Supplemental material, Riccardi et al., “Speech-Language Pathology Assessment of School-Age Children With Traumatic Brain Injury: A Scoping Review,” *LSHSS*, https://doi.org/10.1044/2022_LSHSS-22-00049

Supplemental Material S1. Key findings of the assessments included in this review. Key findings highlight information about assessment reliability, validity, and feasibility relevant to the SLP’s scope of practice.

Key Findings by Assessment			
Assessment (Area of Functioning)	**# Studies Addressing Assessment**	**Key Findings related to Study Purpose and SLP Assessment Practices**	
Structure/Function			
Abbreviated Neuropsychological Testing (Neuro-psychological/ Neuro-cognitive)	1 (Scherwath et al., 2011)	**Validity**	Mild - Scored significantly lower on the Digit Span test compared to normative scores. - Scored significantly lower on Digit Symbol and Digit Span tests compared to controls. - Digit Symbol was not significant compared to normative scores and VLMT was not significant compared to normative scores or controls. - 3 children (30%) were referred for full neuropsychological evaluation based on study’s cut offs. - Overall performance was significantly worse from controls, but not significant with moderate-severe. Mild-Moderate - Scored significantly better on VLMT compared to normative scores. - Digit Span and VLMT scores remained stable over time while Digit Symbol significantly improved at follow-up. - 6 children (12%) were classified as cognitively impaired shortly after injury, but none met the cut off at follow-up. Moderate-Severe - Significantly lower on all tests (Digit Span, Digit Symbol, VLMT) compared to normative scores and controls. - 6 children referred (50%) were referred for full neuropsychological evaluation based on study’s cut offs. - Overall performance was significantly worse from controls, but not significant with mild. Controls (healthy) - No healthy children met the criteria for referral to a full neuropsychological evaluation.
		Other	- Increased feasibility compared to full-length neuropsychological testing with retained sensitivity.

Behavior Assessment System for Children (BASC) (Kamphaus & Reynolds, 2015). (Behavior)	3 (Jones et al., 2018; Kirk et al., 2014; Murphy & Dodd, 2021* also contextual)	**Reliability**	Mild
			- Different patterns in interrater reliability for parent and child reports based on studies that only included mild TBI:
			    - No significant differences in parent and child reports at about 5 mos post-injury (Murphy & Dodd, 2021).
			    - No significant differences at baseline between parent and child reports, but parents reported significantly more child hyperactivity than children at 1 mos, 6 mos, and 12 mos post-injury. Parents tended to report greater child anxiety and child depression than children (Jones et al., 2018).
			- Despite significant differences in scores, parent and child reports showed similar patterns in changes of ratings over time (Jones et al., 2018).

			- Significant change in family stress significantly impacted parent and child discrepancies, such that for families experiencing higher levels of stress, parents reported more internalizing symptoms than children (Murphy & Dodd, 2021).
			  - Family stress change accounted for 20% of the variance in discrepancies (Murphy & Dodd, 2021).
			  - The somatization subscale of internalizing symptoms index was negatively correlated with informant discrepancy scores and was significantly impacted by family stress change. For families with no family stress, parent and child scores were consistent, but, with even a minor level of family stress, discrepancies emerged. Significant increases in parent-reported internalizing symptoms were associated with greater family stress but this pattern was not seen in child reports (Murphy & Dodd, 2021).

		Validity	Mild - No significant correlations between dichotomized BASC-2 validity scales (1- valid, 2- caution or extreme caution) to Medical Symptom Validity Test (MSVT) pass/fail performance (Kirk et al., 2014). - Of the 50 participants who failed the MSVT, 8% generated an invalid profile on the BASC-2. Of the 224 participants who passed the MSVT, 13% generated an invalid profile on the BASC-2 (Kirk et al., 2014). - When using BASC-2 F-index (tally of items where patient gave an overly negative self-report [i.e., "faking bad"]), 6% of participants who failed the MSVT produced an invalid profile, while 2% of participants who passed the MSVT produced an invalid profile (Kirk et al., 2014).

Behavioral Assessment Screening Tool-A (BAST-A) (Juengst et. al., 2019). (Behavior)	1 (B Jeungst et al., 2020)	**Validity**	Mild
			- A 3-factor solution with 46 primary items fit the data best:
			  - The three factors included: Affect & Fatigue, Executive & Social Function, and Risk Behaviors.
			  - Showed good to excellent internal consistency.
			- The BAST-A showed good convergent and discriminant validity when compared to the General Anxiety Disorders 8, Patient Health Questionnaire, Brief Resilience Scale, and the Pittsburgh Sleep Quality Index.
			- Modifications from the BAST to the BAST-A appear to reflect the developmental stage of school-aged adolescents with mild TBI.

Behavior Rating Inventory of Executive Functioning (BRIEF) (Gioia et. al., 2015). (executive functioning)	1 (Donders et al., 2010)	**Validity**	Mild-Severe
			- A two-factor model fit the data best:
			  - Two factors included: 1) Inhibit, Initiate, Working Memory, Plan/Organization, Organization of Materials, and Monitor scales; and 2) Shift and Emotional Control scales.
			  - The Inhibit scale co-varied more with the scales of the Metacognitive Index than the standardization sample where it's included in the Behavioral Regulation Index.
			  - Criterion validity was confirmed with Factor 1 in significant associations with premorbid history variables and injury-related variables (e.g., longer time since injury, premorbid ADHD), but only premorbid history variables were significantly associated with Factor 2, indicating Factor 2 results are not clearly influenced by injury variables.

ChEERS (Children's Exertional Effects Rating Scale) (Sady et al., 2019)	1 (Sady et al., 2019)	**Reliability**	Mild - Sub-optimal internal consistency reliability and almost all symptoms showed significant item-total correlations. - Reliability was stronger in the not-recovered group than the recovered and uninjured groups. - Significant correlations in total score from pre- to post-test ratings for all groups. - Participants 8-12 yo and 13-18 yo showed higher stability than 5-7 yo.

(Symptom monitoring/ assessment)		**Validity**	Mild
			- Convergent validity of the pre-test ChEERS and total PCSI score was significant in all injury and age groups.
			- Divergent validity of the ChEERS EEI (change score) and total PCSI score was only significant in the not recovered 8-12 and 13-18 yo, indicating a stronger relation between pre-test ChEERS and PCSI than ChEERS EEI and PCSI in older children and adolescents across groups.
			- Comparison of convergent and divergent validity was significant for 8-12 and 13-18 yo of all groups.
			- Significant group differences were observed between uninjured and not-recovered and recovered and not-recovered groups but not between uninjured and recovered.
			  - Recovered groups showed a significant increase in scores from pre- to post test of medium effect and their scores were significantly greater compared to both uninjured and recovered.
			  - More individuals in the uninjured and recovered group were categorized as low EEI at baseline, compared to not-recovered, while more not-recovered fell into small or large change in symptoms tiers than recovered for uninjured.

CNS Vital Signs (Gualtieri & Johnson, 2006). (Symptom monitoring/ assessment)	3 (Khetani et al., 2018; Plourde & Brooks, 2017; Brooks et al., 2016)	**Validity**	Mild - Participants who underwent cognitive testing (CNS Vital Signs) did not differ from those who did not undergo acute cognitive testing on mean symptom ratings on the parent-reported and self-reported PCSI or the proportion who were not recovered at 7-10 days, 1 month, 2 months, or 3 months (effect sizes were small for all time points; Brooks et al., 2016). Moderate-Severe - In an outpatient neuropsychology office, there were no significant differences on domain scores according to age for the whole sample (8-12 vs 13-18 yo) or by group (TBI vs control). Scores ranged from low average to average for the TBI group and all average for the control group. The TBI group had significantly lower scores on the Neurocognition Index with very large effect size, compared to control group. Moderate-severe TBI group had sig lower scores of at least medium effect on Memory, Reaction Time, Complex Attention, and Cognitive Flexibility. Some differences emerged in the subtests or sub-components of these domains. - Significantly more moderate-severe TBI participants were classified as having very low cognitive scores (i.e., 3 or more low scores) than controls, including classification based on the Neurocognition Index and Reaction Time individually, but not other domains. DFA analyses indicated that scores correctly classified 75.9% of the sample in their respective group (TBI vs control) and had 87.9% specificity (identifying controls) than sensitivity (60.0%, identifying TBI). Complex attention and visual memory domains were the best predictors. - In only the TBI group, there were no significant differences in domain scores between moderate and severe TBI (Plourde and Brooks, 2017).
		Other	Mild - Testing in the ED was feasible, acceptable, and tolerable to most patients and families. - The full-length battery took an average of 29.7 minutes. The shortened battery took on average 15.2 minutes. When the protocol was changed to the shortened battery, participation of eligible mTBI patients increased from 29 to 74%, although a similar percentage of patients reported time constraints as reason for declining participation. Of the 60 patients who did not enroll, 13 cited being "too injured" (e.g., fatigue, inability to sit upright for testing; Khetani et al., 2018).

			- There also was no difference in symptom outcome for those who underwent a shortened (*n* = 27; four subtests, mean time = 16 min) or full-length (*n* = 50; seven subtests, mean time = 28 min) version of the computerized test (Brooks et al., 2016).
Comprehensive Trail Making Test (Reynolds, 2019). (Neuro-psychological/Neuro-cognitive)	2 (Allen et al., 2009; Armstrong et al., 2008)	**Validity**	Moderate-Severe - For 11-18 years old, the CTMT shows good construct validity, convergent validity with other established neurocognitive tests (i.e., WISC-IV, Woodcock Johnson, Conner's Continuous, Grooved Pegboard), criterion validity such that the TBI group performed significantly worse on all scores compared to normal comparisons/healthy controls, and good predictive discrimination between healthy controls and individuals with TBI (Allen et al., 2009). - Some Trails were significantly correlated with injury severity (GCS; Trails 1, 2, 3) and time post-injury (Trails 3, 4). Given the Complex Sequencing factor of the CTMT had higher correlations with other established neurocognitive tests compared to the Simple Sequencing factor, the Complex Sequencing factor might be more useful in predicting academic performance post-TBI (Allen et al., 2009; Armstrong et al., 2008). - CTMT scores were higher correlated for the TBI group than the normal comparison group, indicating the factor structure might vary based on condition. The two studies found some differences in correlations amongst Trails, warranting further investigation. Some Trails were more sensitive to TBI than others, but overall the CTMT had good sensitivity and specificity (Allen et al., 2009; Armstrong et al., 2008).

Health and Behavior Inventory (Levant et al., 2011) (HBI) (child vs parent ratings) (Symptom monitoring/ assessment)	2 (Johnson et al., 2021; Patsimas et al., 2020)	**Reliability**	Mild
			Interrater reliability:
			- Johnson et al. (2021) found poor-good agreement by inter-class correlations and weak-strong correlations by Spearman Rho across both subscales (i.e., cognitive and somatic) and total score. Patsimas et al. (2020) found all scores (i.e., total, cognitive-symptom subscale, somatic-symptom subscale) were highly correlated between patients and parents.
			- All parent and child scores were significantly correlated for the total sample, by gender, and across age groups (11-13 yo, 14-15 yo, 16-18 yo; Johnson et al., 2021). Correlations between 8-12 yos and their parents were greater than adolescents (13-18 yo) and their parents (Patsimas et al., 2020).
			  - Children generally reported more symptoms than parents (Johnson et al., 2021; Patsimas et al., 2020).
			  - Significant differences were found in total and subscale scores for females, 14-15 yos, 16-18 yos, and total sample (Johnson et al., 2021).
			  - Child-reported scores were associated with symptom durations and return-to-play times but parent-reported scores were not (Patsimas et al., 2020).

ImPACT including Post-Concussion Symptom Scale (PCSS) (Lovell & Collins, 1998; Riverside Insights, n.d.)	4 (Hange et al., 2015; Lau et al., 2011; Lovell et al., 2013; Wiebe et al., 2012)	**Validity**	Mild Results varied from different studies on the sensitivity, specificity, and predictive value of combining neurocognitive testing scores (ImPACT) with symptom scores (PCSS). - The addition of ImPACT testing to PCSS identified 33% more athletes as having a reliable baseline change than PCSS alone (Lovell et al., 2013). - "Poor" ImPACT performance was not particularly useful in predicting athletes with protracted symptoms (at 1 week (positive predictive value, 70.8%; negative predictive value, 43.5%; at 2 weeks: positive predictive value, 47.8%; negative predictive value, 68.9%). In bivariate analysis, there was no association between ImPACT performance and persistent symptoms at 1 or 2 weeks, even when excluding outliers (Hang et al., 2015).

(Neuro-psychological/Neuro-cognitive And Symptom monitoring/ assessment)			- Discriminant function analysis of the 4 symptom clusters (migraine, cognitive, sleep, neuropsychiatric) of the PCSS was significantly able to correctly classify 63.21% of athletes into long or short recovery, with a sensitivity of 46.94% and a specificity of 77.20% in predicting prolonged recovery. The positive predictive value and negative predictive value were 63.90% and 62.86%, respectively. Discriminant function analysis of the 4 ImPACT neurocognitive composite scores (verbal memory, visual memory, processing speed, reaction time) alone showed significance and was able to correctly classify 65.38% of athletes into long or short recovery. It had a sensitivity of 53.20% and a specificity of 75.44% in predicting prolonged recovery. The positive predictive value and negative predictive value were 64.10% and 66.15%, respectively. The combined discriminant function analysis with the 4 symptom clusters and 4 ImPACT neurocognitive composite scores was significant, and was able to correctly classify 73.53% of athletes into long or short recovery. It had a sensitivity of 65.22% and a specificity of 80.36% in predicting long recovery. The positive predictive value and negative predictive value were 73.17% and 73.8%, respectively. A discriminant function analysis for the total symptom score on the 22-item PCSS alone, was significant and showed a sensitivity and specificity of 40.81% and 79.31%, respectively. The positive predictive value was 62.50% and the negative predictive value was 61.33%. Combining ImPACT and specific symptom cluster PCSS scores might provide better predictive value that either alone. - The migraine symptom cluster contributed the most and was assessed in a separate discriminant function analysis. The migraine symptom cluster alone had a sensitivity of 44.89% and specificity of 78.95% in predicting long recovery. The positive predictive value and negative predictive value were 64.71% and 62.50%, respectively (Lau et al., 2011). - Using the mean of the validation cohort patients' 4 neurocognitive deficit composite percentiles at baseline (of the 5 variables evaluated), a cut point of less than 39 percentile had high sensitivity (0.89) and specificity (0.80) and an area under the ROC curve of 0.85 in predicting the presence of any impairment at follow-up for the derivation cohort; it discriminated equally well in the validation cohort. A cut point of less than 27 percentile had good sensitivity (0.67) and specificity (0.67) and area under the ROC curve of 0.67 in predicting the presence of severe impairment in the derivation cohort at follow-up; it discriminated equally well in the validation cohort. Patients' symptoms at baseline did not predict impairment at 2 weeks post hospital discharge (Wiebe et al., 2012).
Lebby-Asbell Neurocognitive Screening Examination for Adolescents (Lebby et. al., 2015) (Neuro-psychological/Neuro-cognitive)	1 (Kahn et al., 2015)	**Validity**	Mild-Severe - Criterion validity was assessed through length of duration to follow verbal commands. Some subtests (5 of 12) were significantly associated with duration of time to follow commands (Orietnation, Expressive Vocabulary, Verbal Associations, Judgment, Visual Memory) with effect sizes ranging from medium to large where longer duration of time to follow commands was associated with worse performance on the subtest. - Concurrent validity was assessed by regression with variance in WeeFIM Cognition scores at discharge, after controlling for relevant variables. The regression model, including demographic, injury, and admission WeeFIM cognition variables, and Visual Memory and Verbal Associations subtest scores from the LANSE-A, accounted for 62% of the variance in WeeFIM Cognition scores at discharge from rehabilitation ($p < .07$).

Memory Validity Profile (Sherman & Brooks, 2012). (Performance validity)	1 (Brooks & Sherman, 2019)	**Validity**	Mild - Receiver operating characteristics indicated good separation of the MVP discriminating valid vs invalid performance by AUC (.80). Only 6.5% of participants had invalid performances on the MVP (31.6% sensitivity; 100% specificity). Different cut-off scores were investigating using the MSVT and TOMM data and a cut off score of a total of less than 31 showed optimal sensitivity and specificity (63% sensitivity, 93% specificity) to detect invalid performance.
Paediatric Awareness Questionnaire (Lloyd et. al., 2017) (Self-awareness)	1 (Lloyd et al., 2018)	**Reliability**	Mild-Moderate - PAQ had excellent internal consistency for parent, child, and clinician versions in the TBI group and for the parent-child discrepancy score and clinician-child discrepancy score. Internal consistency ranged from fair to excellent for the typically developing group. Excellent concordance was found between parent and clinician rating for the TBI sample. - Children with TBI typically overestimated their functioning compared to their parents (poorer self-awareness), while children who were typically developing tended to rate their functioning lower than their parents.
		Validity	Mild-Moderate - All but 2 of the 37 items were well correlated and deletion of any item would have resulted in no or negligible score improvement, indicating good construct validity. - The PAQ and Knowledge of Injury Checklist (KIC) showed good convergent validity for parent and child ratings.
Patient Health Questionnaire 9 (PHQ-9) (Kroenke et. al., 2001). (Symptom monitoring/ assessment)	1 (Johnson et al., 2021)	**Reliability**	Mild - Inter-rater reliability was variable with poor-good agreement by inter-class correlations and weak-strong correlations by Spearman row with all correlations significant. - Children reported greater symptoms than parents with significant differences for all groups except 11-13 yos. The greatest differences were seen for females and 16-18 yo.
Pediatric Quality of Life Multidimensional Fatigue Scale (PedsQL MFS) (Varni, 1998). (fatigue)	1 (Crichton et al., 2017)	**Reliability**	Mild-Severe - Internal consistency was acceptable for parent and child ratings of all domains. - When comparing child and parent reports, poor to fair agreement was noted for general and sleep/rest fatigue and moderate agreement for cognitive fatigue. - Although not tested statistically, children tended to report worse fatigue for sleep/rest and cognitive and better scores for general fatigue.
		Validity	Mild-Severe - When compared with CDI-2 parent and child report, the PedsQL MFS was correlated indicating good construct validity.
		Other	Mild-Severe - The PedsQL MFS met acceptable measurement standards for range of measurement. Two parents and no children reported maximum possible values. No parents or children reported minimum possible values.

Post-Concussion Symptom Inventory (PCSI) (Gioia et al., 2008). (Symptom monitoring/ assessment)	2 (Mayer et al., 2020; Zemek et al., 2016)	**Reliability**	Mild - The TBI group showed poor inter-rater reliability for the PCSI total in retrospective ratings compared to parent-report, but correspondence increased for subacute and early chronic phases of reporting. HCs showed poor inter-rater reliability of symptom burden from self-report to parent-report. - Test-retest reliability was fair between retrospective and both subacute and early chronic but poor between subacute and early chronic, especially for parent-reported PCSI compared to self-report. When used to assess healthy controls, the PCSI shows fair to good test-retest reliability but has evidence of bias (higher symptom rating) during retrospective reporting (Mayer et al., 2020).

		Validity	Mild
			- Six methods were compared to identify PPCS in TBI patients.
			    - Methods: PCSI scores (i.e., ICD-10; Smyth et al. (citation 13 in article); Reliable Change index; regression-based- subacute and early chronic scores regressed separately to determine deviation between observed and predicted ratings; *Z*(log10)- standardized common logarithm; *z*(SA-R)- standardized difference between subacute and early chronic scores) through self-report.
			    - There was a 49% spread at the early chronic visit, with inter-method agreement between simple change and standardized change being notably lower at early chronic visit. Likely due to recovery, agreement was poor to fair for all methods (standardized change, simple change through self-report) across subacute and early chronic assessments in TBI patients.
			    - Parent-report ratings followed the same overall pattern, but inter-method agreement was generally higher except between subacute and early chronic assessments.
			    - HCs were more likely to be misclassified for simple change algorithms. current findings question the utility of the most commonly used Simple Change scores for diagnosis of PPCS in clinical settings (Mayer et al., 2020).
			- A 12-point PPCS risk score model, using the PCSI, was derived.
			    - Model included variables of: female sex, age of 13 years or older, physician-diagnosed migraine history, prior concussion with symptoms lasting longer than one week, headache, sensitivity to noise, fatigue, answering questions slowly, and 4 or more errors on the Balance Error Scoring System.
			    - AUC was 0.71 for the derivation cohort and 0.68 for the validation cohort.
			    - Specificity and sensitivity increased with higher risk scores.
			    - Three cut off points were selected for PPCS risk: low risk = 3 pts or less; medium risk = 4-8 points; high risk = 9 pts or greater. This risk score had modest discrimination to stratify PPCS risk from 48 hours post-injury to 28 days and acceptable calibration between observed and predicted probability.
			    - When adding physician prediction at time of ED visit to the model of the validation cohort, there was a significant improvement in prediction of PPCS than compared to physician prediction alone or the original model (Zemek et al., 2016).

Revised Child Anxiety and Depression Scale-Short Version (RCADS)	1 (Johnson et al., 2021)	**Reliability**	Mild - Interrater reliability was found to have poor-good agreement by inter-class correlations and weak-strong correlations by Spearman Rho. All correlations were sig except for 16-18 yo as measured by inter-class correlations (significant when measured by Spearman Rho).

(Ebesutani et. al., 2012). (Symptom monitoring/ assessment)			- Children reported greater symptoms than parents with the greatest differences observed for females and 16-18 yos.
Self Awareness Assessment[a] (Krasny-Pacini et al., 2015) (self-awareness)	1 (Krasny-Pacini et al., 2015)	**Validity**	Severe - For the two participants (of 5) that could be included in this scoping review due to age of injury, the three measures showed varied levels of anosognosia/awareness deficits. - One participant scored noticeably higher (about 70% points higher; worse awareness) for the on-line awareness measure, compared to the two metacognitive knowledge measures. The other participant had less discrepancy between measures, scoring about 10% higher (worse awareness) on the discrepancy score metacognitive knowledge measure than the other two measures.
		Other	Severe - Within a context-sensitive pediatric Goal Management Training (GMT) intervention, these assessments showed good feasibility with all children being able to understand both measures of metacognitive knowledge and with trainers having a maximum of two items (of 31) that they could not reliably assess at the end of the intervention.
Sport Concussion Assessment Tool v3 (SCAT 3) (Sport Concussion Assessment Tool, 2013) (Symptom monitoring/ assessment)	3 (Bressan et al., 2020; Liu & Hicks, 2021; WIlmoth, 2020)	**Reliability**	Mild - At 14-days post-injury, children reported higher symptom severity for some cognitive symptoms (‘distracted easily’ (adj. p = .015) and ‘confused’ (adj. p = .015)), but there was no difference between child and parent report of the four symptom domains, symptom burden, or symptom severity with moderate to high agreement (Liu & Hicks, 2021). - At 4-weeks post-injury, there were no differences in individual symptoms, four symptom domains, symptom burden, or symptom severity between child and parent report. Parent reports of symptoms tended to be higher than child reports, although not significant. Age had no effect (p > .05) on the discrepancy between child/parent reports and neither did the interactions of time of assessment with age, sex, ethnicity, or number of previous concussions (Liu & Hicks, 2021). - When comparing multiple short forms (i.e., 14, 10, 5, and 3 items) at multiple time points (ranging from 24-48 hours to 45 days), each form showed relatively modest test-retest reliability (Wilmoth et al., 2020).
		Validity	Mild - The SCAT3 and ChildSCAT3, when used in the ED, had poor discriminative validity in not significantly predicting PCSS (as measured by the PCSI) 1-month post injury. Multiple logistic regression models (number and severity of symptoms) were not significant for parent or child report. AUC of receiver operator characteristic curves were all below 0.6 (Bressan et al., 2020). - SCAT short forms with 14, 10, 5, and 3 items were developed according to item information curves for a concussed group 48 hours post injury. Feeling slowed down, difficulty concentrating, and feeling in a fog composed the 3-item form. Across symptom severity, items pertaining to emotions provided the weakest estimates. Internal consistency was excellent for the 3- and 22- item forms. All forms correlated with neurocognitive performance (ImPACT), balance (Balance Error Scoring System) and emotional symptoms (Brief Symptom Inventory: 18). Each SCAT formed showed good sensitivity, with the 14-item short form performing best, low rates of false positives, and good positive predictive value. (Wilmoth et al., 2020)

Activities/Participation			

CLASS (Concussion Learning Assessment and School Survey) as the outcome of interest in an Evidence Based Assessment Approach (Gioia et. al., 2020). (academic functioning)	1 (Ransom et al., 2016)	**Validity**	Mild
			- With latent class analysis (i.e., identifies classes of respondents based on the distribution of the responses of the sample where individuals in one class are more similar to one another), two distinct groups were established based on their responses: 1) high and 2) low self-reported academic problems.
			- When used as part of an evidence-based assessment approach, the CLASS showed good content validity (i.e., items reflect relevant symptoms/domains based on existing research/measures) with 56% of participants being categorized as having a “low” number of academic problems and 44% as “high.”
			- When used as part of an evidence-based assessment approach, the CLASS showed good convergent and divergent validity:
			  - Significant associations with pre-injury history of symptoms.
			  - BRIEF and PCSI self-report ratings and ratings of exertional response (EEI on ChEERS) were significant predictors of academic problems (PCSI more so for high school students)
			  - No associations with cognitive performance measures (i.e., Multimodal Assessment of Cognition & Symptoms for Children; ImPACT).
			- When using the BRIEF and the PCSI to supplement the CLASS, the BRIEF and PCSI were significant predictors of overall CLASS scores. Using ROC analyses, they were shown to have excellent accuracy in distinguishing high vs low academic problems, while cognitive measures did not have high accuracy, except for EEI which was significant.
			  - The BRIEF, PCSI, and cognitive measures did not show good discriminative validity for high vs low academic problems. Further, BRIEF and PCSI parent-report ratings were not significant predictors of academic problems.

Contextual Factors			
McMaster Family Assessment Device (FAD) (Epstein, 1983). (Family functioning)	1 (Barney & Max, 2005)	**Reliability**	Mild-Severe - No significant differences in parent and child report.
McMaster Structured Interview for Families (McSIFF) as scored on the McMaster Clinical Rating Scale (MCRS) (Miller et. al., 1994). (family functioning)		**Validity**	Mild-Severe When comparing the MCRA (interview-rated) and FAD (family self-rated) across all time points: - Significantly correlated (low-moderate) with correlations increasing over time (i.e., more highly correlated at 2 years post-injury than baseline or 1 year post-injury) - Significant but modest agreements (i.e., marginal-good reproducibility; 65.9-73.0% agreement of categorical ratings (i.e., clinically healthy vs unhealthy) - 0.50-0.67 sensitivity (i.e., 50-67% of families identified by both the MCRS and FAD as clinically unhealthy) - 65.9-73.0% of cases where FAD predicted the diagnosis of healthy or unhealthy on the MCRS - High specificity (i.e., .77-.81 or 77-81% of families identified by both the MCRS and FAD as clinically healthy)

			- High likelihood ratios (i.e., ORs = 3.8-8.9, the odds a family with an unhealthy FAD score is more likely to be diagnosed by the MCRS)

Pediatric Quality of Life Scale/Inventory (Varni & Limbers, 2009). (quality of life)	1 (Johnson et al., 2021)	**Reliability**	Mild
			- Inter-rater reliability showed poor to excellent agreement through inter-class correlations and weak-strong correlations through Spearman Rho for all scores. For both inter-class correlations and Spearman Rho tests, the following correlations were significant: 1) all scores for the total sample of 8-12 yo except school functioning; 2) physical functioning, social functioning, and total scores for females 8-12 yo; 3) social functioning for males 8-12 yo; 4) all scores for 13-18 yo except for social functioning for males. Inter-class correlations also revealed significant correlations for males 8-12 yo in emotional and psychosocial functioning.
			  - Younger children (8-12 yo) reported better functioning that their parents, with significant differences for males in social functioning, females in physical functioning, and the total sample for physical, social, school, and psychosocial functioning, and the total score.
			  - For older children (13-18 yo), males reported better functioning than their parents with a significant difference in school functioning. Parents reported significantly better functioning for females in emotional functioning. For the total sample, parents reported better emotional functioning, but children reported better physical functioning.

[a]using: 2 measures of metacognitive knowledge/intellectual awareness (1 = discrepancy score between child's rating of questionnaire and trainer's rating; 2 = percentage of stories the child thought would never happen to them while the trainer saw similar events regularly happening to the child, divided by the total number of stories); one measure of on-line/emergent awareness (percentage of activities judged by the child as easy while they completely failed or required significant support to achieve the goal, divided by total number of intervention sessions)