Journal of Athletic Training 2003;38(2):104–112
© by the National Athletic Trainers' Association, Inc
www.journalofathletictraining.org

Evidence for the Factorial and Construct Validity of a Self-Report Concussion Symptoms Scale

Scott G. Piland*; Robert W. Motl*; Michael S. Ferrara*; Connie L. Peterson†

*The University of Georgia, Athens, GA; †James Madison University, Harrisonburg, VA

Scott G. Piland, MA, LATC, contributed to conception and design; acquisition and analysis and interpretation of the data; and drafting, critical revision, and final approval of the article. Robert W. Motl, PhD, contributed to conception and design; analysis and interpretation of the data; and drafting, critical revision, and final approval of the article. Michael S. Ferrara, PhD, ATC, contributed to conception and design; acquisition and analysis and interpretation of the data; and drafting, critical revision, and final approval of the article. Connie L. Peterson, PhD, ATC, contributed to acquisition of the data and drafting and final approval of the article.

Address correspondence to Scott G. Piland, MA, LATC, Department of Exercise Science, The University of Georgia, 300 River Road, Athens, GA 30602. Address e-mail to spiland@coe.uga.edu.

Objective: To evaluate the factorial and construct validity of the Head Injury Scale (HIS) among a sample of male and female collegiate athletes.

Design and Setting: Using a cross-sectional design, we established the factorial validity of the HIS scale with confirmatory factor analysis and the construct validity of the HIS with Pearson product moment correlation analyses. Using an experimental design, we compared scores on the HIS between concussed and nonconcussed groups with a 2 (groups) $\times$ 5 (time) mixed-model analysis of variance.

Subjects: Participants (N = 279) in the cross-sectional analyses were predominately male (n = 223) collegiate athletes with a mean age of 19.49 $\pm$ 1.63 years. Participants (N = 33) in the experimental analyses were concussed (n = 17) and nonconcussed control (n = 16) collegiate athletes with a mean age of 19.76 $\pm$ 1.49 years.

Measurements: All participants completed baseline measures for the 16-item HIS, neuropsychological testing battery, and posturography. Concussed individuals and paired controls were evaluated on days 1, 2, 3, and 10 postinjury on the same testing battery.

Results: Confirmatory factor analysis indicated that a theoretically derived, 3-factor model provided a good but not excellent fit to the 16-item HIS. Hence, the 16-item HIS was modified on the basis of substantive arguments about item-content validity. The subsequent analysis indicated that the 3-factor model provided an excellent fit to the modified 9-item HIS. The 3 factors were best described by a single second-order factor: concussion symptoms. Scores from the 16-item HIS and 9-item HIS were strongly correlated, but there were few significant correlations between HIS scores and scores from the neuropsychological and balance measures. A significant group-by-day interaction was noted on both the 9-item HIS and 16-item HIS, with significant differences seen between groups on days 1 and 2 postconcussion.

Conclusions: We provide evidence for the factorial and construct validity of the HIS among collegiate athletes. This scale might aid in return-to-play decisions by physicians and athletic trainers.

Key Words: confirmatory factor analysis, symptoms

Sport-related concussion research has increased considerably over the past decade. As a result, numerous definitions, severity scales, and return-to-play guidelines have been developed.[1–6] Team physicians and certified athletic trainers have access to a greater amount of information, but the debate concerning the validity and practicality of concussion-evaluation methods and diagnostic tools continues.

Balance, neuropsychological performance, and self-reported symptoms have been commonly cited as means of identifying concussive symptoms.[7–15] Although some investigators have evaluated the validity of balance or neuropsychological performance (or both) for monitoring concussion resolution,[11,12,14,16] the cost and practicality of some of these measures are prohibitive. Balance-diagnostic equipment and forceplate measures are expensive, time consuming, and limited to a few institutions and professional organizations. Use of neuropsychological tests requires that personnel be trained in test administration and interpretation of scores. Hence, clear barriers exist to the implementation of these measures.[17]

Some researchers have cited self-reported symptoms as a practical method for monitoring concussive symptoms.[13,14,18] Self-reported symptoms after concussion have been documented for many years, and in the 3 most commonly used concussion grading scales (Cantu, Colorado, American Academy of Neurology), self-reported symptoms are primary factors in return-to-play guidelines.[1–3,14,18,19] Additionally, researchers often have reported data regarding symptom type, incidence, prevalence, and severity[8,13] but have paid limited attention to the significance of empirically derived, self-report symptom scales.[8,19,20] However, as with the previously mentioned methods, self-reported scales also have weaknesses and limitations. The most commonly noted weakness of a self-report method relates to the honesty and motivation of the respondent. This major threat to validity is minimized by the

Table 1. Head Injury Scale Self-Report Concussion-Symptoms Scale

16-Item Head Injury Scale							
Symptom	Never			Sometimes			Always
Headache	0	1	2	3	4	5	6
Nausea	0	1	2	3	4	5	6
Vomiting	0	1	2	3	4	5	6
Balance difficulty/dizziness	0	1	2	3	4	5	6
Fatigue	0	1	2	3	4	5	6
Trouble falling asleep	0	1	2	3	4	5	6
Sleeping more than usual	0	1	2	3	4	5	6
Drowsiness	0	1	2	3	4	5	6
Sensitivity to light and noise	0	1	2	3	4	5	6
Sadness	0	1	2	3	4	5	6
Nervousness	0	1	2	3	4	5	6
Numbness/tingling	0	1	2	3	4	5	6
Feeling ''slowed down''	0	1	2	3	4	5	6
Feeling like ''in a fog''	0	1	2	3	4	5	6
Difficulty concentrating	0	1	2	3	4	5	6
Difficulty remembering	0	1	2	3	4	5	6

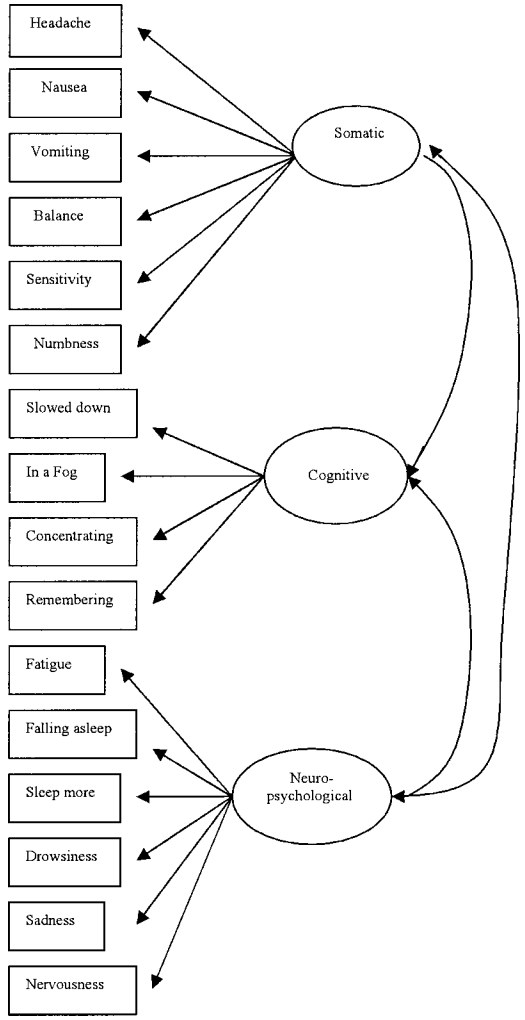


Figure 1. Theoretic measurement model for the Head Injury Scale.

psychometric soundness of the measure and regulated administration.[21,22]

Most investigators examining self-report concussion symptoms have used a variety of simple checklists and summative scales with visually similar designs. The Head Injury Scale (HIS) (Table 1) is a theoretically driven, 16-item, self-report scale derived from those most commonly described in the literature. The items represent symptoms that have been commonly affiliated with sport-related concussion and postconcussion syndrome.[2,18,23,24] To our knowledge, scales in the sport-related concussion literature and their derivations have not been subjected to tests of factorial and construct validity. Evidence of factorial validity is necessary to generate an exact specification or mapping of symptoms (observed variables) with symptom groupings and overall concussive symptoms (latent constructs). Evidence of factorial validity also is important because items on the HIS might not adequately describe concussion symptoms,[18,25,26] and hence, the HIS might need to be modified.

To evaluate the factorial validity of the HIS, it is necessary to first posit and then test a theoretically derived model that describes the latent structure or causal processes underlying responses to the items on the HIS. We adopted a previously described theoretically based model.[17] According to this model, the 16 items on the HIS are described by the 3 latent constructs of somatic, neuropsychological, and cognitive symptoms (Figure 1).[17] These 3 constructs are strongly interrelated and might be represented by an overall latent construct of concussive symptoms (Figure 2). Strongly interrelated factors are often caused by a second-order construct similar to highly related items being caused by first-order constructs.[27] The theoretically derived model is then directly tested for its fit to the 16-item HIS through confirmatory factor analysis (CFA).

Construct validity is essentially the extent to which inferences from scores on a test can be made in relation to the construct of interest.[28] Construct validity is on the basis of the integration of any evidence that bears on the interpretation or meaning of the test scores.[28] Hence, there are multiple methods for evaluating the construct validity of scores from the HIS. Common methods include the examination of correlations with external criteria and the comparison of scores across groups expected to differ in the latent variable of interest.[28] For example, one could examine the expected relationships between baseline HIS scores and scores from composite balance and neuropsychological measures. The HIS scores also could be compared among groups of injured and uninjured samples during a concussive episode.

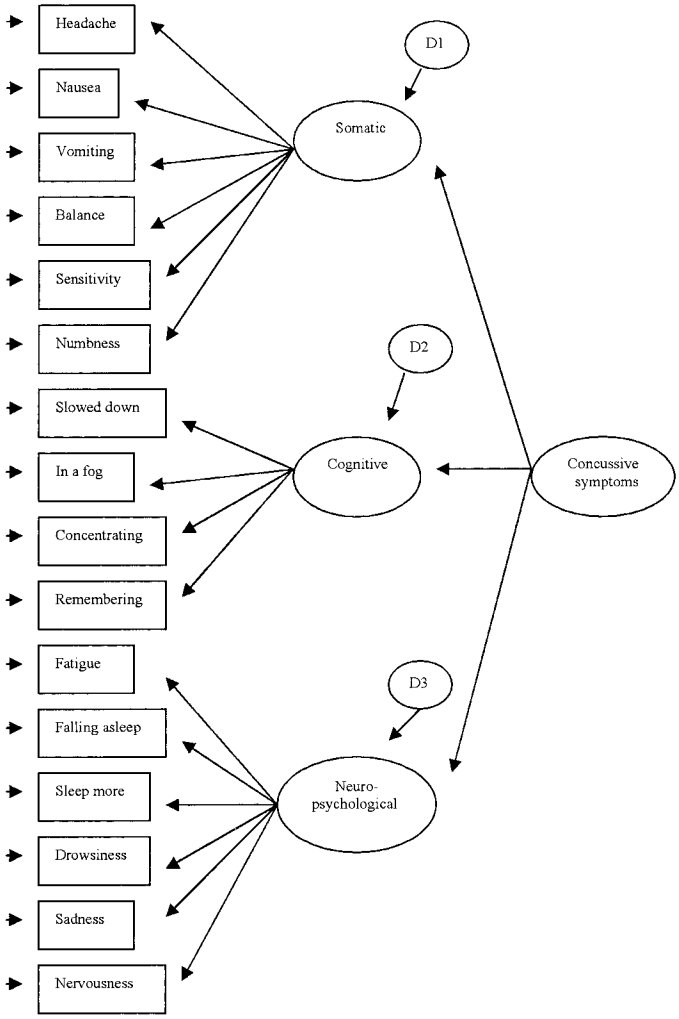


Figure 2. Theoretic second-order measurement model for the Head Injury Scale.

We evaluated the factorial and construct validity of the HIS using CFA and correlation analyses on baseline data from preseason concussion testing of uninjured male and female collegiate athletes. Further evaluations of the construct validity of the original and modified HIS were based on comparisons of scores between groups of concussed and nonconcussed collegiate athletes from before to after a concussive episode.

METHODS

Participants

Participants in the cross-sectional analysis (N = 279) were National Collegiate Athletic Association Division I collegiate sport participants. Subjects' ages ranged from 17 to 27 years (mean = 19.49 ± 1.63 years). The sample was predominantly male (n = 223). All participants signed an informed consent document, and the study was approved by the institutional review board.

Participants in the experimental analyses (N = 33) consisted of concussed (n = 17: 15 males, 2 females) and nonconcussed (n = 16: 14 males, 2 females) Division I collegiate athletes. Subjects' ages ranged from 18 to 23 years (mean = 19.76 ± 1.49 years). Subjects' heights ranged from 66 to 76 in (167.64 to 193.04 cm) (mean height = 71.35 ± 4.77 in [181.23 ± 12.12 cm]). We used the American Academy of Neurology guidelines, and the injured group comprised 1 grade I and 16 grade II concussions.[3] Two concussions were in women's soccer players, with the remaining injuries occurring in football players.

Instrumentation

The 16-item HIS checklist was considered to be representative of the instruments most commonly described in the literature.[17,19,29,30] Discrepancies among other described instruments are mainly reflected in the instructions given to the respondent. Some instruments require a response that is reflective of the symptom severity, whereas others evaluate the duration of each experienced symptom over a set period of time.[8,17,19,20] The HIS finds its resemblance to other instruments in that the 16 items represent symptoms most commonly associated with concussion; also, each item is rated on a 7-point Likert-type scale with response options of ''never'' (0) to ''always'' (6) (see Table 1).

To evaluate the construct validity of HIS scores on the basis of correlation analyses, we used a number of neuropsychological tests commonly employed to evaluate sport-related concussion.[17,29] The neuropsychological testing battery presumably evaluated the domains of attention, short-term memory, speed of information processing, concentration, and verbal fluency. The test battery included the Hopkins Verbal Learning Test (3 trials combined), Halsted-Reitan Trail-Making Test, Symbol Digit Modality Test, Wechsler Digit Span, and Controlled Oral Word Association Test. We expected that scores from those measures might exhibit small to moderate correlations with the overall composite duration score of the HIS.

The NeuroCom Smart Balance Master (NeuroCom Intl Inc, Clackamas, OR) was also used to evaluate the construct validity of the HIS. The test measures an individual's use of sensory inputs, including visual, vestibular, and somatosensory systems, to perform balance tasks. The testing procedure we incorporated has been described previously.[15,31,32] We obtained this measure because the composite balance score may exhibit small to moderate correlation with scores from the HIS.

Procedure

The baseline measures were collected by research staff trained in the administration of a concussion test battery (HIS scale, neuropsychological tests, balance tests). The HIS was administered in conjunction with the neuropsychological and balance assessments during the participant's freshman or preseason physical examination. The instructions for the HIS were given by the test administrator to the subject as follows:

> Here is a list of symptoms that people often feel when they have had a concussion. Please address each symptom based on how you have felt on an average 24-hour period during the past 7 days. Rate the symptoms on a scale of 0 to 6. Zero (0) means that on an average day you have never experienced the symptom, 1 means you experience the symptom very briefly during an average 24-hour period, 3 means the symptom, on an average day, has been present for about half of the preceding 24-hour period, and 6 means the symptom, on an average day, has been continuous through the same time period.

Participants in the experimental part of this study were members of the original sample that underwent baseline testing. The participants either sustained a concussion or were nonconcussed individuals who were paired with concussed

athletes on the basis of sex and age. Upon sustaining a concussion, players were assessed at days 1 (24 hours), 2 (48 hours), 3 (72 hours), and 10 (240 hours) postinjury. Instructions for the HIS were similar to those given by the test administrator at baseline, except symptoms were assessed for the previous 24 hours, not for an average 24-hour period during the past 7 days.

DATA ANALYSIS

Confirmatory Factor Analysis. The factorial validity of the HIS was tested using CFA with maximum likelihood estimation in LISREL 8.50 (Scientific Software Intl, Inc, Chicago, IL). Maximum likelihood estimation assumes that the data represent a multivariate normal distribution; this was not the case in the present sample, based on the ordinal nature of the HIS items and the estimates of multivariate skewness (Mardia normalized estimate = 68.73) and kurtosis (Mardia normalized estimate = 23.90) generated using PRELIS 2.50 (Scientific Software International, Inc). Estimates from normal theory methods such as maximum likelihood are only minimally biased by nonnormal, ordinal data,[33,34] but the standard errors and test statistics can be extremely biased.[35] Consequently, we used the Satorra-Bentler scaled chi-square statistic and standard errors.[36,37] The Satorra-Bentler method corrects the chi-square statistic and the standard errors to account for the nonnormality of the data, and the correction yields more accurate goodness-of-fit statistics and standard errors than other methods developed for nonnormal data.[35,38,39]

Model Specification. Model specification is important for subsequent replication of our results with new data sets. Hence, the initial measurement model underlying the HIS contained 3 latent variables of cognitive, neuropsychological, and somatic symptoms (see Figure 1). The measurement model was specified for input into LISREL using standard procedures for establishing settings in matrices containing factor loadings (λ_x), factor variances and covariances (ϕ), and item uniqueness (θ-δ). The matrix containing factor loadings was specified to reflect simple structure such that items on the HIS loaded on only one of the latent variables. The factor loading for the first indicator on each latent variable was constrained to be 1.0 to establish the metric of each latent variable. The matrix of factor variances and covariances was specified to be symmetric. The matrix of item uniqueness was specified to be diagonal.

We also tested the fit of a higher-order model to the HIS that is similar to Figure 2. The higher-order model was specified such that a single second-order factor (ie, symptoms) described the covariances among the 3 first-order factors (ie, cognitive, neuropsychological, and somatic symptoms). The higher-order model was specified for input into LISREL using standard procedures for establishing settings in matrices containing factor loadings (λ_x), factor variances and covariances (ϕ and ψ), item uniqueness (θ-δ), and path coefficients (γ). The matrix containing factor loadings was specified to reflect simple structure. The factor loading for the first indicator on each latent variable was constrained to be 1.0. The matrices of factor variances and covariances and item uniqueness were specified to be diagonal. The matrix of path coefficients was specified to link a single second-order factor to the 3 first-order factors. The path coefficient for the first latent variable was constrained to be 1.0 to establish the metric of the second-order factor.

Model Fit. We used the chi-square statistic and the Nonnormed Fit Index (NNFI), Comparative Fit Index (CFI), and Root Mean Square Error of Approximation (RMSEA) to evaluate the fit of the models. The chi-square statistic assessed absolute fit of the model to the data, but it is sensitive to sample size and assumes the correct model.[27,40,41] The NNFI is a type II incremental fit index and tests the proportionate improvement in fit by comparing the target model to a baseline model with no correlations among observed variables.[42] The CFI is a type III noncentrality-based fit index and tests the relative improvement in fit by comparing the target model with a baseline model with no correlations among observed variables.[43] The NNFI and CFI values should approximate 0.90 and 0.95 to indicate minimally acceptable[27,42] and good model-data fit.[44] The RMSEA represents closeness of fit or ''error per degree of freedom.''[45] The RMSEA value should be equal to or less than 0.06 to demonstrate a good model-data fit.[44]

Model Modifications. The measurement model was modified to improve the factorial validity of the HIS by identifying a subset of items that best tapped the latent variables. Model modifications were conducted almost exclusively on the basis of substantive information. This substantive information was supplemented only by factor loadings, squared multiple correlations (SMCs), standardized residuals, and modification indices (MIs) provided as output by LISREL. Small factor loadings and SMCs identify items that might not strongly tap the latent constructs of interest. Large standardized residuals (ie, greater than +2.58 or less than −2.58) and MIs in the theta-delta matrix identified pairs of items that were not accurately predicted by the model.[40,46,47] The offending items were removed, and the modified measurement model was retested to determine whether the modifications resulted in an improved fit. The process of identifying and testing model modifications was continued until a reasonable model was generated as indicated by item content and the fit indices, but modifications were made only when substantively appropriate.[40,46–48]

Bivariate Correlation

We used Pearson product moment correlation coefficients to examine the relationships between overall composite scores from the HIS checklist and scores from the other putative measures of concussion (neuropsychological battery and NeuroCom). Composite HIS scores were computed by summing item scores using unity weights. The α value was set at $P \leq .05$ for the correlation analyses.

Experimental Condition

For the second part of this study, we used a 2 (groups: injured and uninjured control) × 5 (time: testing days) mixed-model analysis of variance based on the multivariate F statistic (Pillai-Bartlett) to evaluate differential changes in HIS scores between the injured and control groups across days. Effect sizes associated with the F statistics were expressed as η^2. The Greenhouse-Geisser epsilon (ϵ) was reported when the sphericity assumption was violated (ie, if the Mauchly test of sphericity was statistically significant at $P < .05$). Post hoc analysis of differences between injured and control athletes on individual testing days was performed using independent-samples *t* tests with a 1-tailed *P* value. The α value was adjusted for family-wise comparison using the Bonferroni method. Effect sizes between groups were expressed as the Cohen *d* (ie, injured mean minus uninjured mean divided by the pooled SD).

Table 2. Baseline Responses to the 16-Item Head Injury Scale (N = 279)

Item	Mean	SD	Skewness	Kurtosis
Headache	1.14	1.26	1.06	0.65
Nausea	0.38	0.77	2.32	5.46
Vomiting	0.11	0.42	4.25	20.49
Balance difficulty/dizziness	0.36	0.81	2.99	11.47
Fatigue	0.98	1.22	1.27	1.39
Trouble falling asleep	0.77	1.23	1.78	2.76
Sleeping more than usual	0.75	1.24	1.79	2.86
Drowsiness	0.73	1.14	1.76	2.91
Sensitivity to light and noise	0.33	0.88	3.03	9.36
Sadness	0.40	0.97	3.16	10.92
Nervousness	0.88	1.22	1.41	1.46
Numbness/tingling	0.18	0.60	4.46	24.14
Feeling ''slowed down''	0.50	0.97	2.31	6.19
Feeling like ''in a fog''	0.25	0.73	3.68	15.01
Difficulty concentrating	0.73	1.21	1.91	3.41
Difficulty remembering	0.48	0.94	2.53	7.65

Data analysis was performed using the Statistical Package for the Social Sciences (version 10.1, SPSS Inc, Chicago, IL).

RESULTS

Descriptive Statistics

The means, standard deviations, skewness, and kurtosis values for the individual items on the HIS at baseline are reported in Table 2. The means and standard error of the means for the 16-item HIS and 9-item HIS between concussed and nonconcussed groups across days are represented in Figures 3 and 4.

Confirmatory Factor Analysis on the 16-Item Head Injury Scale

The 3-factor measurement model represented an acceptable fit to the 16-item HIS ($\chi^2_{101} = 195.64$, RMSEA = 0.058 [90% confidence interval = 0.046–0.070], NNFI = 0.962, CFI = 0.968). The chi-square statistic was significant ($P = .0001$), but the RMSEA value did not exceed the acceptable threshold value of 0.06, and the 90% confidence interval around the RMSEA point estimate only minimally exceeded 0.06. The CFI and NNFI values were above the minimally accepted threshold value of 0.90[42] and exceeded the 0.95 criterion established by Hu and Bentler.[44] The standardized covariances between latent constructs were .739 (Somatic and Neuropsychological), .836 (Somatic and Cognitive), and .906 (Neuropsychological and Cognitive). The mean of the standardized factor loadings was .542, with a median of .524; the standardized factor loadings ranged from .285 to .769. Hence support for the 3-factor model for the 16-item HIS was modest, but the model likely could be improved on the basis of substantively informed model modifications.

Model Modifications

Substantive ideas concerning the content validity of individual items were used to identify possible areas in which the HIS could be modified to improve its factorial validity. The substantively based modifications were supplemented on the basis of empirical information from the CFA (ie, factor loadings, SMCs, standardized residuals, and MIs) but only when necessary to identify the offending item. The item ''vomiting'' was considered to be a physical act caused by ''nausea.'' Hence, redundancy existed across items. The vomiting item was removed on the basis of the CFA. ''Sadness'' and ''nervousness'' are peculiar concussion symptoms and appeared to

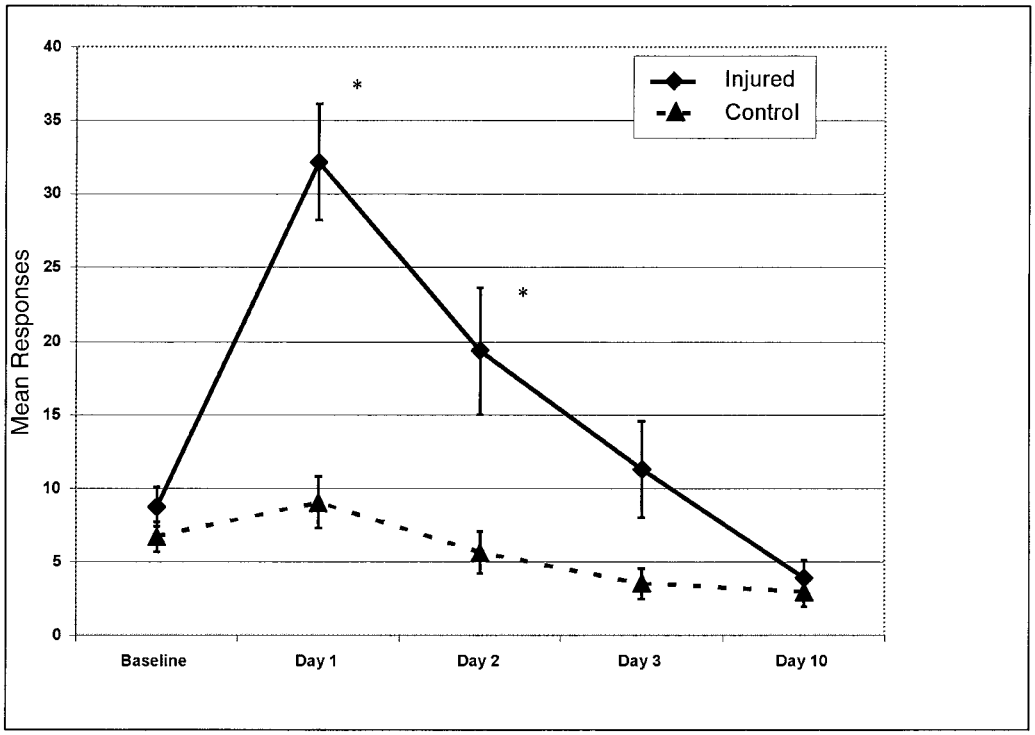


Figure 3. Mean responses and standard error of the mean for injured and control groups using the 16-Item Head Injury Scale across testing days. *$P \leq .05$.

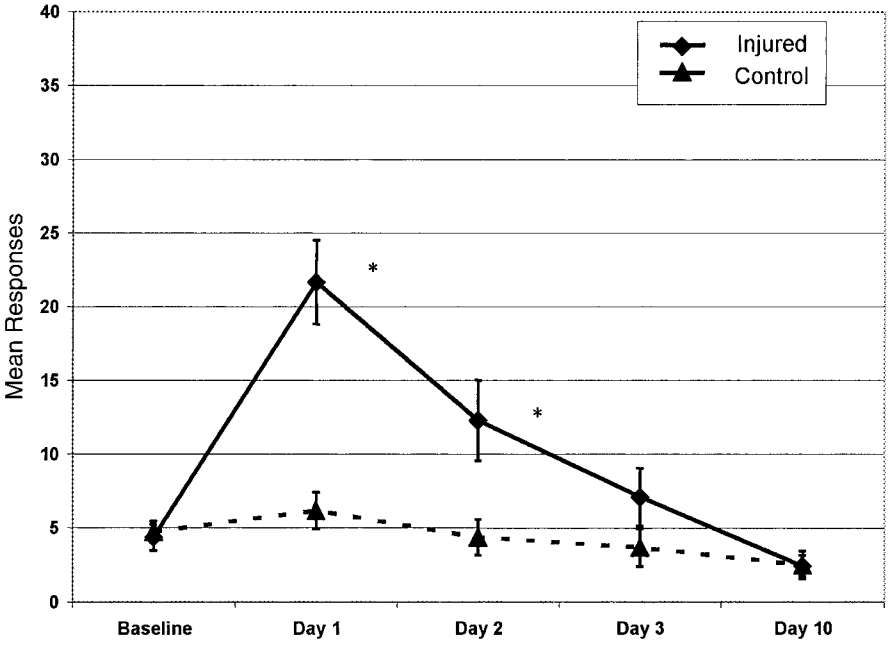


Figure 4. Mean responses and standard error of the mean for injured and control groups using the 9-Item Head Injury Scale across testing days. *$P \leq .05$.

have weak content validity; both items were removed. ''Sleeping more than usual'' was redundant with other neuropsychological symptoms (eg, trouble falling asleep and drowsiness), and thus, was removed. The item ''sensitivity to light and noise'' combined 2 symptoms that might have been independently linked to the construct of concussion symptoms, but in the combined format, the item was deemed problematic and was removed from the model. The item ''feeling numbness and tingling'' also combined 2 symptoms. Further, this item appeared to be more strongly linked to anatomical injury than to neurologic injury. Lastly, the items ''difficulty remembering'' and ''difficulty concentrating'' seemed to be highly related. Empirical information from the CFA demonstrated that the item ''difficulty remembering'' should be removed. The substantively informed modifications in total led to the removal of 7 items, resulting in a 9-item version of the HIS.

Confirmatory Factor Analysis on the 9-Item Head Injury Scale

The 3-factor measurement model represented an excellent fit to the 9-item HIS ($\chi^2_{24} = 30.02$, RMSEA = 0.030 [90% confidence interval = 0.000–0.060], NNFI = 0.993, CFI = 0.995). The chi-square statistic was not significant ($P = .184$). The RMSEA value was 0.030, and its confidence interval included zero, indicating the possibility of exact model-data fit. The NNFI and CFI also provided strong support for the 9-item, 3-factor model. The NNFI and the CFI values greatly exceeded the 0.95 standard.[44] The standardized covariances between latent constructs were .667 (Somatic and Neuropsychological), .717 (Somatic and Cognitive), and .864 (Neuropsychological and Cognitive). The mean of the factor loadings for the 9-item HIS was .608, with a median of .583; the factor loadings ranged from .455 and .882. Consequently, support was strong for the 3-factor model to the 9-item HIS.

We tested the fit of a single second-order factor to describe the covariances among the 3 first-order factors. The higher-order model represented an excellent fit to the 9-item HIS ($\chi^2_{24} = 30.02$, RMSEA = 0.030 [90% CI = 0.000–0.060], NNFI = 0.993, CFI = 0.995), and fit identically compared with the correlated, 3-factor measurement model. This was expected because the higher-order model contained only 3 first-order latent variables.[27] Therefore, support was strong for a single second-order factor underlying the 3 first-order factors underlying the 9-item HIS.

Internal Consistency

The internal consistency of the 16-item HIS and 9-item HIS was estimated using Cronbach coefficient alpha.[49] The estimates of internal consistency for the 16-item HIS and 9-item HIS were 0.84 and 0.78, respectively. The reason for the slight difference in estimates involves the number of test items; coefficient alpha is positively biased by the number of items on a self-report scale.

Correlation Analysis

Few statistically significant relationships were observed between HIS scores and scores from other putative measures of concussion. No statistically significant relationships were found between composite balance scores and scores from either version of the symptom scale. Correlations significantly greater than zero were observed between scores from the 16-item HIS and the Hopkins Verbal Learning Test Combined and between scores from the 16-item HIS and the Controlled Oral Word Association Test (Table 3). Significant correlations were also observed between scores from the 9-item HIS and scores from the Hopkins Verbal Learning Test Combined and between scores from the 9-item HIS and the Controlled Oral Word Association Test (see Table 3). Negative correlations were expected because of the nature of the individual score values (eg, higher scores on the Hopkins Verbal Learning Test Combined and the Controlled Oral Word Association Test are better, whereas a lower score on the HIS is better). A strong,

Table 3. Correlations of Responses to the 16-Item and 9-Item Head Injury Scales with the Neuropsychological Testing Battery and NeuroCom Balance Master (Pearson *R*)

Measure	16-Item Head Injury Scale	9-Item Head Injury Scale
Composite balance	−0.023	−0.043
Hopkins Verbal Learning Test Combined	−0.143*	0.157†
Halsted-Reitan Trail-Making Test A	−0.005	−0.050
Halsted-Reitan Trail-Making Test B	−0.063	−0.065
Symbol Digit Modality Test	0.038	0.031
Wechsler Digit Span Forward Test	0.006	0.009
Wechsler Digit Span Backward Test	−0.038	−0.044
Hopkins Verbal Learning Test Delayed	0.073	0.096
Controlled Oral Word Association Test	0.161†	0.187†

*Statistically significant at $P < .05$.
†Statistically significant at $P < .01$.

statistically significant correlation was noted between scores from the 16-item HIS and the 9-item HIS ($r = .953$).

Experimental Condition Analysis

The 2×5 mixed-model analysis of variance on the 16-item HIS demonstrated a significant group-by-time interaction ($F_{4,28} = 7.16$, $P = .001$, $\eta^2 = 0.521$, $\epsilon = 0.537$). No statistically significant difference was seen between concussed and nonconcussed groups on the baseline HIS scores ($t_{32} = 0.661$, $P = .513$, $d = 0.23$). Large, statistically significant differences were demonstrated between concussed and nonconcussed groups on the 16-item HIS scores on day 1 ($t_{32} = 4.42$, $P = .001$, $d = 1.68$) and day 2 ($t_{32} = 2.55$, $P = .001$, $d = 1.00$). Nonsignificant differences existed between groups on day 3 ($t_{32} = 1.92$, $P = .063$, $d = 0.76$) and day 10 ($t_{32} = 0.458$, $P = .650$, $d = 0.16$) (see Figure 3).

The 2×5 mixed-model analysis of variance on the 9-item HIS demonstrated similar results with a significant group-by-time interaction ($F_{4,28} = 6.38$, $P = .001$, $\eta^2 = 0.542$, $\epsilon = 0.595$). No statistically significant difference was noted between concussed and nonconcussed groups on baseline HIS scores ($t_{32} = -0.322$, $P = .749$, $d = -0.11$). Large, statistically significant differences were seen between groups on the 9-item HIS scores on day 1 ($t_{32} = 4.50$, $P = .001$, $d = 1.81$) and day 2 ($t_{32} = 2.43$, $P = .001$, $d = 0.93$) but not on day 3 ($t_{32} = 1.38$, $P = .177$, $d = .49$) or day 10 ($t_{32} = -0.047$, $P = .963$, $d = -0.02$) (see Figure 4).

DISCUSSION

We evaluated the factorial and construct validity of a commonly accepted concussion-symptoms scale, namely, the HIS. Initially, we tested the fit of a theoretically generated measurement model to the 16-item HIS using CFA. The measurement model demonstrated a good but not perfect fit to the 16-item HIS. We then used substantive information supplemented by empirical indices[40,46,47] to refine and improve the factorial validity of the HIS. The modified version of the HIS consisted of 9 items, and the measurement model exhibited an excellent, nearly perfect fit. We then established the construct validity of scores from the 16-item HIS and the 9-item HIS on the basis of correlations with scores from external criteria and comparisons of scores between injured and uninjured groups.

Factorial Validity

The initial, theoretically derived measurement model consisted of 3 correlated factors and represented a good but not perfect fit to the 16-item HIS. The lack of a perfect fit identified the need to refine the 16-item HIS. This was further supported by our observation that some of the 16 symptoms did not appear to ideally tap one of the 3 common groups of concussion symptoms. Including symptoms that do not tap the latent construct of concussive symptoms within a concussion scale might be problematic. The inclusion of nonrelevant indicators could increase the chances of a false-positive result. Moreover, inclusion of nonrelevant indicators might increase the chance of an athlete reporting a higher number of symptoms at uninjured baseline because of another unknown or unrelated condition. Evaluation of concussion resolution depends upon the presence of a representative baseline measure. If baseline measures are confounded by reports of symptoms that are not specific to concussion, injury severity and resolution could be misdiagnosed and underreported.

We then generated a 9-item version of the HIS. The 9-item version was based primarily on an inspection of item content and then supplemented by empirical evidence from the CFA. We found that a 3-factor correlated measurement model represented an excellent, nearly perfect fit to the 9-item HIS. The 3 first-order factors were best described by a single second-order factor, namely, concussion symptoms. Hence, the 9-item HIS exhibited strong evidence of factorial validity. Headache, nausea, and difficulty in balancing tapped into the somatic symptoms group. Fatigue, trouble falling asleep, and drowsiness tapped into the neuropsychological group. Feeling ''slowed down,'' feeling ''in a fog,'' and difficulty concentrating tapped into the cognitive group of symptoms. We believe that of the 16 original symptoms, the 9 symptoms in the final model represent excellent descriptors of concussion.

Construct Validity

We evaluated the construct validity of scores from the modified 9-item HIS and the original 16-item HIS. Importantly, scores from the 9-item HIS and the 16-item HIS were strongly correlated. The strong relationship between scores from the modified and the original HIS provided convergent evidence for the construct validity of the 9-item HIS scores. The strong relationship indicates that the 9-item HIS taps the construct of concussion symptoms comparably with the 16-item HIS.

We observed few statistically significant relationships between scores from the 9-item HIS and the 16-item HIS and other putative measures of concussion. The relative lack of significant and sizable relationships between scores from the 9-item HIS and the 16-item HIS and other putative measures of concussion symptoms might provide discriminate evidence for the construct validity of HIS scores. Perhaps the HIS taps components of concussion symptoms that are not captured by other putative measures of concussion. Importantly, the truncated distribution of baseline HIS scores and scores from the other measures of concussion symptoms might have limited our ability to adequately evaluate the possible existing correlations. Therefore, we further evaluated the construct validity of HIS scores by comparing scores between injured and uninjured groups across time.

The 9-item HIS and the 16-item HIS were comparable in monitoring self-reported concussion-symptom resolution. With both versions of the HIS, a nonsignificant difference was seen between groups at baseline, but scores were significantly increased on days 1 and 2 in the concussed group compared with the nonconcussed group. No significant differences in HIS scores between groups on day 3 were observed, but the moderate effect size on this day suggests the presence of substantial group differences. This reflects the decreased statistical power of the small number of subjects in this study. Group differences on day 10 were not supported by either statistical group differences or elevated effect sizes. This pattern of change was similar to other research.[13,19] Typically, self-reported concussion-related symptoms have been shown to resolve within 2 to 7 days using either duration or severity as a descriptor.[8,13,18,19,31] Although these types of findings are common in the sport-related literature, it is important to realize that longer durations have been reported in the non-athletic population.[50–52] For this study, we focused on the common effects of grades I and II sport-related head injuries (American Academy of Neurology concussion scale[1]). We observed a duration of less than 10 days for the resolution of self-reported concussion symptoms. Hence, our observed changes in HIS scores are consistent with previous reports in the literature and provide strong support for the construct validity of scores from both the 9-item HIS and the 16-item HIS.

A smaller difference was observed between groups in mean scores on the 9-item HIS and the 16-item HIS at baseline and on day 10. This suggests that with the 16-item HIS, subjects within the concussed group reported a higher number of symptoms that were not specific to concussion at uninjured baseline than did the controls. Group means were more similar on the 9-item HIS. We believe that this observation lends further support to the modified, 9-item HIS (see Figures 3 and 4).

Several factors limited our study. We did not cross-validate the 9-item HIS in an independent sample. We also acknowledge the small sample size used in the experimental condition. However, the number of athletes who incurred a concussion during this study is consistent with the rate of concussion injury reported in the literature.[53] Another limitation to this study was that respondents reported their symptoms in terms of duration. The severity of each symptom was not defined or evaluated. We believe that future researchers should incorporate the use of severity descriptors. This would not only be consistent with current ideals but would also allow greater and more specific elaboration on the meaning of reported symptoms. Finally, data to conduct appropriate reliability analysis were not collected for this study. Reliability and objectivity evidence are integral to the continued development and support of this type of self-report measure.

CONCLUSIONS

This study has provided evidence for the factorial and construct validity of the 9-item HIS instrument. We acknowledge that symptoms included in other scales and symptoms that have been combined within the current form (eg, sensitivity to light and noise) may still function as indicators of concussion; however, scores from the 9-item HIS performed optimally in evaluating self-reported concussion-symptom resolution.

Concern for score interpretation and meaning is rightfully generated with the use of a self-report method. Threats to the validity of such types of measures can be reduced with evidence of the psychometric properties of the instrument. We offer evidence for the soundness of a commonly used self-reported symptom scale. These findings support the use of current instruments and provide directions for their continued development and improvement.

The incorporation of a valid self-report symptom scale into a concussion test battery is necessary in making return-to-play decisions when the presence of concussion symptoms should preclude an athlete's participation in sport. Commonly accepted methods for assessing concussion resolution (complete neuropsychological batteries and posturography) may not be available to most certified athletic trainers and team physicians. Thus, the existence of a valid self-report symptom scale that is simple, efficient, and effective provides a measurement tool that can be used in conjunction with other simple measures to offer the physician and athletic trainer more concrete resolution information on which to base return-to-play decisions.

ACKNOWLEDGMENTS

We thank Stephen N. Macciochi, PhD, of the Shepherd Spinal Center, for his excellent review of, and helpful comments on, the initial version of the manuscript. We also extend our gratitude to Marty Mrazik, PhD, of the Mallard Center, for his work in the development of the HIS instrument and to Ron W. Courson, ATC, PT, CSCS, and Ron Eliot, MD, for their assistance in data collection.

REFERENCES

1. American Academy of Neurology. Practice parameter: the management of concussion in sports [summary statement]. Report of the Quality Standards Subcommittee. *Neurology.* 1997;48:581–585.
2. Barth JT, Alves WM, Ryan TV, et al. Mild head injury in sports: neuropsychological sequelae and recovery of function. In: Levin HS, Eisenberg HA, Benton AL, eds. *Mild Head Injury.* New York, NY: Oxford University Press; 1989.
3. Cantu RC. Return to play guidelines after a head injury. *Clin Sports Med.* 1998;17:45–60.
4. Ferrara MS, McCrea M, Peterson CL, Guskiewicz KM. A survey of practice patterns in concussion assessment and management. *J Athl Train.* 2001;36:145–149.
5. Roos R. Guidelines for managing concussion in sports: a persistent headache. *Physician Sportsmed.* 1996;24(10):31–46.
6. Wojtys EM, Hovda DA, Landry G, et al. Current concepts: concussion in sports. *Am J Sports Medicine.* 1999;27:676–687.
7. Bernstein DM. Recovery from mild head injury. *Brain Inj.* 1999;13:151–172.
8. Echemendia RJ, Putukian M, Mackin RS, Julian L, Shoss N. Neuropsy-

chological test performance prior to and following sports-related mild traumatic brain injury. *Clin J Sport Med.* 2001;11:23–31.
9. Guskiewicz KM, Riemann BL, Perrin DH, Nashner LM. Alternative approaches to the assessment of mild head injury in athletes. *Med Sci Sports Exerc.* 1997;29(suppl 7):S213–S221.
10. Guskiewicz KM, Perrin DH. Effect of mild head injury on cognition and postural stability [abstract]. *J Athl Train.* 1998;33(suppl):S-8.
11. Guskiewicz KM, Ross SE, Marshall SW. Postural stability and neuropsychological deficits after concussion in collegiate athletes. *J Athl Train.* 2001;36:263–273.
12. Lovell MR, Collins MW. Neuropsychological assessment of the college football player. *J Head Trauma Rehabil.* 1998;13:9–26.
13. Macciocchi SN, Barth JT, Alves WM, Rimel RW, Jane JA. Neuropsychological functioning and recovery after mild head injury in collegiate athletes. *Neurosurgery.* 1996;39:510–514.
14. McCrory PR, Ariens MT, Berkovic SF. The nature and duration of acute concussive symptoms in Australian football. *Clin J Sport Med.* 2000;10: 235–238.
15. Riemann BL, Guskiewicz KM. Objective mild head injury evaluation through a battery of clinical postural stability tests [abstract]. *J Athl Train.* 1998;33(suppl):S-18.
16. Bleiberg J, Halpern EL, Reeves D, Daniel JC. Future directions for the neuropsychological assessment of sports concussions. *J Head Trauma Rehabil.* 1998;13:36–44.
17. Grindel SH, Lovell MR, Collins MW. The assessment of sport-related concussion: the evidence behind neuropsychological testing and management. *Clin J Sport Med.* 2001;11:134–143.
18. Maroon JC, Lovell MR, Norwig J, Podell K, Powell JW, Hartl R. Cerebral concussion in athletes: evaluation and neuropsychological testing. *Neurosurgery.* 2000;47:659–672.
19. Macciocchi SN, Barth JT, Littlefield LM. Outcome after mild head injury. *Clin Sports Med.* 1998;17:27–36.
20. Binder LM. Persisting symptoms after mild head injury: a review of the post concussive syndrome. *J Clin Exp Neuropsychol.* 1986;8:323–346.
21. Lezak MD. *Neuropsychological Assessment.* New York, NY: Oxford University Press; 1995.
22. Maruish ME, Moses JA. *Clinical Neuropsychology: Theoretical Foundations for Practitioners.* Mahwah, NJ: Lawrence Erlbaum Assoc; 1997: 132–134.
23. Barth JT, Diamond R, Errico A. Mild head injury and post concussion syndrome: does anyone really suffer? *Clin Electroencephalogr.* 1996;27: 183–186.
24. Mittenberg W, Strauman S. Diagnosis of mild head injury and the postconcussion syndrome. *J Head Trauma Rehabil.* 2000;15:783–791.
25. Dikmen SS. Recovery patterns: natural history. *Recovery.* 1998;9:10–11.
26. Larrabee GJ. Constellation of complaints. *Recovery.* 1998;9:12–14.
27. Bollen KA. *Structural Equations with Latent Variables.* New York, NY: John Wiley; 1989.
28. Messick S. Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *Am Psychol.* 1995;50:741–749.
29. Alves WM, Rimel RW, Nelson WE. University of Virginia prospective study of football-induced minor head injury: status report. *Clin Sports Med.* 1987;6:211–218.
30. Alves WM, Macciocchi SN, Barth JT. Postconcussive symptoms after uncomplicated mild head injury. *J Head Trauma Rehabil.* 1993;8:48–59.
31. Guskiewicz KM, Padua DP, Myers JB. Return to play decisions following a mild head injury in collegiate and high school football players [abstract]. *J Athl Train.* 1998;33(suppl):S-22.
32. Riemann BL, Caggiano NA, Lephart SM. Examination of a clinical method of assessing postural control during a functional performance task. *J Sport Rehabil.* 1999;8:171–183.
33. Mardia KV. Measures of multivariate skewness and kurtosis with applications. *Biometrika.* 1970;57:519–530.
34. Muthen B, Kaplan D. A comparison of some methodologies for the factor analysis of non-normal Likert variables. *Br J Math Stat Psychol.* 1985; 38:347–357.
35. Chou CP, Bentler PM, Satorra A. Scaled test statistic and robust standard errors for non-normal data in covariance structure analysis: a Monte Carlo study. *Br J Math Stat Psychol.* 1991;44:347–357.
36. Satorra A, Bentler PM. Scaling corrections for chi-square statistics in covariance structure analysis. In: *Proceedings of the Business and Economic Statistical Section: Papers Presented at the Annual Meeting of the American Statistical Association.* Alexandria, VA: American Statistical Association; 1988.
37. Satorra A. Alternative test criteria in covariance structure analysis: a unified approach. *Psychometrika.* 1989;54:131–151.
38. Fouladi RT. Performance of modified test statistics in covariance and correlation structure analysis under conditions of multivariate nonnormality. *Struct Equation Model.* 2000;7:356–410.
39. Hu LT, Bentler PM, Kano Y. Can test statistics in covariance structure analysis be trusted? *Psychol Bull.* 1992;112:351–362.
40. Jöreskog KG. Testing structural equation models. In: Bollen KA, Long JS, eds. *Testing Structural Equation Models.* Newbury Park, CA: Sage; 1993.
41. Jöreskog KG, Sörbom D. *LISREL 8: User's Reference Guide.* Chicago, IL: Scientific Software Intl Inc; 1996.
42. Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariance structures. *Psychol Bull.* 1980;88:588–606.
43. Bentler PM. Comparative fit indexes in structural modeling. *Psychol Bull.* 1990;107:238–246.
44. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equation Model.* 1999;6:1–55.
45. Browne MW, Cudeck R. Alternate ways of assessing model fit. In: Bollen KA, Long JS, eds. *Testing Structural Equation Models.* Newbury Park, CA: Sage; 2002:136–162.
46. Motl RW, Conroy DE. Confirmatory factor analysis of the Physical Self-Efficacy Scale with a college-aged sample of men and women. *Measure Phys Educ Exerc Sci.* 2000;4:13–27.
47. Motl RW, Conroy DE. Validity and factorial invariance of the Social Physique Anxiety Scale. *Med Sci Sports Exerc.* 2000;32:1007–1017.
48. Paulhus DL. Measurements and control of response bias. In: Robinson JP, Shaver PR, Wrightsmen LS, eds. *Measures of Personality and Social Psychological Attitudes.* San Diego, CA: Academic Press; 1991:17–59.
49. Cronbach LJ. Coefficient alpha and internal structure tests. *Psychometrika.* 1951;6:297–334.
50. National Collegiate Athletic Association. Available at: www.NCAA.org. Accessed: .
51. Rutherford WH, Merritt JD, McDonald JR. Symptoms at one year following concussion from minor head injuries. *Inquiry.* 1979;10:225–230.
52. Barrett K, Ward AB, Boughey A, Jones M, Mychalkiw W. Sequelae of minor head injury: the natural history of post-concussive symptoms and their relationship to loss of consciousness and follow-up. *J Accid Emerg Med.* 1994;11:79–84.
53. Ingebrigtsen T, Waterloo K, Marup-Jensen S, Attner E, Romner B. Quantification of post-concussion symptoms 3 months after minor head injury in 100 consecutive patients. *J Neurol.* 245:609–612.