Structural Validity of a Self-Report Concussion-Related Symptom Scale

SCOTT G. PILAND[1], ROBERT W. MOTL[2], KEVIN M. GUSKIEWICZ[3], MICHAEL MCCREA[4], and MICHAEL S. FERRARA[5]

[1]School of Human Performance and Recreation, University of Southern Mississippi, Hattiesburg, MS; [2]Department of Kinesiology, University of Illinois at Urbana-Champaign, Champaign, IL; [3]Department of Exercise and Sport Science, University of Carolina at Chapel Hill, Chapel Hill, NC; [4]Neuroscience Center, Waukesha Memorial Hospital, Waukesha, WI; and [5]Department of Exercise Science, St. Mary's Athletic Training Education and Research Laboratory, University of Georgia, Athens, GA

ABSTRACT

PILAND, S. G., R. W. MOTL, K. M. GUSKIEWICZ, M. MCCREA, and M. S. FERRARA. Structural Validity of a Self-Report Concussion-Related Symptom Scale. *Med. Sci. Sports Exerc.*, Vol. 38, No. 1, pp. 27–32, 2006. **Purpose:** This study evaluated the factorial validity of a self-report measure of concussion-related symptom severity among a large sample of male, high-school athletes. **Methods:** Participants ($N = 1089$) were nonconcussed, male, high-school football players. All participants completed a single baseline self-report measure of concussion-related symptom severity, namely the graded symptom checklist (GSC). We tested the factorial validity of the measure with confirmatory factor analysis using LISREL 8.50. **Results:** The analysis indicated that a theoretically derived, three-factor model provided a good, but not excellent, fit for the 16-item GSC. Excellent model-data fit was demonstrated for the three-factor model for a 9-item version of the GSC. In both instances, the three factors were best described by a single second-order factor, namely concussion symptomatology. **Conclusions:** This study provides additional evidence for the factorial validity of a summative self-reported measure of concussion-related symptoms. The factor structure represents a cohesive group of nine symptoms that can be explained by three underlying latent variables, namely somatic symptoms, neurobehavioral symptoms, and "cognitive" symptoms, subsumed under a single higher-order factor, namely concussion symptoms. **Key Words:** CONFIRMATORY FACTOR ANALYSIS, FACTORIAL VALIDITY, SELF-REPORT SYMPTOMS (SRS), MEASUREMENT

Current return to play guidelines specify that an athlete must be asymptomatic or symptom free after a concussive incident, for at least 7 d before returning to sport activity (9–11,20). Historically, clinicians have used a multifaceted approach to concussion assessment and decision-making related to the return of concussed individuals to the field of play. This multifaceted approach involves several categories of evaluation, including self-reported symptoms (SRS), physical examination, posturography, neurocognitive performance, and neuroimaging. Scores from each type of measure presumably reflect a gradient of impairment, which assists clinicians in determining injury severity and measuring postinjury recovery (1,3,11,14–17,21,22,25,26). We are unaware of strong evidence supporting the validity of this inference from scores of self-report measures.

Because decisions about the athlete's condition are drawn, in part, on scores from self-report measures, it is appropriate to assess score validity. Validity is a unitary concept that refers to the correctness, meaningfulness, and efficacy of the inferences from scores on a measure (24). Validity is based on the degree to which theoretical and empirical evidence supports the inferences that are made from test scores. The unified theory of validity outlines six types of evidence for score validity. One is factorial validity, which describes how well an underlying dimension can be used to summarize relationships among items. Although all types of validity evidence are important, factorial validity is singularly distinguishable because it provides a basis for the generation of summative scores, which are then used for subsequent tests of score validity, such as divergent and convergent validity (24).

We are aware of only one investigation that has tested the factorial validity of responses to a summative measure of self-report concussion-related symptoms (26). That study evaluated responses of collegiate athletes ($N = 279$) to a 16-item measure of symptom duration using confirmatory factor analysis (CFA). Results demonstrated that a theoretically derived three-factor model provided a good, but not excellent, fit to scores from the 16-item symptom scale. Based on substantive arguments about item content, the scale was modified to include only 9 of the original 16 symptoms, and the subsequent analysis indicated that the three-factor model provided strong evidence of factorial validity with an excellent fit for scores from the nine items.

Although previous research provides some initial structural evidence for the validity of scores on a self-reported scale among nonconcussed athletes, we note a limitation of that research. The sample of respondents was large enough

Address for correspondence: Scott G. Piland, The University of Southern Mississippi, School of Human Performance and Recreation, 118 College Drive #5142, Hattiesburg, MS., 39406-0001; E-mail: scott.piland@usm.edu.
Submitted for publication February 2005.
Accepted for publication August 2005.

0195-9131/06/3801-0027/0
MEDICINE & SCIENCE IN SPORTS & EXERCISE®
Copyright © 2006 by the American College of Sports Medicine
DOI: 10.1249/01.mss.0000183186.98212.d5

Copyright © American College of Sports Medicine. Unauthorized reproduction of this article is prohibited.

initially to evaluate the underlying structure of the instrument, but it was not sufficient to support adequacy of either model in another sample (13,18). Also, a novel attribute of that study was that the instructions required participants to report the duration of each symptom over a specific time span. Typically, self-reported concussion scales require a response associated with symptom severity (12,15,21).

To address the issues of the aforementioned research, this investigation evaluated the factorial validity of a self-reported symptom severity scale among a large sample of healthy, nonconcussed high-school athletes.

METHODS

Participants. This study was IRB approved. All participants signed parental consent as well as informed assent documents before participation in accordance with the rules and regulations designed for the protection of human subjects. Participants ($N = 1089$) were healthy, nonconcussed, active members of interscholastic high-school football teams from across the eastern United States. The sample was all male with a mean age of 16.3 ± 0.9 yr.

Instrumentation. The instrument used in this study was the graded symptom checklist (GSC) (15,23). The GSC contains 20 concussion-related symptoms that are rated on a seven-point Likert-type scale with anchors ranging from none to mild to moderate to severe. Participants were asked to respond whether or not they were experiencing a particular symptom on the day of the questionnaire administration and, if so, to rate the severity of the symptom on a scale ranging from 0 (no symptom reported), and 1 (very mild) to 6 (very severe). The GSC is similar in design to the post concussion scale (PCS) introduced in 1998 (21). Consistent with previous research (26), we evaluated the responses to 16 of the 20 GSC items that were identical to the symptoms on the head injury scale (HIS) (25,26). The 16 items included headache, nausea, vomiting, balance problems, sensitivity to light or noise, numbness or tingling, sleeping more than usual, drowsiness, fatigue, sadness, nervousness, trouble falling asleep, feeling "slowed down," feeling like "in a fog," difficulty concentrating, and difficulty remembering. We also evaluated a subset of 9 of the 16 items that have been shown to provide the best potential for model-data fit (26). These nine GSC items include headache, nausea, balance problems, fatigue, drowsiness, trouble falling asleep, feeling "slowed down," feeling like "in a fog," and difficulty concentrating.

Data analysis. Based on reported model-data findings from previous research (26), factorial validity was analyzed by incorporating a strictly confirmatory approach (2). The models were tested using confirmatory factor analysis. The confirmatory factor analysis was performed using LISREL (version 8.50; Scientific Software Incorporated, Chicago, IL) with weighted least squares estimation. The input polychoric correlation matrix and asymptotic variance–covariance matrix were computed with PRELIS (version 2.50; Scientific Software Incorporated, Chicago, IL). Weighted least squares estimation with polychoric correlations and asymptotic variances–covariances were selected because of the ordered categorical nature of the data and violations of normality, as evidenced by the skewness and kurtosis estimates in Table 1. The size of the sample was adequate to both estimate and tests the models using weighted least squares estimation (19).

TABLE 1. Baseline responses to the 16-item severity scale ($N = 1089$).

	Mean	SD	Skewness	Kurtosis
Headache	0.47	1.01	2.55	6.75
Nausea	0.18	0.60	4.32	21.19
Vomiting	0.06	0.40	8.59	81.27
Dizziness	0.22	0.70	3.91	16.72
Fatigue	0.61	1.12	2.06	4.00
Trouble falling asleep	0.51	1.13	2.56	6.49
Sleeping more	0.42	1.04	2.91	8.57
Drowsiness	0.72	1.22	1.93	3.39
Sensitivity to light or noise	0.19	0.73	5.06	28.34
Sadness	0.14	0.60	5.67	38.00
Nervousness	0.27	0.77	3.63	15.25
Numbness or tingling	0.13	0.56	5.30	31.07
Feeling slowed down	0.28	0.83	3.90	17.09
Feeling in a fog	0.11	0.49	5.85	39.46
Concentration difficulty	0.37	0.91	2.97	9.26
Difficulty remembering	0.33	0.85	3.13	10.88

Range: 0 = no symptoms to 6 = severe symptoms.

Model specification. We initially tested the fit of a model containing three correlated factors for the 16 items listed previously (Fig. 1), which was tested for the nine GSC items (Fig. 2). The matrix of factor loadings was specified to reflect simple structure (i.e., each item loaded on only a single factor). The factor loading for the first item on each factor was set to be 1.0 to establish the metric of the latent variables. The matrix of item uniqueness was diagonal, and did not include correlated uniqueness. The matrix containing the factor variances and covariances was symmetric.

We then tested the fit of a second-order model for the 16-item (Fig. 3) and 9-item (Fig. 4) versions of the GSC. The second-order model contained three first-order factors subordinate to a single second-order factor. Again, the matrix of factor loadings for the first-order factors was specified to reflect simple structure, and the factor loading for the first item on each first-order factor was set to be 1.0 to establish its metric. The matrix of item uniqueness was diagonal. The matrix of first-order factor variances and covariances was diagonal. The second-order factor loadings were freely estimated, and the second-order factor variance was set to be 1.0 to establish its metric.

Model fit. The χ^2 statistic and subjective indices were used to evaluate the fit of the models. The χ^2 statistic assessed absolute fit of the model to the data, but it is sensitive to sample size and assumes the correct model (4,7,19). Accordingly, subjective indices of fit were employed to judge and compare the fit of the models. The root mean square error of approximation (RMSEA) represents closeness of fit (8). The RMSEA value should approximate or be <0.05 to demonstrate close fit of the model (8). The 90% confidence interval (CI) around the RMSEA point estimate should contain 0.05 to indicate the possibility of close model-data fit (8). Both the non-normed fit index (NNFI) and comparative fit index (CFI) are incremental fit indices, and test the proportionate improvement in fit by comparing the target model with a baseline model with no

Copyright © American College of Sports Medicine. Unauthorized reproduction of this article is prohibited.

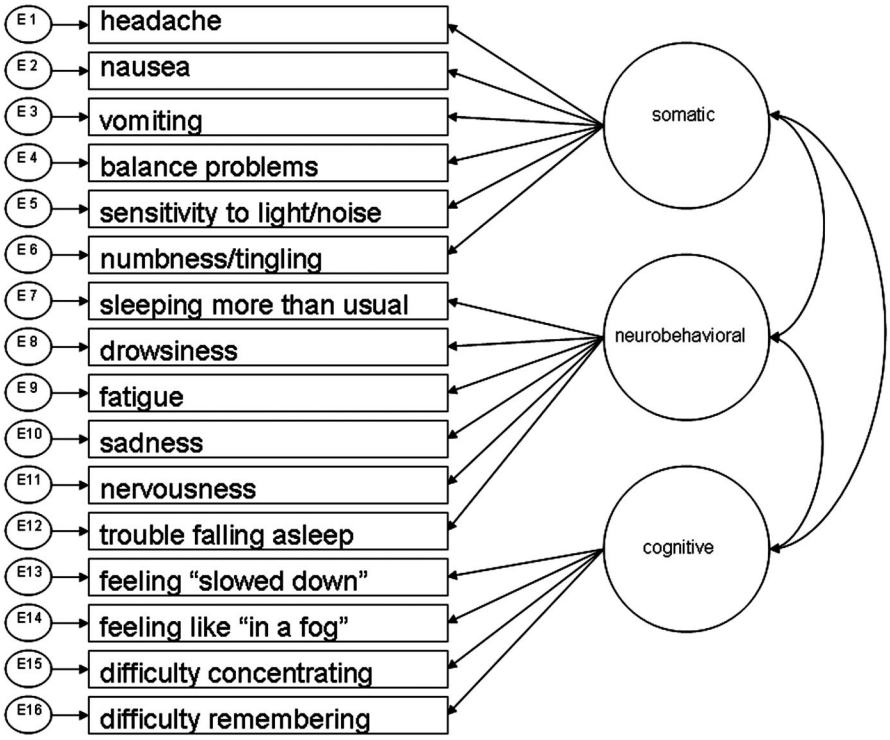


FIGURE 1—Theoretical measurement model for the 16-item GSC.

correlations among observed variables (4,5,8). NNFI and CFI values approximating 0.95 were considered indicative of good model-data fit (6,18).

RESULTS

Descriptive statistics. The means, standard deviations, skewness, and kurtosis values for the individual items on the scale are reported in Table 1. The distribution of responses were both highly skewed and kurtosed, meaning that the data were not normal. This was expected for baseline nonconcussed values

CFA on the 16-item scale. The three-factor measurement model represented a good, but not excellent, fit to the 16-item scale (χ^2 = 315.97, *df* = 101, RMSEA = 0.04 (90% CI = 0.04–0.05), NNFI = 0.96, CFI = 0.97). The χ^2 statistic was significant ($P < 0.0001$), but the RMSEA was less than the acceptable threshold value of 0.05. The CFI and the NNFI values were above the criteria value of 0.95 recommended by Hu and Bentler (18). This provided adequate support for the good fit of the three-factor model to the data, but previous findings indicate that model-data fit can be improved based on reported model modifications (26). The standardized covariances between latent constructs

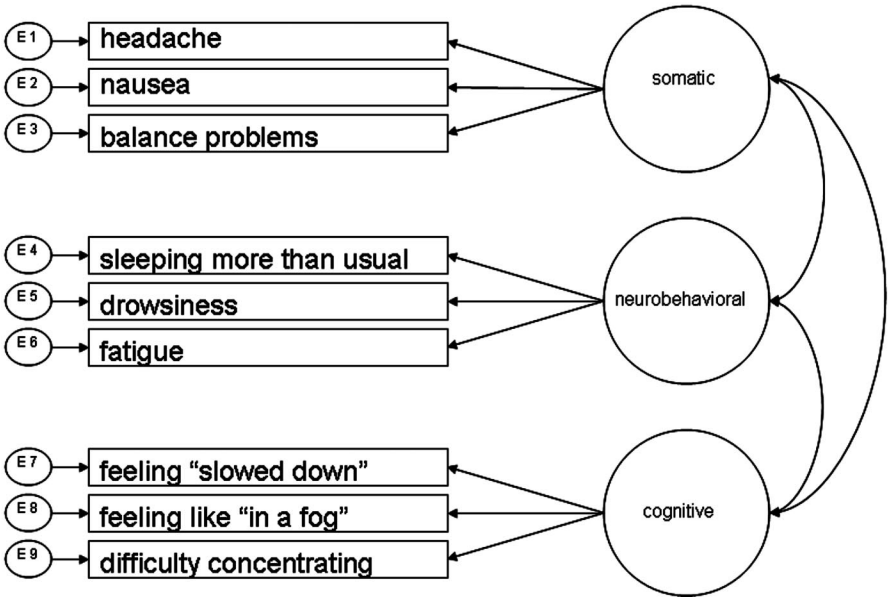


FIGURE 2—Theoretical measurement model for the nine-item GSC.

Copyright © American College of Sports Medicine. Unauthorized reproduction of this article is prohibited.

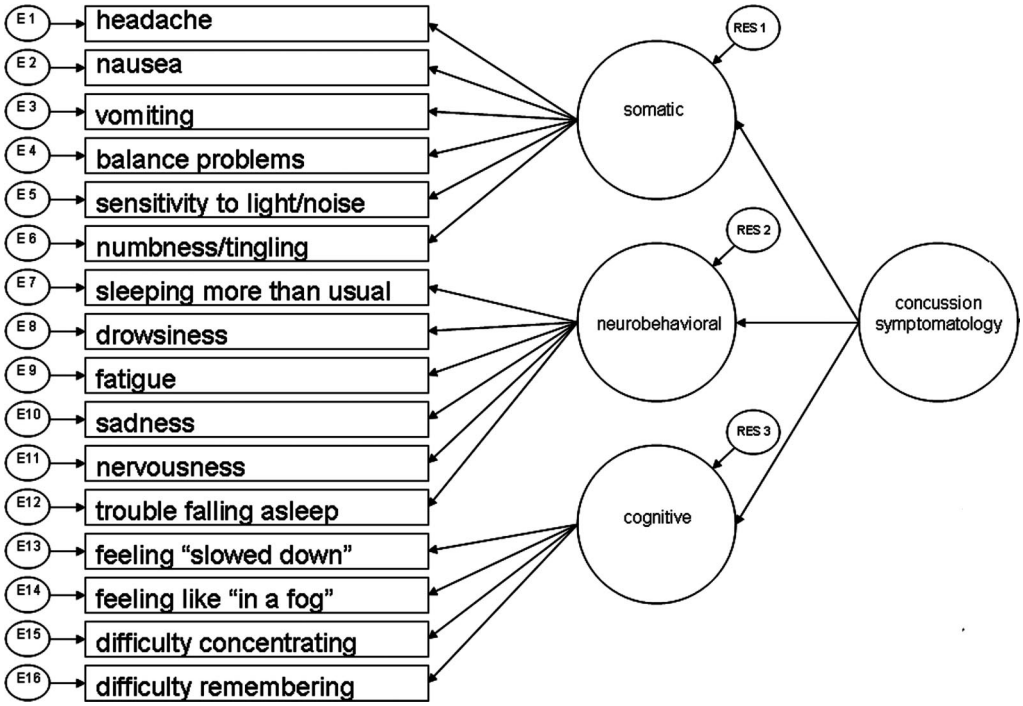


FIGURE 3—Theoretical second-order, measurement model for the 16-item GSC.

were 0.93 (somatic and neurobehavioral), 0.94 (somatic and cognitive), 0.93 (neurobehavioral and cognitive). The mean of the factor loadings for the 16-item scale was 0.85, with a median of 0.86; the factor loadings ranged between 0.68 and 0.96. Consequently, support was seen for the three-factor model to the 16-item scale.

We tested the fit of a single second-order factor to describe the covariances among the three first-order factors. The higher-order model represented a good fit to the 16-item HIS (χ^2) = 315.97, *df* = 101, RMSEA = 0.04 (90% CI = 0.04–0.05), NNFI = 0.96, CFI = 0.97), and fit identically compared with the correlated, three-factor measurement model. This was expected as the higher-order model contained only three first-order latent variables (7). Therefore, support was seen for a single second-order factor underlying the three first-order factors underlying the 16-item GSC.

CFA on the 9-item scale. The structure analysis of the 9-item scale resulted in stronger indices of fit than those reported with the 16-item scale. The three-factor measurement model represented an excellent fit to the 9-item scale (χ^2) = 55.79, *df* = 24, RMSEA = 0.03 (90% CI = 0.02–0.05), NNFI = 0.98, CFI = 0.99). The χ^2 statistic was

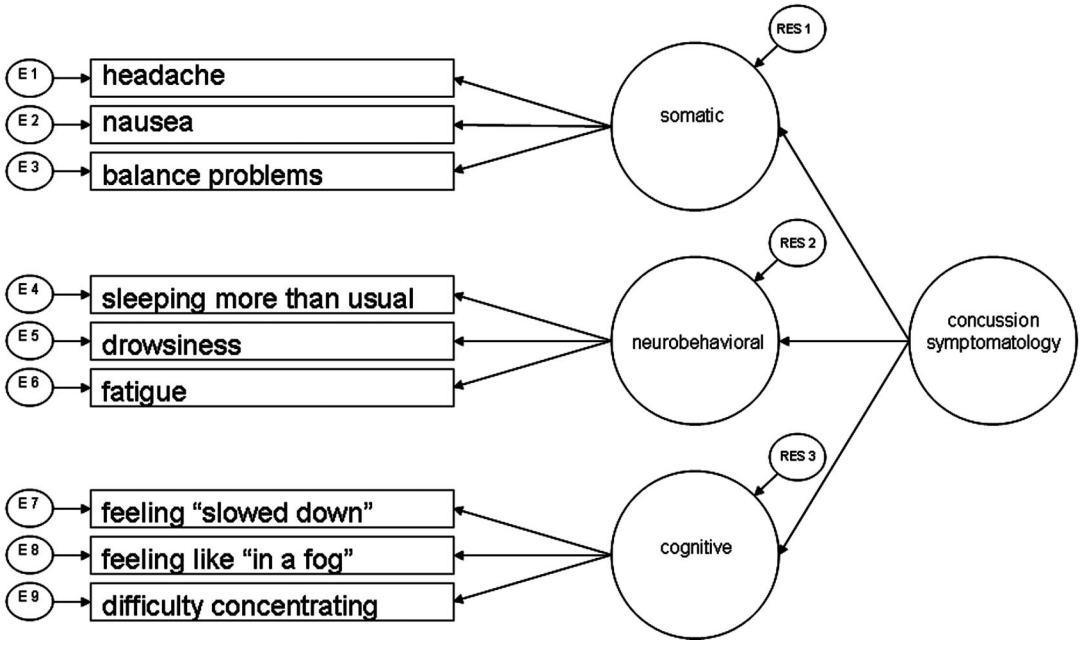


FIGURE 4—Theoretical second-order, measurement model for the nine-item GSC.

Copyright © American College of Sports Medicine. Unauthorized reproduction of this article is prohibited.

significant ($P < 0.0001$), but the RMSEA did not exceed the acceptable threshold value of 0.05 and was lower than the RMSEA of the 16-item scale. The CFI and the NNFI values exceeded acceptable threshold values of 0.95 and were higher than the previous 16-item scale. These improvements in indices of fit provide strong support for the excellent fit of the three-factor model to the data. The standardized covariances between latent constructs were 0.86 (somatic and neurobehavioral), 0.91 (somatic and cognitive), and 0.90 (neurobehavioral and cognitive). The mean of the factor loadings for the nine-item scale was 0.83, with a median of 0.88: the factor loadings ranged between 0.62 and 0.93. Consequently, strong support was seen for the three-factor model to the nine-item scale.

We tested the fit of a single second-order factor to describe the covariances among the three first-order factors. The higher-order model represented an excellent fit to the nine-item HIS (χ^2) = 55.79, df = 24, RMSEA = 0.03 (90% CI = 0.02–0.05), NNFI = 0.98, CFI = 0.99), and fit identically compared with the correlated, three-factor measurement model. As with the 16-item model, this was expected because the higher-order model contained only three first-order latent variables (7). Therefore, support was seen for a single second-order factor underlying the three first-order factors underlying the nine-item GSC.

DISCUSSION

This study provided additional evidence for the factorial validity of a summative self-reported measure of concussion-related symptoms. Although the variance associated with the present data is small, our results support that noninjured athletes report concussion-related symptoms (e.g., headache, nausea, balance difficulty, fatigue, trouble falling asleep, drowsiness, feeling "slowed down," feeling "in a fog," and difficulty concentrating) at nonconcussed preseason baseline. In addition, these symptoms are reported in a predictable scoring structure. This structure represents a tightly bound, cohesive group of nine symptoms that can be explained by three underlying latent variables ("somatic" symptoms, "neurobehavioral" symptoms, and "cognitive" symptoms) and a single higher-order factor (concussion symptomatology). Such evidence provides strong support for the structural validity of the measurement of sport-related concussion.

The present study successfully addressed limitations of previous research that examined the structural aspects of score validity for self-reported concussive symptoms (26). The sample of respondents was large enough to establish the underlying scoring structure of the 16-item instrument, and sufficient to support the data-generated 9-item instrument. Moreover, findings suggest that the underlying scoring structure is not affected by the type of descriptor (duration or severity) required by the instrument, although this issue needs to be evaluated in greater detail by more appropriate means in a sample of athletes with concussion.

A novel aspect of this study was that our sample was composed of high-school student athletes. The implication of using this sample is that younger student-athletes respond to a self-reported measure of concussion-related symptoms in a similar fashion to that of the older college-aged population, and that separate scales are not likely necessary for clinical use in these two populations. The relevance of this finding is also substantiated by the fact that a greater number of physicians and certified athletic trainers actively work with the younger of the two populations. Often, clinicians who serve this level of athlete do not have access to the more expensive and time-consuming measures of sport-related concussion and, thus, rely on more subjective or informal methods of concussion assessment. Finding a similarity of response structure between the different aged groups of student athletes provides strong support for the potential inferences drawn across varying age groups and justifies the need for the continued development of a systematic measure of self-report symptoms following sport-related concussion.

It is critical to realize that across the vast spectrum of the test and measurement validation process, our findings are mere first steps that provide a solid foundation for future investigations. It is commonly realized that methods of self-report are suspect to violations of external and internal validity. Issues related to the generalizability of baseline responses to postconcussion responses, test administration, respondent honesty, socially motivated response bias, clarity of instructions, and relationships of baseline SRS scores to other purported measures of concussion severity all play integral roles into the interpretation and understanding of score meaning. With this understanding, the investigation of these issues is more reliably performed with measures shown to be adequate. The findings from this study build on the adequacy of the investigated measure, which provides support for its use as an investigative tool. It is also important to understand that empirical evidence for the validity of concussed responses to summative SRS scales has not yet been performed with large samples. Such evidence must be presented before clinicians' base treatment decisions on inferences made from scores obtained via postinjury self-report symptom scales.

REFERENCES

1. Alves, W., S. N. Macciocchi, and J. T. Barth. Postconcussive-symptoms after uncomplicated mild head injury. *J. Head Trauma Rehabil.* 8: 48–59, 1993.
2. Anderson, and J., D. Gerbing. Structural equation modeling in practice: a review and recommended two-step approach. *Psychol. Bull.* 103:411–423, 1988.
3. Barth, J. T., S. Macciocchi, A. Giordano, R. W. Rimel, J. A. Jane, and J. Boll. Neuropsychological sequelae of mild head injury. *Neurosurgery* 13:529–533, 1983.
4. Bentler, P. M., and D. G. Bonett. Significance tests and goodness of fit in the analysis of covariance structures. *Psychol. Bull.* 88: 588–606, 1980.
5. Bentler, H. Comparative fit indexes in structural modeling. *Psychol. Bull.* 107:238–246, 1990.

Copyright © American College of Sports Medicine. Unauthorized reproduction of this article is prohibited.

6. BENTLER, H., and D. BONETT. Significance tests and goodness of fit in the analysis of covariance structures. *Psychol. Bull.* 88:588–606, 2003.
7. BOLLEN. K. A. *Testing Structural Equation Models*, Newbury Park, CA: Sage, 1993, pp. 10–39.
8. BROWNE, M. W., AND R. CUDECK. Alternate ways of assessing model fit. In: *Testing Structural Equation Models*, K.A. Bollen and J.S. Long (Eds.). Newbury Park, CA: Sage, 2002, pp. 136–162.
9. CANTU, R. C. Return to play guidelines after a head injury. *Clin. Sports Med.* 17:45–60, 1998.
10. CANTU, R. C. Posttraumatic retrograde and anterograde amnesia: pathophysiology and implications in grading and safe return to play. *J. Athl. Train.* 36:244–248, 2001.
11. COLLINS, M. W., S. H. GRINDEL, M. R. LOVELL, et al. Relationship between concussion and neuropsychological performance in college football players. *JAMA* 282:964–970, 1999.
12. COLLINS, M. W., G. L. IVERSON, M. R. LOVELL, D. B. MCKEAG, J. NORWIG, and J. MAROON. On-field predictors of neuropsychological and symptom deficit following sports-related concussion. *Clin. J. Sports Med.* 13:222–229, 2003.
13. CUDECK, R., and M. BROWNE. Cross validation of covariance structures. *Multivariate Behavior Research* 18:147–167, 1983.
14. GRINDEL, S. H., M. R. LOVELL, and M. COLLINS. The assessment of sport-related concussion: the evidence behind neuropsychological testing and management. *Clin. J. Sports Med.* 11:134–143, 2001.
15. GUSKIEWICZ, K. M., M. MCCREA, S.W. MARSHALL, et al. Cumulative effects associated with recurrent concussion in collegiate football players: the NCAA Concussion Study. *JAMA* 290:2549–2555, 2003.
16. GUSKIEWICZ, K. M. Postural stability assessment following concussion: one piece of the puzzle. *Clin. J. Sports Med.* 11:182–189, 2001.
17. GUSKIEWICZ, K. M., S. E. ROSS, and S. W. MARSHALL. Postural stability and neuropsychological deficits after concussion in collegiate athletes. *J. Athl. Train.* 36:198–203, 2001.
18. HU, L., and P. M. BENTLER. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structure Equation Model* 6:1–55,1999.
19. JORESKOG, K., AND D. SÖRBOM. *LISREL 8: User's Reference Guide.* Chicago, IL: Scientific Software International, Inc. 1996, pp. 271–274.
20. KELLY, J. P., and J. H. ROSENBERG. Practice parameters: the management of concussion in sports: report of the quality standards committee. *Neurology* 48:581–585, 1997.
21. LOVELL, M. R., and M. W. COLLINS. Neuropsychological assessment of the college football player. *J. Head Trauma Rehabil.* 13:9–26, 1998.
22. MACCIOCCHI, S. N., J. T. BARTH, W. ALVES, R. W. RIMEL, and J. A. JANE. Neuropsychological functioning and recovery after mild head injury in collegiate athletes. *Neurosurgery* 39:510–514, 1996.
23. MCCREA, M., K. M. GUSKIEWICZ, S. W. MARSHALL, et al. Acute effects and recovery time following concussion in collegiate football players: the NCAA concussion study. *JAMA* 290:2556–2563, 2003.
24. MESSICK, S. Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *Am. Psychol.* 50:741–749, 1995.
25. PETERSON, C., M. FERRARA, M. MRAZIK, S. PILAND, and R. ELLIOTT. Evaluation of neuropsychological domain scores and postural stability following cerebral concussion in sports. *Clin. J. Sports Med.* 13:230–237, 2003.
26. PILAND, S., R. MOTL, M. FERRARA, and C. PETERSON. Evidence for the factorial and construct validity of a self-report concussion symptoms scale. *J. Athl. Train.* 38:104–112, 2003.

Copyright © American College of Sports Medicine. Unauthorized reproduction of this article is prohibited.