

# Psychometric Evaluation of the Standardized Assessment of Concussion

## *Evaluation of Baseline Score Validity Using Item Analysis*

Brian G. Ragan, PhD, ATC; Stephen D. Herrmann, MS, ATC; Minsoo Kang, PhD; and Mick G. Mack, PhD

### ■ ABSTRACT

The Standardized Assessment of Concussion (SAC) is a concussion screening tool used in identifying concussions by measuring performance decreases compared with baseline scores. This places great importance on valid baseline and postinjury scores. The objective was to examine baseline psychometric properties of SAC items using item analysis. One hundred forty-seven healthy adults ( $n = 72$  men,  $n = 75$  women; age = 18-59 years) completed all 3 SAC versions in random block order. Item analysis examined item difficulty ( $.1 \geq P \leq .92$ ) and discrimination (point-biserial correlation  $\geq .1$ ). Acceptable items met both criteria. All 3 SAC versions had few acceptable items (A = 33%, B = 30%, C = 27%). Most items were too easy (76%). No orientation, 7% of immediate memory, and 60% of concentration were acceptable. Most SAC items have unacceptable psychometric properties when used to establish baseline cognitive abilities, which questions the validity for postinjury score comparisons.

---

**M**ild traumatic brain injuries (MTBI), also known as *concussions*, are a serious concern that affect athletes and the general public.<sup>1</sup> A

---

Dr Ragan is from the Department of Family Medicine, Case Western Reserve University, Cleveland, Ohio; Mr Herrmann is from the Healthy Lifestyles Research Center Program in Exercise & Wellness, Arizona State University, Mesa, Ariz; Dr Kang is from the Department of Health and Human Performance, Middle Tennessee State University, Murfreesboro, Tenn; and Dr Mack is from the School of Health, Physical Education & Leisure Services, University of Northern Iowa, Cedar Falls, Iowa.

Originally submitted February 18, 2009.

Accepted for publication May 18, 2009.

The authors have no financial or proprietary interest in the materials presented herein.

Presented in part at the National Athletic Trainers' Association Annual Meeting; June 2007; Anaheim, California. The authors thank Karen Kirkwood, Whitney Harris, and Ranae Abel for their help with the data collection.

Address correspondence to Brian G. Ragan, PhD, ATC, Department of Family Medicine, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106-7136; e-mail: brian.ragan@case.edu.

doi:10.3928/19425864-20090625-07

*traumatic brain injury* is defined as a blow or penetrating force to the head, which leads to a disruption of normal brain function.<sup>2</sup> Mild traumatic brain injuries account for 85% of the 1.5 million head injuries that occur annually throughout the United States; approximately 300,000 of these occur in sport-related activities.<sup>3,4</sup> Langlois et al<sup>5</sup> have reported that sport-related traumatic brain injuries are severely underestimated. They estimated that approximately 1.6 to 3.8 million sport-related traumatic brain injuries occur each year.

Memory, attention, and information processing are some of the cognitive functions that are affected by MTBI.<sup>6-9</sup> The accurate measurement of these functions is vital for health care professionals to identify and manage MTBI. It is important to use standardized tests for the assessment of an individual's cognitive status and severity after sustaining an MTBI. Several brief diagnostic screening tools have been developed that can be used to evaluate changes in an individual's cognitive status. For example, the Mini-Mental Status Examination (MMSE)<sup>10</sup> is commonly used with older adults and other cognitively impaired populations. In sports medicine, the Standardized Assessment of Concussion (SAC) was created to provide team physicians, athletic trainers, and other medical professionals involved in the care of sport-related MTBI with a quick standardized, objective, screening tool for use on the field.<sup>11-13</sup>

Unlike the MMSE, which uses a criterion-referenced standard<sup>14</sup> (ie, cut-score) for identification of cognitive dysfunction, the SAC uses an individual-centered standard. This approach is used in sport-related MTBI assessment protocols and is recommended by the National Athletic Trainers' Association.<sup>15,16</sup> The individual-centered standard requires testing to document the athlete's base-

line (noninjured or healthy) cognitive skill level, which is then compared to their postinjury cognitive level.<sup>17</sup>

Identifying dysfunction using an individual-centered standard places great importance on a valid and reliable baseline test score.<sup>18</sup> If a baseline test is not accurate, any postinjury comparisons would be unreliable and attenuate the sensitivity and specificity of the instrument.<sup>17</sup> The article by Ragan and Kang<sup>17</sup> provides a more detailed description of the types of testing standards used in concussion testing.

There have been some psychometric evaluations of the SAC, specifically addressing reliability and validity evidence.<sup>19-22</sup> The sensitivity of the SAC has been reported to range between 0.72 and 0.78.<sup>19,20</sup> Conversely, the validity and utility of the SAC has been called into question because of significant ceiling effects and a lack of data to support its clinical use,<sup>23,24</sup> yet it is still identified as a tool to use in MTBI assessment.<sup>15,16</sup>

In addition, all previous psychometric investigations of the SAC have used total test score and subscale scores for analysis.<sup>19-22</sup> Although these summary measurements are needed, one of the most fundamental evaluations is noticeably missing. An item analysis of the SAC, which evaluates the psychometric properties of the items on a test, has not been reported. A valid and defensible test must be constructed with items that have the proper psychometric properties, item difficulty and discrimination. An item analysis evaluates each item on a test or instrument for difficulty (ie, too easy, acceptable, or too difficult) and its ability to discriminate between examinees (ie, high performers versus low performers; concussed versus nonconcussed). Despite some initial success, the validity and serious methodological shortcomings remain problematic while using the SAC to identify sport-related concussion. These issues may adversely influence medical professionals' decision making and management of MTBI. The purpose of this study was to perform an item analysis on the SAC to examine its validity as a baseline measure for use with an individual-centered standard.

## METHOD

### Participants

A total of 147 healthy adult participants ( $n = 72$  men,  $n = 75$  women; age [mean  $\pm$  SD] =  $21.36 \pm 1.7$  years) who were not currently experiencing symptoms of a concussion volunteered for this study. All participants were college students from 2 midwestern public universities

located in small communities. Participants were recruited from various classes and did not receive compensation.

### Instruments

The SAC test was designed to assess the acute cognitive functions of an individual before and after suffering a suspected head injury in  $<5$  minutes. The SAC test evaluates 4 domains, or sections: orientation, immediate memory, concentration, and delayed memory, which are frequently affected by MTBI. There are 3 versions of the SAC that have been reported to be equivalent.<sup>11</sup> The immediate memory, concentration, and delayed memory sections contain unique items on each version, whereas the orientation section of all 3 versions is identical. The test is scored by adding the number of correct responses; 5 points can be scored in each of the orientation, concentration, and delayed memory sections, and 15 points within the immediate memory section. There are a total of 30 points possible.<sup>11</sup>

### Procedure

The procedures for this one-time observational design study were approved by the university's institutional review board prior to testing. Each participant read and signed an informed consent. The SAC test was administered in a quiet laboratory setting by 1 of 3 examiners following the guidelines in the SAC administration manual.<sup>11</sup> Items from the 3 versions of the SAC were combined to create a new longer test, which took approximately 10 minutes to administer. The new test contained 78 unique items from the 3 versions. There are 6 common items (5 orientation items and the last concentration item, "months in reverse order") administered once. In addition, 24 unique items on each of the 3 versions were administered for a total of 78 unique items. Counterbalancing was used to control for sequencing effects by administering the subscales of the SAC in random block order. The random block orders were ABC, BCA, CAB to counter any order effect. For example, Version A (30 items) would be administered, followed by immediate memory in B (15 items), concentration in B (4 items), delayed memory in B (5 items), immediate memory in C (15 items), concentration in C (4 items), and, finally, delayed memory in C (5 items). The Figure illustrates the counterbalance procedures for the 3 versions of the SAC.

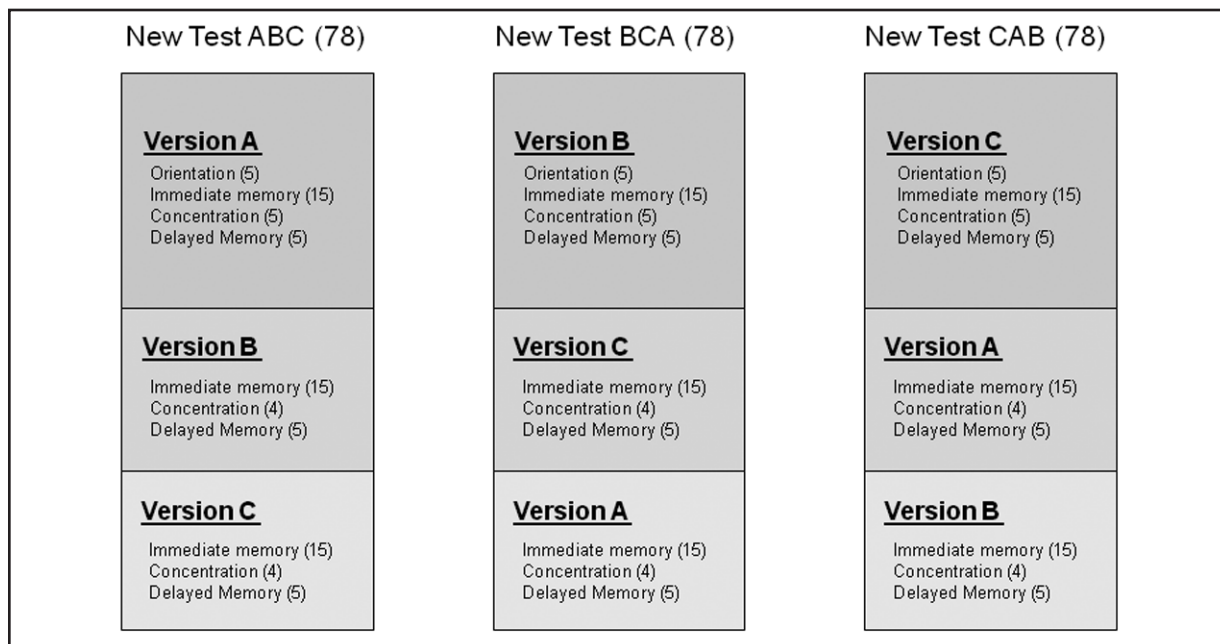


Figure. Standardized Assessment of Concussion random block counterbalancing procedure. The number of items is presented within the parentheses.

**Data Analyses**

Descriptive statistics for each version of the SAC were calculated. An item analysis was performed using 2 psychometric statistics—item difficulty and item discrimination, which were used to ultimately make an item determination. The acceptable range for these psychometric statistics were determined a priori based on the type of testing standard used and the recommendations set forth by the National Organization for Competency Assurance (NOCA) handbook,<sup>25</sup> which sets recommendations and standards for high stakes testing (eg, licensing examinations). It is logical that this type of cognitive screening instrument be treated as a high stakes test considering the consequences of MTBI.

*Item Difficulty.* Item difficulty (P) refers to the proportion of examinees who answered the item correctly. The acceptable range of item difficulty recommended by the NOCA depends on the type of standard used for interpretation. The acceptable range for a criterion-referenced standard is .33 to .92, whereas a wider range (.1 to .92) is recommended for an individual-centered method or norm-reference standard. Because the individual-centered testing standard is recommended,<sup>16</sup> an acceptable P range was set between .1 and .92 to allow for wide range of item difficulties.<sup>25</sup> An item with a P < .1 was considered too difficult. This means that <10% of the examinees answered the item correctly. Conversely,

an item with a P > .92 was considered too easy, which indicates almost all of the examinees responded correctly. Items that are too easy or difficult provide little useful information about the participants’ baseline cognitive ability level.

*Item Discrimination.* Item discrimination is the item’s ability to distinguish between different levels of examinees. An item that discriminates well will have a higher percentage of top performers answer the item correctly. A point-biserial correlation represents the relationship between a dichotomous variable (ie, correct or incorrect) and a continuous variable (ie, test score). A point-biserial correlation  $\geq .1$  is considered an appropriate level of item discrimination.<sup>25</sup> A correlation of < .1 indicates that the item lacks appropriate discrimination. Because the individual-centered method is required and a wide range of item difficulties is needed, the acceptable item discrimination threshold is affected. Point-biserial correlation suffers from attenuation as the split between top and bottom performers deviated further away from 50%.<sup>26</sup> To ensure items would be included that measure the top and bottom performs (ie, extremes) and assessed fairly, the point-biserial threshold was set consistent with NOCA recommendations at 0.1.

The individual-centered standard used in the SAC requires a comparison between postinjury score and the individual’s baseline score.<sup>16</sup> This intra-individual com-

TABLE 1

Item Classification Based on Item Difficulty (Item Number)<sup>a</sup>

SAC SECTION	TEST VERSION	ACCEPTABLE ITEM DIFFICULTY (.1 ≥ P ≤ .92)	UNACCEPTABLE ITEM DIFFICULTY (P < .1 OR P > .92)
Orientation	A		1, 2, 3, 4, 5
	B		1, 2, 3, 4, 5
	C		1, 2, 3, 4, 5
Immediate memory	A	10, 15	6, 7, 8, 9, 11, 12, 13, 14, 16, 17, 18, 19, 20
	B	8, 9	6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20
	C		6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20
Concentration	A	22, 23, 24, 25	21
	B	22, 23, 24, 25	21
	C	22, 23, 24, 25	21
Delayed memory	A	26, 27, 28, 29, 30	
	B	26, 27, 28, 29, 30	
	C	26, 27, 28, 29, 30	

Abbreviation: SAC, Standardized Assessment of Concussion.

<sup>a</sup>The item number used indicates the placement of the items on the SAC. For example, item number 6 would be the 6th item on the particular version of the SAC. It also would be the first immediate memory item.

parison requires the test to accurately measure a wide range of abilities; otherwise the usefulness of comparison (postinjury to baseline) is unwarranted. A test with a wide range of item difficulties is necessary to ensure accurate baseline test scores, which can be used to accurately measure the amount of cognitive dysfunction.

**Item Determination.** Each item was classified as either acceptable or unacceptable. For an item to be considered acceptable, it must have both an appropriate item difficulty and acceptable item discrimination. If one or both of these criteria were not satisfied, the item was considered unacceptable.

## RESULTS

The mean score, standard deviation, and Guttman-Cronbach alpha for each version was calculated. Version A had a mean score of  $26.7 \pm 2.5$  and  $\alpha = 0.65$ ; version B was  $26.4 \pm 2.2$  with  $\alpha = 0.52$ ; and version C was  $27.2 \pm 1.8$  with  $\alpha = 0.44$ .

### Item Difficulty

The item analysis indicated that no items were too difficult. However, a substantial number of items were too

easy (version A = 19, 63%; version B = 19, 63%; and version C = 21, 70%). The SAC items based on item difficulty classifications are presented in Table 1. The mean item difficulty and range for each section is presented for all 3 versions in Table 2. As indicated in Table 2, all of the immediate memory and the orientation sections have a mean difficulty greater than the 0.92 threshold, indicating that the sections are too easy.

### Item Discrimination

The results indicated that each version varied in the number of items that had appropriate discrimination; Version A had 21 items (70%), version B had 23 items (77%), and version C had 14 items (47%). The immediate memory section was the most varied and problematic. Item discrimination results are presented in Table 3 based on point-biserial correlations, listed by SAC version and section.

### Item Determination

Item determination indicated, based on the results, that the 3 SAC versions had very few items that obtained acceptable difficulty and discrimination. Version A had

TABLE 2

**Descriptive Statistics of Item Difficulties (P)**

SAC SECTION	TEST VERSION <sup>a</sup>	NO. OF ITEMS	MEAN SECTION DIFFICULTY	
			(MEAN ± SE)	RANGE
Orientation	A	5	0.97±0.01 <sup>b</sup>	0.93-1.00
	B	5	0.97±0.01 <sup>b</sup>	0.93-1.00
	C	5	0.97±0.01 <sup>b</sup>	0.93-1.00
Immediate memory	A	15	0.95±0.02 <sup>b</sup>	0.69-1.00
	B	15	0.96±0.01 <sup>b</sup>	0.85-0.99
	C	15	0.99±0.00 <sup>b</sup>	0.97-1.00
Concentration	A	5	0.73±0.10	0.33-0.95
	B	5	0.73±0.10	0.33-0.99
	C	5	0.75±0.10	0.36-0.97
Delayed memory	A	5	0.79±0.02	0.75-0.90
	B	5	0.69±0.03	0.58-0.79
	C	5	0.75±0.02	0.71-0.84

Abbreviation: SAC, Standardized Assessment of Concussion.

<sup>a</sup> Version A, mean difficulty = 0.89; version B, mean difficulty = 0.88; version C, mean difficulty = 0.91.

<sup>b</sup> Indicates that the section difficulty was too easy.

10 (33%), version B had 9 (30%), and version C contained only 8 (27%) items that met both criteria. Table 4 depicts the item determinations for all items on the 3 versions. All of the common orientation items were unacceptable. Acceptable items in the immediate memory section varied among the versions: version A had 2 of 15 (13%), version B had 1 of 15 (7%), and version C had 0 of 15 (0%). The concentration section in all versions contained 3 of 5 (60%) acceptable items. All 3 versions had 5 of 5 (100%) acceptable items for the delayed memory section.

**DISCUSSION**

The results of this study question the validity of the SAC as an adequate baseline measure of orientation, immediate memory, and concentration. Many of the items (63% to 70%) were too easy, provided little information, and could be eliminated during baseline testing. Although these items are not useful during baseline testing, they may be appropriate for postinjury testing. This abundance of easy items supports previous reports of ceiling effects, specifically with baseline testing using the SAC.<sup>23</sup> Additional items lacked appropriate discrimination and should be deleted from baseline testing. Thus, only 8 to 10 items remain depending on the version. Perhaps more significantly, all 5 of the orientation items and 87% to 100% of the immediate memory items were deemed unacceptable. These results seriously threaten

the construct validity of the SAC. On the basis of the item analysis, the SAC is essentially a delayed memory test.

The lack of difficult items is also problematic for a different reason. The SAC does not have the ability to accurately measure the true ability of many individuals, as most questions are relatively easy. Thus an individual receiving a maximum score of 30 may actually have a true ability of 36. After experiencing a concussion, the SAC would not be able to identify a 6-point decrease in cognitive ability, whereas a 3-point decrease is supposed to indicate an MTBI.<sup>19,20</sup> This failure to assess the extent of cognitive impairment in many individuals highlights serious validity and methodological issues of the SAC.

The evaluation of each item using an item analysis is critical in establishing test validity prior to clinical use.<sup>25,27</sup> Results using these analyses in the current study indicate that more difficult items must be developed to accurately measure the healthy baseline cognitive function so that clinically relevant comparisons can be made with the individual-centered method. Although the orientation section is not expected to discriminate between examinees with high ability, other cognitive functions measured by the SAC, such as immediate memory, concentration, and delayed memory, are expected to have a wide distribution of abilities. It is these sections that must have appropriate items to assess both baseline and injured cognitive status. If a wide range of abilities like immedi-

TABLE 3

### Item Classification Based on Item Discrimination (Item Number)

SAC SECTION	TEST VERSION	ACCEPTABLE ITEM DISCRIMINATION	UNACCEPTABLE ITEM DISCRIMINATION
		(POINT-BISERIAL CORRELATION $\geq$ .1)	(POINT-BISERIAL CORRELATION $<$ .1)
Orientation	A	2, 3	1, 4, 5
	B	2, 3	1, 4, 5
	C	2, 3	1, 4, 5
Immediate memory	A	7, 8, 9, 10, 12, 13, 14, 15, 17, 20	6, 11, 16, 18, 19
	B	6, 8, 10, 11, 12, 13, 14, 15, 16, 17, 19, 20	7, 9, 18
	C	7, 10, 19	6, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 20
Concentration	A	21, 22, 23, 24	25
	B	21, 22, 23, 24	25
	C	21, 22, 23, 24	25
Delayed memory	A	26, 27, 28, 29, 30	
	B	26, 27, 28, 29, 30	
	C	26, 27, 28, 29, 30	

Abbreviation: SAC, Standardized Assessment of Concussion.

ate memory is not expected, then the need for a baseline measure is unnecessary.

In addition to the classification of items as acceptable or unacceptable, another interesting finding was the examination of the mean item difficulty and range for each section across the 3 versions. For example, the immediate memory section difficulties vary across versions. The most difficult immediate memory section of the SAC was version A (mean = 0.95, range = 0.69-1.0). This was different from the easiest immediate memory section on version C (mean = 0.99, range = 0.97-1.0). In fact, none of the 15 immediate memory items of version C were acceptable for baseline scores. Similar variations are seen in the delayed memory sections in which there was a 10% difference in difficulty level (version A = 0.79; version B = 0.69). The delayed memory section of version B was the most difficult. These differences in section difficulty between the versions calls into question the version (test) equivalence. The test version may not be of equal difficulty.

Differences in item difficulty between the versions also demonstrate the inappropriate summing of the 4 sections to produce a total score without controlling or adjusting for the item difficulty.<sup>28,29</sup> The range of item difficulties varies between domains and versions. For example, in the immediate memory section, the most difficult item on version A had a P of 0.62, whereas the most difficult item of version C had a P of 0.96. This represents a huge

difference in item difficulty. This difference can lead to an incorrect assessment.

For instance, if an individual was administered version C (easiest version) at baseline, he or she most likely would have scored very well. Following a suspected MTBI, version A (most difficult version) was administered resulting in a decrease in test performance score. This decrease in score may not necessarily be a result of cognitive impairment but instead, the difference in version difficulty. Another scenario could occur if the tests were given in the reverse order. It is common clinical practice to reduce testing effects by administering an alternative version of the SAC at the postinjury assessment.<sup>11</sup>

Because of the difference in difficulty between the versions (tests), it would also be inappropriate to use a summed section score for potential comparison without making adjustments. The difference in performance scores could be due to the difference in item difficulty and not changes in cognitive function. Controlling for item difficulty between domains and versions is needed to make valid comparisons. If the difficulty of the items is not carefully controlled, the reliability of the measure may suffer.<sup>17</sup> To clinically adjust for these differences, more research applying advanced psychometric procedures is needed and should be the focus of future research.

Another interesting interpretation of the results is that there could be 2 types of items: items that will work with an individual-centered standard and items that work in

TABLE 4

**Item Determination Based on Both Acceptable Psychometric Properties:  
Item Difficulty and Discrimination (Item Number)**

SAC SECTION	TEST VERSION	ACCEPTABLE ITEM DETERMINATION	UNACCEPTABLE ITEM DETERMINATION
Orientation	A		1, 2, 3, 4, 5
	B		1, 2, 3, 4, 5
	C		1, 2, 3, 4, 5
Immediate memory	A	10, 15,	6, 7, 8, 9, 11, 12, 13, 14, 16, 17, 18, 19, 20
	B	8	6, 7, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20
	C		6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20
Concentration	A	22, 23, 24	21, 25
	B	22, 23, 24	21, 25
	C	22, 23, 24	21, 25
Delayed memory	A	26, 27, 28, 29, 30	
	B	26, 27, 28, 29, 30	
	C	26, 27, 28, 29, 30	

Abbreviation: SAC, Standardized Assessment of Concussion.

a criterion-referenced standard. Table 4 depicts the item classification based on the current baseline and postinjury model. Most of the concentration items and delayed memory have acceptable item statistics, indicating they are appropriate for the individual-centered standard. These items allow for discrimination and variability in individuals' scores. The orientation and immediate memory items do not function appropriately and therefore do not provide useful information at the baseline testing session for follow-up comparisons. However, these items may be useful in identifying MTBI when used as a criterion-referenced standard. This type of standard does not require a baseline score and uses a threshold or *cut-score* to determine group membership. These items may be appropriate when used in this manner. For example, when an individual cannot recall person, place, or time and has trouble immediately reciting 5 words, a clinician can use this straightforward information to conclude that an MTBI or worse has occurred. With different types of items and testing standards possibly contained in the SAC, it is not surprising that the instrument has some inherent measurement issues.

The results from this study demonstrate the need for item analyses to be performed on screening tools and tests. This is especially critical for health-related tests because of the potential implications for clinical diagnosis and management of many health conditions. It also questions the validity of the SAC to be used as a baseline measure for sport-related concussion. These results do not address

the SAC being used with a criterion-referenced standard, but only with an individual-centered standard where a baseline score is needed. However, the results should be interpreted with caution, as no single test should be used as the sole indicator of MTBI assessment and management. Although the sample of participants was reflective of a typical group to which the SAC is administered (ie, college-aged), the validity of the SAC through item analysis in other populations such as high school age (another target population for the SAC) has not been determined. It is possible the item statistics would vary based on age and education level.<sup>30</sup>

The traditional item analysis approach used in this study suggests that additional test development procedures are warranted. More advanced approaches such as Item Response Theory<sup>31</sup> or Rasch modeling<sup>32,33</sup> may be of use. These techniques are often used to overcome psychometric problems commonly associated with traditional methods (see Zhu, Timm, & Ainsworth<sup>29</sup> for an in-depth review) and are currently used in standardized test development and psychological measurement. The suggested models place items and examinees on the same common metric. This allows for the examination of the spread of the items, the location, redundancy, and gapping, which provides better insight into the measurement capabilities of the test. It also is the foundation for computer adaptive testing.<sup>34,35</sup>

Further research should increase the number of acceptable SAC items. Increasing the number of acceptable

items can be accomplished by modifying existing items or by pilot testing additional items. Once developed, more advanced analysis techniques should be used to ensure an adequate number of psychometrically sound items for multiple versions of the SAC. The need for this rigorous development process in medical professions is long overdue. These steps are necessary to provide physicians and athletic trainers with accurate information when making decisions concerning MTBI identification, management, and eventually the return-to-play decision. ■

## REFERENCES

- Centers for Disease Control and Prevention, National Center for Injury Prevention and Control. *Report to Congress on Mild Traumatic Brain Injury in the United States: Steps to Prevent a Serious Public Health Problem*. Atlanta, GA: U.S. Department of Health & Human Services, Centers for Disease Control and Prevention; 2003.
- Centers for Disease Control and Prevention. *National Center for Injury Prevention and Control, 2007*. Centers for Disease Control and Prevention Web site. <http://www.cdc.gov/ncipc/tbi/TBI.htm>. Accessed June 6, 2007.
- Bazarian JJ, McClung J, Shah MN, Cheng YT, Flesher W, Kraus J. Mild traumatic brain injury in the United States, 1998-2000. *Brain Inj*. 2005;19:85-91.
- Notebaert AJ, Guskiewicz KM. Current trends in athletic training practice for concussion assessment and management. *J Athl Train*. 2005;40:320-325.
- Langlois JA, Rutland-Brown W, Wald MW. The epidemiology and impact of traumatic brain injury: A brief overview. *J Head Trauma Rehabil*. 2006;21:375-378.
- Collins M, Grindel S, Lovell M, et al. Relationship between concussion and neuropsychological performance in college football players. *JAMA*. 1999;282:964-970.
- Gronwall D, Wrightson P. Cumulative effect of concussion. *Lancet*. 1975;2(7943):995-997.
- Levin HS, Mattis S, Ruff RM, Eisenberg HM, Marshall LF, Tabaddor K. Neurobehavioral outcome following minor head injury: A 3-center study. *J Neurosurg*. 1987;66:234-243.
- Matser E, Kessels A, Lezak M, Jordan B, Troost J. Neuropsychological impairment in amateur soccer players. *JAMA*. 1999;282:971-973.
- Folstein M, Folstein S, McHugh P. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12:189-198.
- McCrea M, Kelly JP, Randolph C. *The Standardized Assessment of Concussion (SAC): Manual for Administration*. Waukesha, WI: Comprehensive Neuropsychological Services; 1997.
- McCrea M, Kelly JP, Kluge J, Ackley B, Randolph C. Standardized assessment of concussion in football players. *Neurology*. 1997;48:586-588.
- McCrea M, Kelly JP, Randolph C, et al. Standardized assessment of concussion: Onsite mental status evaluation of the athlete. *J Head Trauma Rehabil*. 1998;13:27-35.
- Kuslansky G, Katz M, Verghese J, et al. Detecting dementia with the Hopkins Verbal Learning Test and the Mini-Mental State Examination. *Arch Clin Neuropsychol*. 2004;19:89-104.
- Barth JT, Alves WM, Ryan TV, et al. Mild head injury in sports: neuropsychological sequelae and recovery of function. In: Levin HS, Eisenberg HM, Benton AL, eds. *Mild Head Injury*. New York, NY: Oxford University Press; 1989:257-275.
- Guskiewicz KM, Bruce SL, Cantu RC, et al. National Athletic Trainers' Association position statement: Management of sport-related concussion. *J Athl Train*. 2004;39:280-297.
- Ragan BG, Kang M. Measurement issues in concussion testing. *Athletic Therapy Today*. 2007;12:2-6.
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association; 1999:37-48.
- Barr WB, McCrea M. Sensitivity and specificity of standardized neurocognitive testing immediately following sports concussion. *J Int Neuropsychol Soc*. 2001;7:693-702.
- McCrea M. Standardized mental status assessment of sports concussion. *Clin J Sport Med*. 2001;11:176-181.
- McCrea M. Standardized mental status testing on the sideline after sports-related concussion. *J Athl Train*. 2001;36:274-279.
- McLeod Valovich TC, Barr WB, McCrea M, Guskiewicz KM. Psychometric and measurement properties of concussion assessment tools in youth sports. *J Athl Train*. 2006;41:399-408.
- Echemendia RJ, Julian LJ. Mild traumatic brain injury in sports: Neuropsychology's contribution to a developing field. *Neuropsychol Rev*. 2001;11:69-88.
- Randolf C, McCrea M, Barr WB. Is neuropsychological testing useful in the management of sport-related concussion? *J Athl Train*. 2005;40:139-154.
- Dungan L. Examination development. In: Browning AH, Bugbee AC, Mullins MA, eds. *Certification: A NOCA Handbook*. Washington, DC: National Organization for Competency Assurance; 1996:1-40.
- Cohen J. The cost of dichotomization. *Applied Psychological Measurement*. 1983;7:249-253.
- McDonald RP. *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum; 1999:231-247.
- Kang SJ, Kang M. Rasch calibration of perceived weights of different sports games. *Meas Phys Educ Exerc Sci*. 2006;10:51-66.
- Zhu W, Timm G, Ainsworth B. Rasch calibration and optimal categorization of an instrument measuring women's exercise perseverance and barriers. *Res Q Exerc Sport*. 2001;72:104-116.
- Holsinger T, Deveau J, Boustani M, Williams JW. Does this patient have dementia? *JAMA*. 2007;297:2391-2404.
- Lord FM. *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Erlbaum; 1980.
- Rasch G. *Probabilistic Models for Some Intelligence and Achievement Tests*. Copenhagen, Denmark: Danish Institute for Educational Research; 1960.
- Rasch G. *Probabilistic Models for Some Intelligence and Achievement Tests*. Expanded ed. Chicago, IL: University of Chicago Press; 1980.
- Jette AM. Assessing disability in studies on physical activity. *Am J Prev Med*. 2003;25:122-128.
- Ware JE, Bjorner JB, Kosinski M. Practical implications of item response theory and computerized adaptive testing. *Med Care*. 2000;38:73-82.