Research Article

Factor structure and measurement invariance of post-concussion symptom ratings on the Health and Behaviour Inventory across time, raters, and groups: An A-CAP study

Cherri Zhang[1] ⓘ, Ken Tang[2], Roger Zemek[3,4], Miriam H. Beauchamp[5,6], William Craig[7], Quynh Doan[8], Keith Owen Yeates[1,9] and On Behalf of the Pediatric Emergency Research Canada A-CAP Study Team

[1]Department of Psychology, University of Calgary, Calgary, AB, Canada, [2]Independent Statistical Consultant, Richmond, BC, Canada, [3]Children's Hospital of Eastern Ontario Research Institute, Ottawa, ON, Canada, [4]Department of Pediatrics and Emergency Medicine, University of Ottawa, Ottawa, ON, Canada, [5]Department of Psychology, Université de Montréal, Montreal, QC, Canada, [6]Centre Hospitalier Universitaire Sainte-Justine, Montreal, QC, Canada, [7]Department of Pediatrics, University of Alberta, and Stollery Children's Hospital, Edmonton, AB, Canada, [8]Department of Pediatrics, University of British Columbia, and BC Children's Hospital Research Institute, Vancouver, BC, Canada and [9]Alberta Children's Hospital Research Institute and Hotchkiss Brain Institute, University of Calgary, Calgary, AB, Canada

Abstract

Objectives: To validate the two-factor structure (i.e., cognitive and somatic) of the Health and Behaviour Inventory (HBI), a widely used post-concussive symptom (PCS) rating scale, through factor analyses using bifactor and correlated factor models and by examining measurement invariance (MI). **Methods:** PCS ratings were obtained from children aged 8–16.99 years, who presented to the emergency department with concussion (n = 565) or orthopedic injury (OI) (n = 289), and their parents, at 10-days, 3-months, and 6-months post-injury. Item-level HBI ratings were analyzed separately for parents and children using exploratory and confirmatory factor analyses (CFAs). Bifactor and correlated models were compared using various fit indices and tested for MI across time post-injury, raters (parent vs. child), and groups (concussion vs. OI). **Results:** CFAs showed good fit for both a three-factor bifactor model, consisting of a general factor with two subfactors (i.e., cognitive and somatic), and a correlated two-factor model with cognitive and somatic factors, at all time points for both raters. Some results suggested the possibility of a third factor involving fatigue. All models demonstrated strict invariance across raters and time. Group comparisons showed at least strong or strict invariance. **Conclusions:** The findings support the two symptom dimensions measured by the HBI. The three-factor bifactor model showed the best fit, suggesting that ratings on the HBI also can be captured by a general factor. Both correlated and bifactor models showed substantial MI. The results provide further validation of the HBI, supporting its use in childhood concussion research and clinical practice.

Keywords: psychometrics; factor analysis; brain injuries; traumatic; child; parents; association

(Received 29 October 2021; final revision 12 March 2022; accepted 27 April 2022; First Published online 4 August 2022)

Introduction

Concussions, a form of mild traumatic brain injury (mTBI), occur in approximately 1–2 million children/adolescents yearly in the United States (Bryan et al., 2016), and are frequently associated with cognitive, somatic, behavioral, and emotional symptoms (Ayr et al., 2009). For approximately 1 in 3 children, these post-concussion symptoms (PCS) persist up to 1 month or longer (Barlow et al., 2010; Zemek et al., 2016). Persistent PCS have a negative impact on health-related quality of life and wellness (Beauchamp et al., 2019; Novak et al., 2016). Accurate identification of PCS is essential for appropriate management of concussion.

Current guidelines call for the use of age-appropriate and validated PCS rating scales to assist with diagnosis and monitoring recovery (Lumba-Brown et al., 2018; Reed et al., 2019). The Health and Behaviour Inventory (HBI) is a PCS rating scale that is embedded within the child version of the Sport Concussion Assessment Tool (Davis et al., 2017) and recommended in the NINDS Common Data Elements for pediatric TBI and sport-related concussion (Broglio et al., 2018; McCauley et al., 2012). The HBI, which is intended for use with 8–17 year old children, consists of 20 items separated into two subscales that reflect cognitive and somatic symptoms commonly reported after concussion (Ayr et al., 2009). The 20 items were chosen from a larger pool of 50 items based on the results of common factor analyses using target rotation (Browne, 2001), which identified cognitive and somatic symptom factors that consistently replicated based on both child and parent ratings and over time. The HBI has acceptable psychometric properties, including good-to-excellent internal consistency and moderate-to-good test–retest reliability (O'Brien et al., 2021), is sensitive to pediatric mild TBI and sport-related concussion

(Babl et al., 2017; Patsimas et al., 2020), predicts other outcomes such as quality of life (Moran et al., 2012; Yeates et al., 2012), and is sensitive to treatment of concussion (Hilt et al., 2022; McCarty et al., 2021).

Another approach to validating symptom rating scales such as the HBI is to examine their factor structure and invariance. Factor analyses (exploratory or confirmatory) are commonly used to identify and validate the dimensions of PCS rating scales because they offer empirical evidence for the validity of symptom interpretation and guide the development of subscales (Karr & Iverson, 2020). Identifying the dimensions of concussion symptoms also can offer insights into an individually tailored approach to treatment (Brett et al., 2020). To date, factor analyses of PCS rating scales have yielded somewhat inconsistent results, although factors involving cognitive, somatic, emotional, and sleep symptoms are typically reported (Brett et al., 2020; Nelson et al., 2018). No universally consistent dimensions have been identified, as past studies have found PCS ratings to be both unidimensional and multidimensional (Joyce et al., 2015; Merritt & Arnett, 2014; Piland et al., 2003; Sady, Vaughan, & Giogial, 2014; Waljas et al., 2012). In both clinical and research contexts, total scores are often used in place of more discrete subscales. The bifactor model has become an innovative approach to reconciling these inconsistencies, as it allows for both a dominant general factor and multi-faceted subfactors. In a bifactor model, a broad comprehensive factor directly accounts for each item on the symptom rating scale, and the residual item variance is accounted for by additional subfactors (Nelson et al., 2018).

To validate factor models, measurement invariance (MI) can be assessed to test the applicability of the factor model across time, raters, or groups, by imposing stringent assumptions on the equality of model parameters. Tests of these assumptions can offer insights into the nature of underlying constructs and their measurement (Agtarap et al., 2020; Brett et al., 2020); importantly, confirmation of MI ensures that researchers and clinicians can be confident that scores on PCS rating scales are comparable across time, raters, and different injury types. The four levels of MI are: (1) *configural*, which requires equivalence of factor model structure (i.e., items load onto the same factors); (2) *weak*, which requires equivalence of factor loadings; (3) *strong*, which requires equivalence of item thresholds or intercepts; and (4) *strict*, which requires equivalence of item residuals or unique variances (Putnick & Bornstein, 2016).

The factor structure of the HBI has not been validated since its original derivation (Ayr et al., 2009) and no research has investigated its MI. Therefore, this study had three main goals: (1) to validate the two dimensions of the HBI identified in previous research using both exploratory and confirmatory approaches, (2) to test whether a bifactor model with a general factor and cognitive and somatic subfactors achieves better fit than the correlated two-factor model, and (3) to evaluate MI for all factor models across time (10-days, 3-months, and 6-months post-injury), raters (child vs. parent), and groups (concussion vs. orthopedic injury [OI]). Analyses used data collected from a prospective multicenter, longitudinal cohort study of children aged 8–16.99 years who presented to emergency departments (EDs) at five pediatric hospitals across Canada with either concussion or OI (Yeates et al., 2017). We predicted that analyses would confirm the two-dimensional model of the HBI, but anticipated that a bifactor model might provide a better overall fit than a correlated two-factor model. We also expected to find at least weak MI across all comparisons (time, rater, group) for both bifactor and correlated two-factor models.

Methods

Participants

The study included 967 children and adolescents aged 8 to 16.99 with mTBI or OI. They were recruited for the Advancing Concussion Assessment in Pediatrics (A-CAP) study (Yeates et al., 2017), conducted at five sites from the Pediatric Emergency Research Canada network (Bialy et al., 2018): Alberta Children's Hospital (Calgary, Alberta), Children's Hospital of Eastern Ontario (Ottawa, Ontario), Sainte-Justine Hospital (Montreal, Quebec), Stollery Children's Hospital (Edmonton, Alberta), and British Columbia Children's Hospital (Vancouver, British Columbia).

Children were eligible for the concussion group if they presented to the ED within 48 h of sustaining a blunt head trauma and met at least one of the three following criteria, consistent with the WHO definition of mTBI (Carroll et al., 2004): (1) an observed loss of consciousness (LOC), (2) a Glasgow Coma Scale (GCS) score of 13 or 14, or (3) at least one acute sign or symptom of concussion such as post-traumatic amnesia, focal neurological deficits, skull fracture, post-traumatic seizure, vomiting, headache, dizziness, and other mental status changes. Children were eligible for the OI group if they sustained upper or lower extremity fractures, sprains, or strains due to blunt force/physical trauma, associated with a score of 4 or less on the Abbreviated Injury Scale (Medicine, 1990), within 48 h of presentation to the ED. Exclusion criteria for the concussion group were delayed neurological deterioration as indicated by a GCS score less than 13, LOC > 30 min, or post-traumatic amnesia > 24 h. OI group exclusion criteria included any head trauma or symptoms of concussion reported during screening for recruitment, as well as any injury requiring surgical intervention or procedural sedation. For both groups, additional exclusion criteria were hypoxia, hypotension, or shock during or following the injury; non-English-speaking child or parents (non-English and non-French-speaking in Quebec or Ottawa); previous TBI requiring overnight hospitalization; previous concussion within the past 3 months; previous severe neurological or neurodevelopmental disorder such as epilepsy, intellectual disability, or autism (history of attention deficit hyperactivity disorder, learning disability, or Tourette's syndrome was not an exclusion); hospitalization in the previous year for psychiatric disorder; administration of sedative medication prior to ED data collection (fentanyl was not an exclusion if used for pain management only); obvious alcohol or drug ingestion associated with injury; injury related to abuse or assault; and legal guardian not present or child in foster care. Table 1 shows demographic and injury characteristics of the sample.

Measures

The HBI consists of 20 items (Table 2), with each item rated on a 4-point scale (0 = Never, 1 = Rarely, 2 = Sometimes, 3 = Often). The items form two subscales, with the first 11 constituting the cognitive scale and the remaining 9 constituting the somatic scale (Ayr et al., 2009). The wording of child and parent proxy forms differ slightly to reflect first- versus third-person perspective, but the items are otherwise equivalent.

Table 1. Sample demographic and injury characteristics

	mTBI (N = 565)	OI (N = 289)		mTBI (N = 565)	OI (N = 289)
Demographics	n (%) or mean (SD)		Injury characteristics	n (%) or median (IQR)	
Age (y)	12.38 (2.52)	12.58 (2.21)	Previous concussion		
Sex: female	228 (40.35%)	132 (45.67%)	0	382 (67.97%)	202 (70.38%)
Child's race			1	110 (19.57%)	52 (18.12%)
White	385 (69.75%)	195 (70.4%)	≥2	68 (12.1%)	31(10.8%)
Indigenous	13 (2.36%)	3 (1.08%)	Unknown	2 (0.36%)	2 (0.7%)
Asian	51 (9.24%)	22 (7.94%)	Mechanism of injury		
Hispanic	13 (2.36%)	12 (4.33%)	Assault	4 (0.71%)	1 (0.35%)
Black	21 (3.8%)	7 (2.53%)	Fall	250 (44.48%)	147 (51.04%)
Other/multi-racial	69 (12.5%)	38 (13.72%)	Struck object	177 (31.49%)	60 (20.83%)
Parent's education			Struck person	100 (17.79%)	29 (10.07%)
High school or less	85 (15.54%)	42 (15.44%)	Bicycle related	9 (1.6%)	12 (4.17%)
Trade school/2-year college	164 (29.98%)	81 (29.78%)	MVC	9 (1.6%)	0
University bachelor's degree	205 (37.48%)	93 (34.19%)	Other	6 (1.07%)	21 (7.29%)
Higher than bachelor's degree	93 (17%)	56 (20.59%)	Not documented	7 (1.25%)	18 (6.25%)
Site			Injury setting		
Calgary	136 (24.07%)	62 (21.45%)	Sports/recreation	460 (81.85%)	237 (83.16%)
Edmonton	110 (19.47%)	59 (20.42%)	Non-sport related	102 (18.15%)	48 (16.84%)
Montreal	62 (10.97%)	25 (8.65%)	Reliable change cognitive		
Ottawa	140 (24.78%)	47 (16.26%)	Child PA[1]	226 (41.17%)	23 (8.33%)
Vancouver	117 (20.71%)	96 (33.22%)	Child 3M[2]	58 (12.21%)	18 (7.5%)
			Child 6M[3]	47 (10.4%)	19 (7.79%)
			Parent PA[4]	140 (25.41%)	14 (5.05%)
			Parent 3M[5]	49 (10.47%)	13 (5.37%)
			Parent 6M[6]	48 (10.61%)	18 (8%)
			Reliable change somatic		
			Child PA[1]	294 (53.55%)	17 (6.16%)
			Child 3M[2]	76 (16%)	15 (6.25%)
			Child 6M[3]	72 (15.93%)	19 (8.23%)
			Parent PA[4]	326 (59.17%)	17 (6.14%)
			Parent 3M[7]	61 (12.95%)	11 (4.56%)
			Parent 6M[8]	46 (10.31%)	18 (8%)
			Loss of consciousness[9]	88 (16.79%)	2 (0.69%)
			GCS < 15[10]	9 (1.67%)	0
			Number of acute symptoms in ED (possible range 0–22)	8 (6–11)	1.5 (1–3)

mTBI = mild traumatic brain injury; OI = orthopedic injury; SD = standard deviation; PA = post-acute; 3M = 3 month; 6M = 6 month. ED = emergency department.
[1] mTBI = 549, OI = 276.
[2] mTBI = 475, OI = 240.
[3] mTBI = 452, OI = 231.
[4] mTBI = 551, OI = 277.
[5] mTBI = 468, OI = 242.
[6] mTBI = 443, OI = 225.
[7] mTBI = 471, OI = 241.
[8] mTBI = 446, OI = 225.
[9] mTBI = 524, OI = 288.
[10] mTBI = 538.

Table 2. Health and behavior inventory subscale items

Cognitive	Somatic
1. Has trouble sustaining attention	12. Has headaches
2. Is easily distracted	13. Feels dizzy
3. Has difficulty concentrating	14. Has a feeling that the room is spinning
4. Has problems remembering what he/she is told	15. Feels faint
5. Has difficulty following directions	16. Has blurred vision
6. Tends to daydream	17. Has double vision
7. Gets confused	18. Experiences nausea
8. Is forgetful	19. Gets tired a lot
9. Has difficulty completing tasks	20. Gets tired easily
10. Has poor problem-solving skills	
11. Has problems learning	

Procedures

The study was approved by the research ethics boards of all participating institutions. Parents and capable adolescents provided written informed consent, and all other children provided written assent. Designated staff screened for all eligible participants who presented to the EDs. Recruitment took place from September 2016 to December 2018. Figure 1 shows a total of 3051 participants were eligible, and 967 (32%) consented (644 with concussion and 334 with OI).

HBI ratings were directly entered into REDCap databases by both the parent and the child during three face-to-face follow-up visits post-injury. A post-acute assessment was completed within three weeks of the injury (M = 8.39 days, SD = 3.13) by 828 (86%) parents and 829 (86%) children. Ratings were completed by 722 (75%) parents and 728 (75%) children at 3-months post-injury (M = 96.21 days, SD = 9.74) and by 685 (71%) parents and 701 (72%) children at 6-months post-injury (M = 186.06 days,
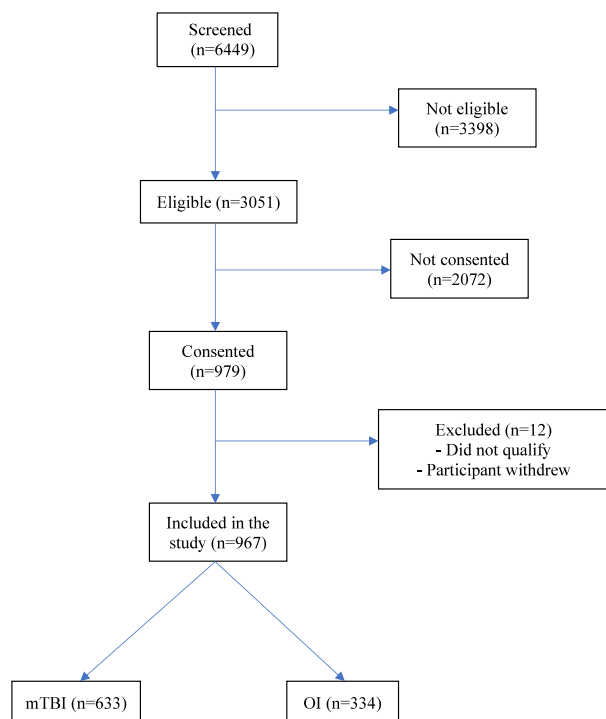
Figure 1. Participant enrollment. mTBI = mild traumatic brain injury, OI = orthopedic injury.

SD = 11.49). Of the parent ratings, 75% were completed by the mother, 21% by the father, and the remaining 4% by adoptive, step, and grandparents. Item-level missingness was less than 5% for the parent's rating at the 3- and 6-month assessments, with no missingness at the first follow-up post-injury visit. No item-level missingness was detected in children's ratings at any time point.

Statistical analysis

Factor analyses

Based on the ordinal nature of the HBI ratings, polychoric correlations were computed to examine inter-item correlation (Agtarap et al., 2020; Karr & Iverson, 2020; Rodriguez et al., 2016) Next, we conducted exploratory factor analyses (EFA) to verify the two factors identified in previous research (Ayr et al., 2009). Because the HBI is intended to measure PCS after concussion, EFAs were conducted separately for mTBI and OI groups and for child and parent ratings. Models adopted Geomin rotation, an oblique method that allows correlations between factors (Agtarap et al., 2020; Brett et al., 2020; Nelson et al., 2018). Factors with eigenvalues > 1, scree plot inflections, and high factor loadings (> 0.7) were all considered to determine candidate models. Confirmatory factor analyses (CFAs) included correlated two-factor and three-factor models, as well as a three-factor bifactor model using two subfactors for the cognitive and somatic items (Figure 2). A four-factor bifactor model could not be examined in confirmatory analyses due to convergence issues. Factors were set to be orthogonal in the bifactor model but were allowed to correlate with one another in the correlated models.

Models were compared using multiple fit indices. Absolute fit was assessed using the root mean square error of approximation (RMSEA). Incremental fit was evaluated using comparative fit index (CFI) and Tucker-Lewis index (TLI). Models with

RMSEA < 0.05 and CFI/TLI > 0.95 were considered good fit, and RMSEA < 0.08 and CFI/TLI > 0.9 were considered acceptable fit. A weighted least squares with mean and variance adjusted (WLSMV) estimator was incorporated to accommodate for the ordinal nature of the HBI items (Agtarap et al., 2020; Karr & Iverson, 2020; Nelson et al., 2018). Theta parameterization was used to allow for the inclusion of residual variances as parameters to permit testing strict MI (Karr & Iverson, 2020). Model-based reliability measures of omega, omega hierarchical (omegaH), and relative omega were computed, with a particular focus on omegaH, to estimate the proportion of observed variance in total and subscale scores that could be attributed to the underlying general and subfactors (Reise, 2012). In addition, Pearson correlations were computed between predicted factor scores and calculated HBI scores to explore their association.

Measurement invariance

MI is a stepwise approach that compares nested models by adding increasingly stringent constraints to model parameters. The analyses test whether a given scale measures the same construct at different time points or between different groups of people. We tested MI across time (post-acute vs. 3-months vs. 6-months), raters (parent vs. child), and groups (concussion vs. OI). MI is defined as four progressively more stringent levels: configural, weak, strong, and strict. In *configural* invariance, the pattern of factor loading is consistent across time/groups/raters, and each item should load most highly onto the same respective factor. *Weak* invariance constrains item loadings on the factors to be equal across time/groups/raters. It allows for the comparison of factor variances and covariances, meaning the proportion of variance in each HBI item accounted for by each factor is similar over time and across groups/raters. *Strong* invariance constrains item thresholds to be equivalent across time/groups/raters. Factor means between groups or within-person mean factor scores across time can be compared at this level, with differences in factor means reflecting true differences in the construct measured. Lastly, *strict* invariance further constrains the residual variances of items to be the same across time/groups/raters. In this case, differences in item parameters (i.e., means, variance, and covariance) are entirely attributable to differences in factor means over time or between groups/raters. Because of model complexity, and to ensure convergence, the test of configural invariance across time for the correlated two-factor model had to be defined differently for parent versus child ratings; no constraints were necessary for the analysis based on child ratings, but one item was constrained to be equal across time for the analysis based on parent ratings. Given the constraint at the configural level is consistent with weak, strong, and strict invariance, our subsequent tests for MI are valid.

MI is established when the fit of the more stringent model (e.g., strong invariance) is no worse compared to the one with more relaxed assumptions (e.g., weak invariance), as indicated by fit statistics. Consistent with prior studies, we rejected the more restricted model if CFI decreased by > 0.01 or RMSEA increased by > 0.015. Chi-square differences were not calculated due to their oversensitivity in large samples and a lack of an appropriate test to handle scaled chi-square differences from the WLSMV estimators. We examined factor scores once strong invariance was achieved to assess differences in mean scores between raters, groups, and time points; factor scores were estimated based on the strict invariance model and expressed in standardized units (i.e., Z scores). To test for group MI, the two highest categories in symptom ratings had to be combined for a few items because the OI group did not select

Correlated two-factor model
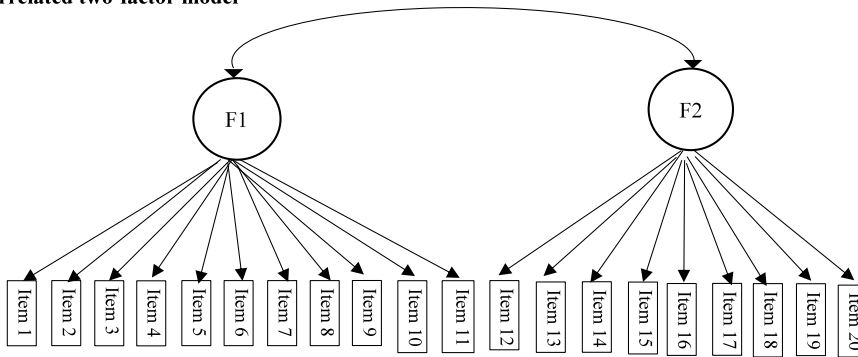
Correlated three-factor model
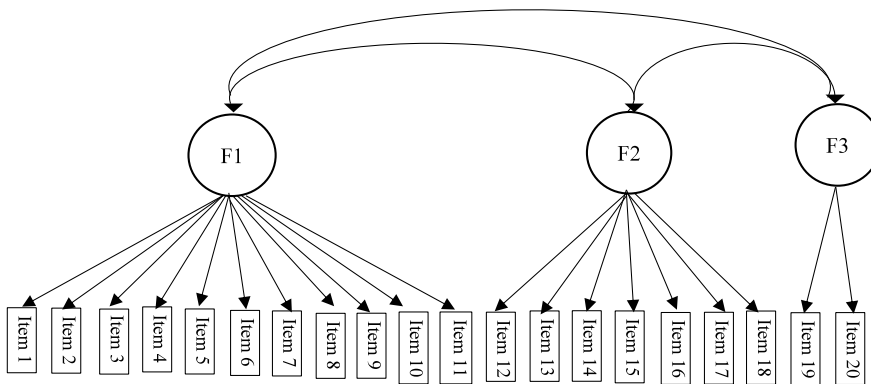
Bifactor model with two specific factors
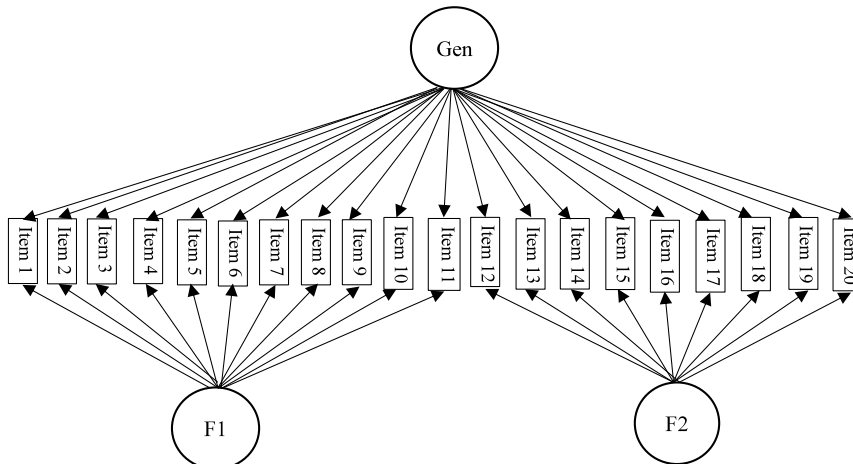
Figure 2. Factor models.

"often" as a response. However, categories were collapsed for only three items for parent ratings at all three time points and two and one item for child ratings at 3 and 6 months, respectively. Factor analyses were implemented in R (version 4.0.3) and descriptive analyses were conducted in Stata (version 15.0/MP, StataCorp, College Station, Texas).

Results

Exploratory factor analyses

EFA analyses showed consistent loadings of the first 11 items, involving cognitive symptoms, onto the first factor across the three time points for both child and parent ratings. This was further evident in the polychoric correlations, which showed strong inter-item correlations among items 1–11 (Figure 3). Items 12–20

involving somatic symptoms loaded onto two different factors when based on eigenvalues > 1, with items 19 and 20 loading separately from the others, suggesting the possibility of a third fatigue factor; however, the eigenvalues for the third factor were relatively low and scree plots suggested a 2-factor solution. When EFA were constrained to 2 factors, items 1–11 loaded onto factor 1 and items 12–20 onto factor 2 across all time points for each group and rater. Occasional loadings of items 19 and 20 onto the first factor were also observed, again suggesting the possibility of a third fatigue factor.

Confirmatory factor analyses

Fit statistics for the correlated two-factor, correlated three-factor, and three-factor bifactor models are presented in Table 3. At each
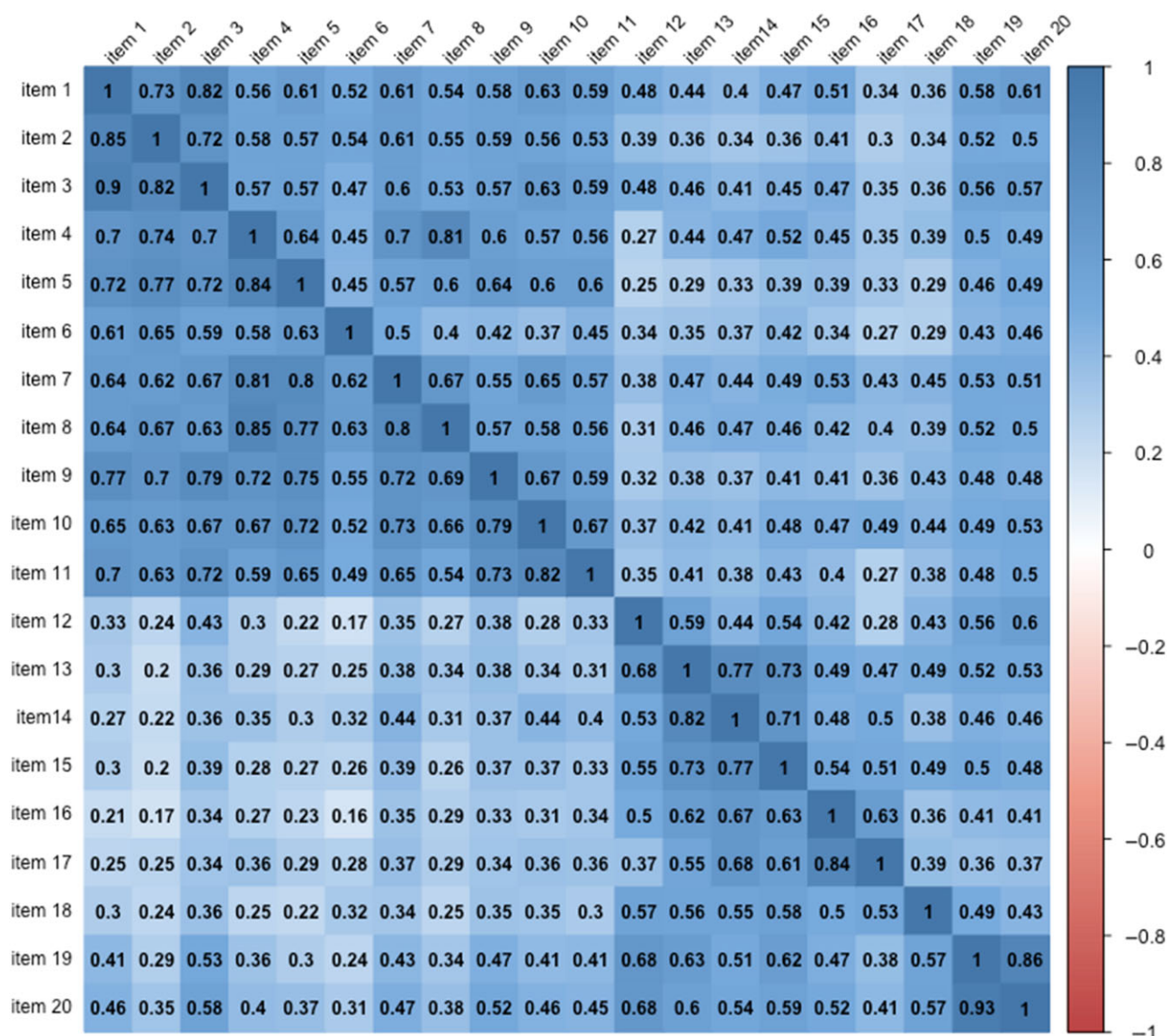
Figure 3. Polychoric correlation of HBI items at post-acute. Correlations above diagonal line are from children's ratings and correlations below the diagonal line are from parents' ratings.

time point, the correlated three-factor model consistently achieved the best absolute and incremental fit, although no model established good fit due to RMSEA > 0.05.

Correlated two-factor model

The correlated two-factor model had the worst fit at post-acute for both raters. All RMSEA were > 0.08 and the fit indices were worse than the other two models at all time points for child's ratings. Fit indices were better for the parent ratings at all time points except for RMSEA post-acutely.

Correlated three-factor model

The correlated three-factor model demonstrated improved fit compared to the more parsimonious two-factor model. Loadings of items 19 and 20 on the third factor, reflecting fatigue, were 0.92 and 0.934, 0.974 and 0.967, and 0.965 and 0.966 at post-acute, 3-months, and 6-months respectively; these loadings were

consistently higher than their loadings on the second factor in the correlated two-factor model, which were 0.883 and 0.894., 0.954 and 0.953, 0.95 and 0.953 at post-acute, 3-months, and 6-months respectively. Parent ratings displayed better comparative fit statistics post-acutely and the child ratings showed better absolute fit than parent ratings (RMSEA 0.081 vs. 0.094) post-acutely, with no apparent differences at other time points.

Bifactor model

The three-factor bifactor model for parent ratings resulted in the worst fit indices at post-acute and 6-months compared to the other two models. For child ratings, absolute and incremental fit were better than the correlated two-factor model at all time points. The bifactor model failed to achieve acceptable fit at any time point for both raters because RMSEA > 0.08; however, comparative fit indices (CFI/TLI) were greater than 0.95. The omegaH for the parent model were 0.89, 0.81, and 0.87 at post-acute, 3-months, and 6-months, respectively. For the child model, omegaH were

Table 3. Fit statistics for confirmatory factor models at each time point across raters

Rater	X^2	df	CFI	TLI	RMSEA (90% CI)
Parent					
Post-acute					
Correlated two-factor	1096.842**	169	0.966	0.962	0.1 (0.094–0.106)**
Correlated three-factor	978.043**	167	0.97	0.966	0.094 (0.088–0.1)**
Bifactor – two-factor	1192.708**	150	0.962	0.952	0.112 (0.107–0.118)**
3 Months					
Correlated two-factor	724.542**	169	0.979	0.976	0.082 (0.076–0.089)**
Correlated three-factor	689.011**	167	0.98	0.978	0.08 (0.074–0.087)**
Bifactor – two-factor	666.706**	150	0.981	0.975	0.084 (0.078–0.091)**
6 Months					
Correlated two-factor	672.473**	169	0.983	0.98	0.081 (0.074–0.087)**
Correlated three-factor	582.083**	167	0.986	0.984	0.074 (0.067–0.08)**
Bi-factor – two-factor	708.42**	150	0.981	0.976	0.09 (0.084–0.097)**
Child					
Post-acute					
Correlated two-factor	962.383**	169	0.945	0.939	0.092 (0.087–0.098)**
Correlated three-factor	768.862**	167	0.959	0.953	0.081 (0.075–0.087)**
Bifactor – two-factor	714.288**	150	0.961	0.951	0.083 (0.077–0.089)**
3 Months					
Correlated two-factor	858.083**	169	0.975	0.972	0.092 (0.086–0.098)**
Correlated three-factor	714.229**	167	0.98	0.978	0.083 (0.077–0.089)**
Bi-factor – two-factor	718.773**	150	0.979	0.974	0.089 (0.083–0.096)**
6 Months					
Correlated two-factor	773.233**	169	0.978	0.975	0.088 (0.082–0.094)**
Correlated three-factor	567.524**	167	0.985	0.983	0.072 (0.066–0.079)**
Bifactor – two-factor	627.243**	150	0.983	0.978	0.083 (0.076–0.09)**

X^2 = chi-square statistic; df = degrees of freedom; RMSEA = root mean square error of approximation statistic; CFI = comparative fit index; TLI = Tucker-Lewis index.
Note: model fitted for mTBI group only.
**p-value $\leq$ 0.001.

Table 4. Fit statistics for measurement invariance of two-factor models across raters within time points

Model	X^2	df	CFI	ΔCFI	TLI	RMSEA (90% CI)	ΔRMSEA
Correlated two-factor							
PA							
Configural	2644.435**	714	0.966	—	0.963	0.057 (0.055–0.059)**	—
Weak	2710.169**	732	0.966	−0.001	0.963	0.057 (0.055–0.059)**	0
Strong	2875.147**	770	0.963	−0.002	0.963	0.057 (0.055–0.059)**	0
Strict	2899.178**	790	0.963	0	0.964	0.057 (0.054–0.059)**	−0.001
3M							
Configural	2269.66**	714	0.971	—	0.968	0.054 (0.052–0.057)*	—
Weak	2315.533***	732	0.97	−0.001	0.968	0.054 (0.051–0.056)*	0
Strong	2404.715**	770	0.969	−0.001	0.969	0.053 (0.051–0.056)*	−0.001
Strict	2442.199**	790	0.969	0	0.97	0.053 (0.051–0.055)*	0
6M							
Configural	2055.799**	714	0.978	—	0.976	0.052 (0.049–0.054)	—
Weak	2094.897**	732	0.978	0	0.976	0.051 (0.049–0.054)	0
Strong	2171.62**	770	0.977	−0.001	0.977	0.051 (0.048–0.053)	−0.001
Strict	2206.606**	790	0.977	0	0.977	0.05 (0.048–0.053)	−0.001
Bifactor-two subfactors							
PA							
Configural	2102.297**	670	0.975	—	0.971	0.051 (0.048–0.053)	—
Weak	2079.032**	707	0.976	0.001	0.974	0.048 (0.046–0.051)	−0.002
Strong	2304.042**	745	0.973	−0.003	0.972	0.05 (0.048–0.052)	0.002
Strict	2305.954**	765	0.973	0	0.973	0.049 (0.047–0.051)	−0.001
3M							
Configural	1921.702**	670	0.977	—	0.973	0.05 (0.047–0.053)	—
Weak	1768.808**	707	0.98	0.004	0.978	0.045 (0.042–0.047)	−0.005
Strong	1918.335**	745	0.978	−0.002	0.977	0.046 (0.043–0.048)	0.001
Strict	1988.057**	765	0.977	−0.001	0.977	0.046 (0.044–0.049)	0
6M							
Configural	1812.246**	670	0.981	—	0.978	0.049 (0.046–0.052)	—
Weak	1626.726**	707	0.985	0.004	0.983	0.043 (0.04–0.046)	−0.006
Strong	1735.248**	745	0.984	−0.001	0.983	0.043 (0.041–0.046)	0
Strict	1726.072**	765	0.984	0	0.984	0.042 (0.04–0.045)	−0.001

X^2 = chi-square statistic; df = degrees of freedom; RMSEA = root mean square error of approximation statistic; CFI = comparative fit index; TLI = Tucker-Lewis index. Note: **p-value <0.001, *0.001 $\leq$ p-value <0.05. Δ: current model subtract the previous (the one above) model.

Table 5. Fit statistics for measurement invariance of three-factor models across raters within time points

Model	X^2	df	CFI	ΔCFI	TLI	RMSEA (90% CI)	ΔRMSEA
Correlated three-factor							
PA							
Configural	2260.503**	705	0.973	—	0.97	0.051 (0.049–0.054)	—
Weak	2321.172**	722	0.972	−0.001	0.97	0.051 (0.049–0.054)	0
Strong	2478.59**	759	0.97	−0.002	0.969	0.052 (0.05–0.054)	0.001
Strict	2549.216**	779	0.969	−0.001	0.969	0.052 (0.05–0.054)	0
3M							
Configural	2019.029**	705	0.975	—	0.973	0.05 (0.047–0.053)	—
Weak	2061.955**	722	0.975	0	0.973	0.05 (0.047–0.052)	0
Strong	2142.312**	759	0.974	−0.001	0.973	0.049 (0.047–0.052)	0
Strict	2158.150**	779	0.974	0	0.974	0.049 (0.046–0.051)	−0.001
6M							
Configural	1691.612**	705	0.984	—	0.982	0.045 (0.042–0.047)	—
Weak	1731.025**	722	0.984	0	0.982	0.045 (0.042–0.047)	0
Strong	1810.533**	759	0.983	−0.001	0.982	0.044 (0.042–0.047)	0
Strict	1848.652**	779	0.983	0	0.983	0.044 (0.042–0.047)	0
Bifactor-three subfactors							
PA							
Configural	2046.508**	663	0.976	—	0.972	0.05 (0.048–0.052)	—
Weak	2108.361**	699	0.976	0	0.973	0.049 (0.047–0.052)	−0.001
Strong	2304.799**	736	0.973	−0.003	0.971	0.05 (0.048–0.053)	0.001
Strict	2315.815**	756	0.973	0	0.972	0.05 (0.047–0.052)	−0.001
3M							
Configural	1694.03**	663	0.981	—	0.977	0.046 (0.043–0.048)	—
Weak	1729.231**	699	0.981	0	0.979	0.044 (0.042–0.047)	−0.001
Strong	1832.276**	736	0.98	−0.001	0.978	0.044 (0.042–0.047)	0
Strict	1836.882**	756	0.98	0	0.979	0.044 (0.041–0.046)	−0.001
6M							
Configural	1624.578**	663	0.984	—	0.982	0.045 (0.043–0.048)	—
Weak	1623.299**	699	0.985	0.001	0.983	0.043 (0.041–0.046)	−0.002
Strong	1705.552**	736	0.984	−0.001	0.983	0.043 (0.041–0.046)	0
Strict	1689.267**	756	0.985	0.001	0.984	0.042 (0.039–0.045)	−0.001

X^2 = chi-square statistic; df = degrees of freedom; RMSEA = root mean square error of approximation statistic; CFI = comparative fit index; TLI = Tucker-Lewis index.
Note: **p-value <0.001, Δ: current model subtract the previous (the one above) model.

Table 6. Fit statistics for measurement invariance of two-factor models within raters across time points

Model	X^2	df	CFI	ΔCFI	TLI	RMSEA (90% CI)	ΔRMSEA
Correlated two-factor							
Parent							
Configural	3815.219**	1635	0.974	—	0.972	0.04 (0.038–0.041)	—
Weak	3865.991**	1667	0.974	0.000	0.973	0.039 (0.038–0.041)	−0.001
Strong	3979.752**	1742	0.974	0.000	0.973	0.039 (0.037–0.041)	0
Strict	4002.196**	1782	0.974	0.001	0.974	0.038 (0.037–0.04)	−0.001
Child							
Configural	3800.63**	1639	0.970	—	0.967	0.039 (0.038–0.041)	—
Weak	3857.096**	1672	0.970	0	0.968	0.039 (0.038–0.042)	0
Strong	3814.95**	1747	0.971	0.002	0.971	0.037 (0.036–0.039)	0
Strict	3776.118**	1787	0.972	0.001	0.973	0.036 (0.035–0.038)	−0.001
Bifactor-two subfactors							
Parent							
Configural	3364.59**	1561	0.979	—	0.976	0.037 (0.035–0.039)	—
Weak	3379.764**	1635	0.98	0.001	0.978	0.035 (0.034–0.037)	−0.001
Strong	3492.985**	1710	0.979	0	0.979	0.035(0.033–0.037)	0
Strict	3465.564**	1750	0.98	0.001	0.98	0.034 (0.032–0.036)	−0.001
Child							
Configural	2865.502**	1561	0.982	—	0.979	0.031 (0.03–0.033)	—
Weak	2978.988**	1635	0.981	−0.001	0.98	0.031 (0.029–0.033)	0
Strong	3087.659**	1711	0.981	0	0.98	0.031 (0.029–0.033)	0
Strict	3095.291**	1751	0.981	0	0.981	0.03 (0.028–0.032)	−0.001

X^2 = chi-square statistic; df = degrees of freedom; RMSEA = root mean square error of approximation statistic; CFI = comparative fit index; TLI = Tucker-Lewis index.
Note: **p-value <0.001, Δ: current model subtract the previous (the one above) model.

Table 7. Fit statistics for measurement invariance of three-factor models within raters across time points

Model	X^2	df	CFI	ΔCFI	TLI	RMSEA (90% CI)	ΔRMSEA
Correlated three-factor							
Parent							
Configural	3508.201**	1614	0.978	—	0.976	0.037 (0.036–0.039)	—
Weak	3560.345**	1648	0.978	0	0.976	0.037 (0.035–0.039)	0
Strong	3676.652**	1721	0.977	−0.001	0.976	0.037 (0.035–0.038)	0
Strict	3671.341**	1761	0.978	0.001	0.977	0.036 (0.034–0.037)	−0.001
Child							
Configural	3123.351**	1614	0.979	—	0.977	0.033 (0.031–0.035)	—
Weak	3186.344**	1648	0.979	0	0.977	0.033 (0.031–0.035)	0
Strong	3296.354**	1722	0.978	−0.001	0.978	0.033 (0.031–0.035)	0
Strict	3300.212**	1762	0.979	0.001	0.979	0.032 (0.03–0.034)	−0.001
Bifactor-three subfactors							
Parent							
Configural	3397.272**	1540	0.978	—	0.975	0.038 (0.036–0.039)	—
Weak	3484.09**	1613	0.978	0	0.976	0.037 (0.035–0.039)	−0.001
Strong	3050.253**	1676	0.984	0.006	0.983	0.031 (0.029–0.033)	−0.006
Strict	3036.812**	1716	0.984	0.001	0.984	0.03 (0.028–0.032)	−0.001
Child							
Configural	2824.293**	1540	0.982	—	0.98	0.031 (0.03–0.033)	—
Weak	2785.086**	1612	0.984	0.002	0.982	0.029 (0.027–0.031)	−0.002
Strong	2884.916**	1686	0.983	0	0.983	0.029 (0.027–0.031)	0
Strict	2952.185**	1726	0.983	0	0.983	0.029 (0.027–0.031)	0

X^2 = chi-square statistic; df = degrees of freedom; RMSEA = root mean square error of approximation statistic; CFI = comparative fit index; TLI = Tucker-Lewis index. Note: ** p-value <0.001, Δ: current model subtract the previous (the one above) model.

0.68, 0.86, and 0.64 at post-acute, 3-months, and 6-months, respectively. These high omega values for the general factor suggest that 64% to 89% of the reliable variance in the HBI can be attributed to this factor. However, some substantive variance was explained by the subfactors, especially for the child ratings. Pearson correlations between the predicted general factor score and the total HBI score for the child and parent ratings were large and significant across all three time points. Correlations between predicted subfactor scores and the calculated cognitive and somatic scores were all significant but ranged from 0.13 to 0.89 (supplementary Table 1).

Measurement invariance
Invariance of the three models in the CFA was tested across raters, time points, and groups. Additionally, a bifactor model with three subfactors was tested across raters and time points as it was able to converge.

Rater invariance

Tables 4 and 5 present the fit statistics for tests of MI across raters for each of the models, with statistics for each of the increasingly more stringent assumptions. All models at all time points demonstrated CFI/TLI > 0.96 and RMSEA < 0.08, indicating acceptable fit. Consistent with the criteria for MI, no CFI decreased more than 0.01 and none of the RMSEA increased more than 0.015; therefore, strict invariance across child and parent ratings was established at each time point for all four models. Analogous to the results in the CFA, the correlated three-factor model had slightly better fit indices than the correlated two-factor model. For the bifactor models, differences in fit between models with three versus two subfactors were less apparent.

Longitudinal invariance

Tables 6 and 7 display the fit statistics for tests of MI across time points separately for child and parent ratings. Model parameters were constrained to be equal across all three time points. Strict invariance was achieved across all time points for all models for both raters. All models retained good fit with CFI/TLI > 0.97 and RMSEA < 0.04. At each level of invariance, fit indices were either similar or better for child than parent ratings for all models except for CFI and TLI for the correlated two-factor model. Model fit was better when comparing the three-factor bifactor model to the correlated two-factor model, particularly for children's ratings (ΔCFI between 0.009 to 0.012 and ΔRMSEA between 0.003 to 0.008). The same difference was apparent but smaller when comparing the three-factor correlated model with the four-factor bifactor model. For each rater, absolute and incremental fit improved as more constraints were imposed on model parameters.

Group invariance

Tables 8 and 9 present the fit statistics for MI across the concussion and OI groups for parents and children, respectively. The bifactor model with three subfactors was unable to converge; hence, fit indices were not derived. For parent ratings, delta parameterization was used at 6-months for the correlated two-factor and bifactor model with two subfactors, precluding testing for strict invariance as residual variances are not included in the model under delta parameterization. For child ratings, delta parameterization was used at post-acute for the bifactor model. Strict invariance was established for all models at each time point separately for child and parent ratings, except for models that used delta parameterization, for which strong invariance was achieved.

The correlated two-factor model at 6-months and correlated three-factor model at 3-months and 6-months achieved acceptable fit at each invariance level for parent ratings. For child ratings, the correlated three-factor and bifactor model achieved acceptable fit at post-acute and 6-months at each invariance level. In turn, no model showed good fit at the configural level due to RMSEA > 0.05. Child ratings demonstrated a consistent pattern of better model fit as more constraints were imposed at all time points. The same pattern was apparent for parent ratings except for the correlated three-factor model at 3-months, with the strong invariance model showing better fit than strict invariance model.

Table 8. Fit statistics for measurement invariance across groups for parent rating

Model	X²	df	CFI	ΔCFI	TLI	RMSEA (90% CI)	ΔRMSEA
Correlated two-factor							
PA							
Configural	1310.509**	338	0.976	—	0.973	0.083 (0.079–0.088)**	—
Weak	1319.058**	356	0.976	0	0.974	0.081 (0.076–0.086)**	−0.003
Strong	1389.518**	391	0.975	0.001	0.976	0.079 (0.074–0.083)**	−0.002
Strict	1421.839**	411	0.975	0	0.977	0.077 (0.073–0.082)**	−0.001
3M							
Configural	1187.129**	338	0.979	—	0.976	0.083 (0.077–0.088)**	—
Weak	1123.128**	356	0.981	0.002	0.98	0.076 (0.071–0.081)**	−0.006
Strong	1213.329**	391	0.98	−0.001	0.98	0.075 (0.071–0.08)**	−0.001
Strict	1177.029**	411	0.981	0.001	0.983	0.071 (0.066–0.076)**	−0.004
6M[1]							
Configural	1028.106**	338	0.991	—	0.99	0.077 (0.071–0.082)**	—
Weak	1038.464**	356	0.991	0	0.991	0.074 (0.069–0.08)**	−0.002
Strong	1072.696**	390	0.991	0	0.992	0.071 (0.066–0.076)**	−0.003
Correlated three-factor							
PA							
Configural	1140.34**	334	0.98	—	0.977	0.076 (0.072–0.081)**	—
Weak	1114.39**	351	0.981	0.001	0.979	0.073 (0.068–0.077)**	−0.004
Strong	1192.557**	385	0.98	−0.001	0.98	0.071 (0.067–0.076)**	−0.001
Strict	1191.943**	405	0.98	0.001	0.98	0.069 (0.064–0.073)**	−0.003
3M							
Configural	993.119**	334	0.984	—	0.981	0.073 (0.068–0.078)**	—
Weak	935.201**	351	0.986	0.002	0.984	0.067 (0.062–0.072)**	−0.006
Strong	1012.136**	385	0.985	−0.001	0.985	0.066 (0.061–0.071)**	−0.001
Strict	1177.029**	411	0.981	0.001	0.983	0.071 (0.066–0.076)**	−0.004
6M							
Configural	847.381**	334	0.993	—	0.993	0.067 (0.061–0.072)**	—
Weak	858.452**	351	0.994	0	0.993	0.065 (0.059–0.07)**	−0.002
Strong	888.411**	384	0.994	0	0.994	0.062 (0.056–0.067)**	−0.003
Strict	921.387**	404	0.993	−0.001	0.994	0.061 (0.056–0.066)**	−0.001
Bifactor-two subfactors							
PA							
Configural	1136.845**	300	0.979	—	0.973	0.082 (0.077–0.087)**	—
Weak	1058.407**	337	0.982	0.003	0.98	0.072 (0.067–0.077)**	−0.01
Strong	1123.573**	371	0.981	−0.001	0.981	0.7 (0.065–0.075)**	−0.002
Strict	1063.071**	391	0.983	0.002	0.984	0.064 (0.06–0.069)**	−0.006
3M							
Configural	1031.037**	300	0.982	—	0.977	0.081 (0.076–0.087)**	—
Weak	830.895**	337	0.988	0.006	0.986	0.063 (0.058–0.068)**	−0.018
Strong	940.273**	371	0.986	−0.002	0.986	0.064 (0.059–0.07)**	0.001
Strict	864.533**	391	0.988	0.002	0.989	0.057 (0.052–0.062)*	−0.007
6M[1]							
Configural	993.805**	300	0.991	—	0.989	0.082 (0.076–0.087)**	—
Weak	769.035**	337	0.995	0.003	0.994	0.061 (0.055–0.067)*	−0.021
Strong	862.634**	370	0.994	−0.001	0.994	0.062 (0.057–0.067)**	0.001

Note: [1]due to convergence issues, delta parameterization was used; therefore, strict invariance cannot be fitted.
**p-value < 0.001.*p-value < 0.05. Δ: current model subtracts the previous (the one above) model.

Incremental fit for parent ratings at 6-months was excellent for each model (CFI > 0.99, TLI > 0.988). Every CFI/TLI in Table 9 (child) was equal to or greater than 0.95, reflecting good incremental fit. The largest improvement in absolute fit was observed in the bifactor model at 3-months and 6-months for both raters; RMSEA decreased between 0.017 to 0.021 moving from configural to weak invariance. The correlated three-factor model showed the best overall fit at the configural level at post-acute, followed by the bifactor model with two subfactors. The three-factor bifactor model, when moving from weak to strict invariance, retained better fit than the correlated three-factor model at 6-months for child ratings.

Comparison of means

The achievement of at least strong invariance indicates that factor means can be compared between different groups and across time.

Table 10 illustrates the mean differences between factors by raters/ groups/time based on the strict invariance three-factor bifactor model, which we chose for this purpose because it had the most acceptable fit indices. When comparing child and parent ratings, the mean of the general factor was significantly higher for child ratings at each time point. The cognitive subfactor showed an opposite pattern, indicating that children rate themselves lower on cognitive symptoms after accounting for general differences in symptom ratings. On the somatic subfactor, ratings differed significantly only at 3-months, with child ratings lower than parent ratings. In group comparisons, the concussion group consistently scored higher than the OI group on the general factor according to both raters. Group differences on the cognitive subfactor were not significant at any time for parent ratings but were higher in the concussion group at post-acute and 3-months for child ratings. The concussion group scored significantly higher on the somatic subfactor at post-acute based on parent ratings and at 3-months

Table 9. Fit statistics for measurement invariance across groups for child rating

Model	X²	df	CFI	ΔCFI	TLI	RMSEA (90% CI)	ΔRMSEA
Correlated two-factor							
PA							
Configural	1258.789**	338	0.955	—	0.95	0.081 (0.076–0.086)**	—
Weak	1138.545**	356	0.962	0.007	0.959	0.073 (0.068–0.078)**	−0.008
Strong	1261.732**	394	0.958	−0.004	0.959	0.073 (0.068–0.078)**	0
Strict	1236.439**	414	0.96	0.002	0.963	0.069 (0.065–0.074)**	−0.004
3M							
Configural	1347.476**	338	0.973	—	0.969	0.091 (0.086–0.096)**	—
Weak	1224.54**	356	0.976	0.004	0.975	0.082 (0.077–0.087)**	−0.009
Strong	1342.339**	392	0.974	−0.002	0.975	0.082 (0.077–0.087)**	0
Strict	1266.805**	412	0.977	0.003	0.979	0.076 (0.071–0.08)**	−0.006
6M							
Configural	1210.892**	338	0.977	—	0.974	0.086 (0.081–0.91)**	—
Weak	1104.475**	356	0.98	0.003	0.979	0.078 (0.072–0.083)**	−0.008
Strong	1203.952**	393	0.979	−0.002	0.979	0.077 (0.072–0.082)**	−0.001
Strict	1143.543**	413	0.981	0.002	0.982	0.071 (0.066–0.076)**	−0.006
Correlated three-factor							
PA							
Configural	974.092**	334	0.969	—	0.965	0.068 (0.063–0.073)**	—
Weak	899.673**	351	0.973	0.004	0.971	0.061 (0.057–0.066)**	−0.007
Strong	959.517**	388	0.972	−0.001	0.973	0.06 (0.055–0.064)**	−0.002
Strict	974.121**	408	0.972	0	0.974	0.058 (0.053–0.063)*	−0.002
3M							
Configural	1149.293**	334	0.978	—	0.975	0.082 (0.077–0.087)**	—
Weak	1049.576**	351	0.981	0.003	0.979	0.074 (0.069–0.079)**	−0.008
Strong	1147.473**	386	0.979	−0.002	0.98	0.074 (0.069–0.079)**	0
Strict	1094.788**	406	0.981	0.002	0.982	0.068 (0.064–0.073)**	−0.005
6M							
Configural	926.156**	334	0.984	—	0.982	0.071 (0.066–0.077)**	—
Weak	862.96**	351	0.987	0.002	0.985	0.065 (0.059–0.07)**	−0.007
Strong	928.633**	387	0.986	−0.001	0.986	0.063 (0.058–0.069)**	−0.001
Strict	900.861**	407	0.987	0.001	0.988	0.059 (0.054–0.064)*	−0.004
Bifactor-two subfactor							
PA[1]							
Configural	973.571**	300	0.967	—	0.959	0.074 (0.069–0.079)**	—
Weak	850.197**	337	0.975	0.008	0.972	0.061 (0.056–0.066)**	−0.013
Strong	928.426**	374	0.973	−0.002	0.973	0.06 (0.055–0.066)**	−0.001
3M							
Configural	1015.505**	300	0.981	—	0.975	0.081 (0.076–0.087)**	—
Weak	765.986**	337	0.988	0.008	0.987	0.059 (0.054–0.065)*	−0.022
Strong	940.792**	372	0.985	−0.004	0.984	0.065 (0.06–0.07)**	0.006
Strict	844.583**	392	0.988	0.003	0.988	0.056 (0.051–0.062)*	−0.009
6M							
Configural	906.302**	300	0.984	—	0.98	0.076 (0.07–0.082)**	—
Weak	747.69**	337	0.989	0.005	0.988	0.059 (0.053–0.065)**	−0.017
Strong	829.308**	373	0.988	−0.001	0.988	0.059 (0.054–0.065)*	0
Strict	784.694**	393	0.99	0.002	0.99	0.053 (0.048–0.059)	−0.006

Note: [1]due to convergence issues, delta parameterization was used; therefore, strict invariance cannot be fitted.
**p-value < 0.001. *p-value < 0.05. Δ: current model subtracts the previous (the one above) model.

based on child ratings. General factor means decreased significantly over time for both raters. Cognitive subfactor means did not change significantly for either rater, while somatic subfactor means decreased significantly over time for both child and parent ratings.

Discussion

The current study validated and supported the two underlying dimensions of the HBI (cognitive and somatic symptoms) established in its original derivation study (Ayr et al., 2009). A third factor, represented by fatigue symptoms, may serve as an additional dimension; however, its inclusion in a four-factor bifactor model was constrained by convergence issues. At the configural level, almost all models achieved either acceptable or good fit, reflecting the robustness of the basic cognitive and somatic factor structure.

When comparing correlated factor models to the bifactor model, however, the bifactor model consistently achieved better fit, and this difference was most apparent when using two subfactors. Fit indices for the bifactor model with two subfactors and the correlated three-factor model were similar; therefore, the three-factor bifactor model may be the better solution based on parsimony, although the negative item-level loadings of items 19 and 20 onto the somatic subfactor suggests the possibility of further dividing this subfactor into vestibular and fatigue components. However, when comparing factor loadings of the correlated two-factor and three-factor bifactor models, the high, consistent, and distinct loadings of the correlated model provide an easier interpretation of the underlying dimensions. In summary, the three-factor bifactor model achieved the most consistent acceptable fit indices across groups, raters, and time, while the correlated two-factor model achieved greater parsimony. Either model can reasonably be used to interpret ratings on the HBI.

Table 10. Mean differences assuming strict measurement invariance for the three-factor bifactor model by rater, group, and time

Model factor	Post-acute			3 months			6 months		
Rater	Parent	Child	*p*-value	Parent	Child	*p*-value	Parent	Child	*p*-value
General factor	0.00	0.745	<0.001	0.00	0.17	0.041	0.00	0.532	<0.001
Cognitive	0.00	−0.224	<0.001	0.00	−0.55	0.006	0.00	−0.666	<0.001
Somatic	0.00	−0.02	0.19	0.00	−0.554	0.006	0.00	0.005	0.879
Group	**OI**	**mTBI**	***p*-value**	**OI**	**mTBI**	***p*-value**	**OI**	**mTBI**	***p*-value**
Parent									
General factor	0.00	2.559	<0.001	0.00	0.847	0.002	0.00	0.112[1]	0.067
Cognitive	0.00	0.312	0.071	0.00	0.019	0.923	0.00	0.065[1]	0.47
Somatic	0.00	1.342	<0.001	0.00	0.197	0.06	0.00	−0.023[1]	0.676
Child									
General factor	0.00	0.398[1]	<0.001	0.00	6.103	<0.001	0.00	0.357	0.038
Cognitive	0.00	0.287[1]	<0.001	0.00	3.371	<0.001	0.00	−0.075	0.695
Somatic	0.00	0.018[1]	0.115	0.00	0.472	<0.001	0.00	0.1	0.079
Time			***p*-value**			***p*-value**			***p*-value**
Parent									
General factor	0.00		—		−0.875	<0.001		−0.945	<0.001
Cognitive	0.00		—		0.127	0.178		0.127	0.263
Somatic	0.00		—		−0.085	0.036		−0.086	0.047
Child									
General factor	0.00		—		−1.251	<0.001		−1.294	<0.001
Cognitive	0.00		—		−0.038	0.683		0.087	0.352
Somatic	0.00		—		−0.048	0.038		−0.048	0.053

Note: [1]strong invariance model used instead.

Both the bifactor and correlated factor models align with other studies regarding the dimensions represented by PCS. The Sport Concussion Assessment Tool-3, the Post-Concussion Symptom Scale, and the Rivermead Postconcussion Symptoms Questionnaire are often used to measure PCS ratings. A specific cognitive factor has been consistently identified in previous research studies that used the bifactor model to examine the structure of those tools (Agtarap et al., 2020; Brett et al., 2020; Joyce et al., 2015; Karr & Iverson, 2020). A specific somatic or physical factor was also identified in four of the studies (Agtarap et al., 2020; Ayr et al., 2009; Joyce et al., 2015; Karr & Iverson, 2020), while fatigue was identified in one study (Brett et al., 2020). Some previous studies also identified a specific vestibular symptom dimension, but that was not apparent on the HBI (Franke et al., 2015; Taylor et al., 2010). The differences in factor structure across rating scales likely depend largely on the types of items included in various instruments (e.g., the HBI does not include emotional symptoms), and perhaps on the nature of the responses required, ranging from yes-no to 7 point Likert-type scales.

Tests of MI demonstrated strict invariance across raters, time, and groups for all comparisons in the absence of convergence issues, and at least strong invariance when convergence could not be achieved. Thus, clinicians and researchers using the HBI can confidently assume that its scores are measuring the same constructs in these different groups, raters, and over time. Indeed, strict invariance across child and parent ratings signifies that the HBI items measure the same constructs on the same scales in both children and parents, and any difference in the means, variances, and covariances of the individual items are solely due to differences in the common factors (Liu et al., 2017). Although past studies have found only moderate agreement between child and parent ratings on the HBI (Ayr et al., 2009; Hajek et al., 2011; Taylor et al., 2010), our findings suggest that the constructs measured by child and parent ratings are the same. Thus, mean differences between child and parent ratings are likely reflections of true differences in child and parent perceptions, rather than differences in what the scales are measuring.

Our study has several limitations. Attrition occurred over time, and the data may have been not missing at random (MAR), which can bias WLSMW estimates. However, we have previously tested for differences between those who returned for follow-up assessments versus those who did not across a broad range of variables and did not find any significant differences, suggesting the data are MAR. Another limitation is that we needed to collapse symptom ratings on several items in the concussion group to test group MI. However, ratings were collapsed for no more than two items for child ratings and no more than three items for parent ratings at any given time point. Finally, the study sample was largely white and well-educated, so the results may not be generalizable to the general population. Future research is needed to determine whether the MI demonstrated in this study extends to more diverse samples.

The good fit of the bifactor model and the strong correlation between the predicted general factor score and the computed total HBI score suggest that clinicians can use a total score when measuring concussion symptoms with the HBI. However, the strong loadings of the correlated two-factor model, as well as the subfactors in the bifactor model, suggest that clinicians may also want to use subscales to isolate specific symptom dimensions to guide individualized treatment and rehabilitation. Indeed, the bifactor model provides a justification for assessing both overall symptom severity and the severity of cognitive and somatic symptoms, which may have distinct diagnostic and prognostic value beyond the total score. Using a total score can obscure the severity of specific symptom types. Two individuals with the same total score could exhibit different scores on the cognitive and somatic subscales; therefore, assessing specific subscales may improve detection of more subtle concussion-related sequelae (Brett et al., 2020). Notably, the general and subfactors of the bifactor model are independent, so the model estimates the unique contribution of the subfactors (Chen et al., 2012). The omegaH values and the factor loadings of the bifactor model signify that a general factor is a good representation of all HBI items, but the subfactors are also needed to account for the remaining variance, providing strong evidence for the use of

both a total score and subscale scores. Use of the bifactor model could allow for a more precise assessment of specific symptom types and allow for more targeted intervention through selection of specific treatment modalities based on subfactor elevations. However, the inconsistent correlations between the predicted factor scores and calculated subscale scores indicate that simply relying on the raw subscale scores as a reflection of the true factor scores might be inadvisable.

Future study is needed to examine the stability of the three-factor bifactor and correlated two-factor models across different post-injury time points and groups (e.g., healthy children, younger children, and children with concussion recruited from settings other than the ED) to validate the HBI's underlying dimensions and to inform clinical applications. A factor consisting of emotional symptoms has sometimes been identified in past factor analytic studies of concussion symptom ratings, but this emerged only for parent ratings in the original derivation study of the HBI, and not for child ratings; validated rating scales of emotional symptoms may be included as a supplement to the HBI for specific research or clinical purposes. Similarly, ratings of behavioral difficulties did not cohere in the derivation of the HBI but may be relevant for younger children (Dupont et al., 2021). Future research also should continue to examine the relative contributions of injury and non-injury factors as predictors of the general and subfactors, or the two correlated factors (cognitive and somatic), of the HBI (O'Neill et al., 2021).

Our study indicates that the HBI structure is best represented either by a three-factor bifactor model that captures an essentially unidimensional general component of concussion symptom severity along with distinct, but related, dimensions of cognitive and somatic symptoms, or by a correlated two-factor model that contains cognitive and somatic subscales. The HBI can be used to measure PCS consistently over time by both parents and children, and in individuals with concussion or other injuries, for comparison across various phases of concussion recovery, at least for children who are originally seen in the ED. Efforts to characterize individuals based on profiles across the dimensions of these models may facilitate a more precise approach to classifying and treating children with concussion.

References

Agtarap, S., Kramer, M. D., Campbell-Sills, L., Yuh, E., Mukherjee, P., Manley, G. T., . . . Nelson, L. D. (2020). Invariance of the Bifactor Structure of Mild Traumatic Brain Injury (mTBI) Symptoms on the Rivermead Postconcussion symptoms questionnaire across time, demographic characteristics, and clinical groups: A TRACK-TBI study. *Assessment*. https://doi.org/10.1177/1073191120913941

Ayr, L. K., Yeates, K. O., Taylor, H. G., & Browne, M. (2009). Dimensions of postconcussive symptoms in children with mild traumatic brain injuries. *Journal of the International Neuropsychological Society*, 15, 19–30. https://doi.org/10.1017/S1355617708090188

Babl, F. E., Dionisio, D., Davenport, L., Baylis, A., Hearps, S. J. C., Bressan, S., . . . Davis, G. A. (2017). Accuracy of components of SCAT to identify children with concussion. *Pediatrics*, 140. https://doi.org/10.1542/peds.2016-3258

Barlow, K. M., Crawford, S., Stevenson, A., Sandhu, S. S., Belanger, F., & Dewey, D. (2010). Epidemiology of postconcussion syndrome in pediatric mild traumatic brain injury. *Pediatrics*, 126, e374–e381. https://doi.org/10.1542/peds.2009-0925

Beauchamp, M. H., Tang, K., Yeates, K. O., Anderson, P., Brooks, B. L., Keightley, M., . . . Zemek, R.L. (2019). Predicting wellness after pediatric concussion. *Journal of the International Neuropsychological Society*, 25, 375–389. https://doi.org/10.1017/S1355617719000286

Bialy, L., Plint, A., Zemek, R., Johnson, D., Klassen, T., Osmond, M., . . . Pediatric Emergency Research, C. (2018). Pediatric emergency research Canada: Origins and evolution. *Pediatric Emergency Care*, 34, 138–144. https://doi.org/10.1097/PEC.0000000000001360

Brett, B. L., Kramer, M. D., McCrea, M. A., Broglio, S. P., McAllister, T. W., Nelson, L. D., . . . Susmarski, A. (2020). Bifactor model of the sport concussion assessment tool symptom checklist: Replication and invariance across time in the CARE consortium sample. *American Journal of Sports Medicine*, 48, 2783–2795. https://doi.org/10.1177/0363546520946056

Broglio, S. P., Kontos, A. P., Levin, H., Schneider, K., Wilde, E. A., Cantu, R. C., . . . Joseph, K. (2018). National institute of neurological disorders and stroke and department of defense sport-related concussion common data elements version 1.0 recommendations. *Journal of Neurotrauma*, 35, 2776–2783. https://doi.org/10.1089/neu.2018.5643

Browne, M. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, 36, 111–150.

Bryan, M. A., Rowhani-Rahbar, A., Comstock, R. D., Rivara, F., & Seattle Sports Concussion Research, C. (2016). Sports- and recreation-related concussions in US youth. *Pediatrics*, 138. https://doi.org/10.1542/peds.2015-4635

Carroll, L. J., Cassidy, J. D., Holm, L., Kraus, J., & Coronado, V. G. (2004). Methodological issues and research recommendations for mild traumatic brain injury: The WHO Collaborating Centre Task Force on Mild Traumatic Brain Injury. *Journal of Rehabilitation Medicine* (43 Suppl), 113–125. https://doi.org/10.1080/16501960410023877

Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J. P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of Personality*, 80, 219–251. https://doi.org/10.1111/j.1467-6494.2011.00739.x

Davis, G. A., Purcell, L., Schneider, K. J., Yeates, K. O., Gioia, G. A., Anderson, V., . . . Kutcher, J. S. (2017). The child sport concussion assessment tool 5th edition (Child SCAT5): Background and rationale. *British Journal of Sports Medicine*, 51, 859–861. https://doi.org/10.1136/bjsports-2017-097492

Dupont, D., Beaudoin, C., Desire, N., Tran, M., Gagnon, I., & Beauchamp, M. H. (2021). Report of early childhood traumatic injury observations & symptoms: Preliminary validation of an observational measure of post-concussive symptoms. *Journal of Head Trauma Rehabilitation*. https://doi.org/10.1097/HTR.0000000000000691

Franke, L. M., Czarnota, J. N., Ketchum, J. M., & Walker, W. C. (2015). Factor analysis of persistent post-concussive symptoms within a military sample with blast exposure. *Journal of Head Trauma Rehabilitation*, 30, E34–E46. https://doi.org/10.1097/HTR.0000000000000042

Hajek, C. A., Yeates, K. O., Taylor, H. G., Bangert, B., Dietrich, A., Nuss, K. E., . . . Wright, M. (2011). Agreement between parents and children on ratings of post-concussive symptoms following mild traumatic brain injury. *Child Neuropsychology*, 17, 17–33. https://doi.org/10.1080/09297049.2010.495058

Hilt, R. J., McCarty, C. A., Rivara, F. P., Wang, J., Marcynyszyn, L. A., Chrisman, S. P. D., . . . Zatzick, D. F. (2022). Exploring heterogeneity of stepped collaborative care treatment response trajectories after adolescent sports injury concussion. *Psychiatry*, 1–12. https://doi.org/10.1080/00332747.2021.2004784

Joyce, A. S., Labella, C. R., Carl, R. L., Lai, J. S., & Zelko, F. A. (2015). The post-concussion symptom scale: Utility of a three-factor structure. *Medicine & Science in Sports & Exercise*, 47, 1119–1123. https://doi.org/10.1249/MSS.0000000000000534

Karr, J. E., & Iverson, G. L. (2020). The structure of post-concussion symptoms in adolescent student athletes: confirmatory factor analysis and measurement invariance. *The Clinical Neuropsychologist*, 1–40. https://doi.org/10.1080/13854046.2020.1850867

Liu, Y., Millsap, R. E., West, S. G., Tein, J. Y., Tanaka, R., & Grimm, K. J. (2017). Testing measurement invariance in longitudinal data with ordered-categorical measures. *Psychological Methods*, 22(3), 486–506. https://doi.org/10.1037/met0000075

Lumba-Brown, A., Yeates, K. O., Sarmiento, K., Breiding, M. J., Haegerich, T. M., Gioia, G. A., . . . Timmons, S. D. (2018). Centers for disease control and prevention guideline on the diagnosis and management of mild traumatic brain injury among children. *JAMA Pediatrics*, 172, e182853. https://doi.org/10.1001/jamapediatrics.2018.2853

McCarty, C. A., Zatzick, D. F., Marcynyszyn, L. A., Wang, J., Hilt, R., Jinguji, T., . . . Rivara, F. P. (2021). Effect of collaborative care on persistent post-concussive symptoms in adolescents: A randomized clinical trial. *JAMA Network Open*, 4, e210207. https://doi.org/10.1001/jamanetworkopen.2021.0207

McCauley, S. R., Wilde, E. A., Anderson, V. A., Bedell, G., Beers, S. R., Campbell, T. F., . . . Yeates, K. O. (2012). Recommendations for the use of common outcome measures in pediatric traumatic brain injury research. *Journal of Neurotrauma*, 29, 678–705. https://doi.org/10.1089/neu.2011.1838

Medicine, AftAoA. (1990). The abbreviated injury scale. Retrieved June 16 from https://www.aaam.org/abbreviated-injury-scale-ais/

Merritt, V. C., & Arnett, P. A. (2014). Premorbid predictors of postconcussion symptoms in collegiate athletes. *Journal of Clinical and Experimental Neuropsychology*, 36, 1098–1111. https://doi.org/10.1080/13803395.2014.983463

Moran, L. M., Taylor, H. G., Rusin, J., Bangert, B., Dietrich, A., Nuss, K. E., . . . Yeates, K. O. (2012). Quality of life in pediatric mild traumatic brain injury and its relationship to postconcussive symptoms. *Journal of Pediatric Psychology*, 37, 736–744. https://doi.org/10.1093/jpepsy/jsr087

Nelson, L. D., Kramer, M. D., Patrick, C. J., & McCrea, M. A. (2018). Modeling the structure of acute sport-related concussion symptoms: A Bifactor approach. *Journal of the International Neuropsychological Society*, 24, 793–804. https://doi.org/10.1017/S1355617718000462

Novak, Z., Aglipay, M., Barrowman, N., Yeates, K. O., Beauchamp, M. H., Gravel, J., . . . Zemek, R. L. (2016). Association of persistent post-concussion symptoms With pediatric quality of life. *JAMA Pediatr*, 170, e162900. https://doi.org/10.1001/jamapediatrics.2016.2900

O'Neill, J. A., Rose, S. C., Davidson, A. M., Shiplett, K. M., Castillo, A., & McNally, K. A. (2021). Predictors of treatment response to multidisciplinary care for persistent symptoms after pediatric concussion. *Developmental Neurorehabilitation*, 1–7. https://doi.org/10.1080/17518423.2021.1917719

O'Brien, H., Minich, N. M., Langevin, L. M., Taylor, H. G., Bigler, E. D., Cohen, D. M., . . . Yeates, K. O. (2021). Normative and psychometric characteristics of the Health and Behavior Inventory among children with mild orthopedic injury presenting to the emergency department: Implications for assessing post-concussive symptoms using the Child Sport Concussion Assessment Tool 5th edition (Child SCAT5). *Clinical Journal of Sport Medicine*, 31, e221–e228. http://doi.org/10.1097/JSM.0000000000000943

Patsimas, T., Howell, D. R., Potter, M. N., Provance, A. J., Kirkwood, M. W., & Wilson, J. C. (2020). Concussion-symptom rating correlation between pediatric patients and their parents. *Journal of Athletic Training*, 55, 1020–1026. https://doi.org/10.4085/1062-6050-200-19

Piland, S. G., Motl, R. W., Ferrara, M. S., & Peterson, C. L. (2003). Evidence for the factorial and construct validity of a self-report concussion symptoms scale. *Journal of Athletic Training*, 38, 104–112. https://www.ncbi.nlm.nih.gov/pubmed/12937520

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Reed, N. Z. R., Dawson, J., Ledoux, AA, et al. (2019). *Living guideline for diagnosing and managing pediatric concussion*. Ontario Neurotrauma Foundation. https://braininjuryguidelines.org/pediatricconcussion/

Reise, S. P. (2012). Invited paper: The rediscovery of Bifactor measurement models. *Multivariate Behavioral Research*, 47, 667–696. https://doi.org/10.1080/00273171.2012.715555

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, 21, 137–150. https://doi.org/10.1037/met0000045

Sady, M. D., Vaughan, C. G., & Gioia, G. A. (2014). Psychometric characteristics of the postconcussion symptom inventory in children and adolescents. *Archives of Clinical Neuropsychology*, 29, 348–363. https://doi.org/10.1093/arclin/acu014

Taylor, H. G., Dietrich, A., Nuss, K., Wright, M., Rusin, J., Bangert, B., . . . Yeates, K. O. (2010). Post-concussive symptoms in children with mild traumatic brain injury. *Neuropsychology*, 24, 148–159. https://doi.org/10.1037/a0018112

Waljas, M., Iverson, G. L., Hartikainen, K. M., Liimatainen, S., Dastidar, P., Soimakallio, S., . . . Ohman, J. (2012). Reliability, validity and clinical usefulness of the BNI fatigue scale in mild traumatic brain injury. *Brain Injury*, 26, 972–978. https://doi.org/10.3109/02699052.2012.660511

Yeates, K. O., Beauchamp, M., Craig, W., Doan, Q., Zemek, R., Bjornson, B., . . . Schneider, K. J. (2017). Advancing concussion assessment in pediatrics (A-CAP): A prospective, concurrent cohort, longitudinal study of mild traumatic brain injury in children: Protocol study. *BMJ Open*, 7, e017012. https://doi.org/10.1136/bmjopen-2017-017012

Yeates, K. O., Kaizar, E., Rusin, J., Bangert, B., Dietrich, A., Nuss, K., . . . Taylor, H. G. (2012). Reliable change in postconcussive symptoms and its functional consequences among children with mild traumatic brain injury. *Archives of Pediatrics and Adolescent Medicine*, 166, 615–622. https://doi.org/10.1001/archpediatrics.2011.1082

Zemek, R., Barrowman, N., Freedman, S. B., Gravel, J., Gagnon, I., McGahern, C., . . . Osmond, M. (2016). Clinical risk score for persistent post-concussion symptoms among children with acute concussion in the ED. *JAMA*, 315, 1014–1025. https://doi.org/10.1001/jama.2016.1203