

# Development of the Mental, Emotional, and Bodily Toughness Inventory in Collegiate Athletes and Nonathletes

Mick G. Mack, PhD; Brian G. Ragan, PhD, ATC

University of Northern Iowa, Cedar Falls, IA

**Context:** The assessment of an individual's mental toughness would assist clinicians in enhancing an individual's performance, improving compliance with the rehabilitation program, and improving the individual treatment program. However, no sound measure of mental toughness exists.

**Objective:** To develop a new measure of mental toughness, the Mental, Emotional, and Bodily Toughness Inventory (MeB-Tough).

**Design:** Participants were invited to complete a 45-item questionnaire.

**Setting:** University research laboratory.

**Patients or Other Participants:** A total of 261 undergraduate students were recruited to complete the questionnaire.

**Main Outcome Measure(s):** The Rasch-calibrated item difficulties, fit statistics, and persons' mental toughness ability estimates were examined for model-data fit of the MeBTough.

**Results:** Forty-three of the 45 items had good model-data fit with acceptable fit statistics. Results indicated that the distribution of items was fittingly targeted to the people and the collapsed rating scale functioned well. The item separation index (6.31) and separation reliability statistic (.98) provided evidence that the items had good variability with a high degree of confidence in replicating placement of the items from another sample.

**Conclusions:** Results provided support for using the new measure of mental, emotional, and bodily toughness.

**Key Words:** Rasch model, psychological skills

## Key Points

- The Mental, Emotional, and Bodily Toughness Inventory demonstrated good model-data fit, indicating that it had good psychometric properties.
- The difficulty of the items had good variability along the measurement scale, meaning respondents' differences could be identified well.

In 1986, sport psychologist James Loehr<sup>1</sup> popularized the term *mental toughness* to describe the ability to consistently maintain an ideal performance state during the heat of competition. Since then, mental toughness has been described as standing tall in the face of adversity and as the ability to rebound from repeated setbacks and failures.<sup>2</sup> Jones et al<sup>3</sup> defined *mental toughness* as having a psychological edge that enables an athlete to cope consistently with the pressures and demands of the sport during competition and in his or her lifestyle and training. More recently, mental toughness has been conceived as an individual's propensity to handle the demands of environmental stressors.<sup>4</sup> Thus, although differences of opinion still exist, there appears to be some agreement that mental toughness is displayed by an individual's ability to cope with the stress and anxiety associated with competitive situations,<sup>3</sup> such as rehabilitation and returning to play after injury.

Because of its perceived importance as a critical psychological skill, mental toughness is recognized as a key ingredient in achieving success.<sup>1,2,5,6</sup> Within the sports medicine context, however, understanding of the effects of being mentally tough is limited and may even hinder rehabilitation behaviors and recovery.<sup>7</sup> Thus, the assessment of an individual's mental toughness would assist clinicians in improving performance and compliance and in

enhancing individual treatment programs. Unfortunately, developing a measure of mental toughness has been problematic,<sup>8,9</sup> and "there currently exists no comprehensively sound measure of mental toughness."<sup>9(p106)</sup> Thus, our first step was to examine the literature to find a comprehensive explanation on which to base a new instrument. Loehr,<sup>6</sup> Jones et al,<sup>3</sup> and Bull et al<sup>10</sup> provided the 3 most methodical and complete explanations or models of mental toughness.

Based on his research, training, and practical experience, Loehr<sup>6</sup> provided a revised description of mental toughness and defined the construct as being able to perform consistently toward the upper range of one's ability, regardless of competitive circumstances. Achieving this level of performance requires an individual to be physically, mentally, and emotionally tough. Loehr's effort to provide concrete strategies for achieving this 3-dimensional concept enables toughness to be further delineated into 9 constructs. The physical dimension involves being well prepared and acting tough. The mental element of toughness encompasses the ability to create an optimal performance state, to access empowering emotions, and to cope. Finally, the emotional aspect of toughness has 4 basic markers: flexibility, responsiveness, strength, and resiliency.

Jones et al<sup>3</sup> used a different approach and attempted to identify the microcomponents of mental toughness by

interviewing 10 international athletes. Using a qualitative method, they identified 12 distinct attributes that were thought to be key dimensions of mental toughness. These attributes then were classified into the 6 general areas of self-belief, desire or motivation, performance-related focus, addressing pressure or anxiety, lifestyle-related focus, and addressing pain or hardship.

In their qualitative study examining the concept of mental toughness, Bull et al<sup>10</sup> interviewed 12 elite English cricketers. Transcript analyses revealed 20 global themes. These themes were arranged into 4 structural categories to represent the mental toughness pyramid and consisted of environmental influence, tough character, tough attitudes, and tough thinking.

When comparing and evaluating these articles, we saw ways in which they were different. For instance, Loehr's writing was based on practical experience, anecdotal stories, and previous "how-to" publications<sup>11,12</sup> suggesting that toughness training may positively influence athletic performance. In contrast, the other 2 articles were based on qualitative research interviews of 10 to 15 athletes each. In some instances, the categories or themes also were disparate. For example, only 1 research group<sup>10</sup> identified environmental influences as instrumental in the development of mental toughness through the athlete's formative years.

However, many of the themes clearly were very similar. For example, the ability to maintain an unshakable, powerful belief in one's ability and the ability to bounce back from setbacks were central themes in each. Similarly, the belief in being well prepared by pushing the boundaries in training and competition was identified by all 3 authors as a critical attribute of a mentally tough athlete.

Because of the many similarities, we performed a more in-depth comparison of the themes in each manuscript. All of the attributes described by Jones et al<sup>3</sup> appeared to fit within the constructs of Loehr,<sup>6</sup> whereas Loehr's inclusion of the ability to create an optimal performance state appeared to be unique to his approach. A subsequent comparison between the constructs of Loehr<sup>6</sup> and Jones et al<sup>3</sup> and the constructs of Bull et al<sup>10</sup> revealed that most of the themes contained within the categories of tough character, tough attitudes, and tough thinking were similar. However, each construct also had themes not found in the other. In particular, Loehr<sup>6</sup> and Jones et al<sup>3</sup> included the specific mental element of being able to cope with competitive anxiety and the physical dimension of acting tough and not being affected by others' performances, whereas Bull et al<sup>10</sup> primarily incorporated a number of exclusive themes under the category of environmental influences. Because practitioners rarely have input into the environmental influences during the childhood years of their patients and because this area is not included in most, if any, definitions of *mental toughness*, we concluded that the inclusion of these themes was not critical to the development of a new mental toughness assessment tool. Conversely, the themes of coping with anxiety and acting tough were viewed as essential components for the mentally tough athlete. Thus, the purpose of our study was to develop a new measure of mental toughness based on Loehr's<sup>6</sup> more comprehensive and practitioner-friendly proposal, which defined *mental toughness* as being able to perform consistently toward the

upper range of one's ability regardless of competitive circumstances.

## METHODS

### Participants

Participants in our study consisted of 261 undergraduate students (165 men, 96 women) at a midwestern university. Demographic data revealed that the sample consisted of 5% (n = 13) first-year, 23% (n = 60) second-year, 43% (n = 112) third-year, and 29% (n = 76) fourth-year or fifth-year students. We did not record the ages of the participants. Of the total sample, 29% (n = 75) indicated that they currently were or had been a member of one of the university's collegiate athletic teams. The remaining 71% (n = 186) were considered nonathletes.

### Measures

**Mental, Emotional, and Bodily Toughness Inventory.** As noted, Loehr<sup>6</sup> defined toughness as a 3-dimensional concept and delineated the mental, emotional, and physical elements of this concept into 9 constructs. Using his explanation of these constructs, we developed a list of 93 potential items (9 to 12 items for each construct). Based on previous research<sup>13,14</sup> and a desire to keep the questionnaire relatively brief, we limited each construct to a maximum of 5 initial items. Thus, the 5 most content-valid items for each construct were selected from this pool, resulting in a total of 45 items. A small sample (n = 6) of athletes pilot tested and evaluated this set of items for item readability and comprehension. Based on their feedback, we modified items judged to be poorly constructed or worded.

Participants were asked to indicate on a 7-point scale (1 = *almost never*, 4 = *sometimes*, 7 = *almost always*) how often they experienced each item. To help reduce the halo effect and to break down potential response sets,<sup>15</sup> 11 of the items were designed to be reverse scored.

**Mental Toughness Scale.** To assess an overall 1-dimensional measure of mental toughness, we asked participants to rate their toughness on a scale of 1 to 20 (1 = *low*, 20 = *high*). Mental toughness was explained to participants as the ability to consistently perform toward the upper range of their talents and skills, regardless of competitive circumstances.

### Procedures

Following approval of the Human Participants Review Committee, undergraduate students were recruited from coaching-education classes. The same individual taught all classes in both fall and spring semesters. Participation in this study was one of several options that students could select for extra credit. Students could determine the type and extent of their participation in extra-credit activities and were informed that they could cease participation at any time without fear of reprisal. Students willing to participate were given surveys and were instructed to complete the items within 2 to 4 days and to return the surveys to their instructor.

## Data Analyses

We selected and used the Rasch analysis model, because it enables comparisons across studies and because traditional approaches for test development based on classic test theory methods have several known psychometric limitations.<sup>16</sup> Some of these limitations are that the calibrations are sample dependent and item dependent and that the items and participants are placed on different scales.<sup>17</sup> Another common psychometric problem is the inability to critically examine the selected response categories.<sup>16</sup> The Rasch model<sup>18,19</sup> is similar to item response theory,<sup>20</sup> which is an advanced measurement practice currently used in standardized test development and psychological measurement.<sup>21</sup> With Rasch analyses, the psychometric properties are evaluated with different methods, making the classic techniques, such as factor analysis and reliability, not applicable.

From a test-construction perspective, the Rasch model enables better evaluation of the items and more precise measurement of the test. After the calibration, both items and people are placed on the same common metric. This enables the items to be examined for spread, redundancy, and gapping across a wide ability range and for their location relative to other items and to the ability estimates of the people. Rasch analyses also enable the development of additional items to target areas on the measurement scale that need improvement. These newly developed items can be placed on the same common metric (scale) as the initial items and people. Thus, additional questions can be created and incorporated into the existing measure. Item parameters also can be used to control and improve the quality of the test. These parameters have a very important feature: invariance. This particular strength of the Rasch model means that the estimates of parameters are stable across samples, so comparisons can be made from different groups tested at different times.

Another major strength of the Rasch model in test construction is the ability to determine empirically whether the categorization selected by the instrument developer is the most appropriate one.<sup>16</sup> Through Rasch analysis and category collapsing, identification of the optimal categorization is attainable. This process ensures that the categories are ordered and that each is the most probable response somewhere on the logit scale.

Although the Rasch model has many advantages, it is not without drawbacks. Two drawbacks are that the data have to fit the model selected (ie, a unidimensional model must have unidimensional data) and that the model requires larger sample sizes than classic testing procedures do.<sup>16</sup>

We used the Facets program (Winsteps version 3.4; Chicago, IL) to analyze the data. Both persons' abilities and items' difficulties were calculated. The Rasch Rating Scale Model,<sup>22,23</sup> which is an extension of the basic Rasch model<sup>18,19</sup> and its mathematical properties, was used for the analysis. This model defines the probability of examinee  $n$  responding in category  $x$  to item  $i$ :

$$\log[P_{nik}/P_{ni(k-1)}] = B_n - D_{gi} - F_{gk},$$

where  $P_{nik}$  is the probability of observing category  $k$  for person  $n$  encountering item  $i$ ,  $P_{ni(k-1)}$  is the probability of observing category  $k-1$ ,  $F_{gk}$  is the difficulty of being

observed in category  $k$  relative to category  $k-1$ , for an item in group  $g$ ,  $B_n$  is the ability of person  $n$ , and  $D_{gi}$  is the difficulty of item  $i$ . Higher mental toughness and more difficult items correspond to higher logit scores.

In the Rasch construction process, both item difficulties and the item-person map are very powerful quantitative and visual advantages. The Wright item-person map displays the location of the people and items on the same metric. It also depicts item spread and a gap between items where no discrimination exists. As such, the rating scale was evaluated for proper functioning using 3 criteria: (1) Was the mean square residual appropriate for each category? (2) Did the average measure (ie, a mean of logit measures in category) per category increase as the category score increased? (3) Were category thresholds (ie, boundaries between categories, which also were stated as the step-difficulty values) ordered?<sup>24,25</sup>

Infit and outfit statistics were used to evaluate the model-data fit for each item in the Rasch analysis. Infit statistics represent the information-weighted mean square residuals between observed and expected responses; outfit statistics are similar to infit statistics but are more sensitive to the outliers. Infit and outfit statistics values that were close to 1 were considered satisfactory model-data fit, whereas values that were more than 1.5 or less than 0.5 were considered a misfit of the model.<sup>26</sup> Values that were more than 1.5 indicated inconsistent performance and values that were less than 0.5 showed too little variation.

In addition to the model-data fit procedure, separation and separation reliability indices were evaluated. The separation statistics indicate how well items are spread along the measurement scale. Separation reliability indicates how confident one can be that the items would have the same respective order from another sample of participants. A higher separation index indicates a greater degree of confidence. In addition, a separation reliability value close to 1.00 indicates that there is good discrimination for a facet along the measurement scale with a high degree of confidence.<sup>27</sup> To establish evidence of validity for the mental toughness scores, the relationship between the ability estimate and the participants' overall rating of their toughness on a 20-point Likert scale was calculated with a Pearson product moment correlation. To compare the toughness estimates between athletes and nonathletes, we used a 2-tailed independent  $t$  test. The  $\alpha$  level was set at .05. To determine whether we had achieved optimal categorization, we conducted a post hoc cross-validation of the choices in another independent sample.

## RESULTS

### Optimal Categorization

The optimal categorization process was performed to achieve ordered categories (choices), where each option category is the most probable response somewhere on the logit scale. The probability of each response category across the scale is displayed in Figure 1. Using this process, the number of choices was collapsed from 7 to 4 options. Seven categories were collapsed from 1234567 to 1112345 and finally to 11234. Thus, the categories and characteristic curves were optimized through collapsing categories. This

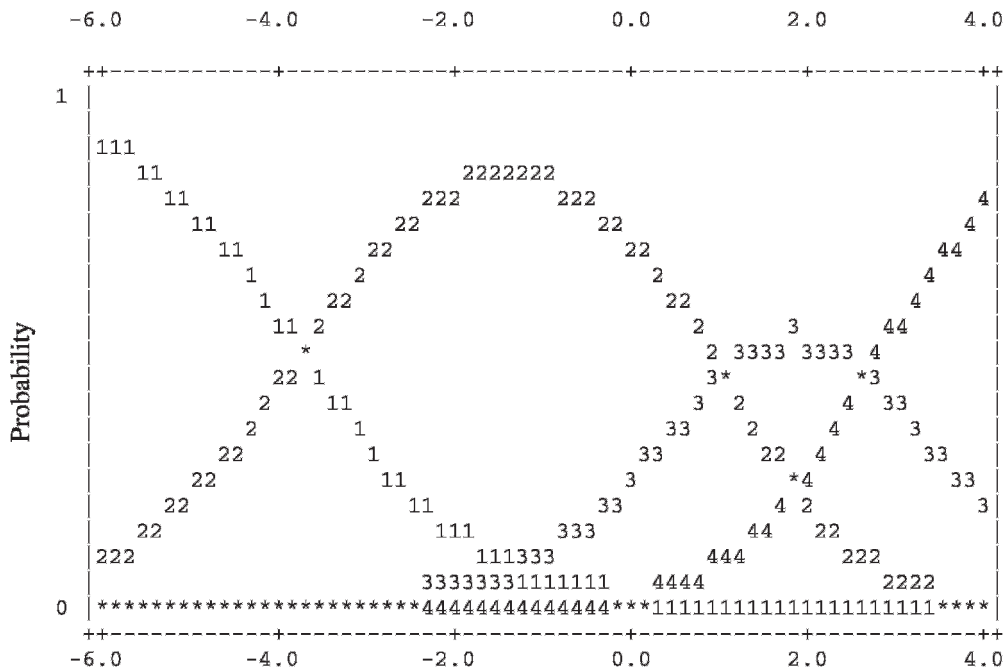


Figure 1. Probability curves for the 4-category scale.

response category collapsing was done because the respondents were not using the lower-end responses.

The optimal categorization of the 4 choices was cross-validated in another independent sample. The cross-validation sample demonstrated that optimal categorization was achieved. The step calibrations were ordered, and each response was the most probable somewhere on the logit scale. Based on the success of the cross-validation, the calibrated data were analyzed.

**Final Calibration.** A Wright item-person map (Figure 2) displays the location and distribution of both the items and the people's abilities on the same common metric. The scale in logit is shown on the left side of the map; people and items are shown on the right side. The map reveals that the distribution of the items was well targeted to the people. Both items and people's abilities were dispersed across a broad spectrum of the logit measurement scale. A summary of rating-scale steps for the 4 weighted categories is reported in Table 1. Overall, the rating scale functioned well. The average measure per category increased as the category score increased. Finally, step-difficulty values (ie, the category thresholds) were ordered as expected with optimization.

### Item Difficulty

The items on the instrument, including calibrated logit scores with SEs, and the infit and outfit statistics are reported in Table 2. The item separation was 6.31, and the item separation reliability was .98. Rasch analysis results indicated that the model fit the data well. Forty-three items had fit statistics within the acceptable range (mean infit mean square =  $1.0 \pm 0.5$ , mean outfit mean square =  $1.0 \pm 0.5$ ). Two items, *I am emotionally unresponsive in competition* and *In competition, I show whatever emotions I am feeling at the moment*, had unacceptable fit statistics of 1.9 and 2.6, respectively, and were eliminated from the final calibration depicted in Table 2.

Overall, the remaining items ranged in difficulty from 1.55 logits (most difficult) to  $-2.22$  logits (least difficult) with a mean of  $0.00 \pm 0.70$ . The item, *I sometimes allow my negative emotions and feelings to lead me into negative thinking*, had the highest difficulty measure (logit = 1.55), followed by *Surprising emotional shifts don't bother me* (logit = 1.47) and *Emotional setbacks are difficult for me to overcome* (logit = 0.70). The item that was the least difficult was *I love the heat of battle* (logit =  $-2.22$ ). The item separation index and separation reliability statistic were 6.31 and .98, respectively. This was evidence that the items had good variability along the measurement scale with a high degree of confidence in replicating placement of the items within the measurement error from another sample of participants.

### Participants' Mental, Emotional, and Bodily Toughness Inventory Abilities

Descriptive information concerning the ability estimates of the participants' mental, emotional, and bodily toughness was examined. The overall mean of the participants' toughness was  $0.53 \pm 1.06$  logits. Conditional SEs of measure (CSEMs) for the ability estimates ranged between  $\pm 0.24$  logits for ability estimates that were more than 2.0 logits and  $\pm 0.34$  for ability estimates that were close to 0 logits. The mean CSEM for the ability estimates was 0.28 logits. The mental MeBTough estimates were further analyzed between collegiate athletes and nonathletes. The nonathletes ( $n = 186$ ) had a mean toughness of  $0.34 \pm 1.05$  logits. The athletes' ( $n = 75$ ) mean toughness estimate of  $0.99 \pm 0.93$  logits was significantly higher than that of nonathletes ( $t_{259} = -4.6, P < .001$ ). Ability estimates also were correlated with the participants' overall opinions of their mental toughness. The correlation between their perceived mental toughness (mean =  $15.9 \pm 2.1$ ) and the Rasch-calibrated ability estimates of MeBTough revealed a moderate positive relationship ( $r = .60, P < .001$ ).



**Table 2. Properties for the 43 Mental, Emotional, and Bodily Toughness Inventory Items by Difficulty**

Summary of Item	Calibration Logit	SE Logit	Infit Mean Square Residuals	Outfit Mean Square Residuals
Allow negative emotions/feelings	1.55	0.13	1.2	1.2
Emotional shifts don't bother me	1.47	0.13	1.2	1.2
Negative emotions hard to change	1.04	0.13	1.0	1.0
Competitive circumstances affect me	0.83	0.12	1.0	1.0
Emotional setbacks are difficult	0.70	0.12	1.0	1.0
Change from negative to positive	0.62	0.12	0.9	0.8
Difficult to maintain balance	0.59	0.12	1.1	1.1
Hard to trigger right emotion	0.57	0.12	1.2	1.1
Maintain my emotional balance	0.57	0.12	0.9	0.9
Emotional demands exceed my ability	0.49	0.12	1.0	1.0
Absorb unexpected emotional changes	0.49	0.12	0.7	0.6
Ability to balance stress/recovery	0.43	0.11	0.8	0.8
Physical demands exceed my capacity	0.38	0.11	1.2	1.2
Remain calm and collected	0.32	0.11	1.0	1.0
Understand my ideal state	0.32	0.11	1.0	1.0
Take a punch emotionally	0.19	0.11	1.2	1.2
Ability to bounce back quickly	0.17	0.11	0.9	0.9
Keep fighting good fight	0.09	0.11	0.9	0.8
Can handle mistakes/failures	0.09	0.11	0.9	0.8
Trigger optimal performance state	0.06	0.11	0.8	0.7
Sustain powerful fighting spirit	0.05	0.11	1.2	1.1
Body language helps me	0.04	0.11	1.1	1.1
Emotionally responsive and engaged	-0.04	0.11	0.8	0.8
Ability to cope	-0.07	0.11	0.8	0.8
Become withdrawn emotionally	-0.12	0.11	1.4	1.4
Can control feelings/emotions	-0.15	0.11	0.7	0.6
Optimal physiologic arousal state	-0.21	0.10	0.9	0.9
I display confidence/energy	-0.21	0.10	1.0	0.9
Under pressure think positively	-0.26	0.10	0.9	0.9
Emotional strength under pressure	-0.31	0.10	0.7	0.7
Sense of challenge/determination	-0.32	0.10	0.9	0.9
Can jump back into fight	-0.33	0.10	0.8	0.8
Tolerance for physical stress	-0.35	0.10	1.0	1.0
Emotional strength to fight back	-0.36	0.10	0.8	0.7
Can sustain physical stress	-0.36	0.10	1.0	1.0
Ability to act tough	-0.42	0.10	1.1	1.1
Access to powerful positive emotions	-0.51	0.10	0.9	0.8
Physically project determination	-0.55	0.10	0.9	0.8
I compete fully recovered	-0.56	0.10	1.3	1.2
Willing to risk losing	-0.75	0.10	1.4	1.4
Commit 100% emotionally	-1.43	0.10	1.3	1.2
Feelings of confidence/energy	-1.55	0.10	1.0	1.1
Love heat of battle	-2.22	0.10	1.0	0.9
Mean	0.00	0.11	1.0	1.0
SD	0.70	0.01	0.2	0.2

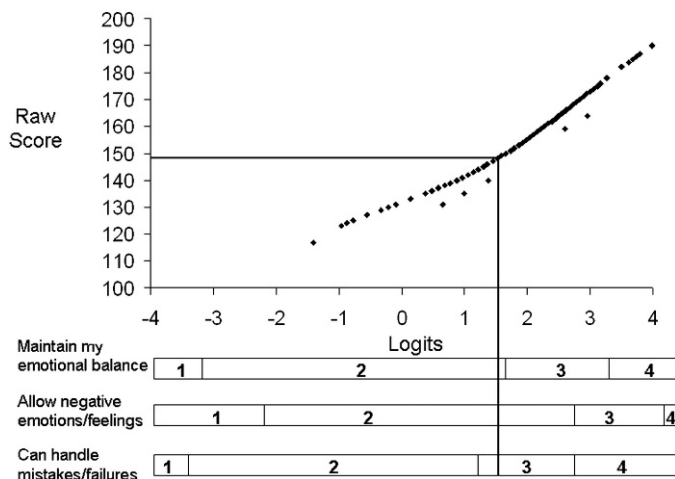
people's ability levels (Figure 2). This finding indicated that easier items are available for individuals with lower abilities, as well as more difficult items for people with higher abilities. The map also showed some gapping at the end ranges, where attention to future item development may be focused.

The item difficulties and person's ability estimates were calculated with CSEM. The CSEM is a measure that depicts the precision of the tool at a specific ability level ( $\theta$ ). A smaller CSEM indicates a better measurement and less error. This is directly related to the location of the items on the logit scale and the categories used. As shown in Table 2, the CSEMs for the items were small, with little variability for all 43 items.

The CSEM for the ability estimates provided valuable information about the precision of the tool. Relatively

equal precision across a large ability range was desired.<sup>17</sup> This was important because a norm-reference-based mental toughness measure was the goal. Tools that can measure better at an ability level provide a lower CSEM. In our study, the CSEMs of the ability estimates were fairly consistent across the ability range (0.24 to 0.34 logits). For practitioners using the instrument, this finding means that, regardless of mental toughness ability, the tool measures most people with the same precision.

From a practical diagnostic view for the athletic trainer, the tool could be used with a scoring sheet that converts the raw scores into logit scores. Researchers can create a scoring sheet that includes an evaluation standard, the items, and a conversion curve. The total score of the respondents then can be converted into logits. Thus, the practitioner can identify the athlete's relative strengths and



**Figure 3.** Potential scoring sheet with 3 sample items and a conversion curve.

weaknesses.<sup>29</sup> Perhaps the athlete has poor coping skills, which may be detrimental to normal rehabilitation. Conversely, clinicians can identify individuals who exhibit signs of acting tough and may not report minor injuries or setbacks, which, if left untreated, could develop into injuries that result in significant time lost from participation. Note that the scoring sheet and conversion curve in this form would require complete data (ie, a response to all 43 items).

For example, one of the participants in our study had a raw score of 147. Using the scoring sheet and conversion curve (see Figure 3), a straight line could be drawn from the raw score (y-axis) to the curve, then a perpendicular line could be drawn down to the logit scale (x-axis). This process converts the raw score into the logit score (ability), which has interval scale properties. This line down from the curve to the logit scale then could be continued through the items and responses. Each item can be placed on the scale, and each response can be represented. Each response is located on the logit scale where it is the most probable response for a particular ability level. This process of item evaluation has a potentially diagnostic benefit. Based on the ability of the athlete, we know the response to the question and the most likely response to the item. This may help evaluate strengths, weaknesses, and abnormal responses.

As shown, this participant's expected response for ability to handle competitive mistakes and failures would be a 3 on the 4-point scale. The participant's ability to maintain emotional balance when stressed and his or her ability to let negative emotions and feelings lead to negative thinking would be expected to be a 2. The line that bisects these regions illustrates this. If the respondent had chosen a 3 for maintaining emotional balance, this choice might indicate a possible strength of that respondent at that ability level (θ). Likewise, if the respondent had chosen a 1, the practitioner might want to intervene to determine why this occurred. Perhaps it is a weakness, a misunderstanding of the item, or something else. Obviously, this type of diagnostic profile would be extremely beneficial when assessing patients' mental toughness, which can provide insight into compliance and recovery.<sup>7</sup> Based on the strengths and weaknesses of the individual, specific

psychological skills training potentially could enhance rehabilitation and recovery.

Another major strength of the Rasch model in test construction is the ability to determine empirically whether the instrument developer selected the most appropriate categorization. Our results showed that the 7-category scale did not perform as predicted. Some categories (options) were not selected or needed. As illustrated in Figure 1, collapsing the number of categories from 7 to 4 provided a distinct separation of probability regions. Rating-scale results reported in Table 1 for the 4 weighted categories provided additional confirmation and indicated that the new scale performed well. Thus, the 4-category scale was deemed most appropriate and was used for the analyses. An independent cross-validation sample provided further evidence that optimal categorization was achieved.

Evidence that the MeBTough is valid is also important. Initial evidence was demonstrated with the differences found between the collegiate athletes and nonathletes in this sample. The moderate positive correlation between the mental toughness estimate and the participant's overall rating of his or her perceived mental toughness provided additional evidence of validity. Future researchers should focus on developing more support for the validity of the MeBTough and on developing more specific items that are tailored to the early stages of rehabilitation to better measure and ultimately serve our patients.

## CONCLUSIONS

In summary, the creation of an instrument capable of assessing one's mental toughness abilities would be of great interest to those helping individuals perform and comply with rehabilitation. Because of the advantages that the Rasch model has for providing powerful information to investigators in the development of a new tool, we selected and used it to analyze our recently created instrument based on Loehr's<sup>6</sup> toughness training strategies. Through item parameter invariance and through independent calibrations of items and people placed on a common metric, the Rasch model provided valuable insight into the tool construction. Results of our study provide support for using the new 43-item measure of mental, emotional, and physical toughness.

## REFERENCES

1. Loehr JE. *Mental Toughness Training for Sports: Achieving Athletic Excellence*. Lexington, MA: Stephen Greene Press; 1986.
2. Goldberg AS. *Sports Slump Busting: 10 Steps to Mental Toughness and Peak Performance*. Champaign, IL: Human Kinetics; 1998.
3. Jones G, Hanton S, Connaughton D. What is this thing called mental toughness? An investigation of elite sport performers. *J Appl Sport Psychol*. 2002;14(3):205-218.
4. Fletcher D, Fletcher J. A meta-model of stress, emotions and performance: conceptual foundations, theoretical framework, and research directions. *J Sports Sci*. 2005;23(2):157-158.
5. Gould D, Dieffenbach K, Moffatt A. Psychological characteristics and their development in Olympic champions. *J Appl Sport Psychol*. 2002;14(3):172-204.
6. Loehr JE. *The New Toughness Training for Sports: Mental, Emotional, and Physical Conditioning From One of the World's Premier Sports Psychologists*. New York, NY: Penguin Putnam; 1994.

7. Levy A, Polman RCJ, Clough PJ, Marchant DC, Earle K. Mental toughness as a determinant of beliefs, pain, and adherence in sport injury rehabilitation. *J Sport Rehabil.* 2006;15(3):246–254.
8. Murphy S, Tammen V. In search of psychological skills. In: Duda JL, ed. *Advances in Sport and Exercise Psychology Measurement.* Morgantown, WV: Fitness Information Technology; 1998:195–209.
9. Middleton SC, Marsh HW, Martin AJ, et al. The Psychological Performance Inventory: is the mental toughness test tough enough? *Int J Sport Psychol.* 2004;35(2):91–108.
10. Bull SJ, Shambrook CJ, James W, Brooks JE. Towards an understanding of mental toughness in elite English cricketers. *J Appl Sport Psychol.* 2005;17(3):209–227.
11. Loehr JE, Stites P. *The Mental Game.* Lexington, MA: Stephen Greene Press; 1990.
12. Loehr JE. The development of a cognitive-behavioral between-point intervention strategy for tennis. In: Serpa S, Alves J, Pataco V, eds. *International Perspectives on Sport and Exercise Psychology.* Morgantown, WV: Fitness Information Technology; 1994:219–233.
13. Jackson SA, Marsh HW. Development and validation of a scale to measure optimal experience: the Flow State Scale. *J Sport Exerc Psychol.* 1996;18(1):17–35.
14. Watson D, Clark LA. Measurement and mismeasurement of mood: recurrent and emergent issues. *J Pers Assess.* 1997;68(2):267–296.
15. Patten ML. *Questionnaire Research: A Practical Guide.* 2nd ed. Los Angeles, CA: Pyrczak Publishing; 2001.
16. 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(2):104–116.
17. Hambleton RK, Swaminathan H, Rogers HJ. *Fundamentals of Item Response Theory.* Newbury Park, CA: Sage Publications; 1991.
18. Rasch G. *Probabilistic Models for Some Intelligence and Achievement Tests.* Copenhagen, Denmark: Danish Institute for Educational Research; 1960.
19. Rasch G. *Probabilistic Models for Some Intelligence and Achievement Tests.* 2nd ed. Chicago, IL: University of Chicago Press; 1980.
20. Lord FM. *Applications of Item Response Theory to Practical Testing Problems.* Hillsdale, NJ: Lawrence Erlbaum; 1980.
21. American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. *Standards for Educational and Psychological Testing.* Washington, DC: American Educational Research Association; 1999:44–45.
22. Andrich D. A rating formulation for ordered response categories. *Psychometrika.* 1978;43(4):561–573.
23. Wright BD, Masters GN. *Rating Scale Analysis.* Chicago, IL: Mesa Press; 1982.
24. Linacre JM. Investigating rating scale category utility. *J Outcome Meas.* 1999;3(2):103–122.
25. Linacre JM. Optimizing rating scale category effectiveness. *J Appl Meas.* 2002;3(1):85–106.
26. Linacre JM. What do infit and outfit, mean-square and standardized mean? *Rasch Meas Trans.* 2002;16(2):878.
27. Fisher WP Jr. Reliability statistics. *Rasch Meas Trans.* 1992;6(3):238.
28. Golby J, Sheard M, Lavellee D. A cognitive-behavioural analysis of mental toughness in national rugby league football teams. *Percept Mot Skills.* 2003;96(2):455–462.
29. Woodcock RW. What can Rasch-based scores convey about a person's test performance? In: Embretson SE, & Hershberger SL, eds. *The New Rules of Measurement: What Every Psychologist and Educator Should Know.* Mahwah, NJ: Lawrence Erlbaum; 1999:105–127.

---

*Mick G. Mack, PhD, contributed to conception and design; acquisition and analysis and interpretation of the data; and drafting, critical revision, and final approval of the article. Brian G. Ragan, PhD, ATC, contributed to analysis and interpretation of the data and drafting, critical revision, and final approval of the article.*

*Address correspondence to Mick G. Mack, PhD, School of Health, Physical Education, and Leisure Services, University of Northern Iowa, Cedar Falls, IA 50614-0241. Address e-mail to mickey.mack@uni.edu.*