et al. (2018)	Group 1: 19.2 (1.4), Group 2: 18.9 (1.5)	361	NA
Hinton-Bayre and Geffen (2005)	20.6 (3.6)	112	NA
Hinton-Bayre et al. (1997)	Concussed: 22.1 (4.4), Control: 19.4 (2.1)	54	10
Hinton-Bayre et al. (1999)	Not concussed: 19.6 (3.3), Concussed after baseline measure: 21.1 (3.6)	222	20
Hinton-Bayre (2012)	Concussed: 21.2 (3.5), Control: 19.9 (2.5)	53	27
Hinton-Bayre (2015)	Concussed: 21.2 (3.5), Controls: 19.9 (2.5)	194	27
Houck et al. (2019)	Median 19 (IQR = 18–20)	15,909	NA

Table 1 (continued)

Study	Age of sample in years, mean (SD or range)	Total sample size	Size of concussed group
Houston et al. (2021)	NR	4,875	NA
Iverson et al. (2003)	16.8 (2.4)	41	41
Iverson et al. (2005)	17.1 (1.9)	72	72
Iverson et al. (2006)	16.1 (2.1)	30	30
Lau et al. (2011)	Short recovery: 16.1, Long recovery: 15.9	108	108
Lau et al. (2012)	Short recovery: 16.1 (1.2), Long recovery: 15.9 (1.2)	108	108
Louey et al. (2014)	Normative: 25.4 (4.5), Retest: 25.8 (4.8), Concussed: 22.3 (2.9)	624	29
Lovell and Solomon (2011)	NR	513	NA
Lovell and Solomon (2013)	Flyers: 14.9 (13–18), Base: 15.4 (12–18), Other position: 14.0 (10–18)	138	138
Lovell et al. (2003)	NR	88	64
Lovell et al. (2004)	15.6 (13–18)	43	43
Macciocchi et al. (1996)	18–26	231	183
MacDonald and Duerson (2015)	16.1 (1.0)	117	NA
Maddocks and Saling (1996)	NR	130	10
Maerlender and Molfese (2015)	18.1	68	NA
Maerlender et al. (2010)	19.1 (17–22)	54	NA
Maerlender et al. (2013)	19.1 (17–22)	54	NA
Maerlender et al. (2016)	19.9	231	NA
Maietta et al. (2021)	13–19	36,091	36,091
Makdissi et al. (2001)	Concussed: 20.5 (3.1), Control: 20.3 (4.2)	13	6
Makdissi et al. (2010)	16–35	1,015	78
Mason et al. (2020)	Male: 19.9 (1.4), Female: 19.6 (1.5)	48	NA
Masterson et al. (2019)	18.5 (1.0)	86	NA
McClincy et al. (2006)	16.1 (2.2)	104	104
McCrea et al. (2003)	Concussed: 20.0 (1.4), Control: 19.2 (1.5)	150	94
McCrea et al. (2005)	Concussion: 20.0 (1.4), Control: 19.2 (1.5)	150	94
McCrea et al. (2013)	Typical recovery: 17.5 (17.3–17.7), Prolonged recovery: 17.0 (16.4–17.7), Control: 17.4 (17.1–17.7)	736	570
Meyer and Arnett (2015)	Group 1: 18.4 (0.8), Group 2: 18.2 (0.7), Group 3 18.3 (0.7), Group 4: 18.7 (1.1)	420	NA
Miller et al. (2007)	NR	58	NA
Moser and Schatz (2002)	16.7 (1.2)	35	14
Moser et al. (2017)	11.6 (0.6)	30	NA
Nelson et al. (2016)	Concussed: 17.5 (2.0), Control: 17.7 (1.8)	331	166
O'Brien et al. (2018)	12.4 (11–14)	53	NA
Onate et al. (2000)	Men: 19.8 (1.2), Women: 19.3 (1.3)	59	NA
Pearce et al. (2015)	25.1 (4.5)	23	8
Pedersen et al. (2014)	23	74	14
Peterson et al. (2003)	Concussed: 20.2 (1.6), Control: 19.3 (1.3)	42	24
Register-Mihalik et al. (2012)	College: 20.0 (0.8), High School: 16.0 (0.9)	40	NA
Register-Mihalik et al. (2013)	Concussed: 18.6 (1.1), Controls: 20.6 (1.6)	132	38
Resch et al. (2016)	Concussed: 20.2 (1.6), Control: 19.0 (0.9)	80	40
Salvatore et al. (2017)	Concussed: 18–23, Control: 17–26	60	30
Schatz and Robertshaw (2014)	15.8 (1.9)	250	250
Schatz and Sandel (2012)	13–21	162	81
Schatz et al. (2006)	Concussed: 16.5 (2.3), Control 17.3 (1.7)	138	66
Schatz (2010)	18.8 (0.6)	95	NA
Schatz and Maerlender (2013)	Concussed: 15.6 (2.0), Control: 15.5 (1.9)	21,917	560

Table 1 (continued)

Study	Age of sample in years, mean (SD or range)	Total sample size	Size of concussed group
Sherry et al. (2019)	15.1 (2.2)	123	64
Sim et al. (2008)	15.7 (1.2)	419	14
Simon et al. (2017)	20.1 (1.4)	38	NA
Sosnoff et al. (2007)	19.8 (2.2)	44	22
Sufrinko et al. (2017)	16.8 (2.2)	125	125
Thoma et al. (2018)	20.3 (1.3)	93	NA
Tsushima et al. (2013)	15.2 (1.7)	51	26
Tsushima et al. (2016)	15.3 (0.5)	212	NA
Valovich McLeod et al. (2006)	9–14	49	NA
Van Kampen et al. (2006)	Concussed: 16.6 (12–27), Controls: 17.3 (14–22)	192	122
Volberding and Melvin (2014)	20.8 (1.8)	23	NA
Womble et al. (2016)	14.6 (1.9)	200	NA

Mean (SD) age provided where reported, else range or other descriptive statistic denoted

ImPACT Immediate Post-Concussion Assessment and Cognitive Test, *NR* not reported, *NA* not applicable (i.e., no concussed athletes in study design)

tests. Psychometric data from all included studies are listed in the supplementary materials (Table S2).

Brief Summary, Risk of Bias, and Level of Evidence

Modified QUADAS-2 ratings for most studies (n = 76) indicated a moderate risk of bias. Common methodological limitations included use of convenience samples or missing patient flow diagrams (e.g., CONSORT), limited control for confounding factors, and unreported effect sizes. High risk of bias (n = 2) was due to protocol deviations (Echemendia et al., 2001, administration in hallways and buses; Maerlender & Molfese, 2015, large group administration with non-standardized instructions). See Table S2 for individual study ratings.

Test–retest reliability coefficients varied highly for both computerized tools (0.14 to 0.93) and standard neuropsychological tests (0.02 to 0.95). The level of evidence for reliability was rated a SORT grade A (consistent good-quality evidence) for computerized tools and standard neuropsychological tests. Sensitivity to acute concussion ranged from 45%–93% for computerized tools and 18%–80% for neuropsychological test batteries. Given estimates were inconsistent and below chance levels at times, the level of evidence for sensitivity was a SORT grade B for computerized tools and standard neuropsychological tests. Table 2 provides the range of test–retest reliability and sensitivity estimates by measure. Figures 2 and 3 depict these ranges with heat maps for computerized tools and standard neuropsychological tests, respectively.

Reliability

Reliability coefficients such as Pearson r or intra-class coefficients (ICCs) assess the reliability or consistency of test scores. Slick (2006) define acceptable reliability as coefficients ≥ 0.70. Nunnally and Bernstein (1994, p. 265) define adequate reliability to be 0.80, or a more stringent 0.90 if important decisions are to be based on specific test scores. Reliable change indices (RCIs; Chelune, 2003) are calculated using reliability coefficients and represent a meaningful change between assessments. Regression-based methods (RBM; McSweeny et al., 1993) calculate a predicted change score using baseline data. Dependent samples t-tests or repeated measures analysis of variance (ANOVA) evaluate practice effects with a significance test.

Test–Retest Reliability

Test–retest reliability was not consistently better for one testing modality over the other. Coefficients had a wide range for standard neuropsychological tests, from 0.02 to 0.95, and for computerized cognitive tools, from 0.14 to 0.93 (see Table 2 and Figs. 2 and 3). Test–retest reliability was most often assessed in ImPACT, typically using annual preseason baseline testing, with shorter (six months, 0.35–0.86) and longer reliability estimates (four years, 0.29–0.69) having a similar span (Echemendia et al., 2016; Mason et al., 2020; Womble et al., 2016). Processing speed tasks had strongest correlations for both modalities, although were most susceptible to practice effects upon three-day retest (Register-Mihalik et al.,

Table 2 Range of reliability and sensitivity estimates for computerized cognitive tools and standard neuropsychological tests

Computerized cognitive tools

Test–retest reliability*			Sensitivity to acute concussion**		
ANAM			**ANAM**	48%[a]	*below acceptable*
code substitution	.37–.79	*low to adequate*			
code substitution delayed	.33–.77	*low to adequate*			
go-no-go	.34–.46	*low*			
match to sample	.53–.77	*low to adequate*			
mathematical processing	.58–.86	*low to adequate*			
procedural reaction time	.30–.62	*low to marginal*			
simple reaction time	.29–.57	*low*			
simple reaction time delayed	.14–.66	*low to marginal*			
CNS Vital Signs			**CNS Vital Signs**	50%[c]	*below acceptable*
complex attention	.49[b]	*low*			
cognitive flexibility	.40[b]	*low*			
composite memory	.43[b]	*low*			
executive function	.44[b]	*low*			
motor speed	.53[b]	*low*			
neurocognition index	.33[b]	*low*			
psychomotor speed	.58[b]	*low*			
processing speed	.61[b]	*marginal*			
reaction time	.53[b]	*low*			
simple attention	.30[b]	*low*			
verbal memory	.41[b]	*low*			
visual memory	.31[b]	*low*			
CogState			**CogState**	45%–93%	*reaches acceptable*
attention	.56–.73	*low to adequate*			
detection	.52–.85	*low to high*			
identification	.62–.86	*marginal to high*			
learning	.34–.74	*low to adequate*			
processing speed	.32–.59	*low*			
one back	.66–.83	*marginal to high*			
one card learning	.40–.93	*low to very high*			
working memory	.59–.81	*low to high*			
ImPACT			**ImPACT**	47%–93%	*reaches acceptable*
reaction time	.24–.76	*low to adequate*			
verbal memory	.15–.83	*low to high*			
visual memory	.35–.78	*low to adequate*			
visual-motor speed	.57–.90	*low to very high*			

Standard, pencil-and-paper neuropsychological tests

Test–retest reliability*			Sensitivity to acute concussion**		
BVMT-R			**BVMT-R**	**	
delayed recall	.02–.35[d]	*low*			
immediate recall	.36–.69[d]	*low to marginal*			
COWAT	.68[e]	*marginal*	**COWAT**	35%[f]	*below acceptable*
HVLT			**HVLT**	39%[f]	*below acceptable*
delayed recall	.51–.64	*low to marginal*			
immediate recall	.49–.68	*low to marginal*			
recognition discriminability	.21–.54	*low*			
SDMT	.39–.77	*low to adequate*	**SDMT**	4%–52%	*below acceptable*
SOCT	.55–.95	*low to very high*	**SOCT**	7%–55%	*below acceptable*

Table 2 (continued)

Standard, pencil-and-paper neuropsychological tests

Test–retest reliability*			Sensitivity to acute concussion**		
Stroop Color-Word	.76–.92	*adequate to very high*	**Stroop Color-Word**	0%–33%[g]	*below acceptable*
Trail-Making Test			**Trail-Making Test**	52%[f]	*below acceptable*
part A	.41–.63	*low to marginal*			
part B	.42–.76	*low to adequate*			
Wechsler Intelligence Scales			**Wechsler Intelligence Scales**		
digit span	.60–.70[e]	*marginal to adequate*	digit span	30%[f]	*below acceptable*
digit-symbol coding	.65–.82	*marginal to high*	digit-symbol coding	26%–59%	*below acceptable*
processing speed index	.78[e]	*adequate*	processing speed index	66%[h]	*marginal*
symbol search	.58–.83	*low to high*	symbol search	**	
			Whole test battery	18–80%	*reaches acceptable*

ANAM Automated Neuropsychological Assessment Metrics, *BVMT-R* Brief Visuospatial Memory Test–Revised, *CogState* CogState Ltd.'s Cog-Sport/Axon Sport, *COWAT* Controlled Oral Word Association Test, *HVLT* Hopkins Verbal Learning Test or Hopkins Verbal Learning Test–Revised, *ImPACT* Immediate Post-Concussion Assessment and Cognitive Test, *SDMT* Symbol Digit Modalities Test, *SOCT* Speed of Comprehension Test, *Wechsler Intelligence Scales* Wechsler Adult Intelligence Scale–Revised, Wechsler Adult Intelligence Scale–III, or Wechsler Intelligence Scale for Children–III

*Test–retest reliability among healthy athletes. Reliability coefficients of < .60 are "low," .60–.69 are "marginal," .70–.79 "adequate," .80–.89 "high," and ≥ .90 "very high" (Slick, 2006, p. 14)

**Percent of athletes demonstrating impaired performance within 72 h of concussion (no studies reported for BVMT-R or symbol search). Acceptable (≥ 80%) and marginal (60–79%) sensitivity thresholds were based on convention commonly used in detection of cognitive impairment (e.g., Blake et al., 2002; Weissberger et al., 2017)

Where indicated, estimates were based on a single study: [a]Nelson et al. (2016) (low risk of bias), [b]Broglio et al. (2018) (medium risk of bias), [c]Czerniak et al. (2021) (low risk of bias), [d]Register-Mihalik et al. (2012) (medium risk of bias), [e]Barr (2003) (medium risk of bias), [f]Broglio et al. (2007b) (medium risk of bias), [g]Hinton-Bayre (2015) (medium risk of bias), [h]Gardner et al. (2012) (low risk of bias)

2012). When retested after one year or longer, estimates for verbal memory tasks were the weakest except for one included study of athletes aged 10–12 years in which verbal memory performed best (Moser et al., 2017). This may reflect the developmental needs of this sample, for which a separate pediatric ImPACT has recently been developed.

Multiple attempts have been made to improve the test–retest reliability of ImPACT, including aggregating multiple baseline (Bruce et al., 2016, 2017) and composite scores (memory and speed; Schatz & Maerlender, 2013). These two methods consistently improved reliability estimates (Brett et al., 2018; Bruce et al., 2017; Echemendia et al., 2016; Schatz & Maerlender, 2013). Overall, test–retest reliability was not improved with increasingly strict invalidity or exclusion criteria (Brett et al., 2016; Register-Mihalik et al., 2012) or across specific populations (i.e., learning disabilities and headache or migraine treatment history; Brett et al., 2018).

Other Forms of Reliability

Cronbach's alpha, a measure of internal consistency, ranged from 0.64 to 0.84 for ImPACT (Gerrard et al., 2017; Higgins et al., 2018). McDonald's omega, a less biased estimate of reliability requiring factor analysis, ranged from 0.52 to 0.72

for ImPACT (Gerrard et al., 2017) and 0.03 to 0.85 for Automated Neuropsychological Assessment Metrics (ANAM; Glutting et al., 2020). Hinton-Bayre and Geffen (2005) examined alternate forms of standard neuropsychological tests and found moderate to large equivalence for the Symbol Digit Modalities Test (SDMT) and the Wechsler Adult Intelligence Scale–Revised (WAIS-R) Digit Symbol subtest. Subtests from the abbreviated ImPACT Quick Test had variable correlations ($r = 0.14$–0.40) with complete ImPACT administration (Elbin et al., 2020). No other studies identified examined internal consistency or alternate forms reliability.

Sensitivity and Specificity

Sensitivity and specificity evaluate how well test performance differentiates concussed and healthy athletes. Percentages represent the proportion of concussed (sensitivity) and healthy athletes (specificity) correctly identified. Thresholds for acceptable (≥ 80%) and marginal (60%–79%) sensitivity have previously been used to describe cognitive impairment (e.g., Blake et al., 2002; Weissberger et al., 2017).

Logistic regression, discriminant function analyses, RCI, and RBM can evaluate sensitivity and specificity. RCI and RBM are able to assess clinically meaningful changes in
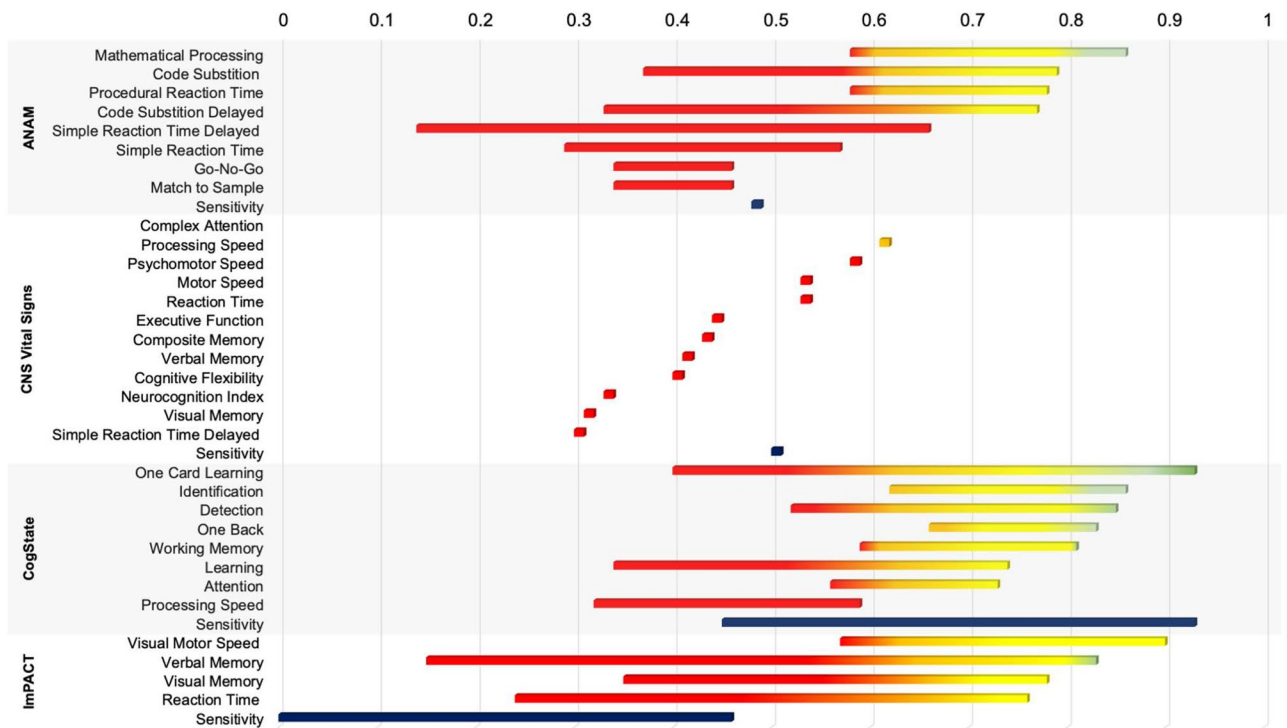
Fig. 2 Heat map of test–retest reliability and sensitivity estimates for computerized cognitive tools.
Note. Bars represent ranges of test–retest reliability estimates by measure or, where noted, sensitivity (0%–100%) to acute concussion
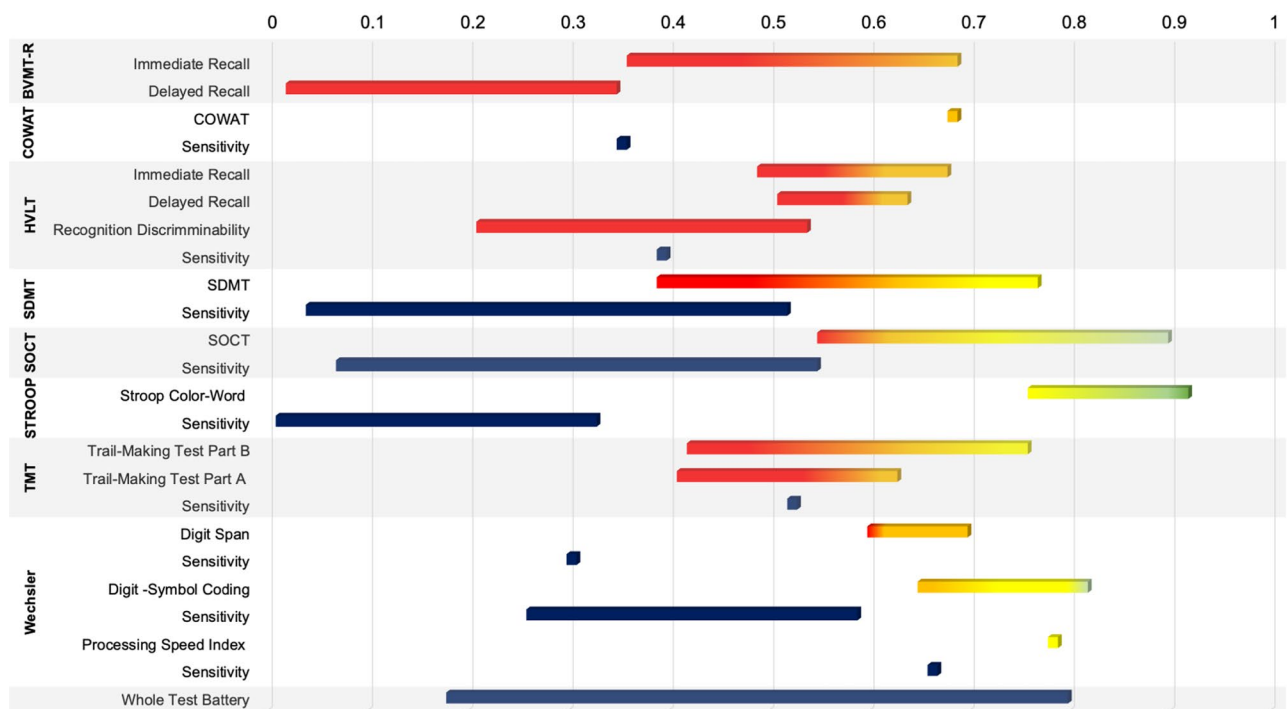

Fig. 3 Heat map of test–retest reliability and sensitivity estimates for standard neuropsychological tests.
Note. Bars represent ranges of test–retest reliability estimates by measure or, where noted, sensitivity (0%–100%) to acute concussion

raw scores (Hinton-Bayre, 2015; Louey et al., 2014; Schatz & Robertshaw, 2014). Echemendia et al. (2012) found RBM to perform better than RCI methods (although similar to normative comparison), whereas examination by Erlanger et al. (2003a, b) and Merz et al. (2021) yielded comparable findings for RCI and RBM.

Sensitivity of computerized tools varied across studies, often below acceptable limits (see Table 2 and Fig. 2). From 80 to 93% sensitivity has been reported for ImPACT and CogState Ltd's computerized concussion tool when administered to athletes within 72 h post-injury (Gardner et al., 2012; Louey et al., 2014; Schatz & Sandel, 2012; Van Kampen et al., 2006). Although, similar studies evaluating reliable impairment in acutely concussed athletes observed only marginal sensitivity for ImPACT and CogState (Abbassi & Sirmon-Taylor, 2017; Broglio et al., 2007a; Czerniak et al., 2021; Nelson et al., 2016; Sufrinko et al., 2017). As athletes progressed from acute to subacute concussion, the percent of athletes with reliable worsening on one or more subtest (23%–55%) was below acceptable levels for ANAM, CogState, and ImPACT (Iverson et al., 2006; Nelson et al., 2016; Sufrinko et al., 2017).

Sensitivity for test batteries and standard neuropsychological measures also varied across studies (see Table 2 and Fig. 3). Hopkins Verbal Learning Test (HVLT), Trail-Making Test (TMT), SDMT, Stroop, and Controlled Oral Word Association Test (COWAT) were at chance levels of detecting acute impairment (McCrea et al., 2005), similar to subtest-level sensitivity for computerized tools (Louey et al., 2014; Nelson et al., 2016). Sensitivity to acute concussion reached 80% for a battery comprised of the SDMT, WAIS-R Digit Symbol, and Speed of Comprehension Test (SOCT; Baddeley et al. as cited in Hinton-Bayre et al., 1999), although this finding was not replicated (sensitivity 18%–44%) in similar test batteries (Broglio et al., 2007b; Makdissi et al., 2010; McCrea et al., 2005).

Assessing Cognitive Effects of Acute Concussion

Significance testing (e.g., *t*-tests, ANOVA, logistic regression) can examine the sensitivity of cognitive performance to the effects of concussion. Athletes' post-injury scores can be compared to healthy controls (between-subjects design) or to baseline (within-subjects design). Cohen's *d* (size of group differences) and η^2 (amount of variance explained) are two widely used statistics that depict effect sizes, or the magnitude of findings. Cohen (1988) defines *d* as small (0.2), medium (0.5), and large (0.8). Effect sizes can be calculated for ANOVA and regression, although comparison is difficult when models contain varying combinations of factors or predictors.

Both testing modalities have consistently been shown to be sensitive to acute concussion (i.e., within 72-h

postconcussion; Joseph et al., 2018), with effect sizes that are generally medium to large. Specifically, effect sizes were 0.54 to 0.86 for ANAM (within-subjects Cohen's *d*), 0.80 to 1.03 for the Cambridge Neuropsychological Test Automated Battery (CANTAB; Cambridge Cognition, Ltd.) (between- and within-subjects Cohen's *d*), -0.88 to -0.18 for CogState (between-subjects effect size), up to 0.36 for the HeadMinder, Inc. Concussion Resolution Index (between- and within-subjects η^2) and variable for the ImPACT (between-subjects maximum $\eta^2 = 0.99$, within-subjects maximum $d = 0.45$, between-subjects maximum $d = 0.95$) (Abbassi & Sirmon-Taylor, 2017; Collins et al., 1999; Gardner et al., 2012; Iverson et al., 2003; Louey et al., 2014; Lovell et al., 2004; Pearce et al., 2015; Register-Mihalik et al., 2013; Schatz & Sandel, 2012; Schatz et al., 2006; Sosnoff et al., 2007). Standard neuropsychological tests including COWAT, Digit Span, HVLT, O'Connor Finger Dexterity test, Repeatable Battery for the Assessment of Neuropsychological Status, Stroop, SDMT, TMT, and Vigil Continuous Performance Test demonstrated large effect sizes (Cohen's $d = 0.80$–1.03; $\eta^2 = 0.91$) (Echemendia et al., 2001; Moser & Schatz, 2002; Pearce et al., 2015). Acute evaluations using COWAT, Digit Span, Grooved Pegboard, HVLT, Stroop, SDMT, and TMT were significant, although effect sizes were not reported (Collins et al., 1999; Guskiewicz et al., 2001).

Validity

Convergent and Discriminant Validity

Convergent validity evaluates the similarity of the assessment with other measures of a similar cognitive domain. Discriminant validity examines the relationship between assessments measuring differing domains. Most studies have evaluated convergent and discriminant validity using correlation coefficients, which can be interpreted using Cohen (1988)'s guidelines for small (0.10), medium (0.30) and large (0.50) effect size. Confirmatory factor analysis and multitrait-multimethod analysis have also been used and have the advantage of taking into account the measurement error inherent in single observed test scores (Floyd & Widaman, 1995; Strauss & Smith, 2009).

Convergent and discriminant validity were examined most frequently for ImPACT. Maerlender et al. (2010) found ImPACT correlated with the California Verbal Learning Test-2 (ImPACT Verbal Memory $r = 0.40$), Brief Visuospatial Memory Test–Revised (BVMT-R) (ImPACT Visual Memory $r = 0.59$), Connor's Continuous Performance Test (ImPACT Reaction Time $r = -0.39$), Delis-Kaplan Executive Function System subtests (ImPACT Visual Motor $r = 0.41$), and Gronwall's Paced Auditory Serial Attention Test (ImPACT Visual Motor $r = 0.39$, Reaction Time $r = -0.31$).

SDMT correlations were stronger for ImPACT speeded subtests ($r = -0.60$ and 0.70) and weaker for ImPACT memory subtests ($r = 0.37$ and 0.46), providing support for convergent and discriminant validity (Iverson et al., 2005).

Studies using C3 Logix and CogState's speeded subtests have reported small to medium correlations with standard neuropsychological tests of processing speed. For C3 Logix, correlations with TMT and SDMT ranged from 0.10–0.78 (Simon et al., 2017). CogState subtests correlated with TMT and WAIS-R Digit Symbol (strongest $r = 0.48$) (Makdissi et al., 2001).

Medium to large correlations were observed among standard neuropsychological tests assessing similar cognitive domains. Valovich McLeod et al. (2006) reported correlations up to 0.92 among Buschke selective reminding test scores and speeded tasks (TMT B, Wechsler Intelligence Scale for Children–Third Edition processing speed subtests). Similarly, Hinton-Bayre et al. (1997) found correlations among speeded tasks ranged from 0.44 to 0.77 (WAIS-R Digit Symbol, SDMT, SOCT).

Factor analysis generally supports factor validity (i.e., that performance reflects the aspect of cognition theoretically underlying the test domain). ImPACT memory and processing speed factors have been identified in samples of healthy and concussed athletes using exploratory (Gerrard et al., 2017; Iverson et al., 2005; Thoma et al., 2018) and confirmatory factor analysis (Maietta et al., 2021; Masterson et al., 2019; Schatz & Maerlender, 2013) as well as multitrait-multimethod analysis with traditional pencil-and-paper assessments (Thoma et al., 2018). Factors of memory and executive functioning have emerged within baseline pencil-and-paper assessments comprising the HVLT–Revised (HVLT-R), BVMT-R, TMT, and COWAT using exploratory factory analysis (Lovell & Solomon, 2011). Czerniak et al. (2021) identified support for a hierarchical factor structure comprised of ANAM's general composite and two lower-level factors. Unlike other computerized tools, C3 Logix did not conform to a two-factor model (Masterson et al., 2019).

Other Forms of Validity

Correlation of cognitive test scores with concussion symptom duration supports predictive validity. For example, ImPACT and Concussion Resolution Index scores in acutely concussed student athletes were found to predict recovery time (Erlanger et al., 2003a, b; Lau et al., 2011, 2012). Worse acute post-injury memory performance has also been associated with subjective symptom reports (Broglio et al., 2009; Erlanger et al., 2003a, b; Lovell et al., 2003).

Discussion

We conducted a systematic review of the psychometric properties of computerized cognitive tools and conventional neuropsychological tests for the assessment of concussion in athletes. There were 103 studies published between 1996 and 2021 that met the inclusion criteria. ImPACT was the most robustly evaluated test. HVLT-R, TMT, and SDMT were most common in standard neuropsychological batteries. Only 13 studies employed hybrid test batteries using both modalities.

Consistent with prior reviews (Farnsworth et al., 2017; Randolph et al., 2005; Resch et al., 2013a, b), reliability coefficients were highly variable for both computerized tools (0.14 to 0.93, see Table 2 and Fig. 2) and conventional neuropsychological tests (0.02 to 0.95, see Table 2 and Fig. 3). Reliability coefficients were generally stronger for speeded tasks. Aggregating multiple baseline and composite scores (Brett et al., 2018; Bruce et al., 2017; Echemendia et al., 2016; Schatz & Maerlender, 2013), as well as detecting suboptimal effort (Walton et al., 2018), can improve reliability estimates in athletes.

Test–retest reliability for two widely used conventional neuropsychological tests was weaker in athletes (HVLT $r = 0.49$–0.68, TMT $r = 0.41$–0.79) (Barr, 2003; Register-Mihalik et al., 2012; Valovich McLeod et al., 2006) than estimates derived from non-athlete samples (HVLT 0.66–0.74, TMT 0.79–0.89) (Benedict et al., 1998; Dikmen et al., 1999). Homogeneity in athlete samples may place a limit on reliability estimates through restricted score variance. Further, the cognitive functions typically assessed in concussion (e.g., psychomotor processing speed, attention, verbal and nonverbal memory) continue to develop throughout adolescence (Carson et al., 2016; Casey et al., 2005; Giedd et al., 1999) and may impact reliability for long test–retest intervals.

Sensitivity varied widely, and was at times below chance levels, as depicted in Figs. 2 and 3. When administered within 72 h post-injury, ImPACT and CogState reached acceptable sensitivity (80% or better). However, for just as many acute samples, marginal ImPACT and CogState sensitivity was observed. For standard neuropsychological test batteries, sensitivity was difficult to compare across studies given the differing numbers of measures and criteria for impairment used. The method used to classify impairment appears to modify sensitivity (Echemendia et al., 2012; Schatz & Robertshaw, 2014). RCI and RBM are superior to using normative comparison to calculate sensitivity (Brett et al., 2016; Maerlender & Molfese, 2015; Moser et al., 2017; O'Brien et al., 2018; Schatz, 2010).

Further study of the psychometric properties of cognitive assessments should include quantitative synthesis. We did not pool systematic review results beyond the provided table and figures given the wide range of measures and statistical methods included, for example, variable test–retest time frames. While broadly inclusive, more specific conclusions were limited by this approach. Future meta-analysis may be possible with tests used across numerous studies (e.g., ImPACT, HVLT-R, TMT, and SDMT).

Continued Gaps in the Literature and Assessment Practices

Attempts to improve the psychometric properties of tests were sparse within the literature and limited to a single test (i.e., aggregating ImPACT baseline and composite scores). Current cognitive assessment tools could leverage modern technologies and procedures such as telehealth, machine learning, and virtual reality to improve ecological validity (Marcopulos & Lojek, 2019; Parsons, 2011). Structural modeling techniques have recently been used to improve a sport concussion symptom inventory (Brett et al., 2020; Wilmoth et al., 2020). Alternate forms are a known method of eliminating practice effects in both traditional neuropsychological and computerized cognitive assessments, and nonequivalence affects test–retest reliability (Echemendia et al., 2009; Resch et al., 2013a, b).

Though not a pre-specified objective, as we excluded studies of non-English speakers, we informally observed during our review that few studies considered sociocultural and ethnoracial differences in their methodologies. Preliminary findings indicate differences on baseline testing based on sociocultural factors including maternal education as well as racial and linguistic diversity (Houck et al., 2018; Maerlender et al., 2020; Wallace et al., 2020), which has very important implications for use of normative data and clinical interpretation. Given the current paucity of data, more research is needed to understand how intersectionality influences cognitive testing performance pre- and post-injury in athletes.

Further research is needed to better understand which aspects of test psychometrics (e.g., test–retest reliability) are affected by cognitive development within adolescent and college-age athletes (Brett et al., 2016). Assessment performance in older athletes also warrants further examination. Similar surveys of non-sport concussion literature should first be conducted in civilian and military populations prior to generalizing these findings. Although no reliable change has been observed in performance from baseline following recent concussion (Lynall et al., 2016), more studies are needed to examine this possibility in athletes who have sustained multiple prior concussions. Future studies should

screen for invalid performance, which is common in this population (Abeare et al., 2018), and examine the relationship between suboptimal effort and psychometric properties (Brett & Solomon, 2017).

Best Practice Recommendations

Ultimately, only a healthcare provider with expertise in assessment and knowledge of psychometric principles who is qualified to interpret test results should be doing so (Bauer et al., 2012). Computerized neurocognitive tests do not eliminate the need for a clinical expert, or at minimum, an appropriately trained clinician to evaluate and interpret test results (Covassin et al., 2009; Echemendia et al., 2009; Meehan et al., 2012; Moser et al., 2015). Demographic and sociocultural factors as well as prior medical and psychiatric history have been observed to influence baseline scores (Cook et al., 2017; Cottle et al., 2017; Gardner et al., 2017; Houck et al., 2018; Wallace et al., 2020; Zuckerman et al., 2013) while pre-injury cognitive status (Merz et al., 2021; Schatz & Robertshaw, 2014), history of attention-deficit hyperactivity disorder (Gardner et al., 2017), and concussion history (Covassin et al., 2013) have been shown to influence post-injury test results. Sport clinicians should utilize the most appropriate normative reference data available (Mitrushina et al., 2005). Normative data are becoming increasingly available for athletes (e.g., Solomon et al., 2015's NFL player normative data). Whether using normative data or individuals' own pre-injury baseline performance, measurement error and multivariate base rates of impairment should be considered in interpretation (Nelson, 2015).

Conclusions

There remains no clear psychometric evidence to support one testing modality over another in the evaluation of sport concussion. Test–retest reliability for speeded tasks was generally stronger overall, although susceptible to variability over time. Sensitivity to acute concussion was greatest for ImPACT, CogState, SDMT, SOCT, and WAIS-R Digit Symbol. A hybrid model combining test modalities would seem to counterbalance and optimize methods for increased accuracy and efficiency. To this end, more robust formal study of psychometric properties should be pursued (Echemendia et al., 2020). It is critical for sport clinicians to have expertise in interpreting test results and to implement a multidimensional assessment that synthesizes an individual's performance within the context of their history and situational factors. Future test revisions should capitalize on advances in psychometric and analytic approaches.

Declarations

Ethics Approval/Informed Consent Not applicable.

Conflict of Interest Dr. Solomon is a paid consultant to the National Football League's Player Health and Safety Department.

References

Abbassi, E., & Sirmon-Taylor, B. (2017). Recovery progression and symptom resolution in sport-related mild traumatic brain injury. *Brain Injury, 31*(12), 1667–1673.

Abeare, C. A., Messa, I., Zuccato, B. G., Merker, B., & Erdodi, L. (2018). Prevalence of Invalid Performance on Baseline Testing for Sport-Related Concussion by Age and Validity Indicator. *JAMA Neurology, 75*(6), 697–703.

Barr, W. B. (2003). Neuropsychological testing of high school athletes. Preliminary norms and test-retest indices. *Archives of Clinical Neuropsychology, 18*(1), 91–101.

Bauer, R. M., Iverson, G. L., Cernich, A. N., Binder, L. M., Ruff, R. M., & Naugle, R. I. (2012). Computerized neuropsychological assessment devices: Joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. *Archives of Clinical Neuropsychology, 27*(3), 362–373.

Belanger, H. G., & Vanderploeg, R. D. (2005). The neuropsychological impact of sports-related concussion: A meta-analysis. *Journal of the International Neuropsychological Society, 11*(4), 345–357.

Benedict, R. H. B., Schretlen, D., Groninger, L., & Brandt, J. (1998). Hopkins Verbal Learning Test – Revised: Normative Data and Analysis of Inter-Form and Test-Retest Reliability. *The Clinical Neuropsychologist, 12*(1), 43–55.

Bernstein, J. P., Calamia, M., Pratt, J., & Mullenix, S. (2019). Assessing the effects of concussion using the C3Logix Test Battery: An exploratory study. *Applied Neuropsychology: Adult, 26*(3), 275–282.

Bialunska, A., & Salvatore, A. P. (2017). The auditory comprehension changes over time after sport-related concussion can indicate multisensory processing dysfunctions. *Brain and Behavior, 7*(12), e00874.

Blake, H., McKinney, M., Treece, K., Lee, E., & Lincoln, N. B. (2002). An evaluation of screening measures for cognitive impairment after stroke. *Age and Ageing, 31*(6), 451–456.

Brett, B. L., Kramer, M. D., McCrea, M. A., Broglio, S. P., McAllister, T. W., Nelson, L. D., Investigators, C. C., Hazzard, J. B., Jr., Kelly, L. A., Ortega, J., Port, N., Pasquina, P. F., Jackson, J., Cameron, K. L., Houston, M. N., Goldman, J. T., Giza, C., Buckley, T., Clugston, J. R., … Susmarski, A. (2020). Bifactor Model of the Sport Concussion Assessment Tool Symptom Checklist: Replication and Invariance Across Time in the CARE Consortium Sample. *The American Journal of Sports Medicine, 48*(11), 2783–2795.

Brett, B. L., Smyk, N., Solomon, G., Baughman, B. C., & Schatz, P. (2016). Long-term Stability and Reliability of Baseline Cognitive Assessments in High School Athletes Using ImPACT at 1-, 2-, and 3-year Test-Retest Intervals. *Archives of Clinical Neuropsychology, 18*, 18.

Brett, B. L., & Solomon, G. S. (2017). The influence of validity criteria on Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT) test-retest reliability among high school athletes. *Journal of Clinical and Experimental Neuropsychology, 39*(3), 286–295.

Brett, B. L., Solomon, G. S., Hill, J., & Schatz, P. (2018). Two-year Test-Retest Reliability in High School Athletes Using the Four- and Two-Factor ImPACT Composite Structures: The Effects of Learning Disorders and Headache/Migraine Treatment History. *Archives of Clinical Neuropsychology, 33*(2), 216–226.

Broglio, S. P., Katz, B. P., Zhao, S., McCrea, M., McAllister, T., Reed Hoy, A., Hazzard, J., Kelly, L., Ortega, J., Port, N., Putukian, M., Langford, D., Campbell, D., McGinty, G., O'Donnell, P., Svoboda, S., DiFiori, J., Giza, C., Benjamin, H., … Lintner, L. (2018). Test-Retest Reliability and Interpretation of Common Concussion Assessment Tools: Findings from the NCAA-DoD CARE Consortium. *Sports Medicine, 48*(5), 1255–1268.

Broglio, S. P., Macciocchi, S. N., & Ferrara, M. S. (2007a). Neurocognitive performance of concussed athletes when symptom free. *Journal of Athletic Training, 42*(4), 504–508.

Broglio, S. P., Macciocchi, S. N., & Ferrara, M. S. (2007b). Sensitivity of the concussion assessment battery. *Neurosurgery, 60*(6), 1050–1058.

Broglio, S. P., & Puetz, T. W. (2008). The effect of sport concussion on neurocognitive function, self-report symptoms and postural control : A meta-analysis. *Sports Medicine, 38*(1), 53–67.

Broglio, S. P., Sosnoff, J. J., & Ferrara, M. S. (2009). The relationship of athlete-reported concussion symptoms and objective measures of neurocognitive function and postural control. *Clinical Journal of Sport Medicine, 19*(5), 377–382.

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

Bruce, J., Echemendia, R., Tangeman, L., Meeuwisse, W., Comper, P., Hutchison, M., & Aubry, M. (2016). Two baselines are better than one: Improving the reliability of computerized testing in sports neuropsychology. *Applied Neuropsychology: Adult, 23*(5), 336–342.

Bruce, J., Echemendia, R. J., Meeuwisse, W., Hutchison, M. G., Aubry, M., & Comper, P. (2017). Measuring cognitive change with ImPACT: The aggregate baseline approach. *The Clinical Neuropsychologist, 31*(8), 1329–1340.

Carson, V., Hunter, S., Kuzik, N., Wiebe, S. A., Spence, J. C., Friedman, A., Tremblay, M. S., Slater, L., & Hinkley, T. (2016). Systematic review of physical activity and cognitive development in early childhood. *Journal of Science and Medicine in Sport, 19*(7), 573–578.

Casey, B. J., Tottenham, N., Liston, C., & Durston, S. (2005). Imaging the developing brain: What have we learned about cognitive development? *Trends in Cognitive Sciences, 9*(3), 104–110.

Chelune, G. J. (2003). Assessing reliable neuropsychological change. In R. D. Franklin (Ed.), *Prediction in Forensic and*

Neuropsychology: Sound Statistical Practices (pp. 123–147). Lawrence Erlbaum Associates.

Clay, M. B., Glover, K. L., & Lowe, D. T. (2013). Epidemiology of concussion in sport: A literature review. *Journal of Chiropractic Medicine, 12*(4), 230–251.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates.

Collie, A., Darby, D., & Maruff, P. (2001). Computerised cognitive assessment of athletes with sports related head injury. *British Journal of Sports Medicine, 35*(5), 297–302.

Collie, A., Makdissi, M., Maruff, P., Bennell, K., & McCrory, P. (2006). Cognition in the days following concussion: Comparison of symptomatic versus asymptomatic athletes. *Journal of Neurology, Neurosurgery & Psychiatry, 77*(2), 241–245.

Collins, M., Lovell, M. R., Iverson, G. L., Ide, T., & Maroon, J. (2006). Examining concussion rates and return to play in high school football players wearing newer helmet technology: a three-year prospective cohort study. *Neurosurgery, 58*(2), 275–286; discussion 275–286.

Collins, M. W., Grindel, S. H., Lovell, M. R., Dede, D. E., Moser, D. J., Phalin, B. R., Nogle, S., Wasik, M., Cordry, D., Daugherty, K. M., Sears, S. F., Nicolette, G., Indelicato, P., & McKeag, D. B. (1999). Relationship between concussion and neuropsychological performance in college football players. *JAMA, 282*(10), 964–970.

Cook, N. E., Huang, D. S., Silverberg, N. D., Brooks, B. L., Maxwell, B., Zafonte, R., Berkner, P. D., & Iverson, G. L. (2017). Baseline cognitive test performance and concussion-like symptoms among adolescent athletes with ADHD: Examining differences based on medication use. *The Clinical Neuropsychologist, 31*(8), 1341–1352.

Cottle, J. E., Hall, E. E., Patel, K., Barnes, K. P., & Ketcham, C. J. (2017). Concussion Baseline Testing: Preexisting Factors, Symptoms, and Neurocognitive Performance. *The American Journal of Sports Medicine, 52*(2), 77–81.

Covassin, T., Elbin, R. J., 3rd., Stiller-Ostrowski, J. L., & Kontos, A. P. (2009). Immediate post-concussion assessment and cognitive testing (ImPACT) practices of sports medicine professionals. *Journal of Athletic Training, 44*(6), 639–644.

Covassin, T., Elbin, R. J., & Nakayama, Y. (2010). Tracking neurocognitive performance following concussion in high school athletes. *The Physician and Sportsmedicine, 38*(4), 87–93.

Covassin, T., Moran, R., & Wilhelm, K. (2013). Concussion symptoms and neurocognitive performance of high school and college athletes who incur multiple concussions. *The American Journal of Sports Medicine, 41*(12), 2885–2889.

Covassin, T., Schatz, P., & Swanik, C. B. (2007). Sex differences in neuropsychological function and post-concussion symptoms of concussed collegiate athletes. *Neurosurgery, 61*(2), 345–351.

Czerniak, L. L., Liebel, S. W., Garcia, G. P., Lavieri, M. S., McCrea, M. A., McAllister, T. W., Broglio, S. P., & Investigators, C. C. (2021). Sensitivity and Specificity of Computer-Based Neurocognitive Tests in Sport-Related Concussion: Findings from the NCAA-DoD CARE Consortium. *Sports Medicine, 51*(2), 351–365.

Dikmen, S. S., Heaton, R. K., Grant, I., & Temkin, N. R. (1999). Test-retest reliability and practice effects of expanded Halstead-Reitan Neuropsychological Test Battery. *Journal of the International Neuropsychological Society, 5*(4), 346–356.

Ebell, M. H., Siwek, J., Weiss, B. D., Woolf, S. H., Susman, J. L., Ewigman, B., & Bowman, M. (2004). Simplifying the language of evidence to improve patient care: Strength of recommendation taxonomy (SORT): A patient-centered approach to grading evidence in medical literature. *The Journal of Family Practice, 53*(2), 111–120.

Echemendia, R. J., Bruce, J. M., Bailey, C. M., Sanders, J. F., Arnett, P., & Vargas, G. (2012). The utility of post-concussion neuropsychological data in identifying cognitive change following sports-related MTBI in the absence of baseline data. *The Clinical Neuropsychologist, 26*(7), 1077–1091.

Echemendia, R. J., Bruce, J. M., Meeuwisse, W., Comper, P., Aubry, M., & Hutchison, M. (2016). Long-term reliability of ImPACT in professional ice hockey. *The Clinical Neuropsychologist, 30*(2), 328–337.

Echemendia, R. J., Herring, S., & Bailes, J. (2009). Who should conduct and interpret the neuropsychological assessment in sports-related concussion? *British Journal of Sports Medicine, 43*(Suppl 1), i32–i35.

Echemendia, R. J., Putukian, M., Mackin, R. S., Julian, L., & Shoss, N. (2001). Neuropsychological test performance prior to and following sports-related mild traumatic brain injury. *Clinical Journal of Sport Medicine, 11*(1), 23–31.

Echemendia, R. J., Thelen, J., Meeuwisse, W., Comper, P., Hutchison, M. G., & Bruce, J. M. (2020). Testing the hybrid battery approach to evaluating sports-related concussion in the National Hockey League: A factor analytic study. *The Clinical Neuropsychologist, 34*(5), 899–918.

Echlin, P. S., Skopelja, E. N., Worsley, R., Dadachanji, S. B., Lloyd-Smith, D. R., Taunton, J. A., Forwell, L. A., & Johnson, A. M. (2012). A prospective study of physician-observed concussion during a varsity university ice hockey season: incidence and neuropsychological changes. Part 2 of 4. *Neurosurgical Focus, 33*(6), E2: 1–11.

Elbin, R. J., D'Amico, N. R., McCarthy, M., Womble, M. N., O'Connor, S., & Schatz, P. (2020). How Do ImPACT Quick Test Scores Compare with ImPACT Online Scores in Non-Concussed Adolescent Athletes? *Archives of Clinical Neuropsychology, 35*(3), 326–331.

Elbin, R. J., Schatz, P., & Covassin, T. (2011). One-year test-retest reliability of the online version of ImPACT in high school athletes. *American Journal of Sports Medicine, 39*(11), 2319–2324.

Erlanger, D., Feldman, D., Kutner, K., Kaushik, T., Kroger, H., Festa, J., Barth, J., Freeman, J., & Broshek, D. (2003a). Development and validation of a web-based neuropsychological test protocol for sports-related return-to-play decision-making. *Archives of Clinical Neuropsychology, 18*(3), 293–316.

Erlanger, D., Kaushik, T., Cantu, R., Barth, J. T., Broshek, D. K., Freeman, J. R., & Webbe, F. M. (2003b). Symptom-based assessment of the severity of a concussion. *Journal of Neurosurgery, 98*(3), 477–484.

Erlanger, D., Saliba, E., Barth, J., Almquist, J., Webright, W., & Freeman, J. (2001). Monitoring Resolution of Postconcussion Symptoms in Athletes: Preliminary Results of a Web-Based Neuropsychological Test Protocol. *Journal of Athletic Training, 36*(3), 280–287.

Espinoza, T. R., Hendershot, K. A., Liu, B., Knezevic, A., Jacobs, B. B., Gore, R. K., Guskiewicz, K. M., Bazarian, J. J., Phelps, S. E., Wright, D. W., & LaPlaca, M. C. (2021). A Novel Neuropsychological Tool for Immersive Assessment of Concussion and Correlation with Subclinical Head Impacts. *Neurotrauma Reports, 2*(1), 232–244.

Farnsworth, J. L., 2nd., Dargo, L., Ragan, B. G., & Kang, M. (2017). Reliability of Computerized Neurocognitive Tests for Concussion Assessment: A Meta-Analysis. *Journal of Athletic Training, 52*(9), 826–833.

Fazio, V. C., Lovell, M. R., Pardini, J. E., & Collins, M. W. (2007). The relation between post concussion symptoms and neurocognitive performance in concussed athletes. *NeuroRehabilitation, 22*(3), 207–216.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*(3), 286–299.

Gardner, A., Shores, E. A., Batchelor, J., & Honan, C. A. (2012). Diagnostic efficiency of ImPACT and CogSport in concussed rugby union players who have not undergone baseline neurocognitive testing. *Applied Neuropsychology: Adult, 19*(2), 90–97.

Gardner, R. M., Yengo-Kahn, A., Bonfield, C. M., & Solomon, G. S. (2017). Comparison of baseline and post-concussion ImPACT test scores in young athletes with stimulant-treated and untreated ADHD. *The Physician and Sportsmedicine, 45*(1), 1–10.

Gerrard, P. B., Iverson, G. L., Atkins, J. E., Maxwell, B. A., Zafonte, R., Schatz, P., & Berkner, P. D. (2017). Factor Structure of ImPACT in Adolescent Student Athletes. *Archives of Clinical Neuropsychology, 32*(1), 117–122.

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., Paus, T., Evans, A. C., & Rapoport, J. L. (1999). Brain development during childhood and adolescence: A longitudinal MRI study. *Nature Neuroscience, 2*(10), 861–863.

Glutting, J. J., Davey, A., Wahlquist, V. E., Watkins, M., & Kaminski, T. W. (2020). Internal (Factorial) Validity of the ANAM using a Cohort of Woman High-School Soccer Players. *Archives of Clinical Neuropsychology, 29*, 29.

Goh, S. C., Saw, A. E., Kountouris, A., Orchard, J. W., & Saw, R. (2021). Neurocognitive changes associated with concussion in elite cricket players are distinct from changes due to post-match with no head impact. *Journal of Science and Medicine in Sport, 24*(5), 420–424.

Gronseth, G. S., Woodroffe, L. M., & Getchius, T. S. (2011). Clinical practice guideline process manual. *American Academy of Neurology.*

Guskiewicz, K. M., Riemann, B. L., Perrin, D. H., & Nashner, L. M. (1997). Alternative approaches to the assessment of mild head injury in athletes. *Medicine & Science in Sports & Exercise, 29*(7), S213-221.

Guskiewicz, K. M., Ross, S. E., & Marshall, S. W. (2001). Postural Stability and Neuropsychological Deficits After Concussion in Collegiate Athletes. *Journal of Athletic Training, 36*(3), 263–273.

Hang, B., Babcock, L., Hornung, R., Ho, M., & Pomerantz, W. J. (2015). Can Computerized Neuropsychological Testing in the Emergency Department Predict Recovery for Young Athletes With Concussions? *Pediatric Emergency Care, 31*(10), 688–693.

Higgins, K. L., Caze, T., & Maerlender, A. (2018). Validity and Reliability of Baseline Testing in a Standardized Environment. *Archives of Clinical Neuropsychology, 33*(4), 437–443.

Hinton-Bayre, A. D. (2012). Choice of reliable change model can alter decisions regarding neuropsychological impairment after sports-related concussion. *Clinical Journal of Sport Medicine, 22*(2), 105–108.

Hinton-Bayre, A. D. (2015). Normative versus baseline paradigms for detecting neuropsychological impairment following sports-related concussion. *Brain Impairment, 16*(2), 80–89.

Hinton-Bayre, A. D., & Geffen, G. (2005). Comparability, Reliability, and Practice Effects on Alternate Forms of the Digit Symbol Substitution and Symbol Digit Modalities Tests. *Psychological Assessment, 17*(2), 237–241.

Hinton-Bayre, A. D., Geffen, G., & McFarland, K. (1997). Mild head injury and speed of information processing: A prospective study of professional rugby league players. *Journal of Clinical and Experimental Neuropsychology, 19*(2), 275–289.

Hinton-Bayre, A. D., Geffen, G. M., Geffen, L. B., McFarl, K. A., & Friis, P. (1999). Concussion in contact sports: Reliable change indices of impairment and recovery. *Journal of Clinical and Experimental Neuropsychology, 21*(1), 70–86.

Houck, Z. M., Asken, B., Clugston, J., Perlstein, W., & Bauer, R. (2018). Socioeconomic Status and Race Outperform Concussion History and Sport Participation in Predicting Collegiate Athlete Baseline Neurocognitive Scores. *Journal of the International Neuropsychological Society, 24*(1), 1–10.

Houck, Z. M., Asken, B. M., Bauer, R. M., Kontos, A. P., McCrea, M. A., McAllister, T. W., Broglio, S. P., & Clugston, J. R. (2019). Multivariate Base Rates of Low Scores and Reliable Decline on ImPACT in Healthy Collegiate Athletes Using CARE Consortium Norms. *Journal of the International Neuropsychological Society*, 1–11.

Houston, M. N., Van Pelt, K. L., D'Lauro, C., Brodeur, R. M., Campbell, D. E., McGinty, G. T., Jackson, J. C., Kelly, T. F., Peck, K. Y., Svoboda, S. J., McAllister, T. W., McCrea, M. A., Broglio, S. P., & Cameron, K. L. (2021). Test-Retest Reliability of Concussion Baseline Assessments in United States Service Academy Cadets: A Report from the National Collegiate Athletic Association (NCAA)-Department of Defense (DoD) CARE Consortium. *Journal of the International Neuropsychological Society, 27*(1), 23–34.

Iverson, G. L., Brooks, B. L., Collins, M. W., & Lovell, M. R. (2006). Tracking neuropsychological recovery following concussion in sport. *Brain Injury, 20*(3), 245–252.

Iverson, G. L., Lovell, M. R., & Collins, M. W. (2003). Interpreting change on ImPACT following sport concussion. *Clinical Neuropsychologist, 17*(4), 460–467.

Iverson, G. L., Lovell, M. R., & Collins, M. W. (2005). Validity of ImPACT for measuring processing speed following sports-related concussion. *Journal of Clinical and Experimental Neuropsychology, 27*(6), 683–689.

Iverson, G. L., & Schatz, P. (2015). Advanced topics in neuropsychological assessment following sport-related concussion. *Brain Injury, 29*(2), 263–275.

Joseph, K. R., Mendoza-Puccini, C., Esterlitz, J. R., Gay, K. E., Sheikh, M., & Bellgowan, P. (2018). Streamlining clinical research: The National Institute of Neurological Disorders and Stroke (NINDS), National Institutes of Health (NIH) and Department of Defense (DoD) sport-related concussion common data element (CDE) recommendations. *Neurology, 91*(23 Supplement 1), S24-S24.

Kerr, Z. Y., Wilkerson, G. B., Caswell, S. V., Currie, D. W., Pierpoint, L. A., Wasserman, E. B., Knowles, S. B., Dompier, T. P., Comstock, R. D., & Marshall, S. W. (2018). The First Decade of Web-Based Sports Injury Surveillance: Descriptive Epidemiology of Injuries in United States High School Football (2005–2006 Through 2013–2014) and National Collegiate Athletic Association Football (2004–2005 Through 2013–2014). *Journal of Athletic Training, 53*(8), 738–751.

Lau, B. C., Collins, M. W., & Lovell, M. R. (2011). Sensitivity and specificity of subacute computerized neurocognitive testing and symptom evaluation in predicting outcomes after sports-related concussion. *American Journal of Sports Medicine, 39*(6), 1209–1216.

Lau, B. C., Collins, M. W., & Lovell, M. R. (2012). Cutoff scores in neurocognitive testing and symptom clusters that predict protracted recovery from concussions in high school athletes. *Neurosurgery, 70*(2), 371–379.

Louey, A. G., Cromer, J. A., Schembri, A. J., Darby, D. G., Maruff, P., Makdissi, M., & McCrory, P. (2014). Detecting cognitive impairment after concussion: Sensitivity of change from baseline and normative data methods using the CogSport/Axon cognitive test battery. *Archives of Clinical Neuropsychology, 29*(5), 432–441.

Lovell, M. R., Collins, M. W., Iverson, G. L., Field, M., Maroon, J. C., Cantu, R., Podell, K., Powell, J. W., Belza, M., & Fu, F. H. (2003). Recovery from mild concussion in high school athletes. *Journal of Neurosurgery, 98*(2), 296–301.

Lovell, M. R., Collins, M. W., Iverson, G. L., Johnston, K. M., & Bradley, J. P. (2004). Grade 1 or "ding" concussions in high school athletes. *American Journal of Sports Medicine, 32*(1), 47–54.

Lovell, M. R., & Solomon, G. S. (2011). Psychometric data for the NFL neuropsychological test battery. *Applied Neuropsychology, 18*(3), 197–209.

Lovell, M. R., & Solomon, G. S. (2013). Neurocognitive test performance and symptom reporting in cheerleaders with concussions. *Journal of Pediatrics, 163*(4), 1192-1195.e1191.

Lynall, R. C., Schmidt, J. D., Mihalik, J. P., & Guskiewicz, K. M. (2016). The Clinical Utility of a Concussion Rebaseline Protocol After Concussion Recovery. *Clinical Journal of Sport Medicine, 26*(4), 285–290.

Macciocchi, S. N., Barth, J. T., Alves, W., Rimel, R. W., & Jane, J. A. (1996). Neuropsychological functioning and recovery after mild head injury in collegiate athletes. *Neurosurgery, 39*(3), 510–514.

MacDonald, J., & Duerson, D. (2015). Reliability of a Computerized Neurocognitive Test in Baseline Concussion Testing of High School Athletes. *Clinical Journal of Sport Medicine, 25*(4), 367–372.

Maddocks, D., & Saling, M. (1996). Neuropsychological deficits following concussion. *Brain Injury, 10*(2), 99–103.

Maerlender, A., Flashman, L., Kessler, A., Kumbhani, S., Greenwald, R., Tosteson, T., & McAllister, T. (2010). Examination of the Construct Validity of Impact Computerized Test, Traditional, and Experimental Neuropsychological Measures. *Clinical Neuropsychologist, 24*(8), 1309–1325.

Maerlender, A., Flashman, L., Kessler, A., Kumbhani, S., Greenwald, R., Tosteson, T., & McAllister, T. (2013). Discriminant construct validity of ImPACT™: A companion study. *Clinical Neuropsychologist, 27*(2), 290–299.

Maerlender, A., & Molfese, D. L. (2015). Repeat baseline assessment in college-age athletes. *Developmental Neuropsychology, 40*(2), 69–73.

Maerlender, A., Smith, E., Brolinson, P. G., Urban, J., Rowson, S., Ajamil, A., Campolettano, E. T., Gellner, R. A., Bellamkonda, S., Kelley, M. E., Jones, D., Powers, A., Beckwith, J., Crisco, J., Stitzel, J., Duma, S., & Greenwald, R. M. (2020). Psychometric properties of the standardized assessment of concussion in youth football: Validity, reliability, and demographic factors. *Applied Neuropsychology: Child*, 1–7.

Maerlender, A. C., Masterson, C. J., James, T. D., Beckwith, J., Brolinson, P. G., Crisco, J., Duma, S., Flashman, L. A., Greenwald, R., Rowson, S., Wilcox, B., & McAllister, T. W. (2016). Test–retest, retest, and retest: Growth curve models of repeat testing with Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT). *Journal of Clinical and Experimental Neuropsychology, 38*(8), 869–874.

Maietta, J. E., Ahmed, A. O., Barchard, K. A., Kuwabara, H. C., Donohue, B., Ross, S. R., Kinsora, T. F., & Allen, D. N. (2021). Confirmatory factor analysis of imPACT cognitive tests in high school athletes. *Psychological Assessment, 13*, 13.

Makdissi, M., Collie, A., Maruff, P., Darby, D. G., Bush, A., McCrory, P., & Bennell, K. (2001). Computerised cognitive assessment of concussed Australian Rules footballers. *British Journal of Sports Medicine, 35*(5), 354–360.

Makdissi, M., Darby, D., Maruff, P., Ugoni, A., Brukner, P., & McCrory, P. R. (2010). Natural History of Concussion in Sport. *American Journal of Sports Medicine, 38*(3), 464–471.

Marcopulos, B., & Lojek, E. (2019). Introduction to the special issue: Are modern neuropsychological assessment methods really "modern"? Reflections on the current neuropsychological test armamentarium. *The Clinical Neuropsychologist, 33*(2), 187–199.

Mason, S. J., Davidson, B. S., Lehto, M., Ledreux, A., Granholm, A. C., & Gorgens, K. A. (2020). A Cohort Study of the Temporal Stability of ImPACT Scores Among NCAA Division I Collegiate Athletes: Clinical Implications of Test-Retest Reliability for Enhancing Student Athlete Safety. *Archives of Clinical Neuropsychology, 26*, 26.

Masterson, C. J., Tuttle, J., & Maerlender, A. (2019). Confirmatory factor analysis of two computerized neuropsychological test batteries: Immediate post-concussion assessment and cognitive test

(ImPACT) and C3 logix. *Journal of Clinical and Experimental Neuropsychology, 41*(9), 925–932.

McClincy, M. P., Lovell, M. R., Pardini, J., Collins, M. W., & Spore, M. K. (2006). Recovery from sports concussion in high school and collegiate athletes. *Brain Injury, 20*(1), 33–39.

McCrea, M., Barr, W. B., Guskiewicz, K., Randolph, C., Marshall, S. W., Cantu, R., Onate, J. A., & Kelly, J. P. (2005). Standard regression-based methods for measuring recovery after sport-related concussion. *Journal of the International Neuropsychological Society, 11*(1), 58–69.

McCrea, M., Guskiewicz, K., Randolph, C., Barr, W. B., Hammeke, T. A., Marshall, S. W., Powell, M. R., Woo Ahn, K., Wang, Y., & Kelly, J. P. (2013). Incidence, clinical course, and predictors of prolonged recovery time following sport-related concussion in high school and college athletes. *Journal of the International Neuropsychological Society, 19*(1), 22–33.

McCrea, M., Guskiewicz, K. M., Marshall, S. W., Barr, W., Randolph, C., Cantu, R. C., Onate, J. A., Yang, J., & Kelly, J. P. (2003). Acute effects and recovery time following concussion in collegiate football players: The NCAA Concussion Study. *JAMA, 290*(19), 2556–2563.

McCrea, M., Meier, T., Huber, D., Ptito, A., Bigler, E., Debert, C. T., Manley, G., Menon, D., Chen, J. K., Wall, R., Schneider, K. J., & McAllister, T. (2017). Role of advanced neuroimaging, fluid biomarkers and genetic testing in the assessment of sport-related concussion: A systematic review. *British Journal of Sports Medicine, 51*(12), 919–929.

McCrory, P., Meeuwisse, W., Dvorak, J., Aubry, M., Bailes, J., Broglio, S., Cantu, R. C., Cassidy, D., Echemendia, R. J., Castellani, R. J., Davis, G. A., Ellenbogen, R., Emery, C., Engebretsen, L., Feddermann-Demont, N., Giza, C. C., Guskiewicz, K. M., Herring, S., Iverson, G. L., … Vos, P. E. (2017). Consensus statement on concussion in sport-the 5(th) international conference on concussion in sport held in Berlin, October 2016. *British Journal of Sports Medicine, 51*(11), 838–847.

McSweeny, A. J., Naugle, R. I., Chelune, G. J., & Luders, H. (1993). "T Scores for Change": An illustration of a regression approach to depicting change in clinical neuropsychology. *The Clinical Neuropsychologist, 7*(3), 300–312.

Meehan, W. P., 3rd., d'Hemecourt, P., Collins, C. L., Taylor, A. M., & Comstock, R. D. (2012). Computerized neurocognitive testing for the management of sport-related concussions. *Pediatrics, 129*(1), 38–44.

Merz, Z. C., Lichtenstein, J. D., & Lace, J. W. (2021). Methodological considerations of assessing meaningful/reliable change in computerized neurocognitive testing following sport-related concussion. *Applied Neuropsychology: Child*, 1–9.

Meyer, J. E., & Arnett, P. A. (2015). Validation of the Affective Word List as a measure of verbal learning and memory. *Journal of Clinical and Experimental Neuropsychology, 37*(3), 316–324.

Miller, J. R., Adamson, G. J., Pink, M. M., & Sweet, J. C. (2007). Comparison of preseason, midseason, and postseason neurocognitive scores in uninjured collegiate football players. *American Journal of Sports Medicine, 35*(8), 1284–1288.

Mitrushina, M. N., Boone, K. B., Razani, J., & D'Ellia, L. F. (2005). *Handbook of normative data for neuropsychological assessment* (2nd ed.). Oxford University Press.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *British Medical Journal, 339*, b2535.

Moser, R. S., & Schatz, P. (2002). Enduring effects of concussion in youth athletes. *Archives of Clinical Neuropsychology, 17*(1), 91–100.

Moser, R. S., Schatz, P., Grosner, E., & Kollias, K. (2017). One year test-retest reliability of neurocognitive baseline scores in 10- to 12-year olds. *Applied Neuropsychology: Child, 6*(2), 166–171.

Moser, R. S., Schatz, P., & Lichtenstein, J. D. (2015). The importance of proper administration and interpretation of neuropsychological baseline and postconcussion computerized testing. *Archives of Clinical Neuropsychology, 4*(1), 41–48.

Nelson, L. D. (2015). False-Positive Rates of Reliable Change Indices for Concussion Test Batteries: A Monte Carlo Simulation. *Journal of Athletic Training, 50*(12), 1319–1322.

Nelson, L. D., LaRoche, A. A., Pfaller, A. Y., Lerner, E. B., Hammeke, T. A., Randolph, C., Barr, W. B., Guskiewicz, K., & McCrea, M. A. (2016). Prospective, Head-to-Head Study of Three Computerized Neurocognitive Assessment Tools (CNTs): Reliability and Validity for the Assessment of Sport-Related Concussion. *Journal of the International Neuropsychological Society, 22*(1), 24–37.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3rd ed.). McGraw-Hill.

O'Brien, A. M., Casey, J. E., & Salmon, R. M. (2018). Short-term test-retest reliability of the ImPACT in healthy young athletes. *Applied Neuropsychology: Child, 7*(3), 208–216.

Onate, J. A., Guskiewicz, K. M., Riemann, B. L., & Prentice, W. E. (2000). A comparison of sideline versus clinical cognitive test performance in collegiate athletes. *Journal of Athletic Training, 35*(2), 155–160.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hrobjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*(n71), 1–9.

Parsons, T. D. (2011). Neuropsychological Assessment Using Virtual Environments: Enhanced Assessment Technology for Improved Ecological Validity. In S. Brahnam, & L. C. Jain (Eds.), *Advanced Computational Intelligence Paradigms in Healthcare* (pp. 6). Springer.

Pearce, A. J., Hoy, K., Rogers, M. A., Corp, D. T., Davies, C. B., Maller, J. J., & Fitzgerald, P. B. (2015). Acute motor, neurocognitive and neurophysiological change following concussion injury in Australian amateur football. A prospective multimodal investigation. *Journal of Science and Medicine in Sport, 18*(5), 500–506.

Pedersen, H. A., Ferraro, F. R., Himle, M., Schultz, C., & Poolman, M. (2014). Neuropsychological factors related to college ice hockey concussions. *American Journal of Alzheimer's Disease & Other Dementias, 29*(3), 201–204.

Peterson, C. L., Ferrara, M. S., Mrazik, M., Piland, S., & Elliott, R. (2003). Evaluation of neuropsychological domain scores and postural stability following cerebral concussion in sports. *Clinical Journal of Sport Medicine, 13*(4), 230–237.

Randolph, C., McCrea, M., & Barr, W. B. (2005). Is neuropsychological testing useful in the management of sport-related concussion? *Journal of Athletic Training, 40*(3), 139–152.

Register-Mihalik, J. K., Guskiewicz, K. M., Mihalik, J. P., Schmidt, J. D., Kerr, Z. Y., & McCrea, M. A. (2013). Reliable change, sensitivity, and specificity of a multidimensional concussion assessment battery: Implications for caution in clinical practice. *Journal of Head Trauma Rehabilitation, 28*(4), 274–283.

Register-Mihalik, J. K., Kontos, D. L., Guskiewicz, K. M., Mihalik, J. P., Conder, R., & Shields, E. W. (2012). Age-related differences and reliability on computerized and paper-and-pencil neurocognitive assessment batteries. *Journal of Athletic Training, 47*(3), 297–305.

Resch, J. E., Brown, C. N., Schmidt, J., Macciocchi, S. N., Blueitt, D., Cullum, C. M., & Ferrara, M. S. (2016). The sensitivity and specificity of clinical measures of sport concussion: Three tests are better than one. *BMJ Open Sport & Exercise Medicine, 2*(1), e000012.

Resch, J. E., Macciocchi, S., & Ferrara, M. S. (2013a). Preliminary evidence of equivalence of alternate forms of the ImPACT. *The Clinical Neuropsychologist, 27*(8), 1265–1280.

Resch, J. E., McCrea, M. A., & Cullum, C. M. (2013b). Computerized neurocognitive testing in the management of sport-related concussion: An update. *Neuropsychology Review, 23*(4), 335–349.

Salvatore, A. P., Cannito, M., Brassil, H. E., Bene, E. R., & Sirmon-Taylor, B. (2017). Auditory comprehension performance of college students with and without sport concussion on Computerized-Revised Token Test Subtest VIII. *Concussion, 2*(2), CNC37.

Schatz, P. (2010). Long-term test-retest reliability of baseline cognitive assessments using ImPACT. *American Journal of Sports Medicine, 38*(1), 47–53.

Schatz, P., & Maerlender, A. (2013). A two-factor theory for concussion assessment using ImPACT: Memory and speed. *Archives of Clinical Neuropsychology, 28*(8), 791–797.

Schatz, P., Pardini, J. E., Lovell, M. R., Collins, M. W., & Podell, K. (2006). Sensitivity and specificity of the ImPACT Test Battery for concussion in athletes. *Archives of Clinical Neuropsychology, 21*(1), 91–99.

Schatz, P., & Robertshaw, S. (2014). Comparing post-concussive neurocognitive test data to normative data presents risks for underclassifying "above average" athletes. *Archives of Clinical Neuropsychology, 29*(7), 625–632.

Schatz, P., & Sandel, N. (2012). Sensitivity and specificity of the online version of ImPACT in high school and collegiate athletes. *American Journal of Sports Medicine, 41*(2), 321–326.

Sherry, N. S., Fazio-Sumrok, V., Sufrinko, A., Collins, M. W., & Kontos, A. P. (2019). Multimodal Assessment of Sport-Related Concussion. *Clinical Journal of Sport Medicine, 18*, 18.

Sim, A., Terryberry-Spohr, L., & Wilson, K. R. (2008). Prolonged recovery of memory functioning after mild traumatic brain injury in adolescent athletes. *Journal of Neurosurgery, 108*(3), 511–516.

Simon, M., Maerlender, A., Metzger, K., Decoster, L., Hollingworth, A., & McLeod, T. V. (2017). Reliability and Concurrent Validity of Select C3 Logix Test Components. *Developmental Neuropsychology, 42*(7), 446–459.

Slick, D. J. (2006). Psychometrics in neuropsychological assessment. In E. Strauss, E. M. S. Sherman, & O. Spreen (Eds.), *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed., pp. 3–43). Oxford University Press.

Solomon, G. S., Lovell, M. R., Casson, I. R., & Viano, D. C. (2015). Normative neurocognitive data for National Football League players: An initial compendium. *Archives of Clinical Neuropsychology, 30*(2), 161–173.

Sosnoff, J. J., Broglio, S. P., Hillman, C. H., & Ferrara, M. S. (2007). Concussion does not impact intraindividual response time variability. *Neuropsychology, 21*(6), 796–802.

Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology, 5*, 1–25.

Sufrinko, A., McAllister-Deitrick, J., Womble, M., & Kontos, A. (2017). Do Sideline Concussion Assessments Predict Subsequent Neurocognitive Impairment After Sport-Related Concussion? *Journal of Athletic Training, 52*(7), 676–681.

Thoma, R. J., Cook, J. A., McGrew, C., King, J. H., Pulsipher, D. T., Yeo, R. A., Monnig, M. A., Mayer, A., Pommy, J., & Campbell, R. A. (2018). Convergent and discriminant validity of the ImPACT with traditional neuropsychological measures. *Cogent Psychology, 5*(1).

Tsushima, W. T., Shirakawa, N., & Geling, O. (2013). Neurocognitive functioning and symptom reporting of high school athletes following a single concussion. *Applied Neuropsychology: Child, 2*(1), 13–16.

Tsushima, W. T., Siu, A. M., Pearce, A. M., Zhang, G., & Oshiro, R. S. (2016). Two-year Test-Retest Reliability of ImPACT in High

School Athletes. *Archives of Clinical Neuropsychology, 31*(1), 105–111.

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.

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. *American Journal of Sports Medicine, 34*(10), 1630–1635.

Volberding, J. L., & Melvin, D. (2014). Changes in ImPACT and Graded Symptom Checklist Scores During a Division I Football Season. *Athletic Training & Sports Health Care: THe Journal for the Practicing Clinician, 6*(4), 155–160.

Wallace, J., Moran, R., Beidler, E., McAllister Deitrick, J., Shina, J., & Covassin, T. (2020). Disparities on Baseline Performance Using Neurocognitive and Oculomotor Clinical Measures of Concussion. *The American Journal of Sports Medicine, 48*(11), 2774–2782.

Walton, S. R., Broshek, D. K., Freeman, J. R., Cullum, C. M., & Resch, J. E. (2018). Valid but Invalid: Suboptimal ImPACT Baseline Performance in University Athletes. *Medicine and Science in Sports and Exercise, 50*(7), 1377–1384.

Weissberger, G. H., Strong, J. V., Stefanidis, K. B., Summers, M. J., Bondi, M. W., & Stricker, N. H. (2017). Diagnostic Accuracy of Memory Measures in Alzheimer's Dementia and Mild Cognitive Impairment: A Systematic Review and Meta-Analysis. *Neuropsychology Review, 27*(4), 354–388.

Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., Leeflang, M. M., Sterne, J. A., & Bossuyt, P. M. (2011). QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine, 155*(8), 529–536.

Wilmoth, K., Magnus, B. E., McCrea, M. A., & Nelson, L. D. (2020). Preliminary Validation of an Abbreviated Acute Concussion Symptom Checklist Using Item Response Theory. *The American Journal of Sports Medicine, 48*(12), 3087–3093.

Womble, M. N., Reynolds, E., Schatz, P., Shah, K. M., & Kontos, A. P. (2016). Test-Retest Reliability of Computerized Neurocognitive Testing in Youth Ice Hockey Players. *Archives of Clinical Neuropsychology, 31*(4), 305–312.

Zuckerman, S. L., Lee, Y. M., Odom, M. J., Solomon, G. S., & Sills, A. K. (2013). Baseline neurocognitive scores in athletes with attention deficit-spectrum disorders and/or learning disability. *Journal of Neurosurgery: Pediatrics, 12*(2), 103–109.