7-1-2012

# Differential Item Functioning Analysis of the Mental, Emotional, and Bodily Toughness Inventory

Yong Gao
*Boise State University*

Mick G. Mack
*University of North Iowa*

Moira A. Ragan
*Measuremental LLC*

Brian Ragan
*University of Ohio*

This is an electronic version of an article published in *Measurement in Physical Education and Exercise Science*, Volume 16, Issue 3, 2012.
*Measurement in Physical Education and Exercise Science* is available online at: http://www.tandfonline.com.
DOI: 10.1080/1091367X.2012.693349

# Differential Item Functioning Analysis of the Mental, Emotional, and Bodily Toughness Inventory

**Yong Gao**
Boise State University

**Moira A. Ragan**
Measuremental LLC

**Mick G. Mack**
University of North Iowa

**Brian Ragan**
University of Ohio

**Abstract**

This study was to use differential item functioning (DIF) analysis to examine if there were items in the Mental, Emotional, and Bodily Toughness Inventory (MeBTough) functioning differently across gender and athletic membership. A total of 444 male (56.3%) and female (43.7%) participants (30.9% athletes and 69.1% non-athletes) responded to the MeBTough items. Using Mantel-Haenszel and SIBTEST methods, 43 items were analyzed for DIF. Four MeBTough items were identified as large DIF items by both Mantel-Haenszel and SIBTEST methods, where item 21 favored non-athletes, item 40 favored athletes, item 2 favored males, and item 17 favored females. Athletic membership DIF disappeared whereas gender DIF still existed at the scale level. Overall, there are gender and athletic membership DIF items in the MeBTough, but only gender DIF still exists at the scale level. Thus, conclusions regarding gender differences in mental toughness should be made with caution when using total MeBTough scores.

**Keywords:** differential item functioning; mental toughness; sport competition; Mantel-Haenszel method; SIBTEST

The ability to perform under pressure is critical for competitive athletes (Gould, Dieffenbach, & Moffatt, 2002). Individuals able to play well when it matters the most are celebrated while those that fail are severely judged and criticized. The common term for this ability to consistently perform toward the upper range of one's skills and talents regardless of competitive circumstances is mental toughness (Loehr, 1994). More recently, the definition of mental toughness has been expanded as having a psychological edge that enables one to cope with the many on and off field demands of sport and to be more consistent and in control under pressure (Jones, Hanton, & Connaughton, 2007). Because of its diverse role as a key psychological component for successful athletes, it is important that mental toughness be adequately measured. The Mental, Emotional, and Bodily Toughness Inventory (MeBTough) is one such measure that has been increasingly used to assess mental toughness in an athletic population (Mack & Ragan, 2008).

The MeBTough is a 43-item questionnaire designed to assess the mental, physical, and emotional aspects of mental toughness (Mack & Ragan, 2008). The mental dimension encompasses the ability to create an optimal performance state, to access empowering emotions, and to cope. The physical dimension consists of being well-prepared and acting tough while the emotional aspect has four markers: flexibility, responsiveness, strength, and resiliency. Each item is answered using a 4-point scale with anchors ranging from 1 (Almost Never) to 4 (Almost Always). The MeBTough has been evaluated using Rasch analysis model (Rasch, 1980) and the results have indicated that the MeBTough has good model-data fit, were fittingly targeted to the studied population, had good variability along the measurement scale, and the use of the four categories (i.e., 4-point scale) for the items was optimal (Mack & Ragan, 2008).

In creating the MeBTough, one of the goals was to develop a universally applicable norm-referenced-based measure so that group differences (e.g., gender difference) can be examined using the aggregate MeBTough scores (Mack & Ragan, 2008). The validity of differences is based on items of a measure performing similarly across different groups (e.g., males and females of the same ability interpret and respond to items equally). This assumption of equality across groups potentially threatens the interpretation of scores, as group differences may be a combination of "true" differences in the primary trait and false differences in secondary traits or "bias." Using the MeBTough, Mack and Ragan (2008) found significant differences in mental toughness between gender and between athletic membership (e.g., athlete or non-athlete). Practical experience seems to also support such observations. However,

This is an electronic version of an article published in *Measurement in Physical Education and Exercise Science*, Volume 16, Issue 3, 2012.
*Measurement in Physical Education and Exercise Science* is available online at: http://www.tandfonline.com.
DOI: 10.1080/1091367X.2012.693349

whether the observed differences represent real mental toughness differences between males and females, and/or between athletes and non-athletes, or the differences are (or may be partially) the results of the presence of biased item(s) in the MeBTough have never been examined. Thus, there is a need to examine potential item bias in the MeBTough attributable to gender and athletic membership before a valid comparison in mental toughness regarding gender and athletic membership can be made.

To detect potentially biased items in the MeBTough (or any other measure), a set of statistical methods known as differential item functioning (DIF) analysis should be used. DIF refers to unequal probabilities of endorsement on an item when two groups are at the same ability level (Dorans & Holland, 1993). Ability typically is defined as the construct a test or an instrument is intended to measure (Roussos & Stout, 1996). Simply, an item may demonstrate DIF when two ability-matched groups of respondents react to the item differently. The presence of DIF indicates a test item may be potentially biased toward a subgroup (e.g., female, minority, etc.), which could pose a threat to the validity of a test or an instrument, and leads to incorrect explanations of test results and interferes with the selection or classification criterion (Camilli & Shepard, 1994).

DIF analysis is a routine procedure for instrument development and validation in educational and psychological testing (Hambleton, 2006). The importance of DIF analysis in survey construction has also been recognized in Kinesiology (e.g., Cohen, 2006; Looney, Spray, & Castelli, 1996; Zhu & Kurz, 1996). For example, Myers and his colleagues provided a nice conceptual introduction to DIF and demonstrated the usefulness of DIF analysis in refining and further validating the coach efficacy scale (Myers, Wolfe, Feltz, & Penfield, 2006). More recently, Gao and Zhu (2011a; 2011b) used the DIF analysis in evaluating the National Health and Nutrition Examination Survey (NHANES) physical activity (PA) questionnaire and found DIF items in the questionnaire. Their findings caution the conclusions regarding subpopulation differences in PA participation using the NHANES PA questionnaire. Thus, because DIF analysis is helpful in constructing unbiased measures, the purpose of this study was to examine the MeBTough items for the presence of DIF by examining variations in responses to each item by individual groups when the overall attribute (mental toughness in this study) was controlled. Two key grouping variables examined were gender and athletic membership (competitive versus recreational).

## Methods

### Participants

Four hundred forty-four college students volunteered for this study (43.7% females; 56.3% males). About 30.9% participants indicated that they were currently or had been a member of one of the university's collegiate athletic teams while 69.1% responded that they were not athletes. Participants reviewed and signed a consent form approved by the Institutional Review Board prior to participating in the current study.

### Measure

The MeBTough (Mack & Ragan, 2008) was used to measure participants' mental toughness. Briefly, the MeBTough includes 43 items with a 1 to 4 category response, asking respondents to rate how often they experienced each item. An example item is "I am willing to put myself totally on the line and risk losing." A category response of "1" represents "almost never", and "4" denotes "almost always". Eleven items are scored in reverse so that lower scores correspond to being mentally tougher. Total scores can range from 43 to 168 with higher scores indicating higher levels of mental toughness. Consistency reliability of the MeBTough was established previously with Cronbach's alpha coefficient equal to .95. Evidence of criterion validity was found with the moderately high correlation ($r = 0.67$) between participants' total MeBTough scores with their self-rated mental toughness scores (Mack & Ragan, 2008).

### Data Analysis

The goal of the study was to use DIF analysis to examine whether there were items in the MeBTough functioning differently between groups. If so, it usually indicates individuals' membership (e.g., being a male or a female, and being an athlete or a non-athlete) affects their responses to a specific item in the MeBTough, implying that item may be potentially biased against a subgroup. DIF analysis has been widely used in educational and psychological measurement practice to locate potentially biased items (Chang, Mazzeo, & Roussos, 1996; Holland & Thayer,

This is an electronic version of an article published in *Measurement in Physical Education and Exercise Science*, Volume 16, Issue 3, 2012.
*Measurement in Physical Education and Exercise Science* is available online at: http://www.tandfonline.com.
DOI: 10.1080/1091367X.2012.693349

1988; Lord, 1980; Mantel & Haenszel, 1959; Shealy & Stout, 1993a; 1993b; Zumbo, 1999). The presence of DIF between two or more comparable groups indicates that members of the respective groups have different likelihoods of endorsing particular items. When DIF exists, it may be inappropriate to utilize aggregate scores (i.e., summed scores) for group comparisons. Judgmental review needs to be carried out for the purpose toward either removing a DIF item from the instrument or correcting the part(s) of the item that may be causing the bias (Camilli & Shepard, 1994).

Many DIF techniques have been developed for DIF detections, among which, the Mantel-Haenszel (MH; Holland & Thayer, 1988; Mantel & Haenszel, 1959) and simultaneous item bias test (SIBTEST; Chang et al., 1996; Shealy & Stout, 1993a) are two of the most popular ones. MH and SIBTSET DIF procedures differ in various ways; however, they all need to match test respondents from the different groups on their ability levels according to the matching criterion. The ability level matching is usually accomplished using total or subtotal test score (i.e., aggregate score). The compared two groups are called the reference group and the focal group, respectively. The focal group usually is referred to as the group of interest (e.g., female) while the reference group is the group of standard with whom the focal group is to be compared (e.g., male; Roussos & Stout, 1996).

In the MH procedure, the focal and reference groups are first matched on their total or subtotal test scores, where respondents at each test score level are considered to be at the same ability level. Then, if the odds of getting an item endorsed at each test score level are the same for both groups will be determined, across all levels of the matching scores. The original MH method can only conduct DIF analysis with dichotomous responses (e.g., 0 or 1). The extension of MH method, also called the generalized Mantel-Haenszel procedure or Mantel procedure (Agresti, 1990; Mantel, 1963) can be applied to both dichotomous and polytomous responses such as the response format (e.g., 1 to 4) that was used in the MeBTough. In SIBTEST procedure, DIF is detected by identifying a secondary dimension in an item that is not part of what the test intends to measure (Shealy & Stout, 1993a; 1993b). Specifically, test items are first split into two subsets: one subset includes items for DIF investigation (i.e., "studied items"), and another subset includes the rest of the items in the test, which are often called the "matching items". The subtotal scores from the matching items are used to put test respondents into different ability levels. Within each ability level, test respondents in the reference and focal groups are considered to have the equivalent ability of being measured. Then, ability differences between the reference and focal groups on a studied item are compared to detect DIF whereas the two groups of respondents are matched at the same matching scores (Roussos & Stout, 1996).

In the current study, total/subtotal scores from the MeBTough were used to match ability levels of the focal and reference groups. The "ability" in this case refers to mental toughness. The focal groups included female and athlete, and the corresponding reference groups were male and non-athlete. DIF analyses were conducted for gender and athlete membership separately. Given that each DIF analysis approach has its own advantages and disadvantages, examining DIF using more than one approach is highly recommended (Hambleton, 2006). Significance level for DIF analyses was set at .001 to account for potential inflation of $\alpha$ from multiple comparisons. A positive Beta from SIBTEST indicates DIF favoring the reference group; that is, the reference group has higher probability to endorse an item than the focal group when they are at the same ability level, and a negative Beta value indicates DIF favoring the focal group. In this study, an item was flagged as a DIF item when it was identified by both MH and SIBTEST methods; and an item was not flagged as a DIF item when it was identified only by a single method. When DIF items were identified, an effort was made also to examine the impact of DIF at the instrument scale level by applying SIBTEST on a bundle of DIF items (Douglas, Roussos, & Stout, 1996). MH DIF analyses were conducted using SAS 9.1 (SAS Institute, 2008) and SIBTEST DIF analyses were conducted using DIFPACK 1.7 (William Stout Institute for Measurement, 2007).

## Results

### Descriptive Statistics

Among all participants, the average mental toughness score was 135.8 (SD = 17.0). Females had lower mental toughness scores with a score of 132.1 (SD = 17.6) when compared to 138.7 (SD = 16.0) for males. Non-athletes had lower mental toughness scores with a score of 134.1 (SD = 16.5), compared to 139.5 (SD = 17.7) for athletes. Significant differences in the average mental toughness scores were observed between both males and females ($t = 4.13$, $df = 442$, $p = 0.001$), and athletes and non-athletes ($t = 3.11$, $df = 442$, $p = 0.002$). There was no interaction between gender and athlete membership on the total mental toughness scores.

This is an electronic version of an article published in *Measurement in Physical Education and Exercise Science*, Volume 16, Issue 3, 2012.
*Measurement in Physical Education and Exercise Science* is available online at: http://www.tandfonline.com.
DOI: 10.1080/1091367X.2012.693349

Descriptive statistics, including item means and standard deviations, item-total correlation corrected, and the coefficient alpha with item deletion, by gender and athletic membership, are provided in Table 1 and Table 2.

**Athlete Membership DIF**

Table 3 presents the results of DIF analyses from SIBTEST and MH approaches with athletes being the reference group and non-athletes the focal group. Items 21 and 40 are identified as DIF items by both MH and SIBTEST DIF methods with all relevant statistics were statistically significant with p-values less than 0.001, where item 21 (Figure 1A) favored non-athletes and item 40 (Figure 1B) favored athletes. The absolute Beta values from SIBTEST for these two items were 0.227 and 0.215, respectively, which were larger than 0.088 (SIBTEST criterion for large DIF identification; Roussos & Stout 1996), indicating the presence of large DIF. The two items functioned differently between athlete and non-athlete groups. More specifically, at the same mental toughness levels, athletes consistently scored higher on item 40 (Figure 1B) than non-athletes while non-athletes tended to have higher scores on item 21 (Figure 1A).

**Gender DIF**

Table 4 presents the results of DIF analyses from SIBTEST and MH approaches with male being the reference group and female the focal group. Items 1, 2, 12, 13 and 17 were identified as DIF items by SIBTEST only, and items 2 and 17 were flagged as DIF items by both SIBTEST and MH methods, where item 2 (Figure 2A) favored males and item 17 (Figure 2B) favored females. The absolute Beta values from SIBTEST for items 2 and 17 were 0.354 and 0.199, respectively, which were larger than 0.088, indicating the presence of large DIF. Items 2 and 17 functioned differently between male and female groups. More specifically, at the same mental toughness levels, males consistently scored higher on item 2 (Figure 2A) than females while females tended to have higher scores on item 17 (Figure 2B).

**Effect of DIF Items on MeBTough**

When DIF exists in a test/instrument, the effect of DIF on the instrument is of great interest because decisions about the measured ability/trait are often made at the scale or test level (Roznowski, 1988) among many test/instrument users. It is possible that DIF exits at the item level, but disappears at the scale/test level due to DIF cancellation (i.e., some DIF items favors the reference group and some favors the focal group so that DIF was cancelled out at the scale level). It is possible also that the amount of DIF in any single item is small but over several such items small amount of DIF produces an unacceptable amount of DIF for a test, which is called DIF amplification (Douglas et al., 1996).

By combining the two DIF items by the athlete membership and testing DIF for the bundle of items, SIBTEST result showed the absolute Beta value was equal to 0.032 with p larger than 0.05, indicating there was no DIF any more. At the item level, item 21 favored non-athlete group and item 40 favored athlete group, the total amount of DIF for these two items, however, was cancelled out for the test. The total score from the MeBTough, as it relates to underlying mental toughness measure, is nearly the same for the two groups.

Similarly, the two gender DIF items were combined and tested for DIF using SIBTEST. The result showed the Beta value was equal to 0.200 with p less than 0.05, indicating the presence of large DIF favoring the male group. Total score from the MeBTough, as it relates to underlying mental toughness measure, is not the same for the two gender groups. Males tended to have higher total scores than females even when they actually were at the same mental toughness levels.

## Discussion

In the current study, a few MeBTough items with the presence of significant DIF have been identified across gender and athletic membership. The results were consistent between SIBTEST and the MH DIF methods although the MH method provided more conservative results than SIBTEST. Items 2, 17, 21 and 40 were identified as DIF items, with item 21 favoring non-athletes, item 40 favoring athletes, item 2 favoring males, and item 17 favoring females.

This is an electronic version of an article published in *Measurement in Physical Education and Exercise Science*, Volume 16, Issue 3, 2012.
*Measurement in Physical Education and Exercise Science* is available online at: http://www.tandfonline.com.
DOI: 10.1080/1091367X.2012.693349

Once identified, the next step is to examine the possible cause for the presence of DIF in each item with a judgmental review process. This judgmental review should determine whether the DIF in that item is an indicator of a relevant or an irrelevant secondary dimension to the construct in question. Typically, an item affected by a secondary dimension relevant to the intended construct is not considered to be a biased item and is not recommended for removal from a test although the proportion of such items in the test should be carefully controlled (Camilli & Shepard, 1994). An item judged to reveal a trait irrelevant to the intended construct is believed to be a biased item against a particular group and is commonly recommended to be removed from the test (e.g., Myers et al., 2006). Finally, the impacts of the identified DIF items on the MeBTough must also be described.

**Causes and Influences of DIF Items**

The presence of athlete membership DIF in items 21 and 40 has similarities because both items are asking about the ability to withstand the emotional strains of competition. Item 21 ("I sometimes allow my negative emotions and feelings to lead me into negative thinking."), which is reversely scored, favored the non-athletes. As one of the 3 most difficult items (1.04 logits; Mack & Ragan, 2008), non-athletes may not have actually experienced the detrimental effects that negative emotions can have on athletic performance while all competitive athletes have probably experienced the effect. Item 40 ("I can sustain a powerful fighting spirit against almost impossible odds.") favored the athletes. Having experienced the battle of competition on many occasions, athletes may have the coping skills and be more confident in their ability to maintain their composure in these types of unfavorable competitive scenarios. Therefore, items 21 and 40 may also measure a secondary trait (i.e., experience in sport competition) that may be related to the primary trait (i.e., mental toughness) by the MeBTough. When having the same mental toughness, the athletes and non-athletes responded to the investigated item differently because of the difference in competitive experience between the two groups. The non-athletes lack experience in sport competition compared with the athletes. Sport competition experience is necessary for an appropriate response to the investigated items because the MeBTough was originally developed to measure mental toughness, the ability to successfully perform under sport-related competitive scenarios. Thus, items 21 and 40 are not considered to be biased items for the MeBTough. Further analysis found that the DIF exists only at the item level and the effects of DIF in the two items were cancelled out at the scale level. Therefore, total scores were not affected by DIF, which indicates that conclusions (of group difference) about mental toughness could be made based on total MeBTough scores for the athlete and non-athlete groups.

The results suggesting that athlete and non-athlete DIF were cancelled and, thus not significant, are very promising. This would suggest that the MeBTough has a much wider range of application than was originally envisioned. Perhaps the MeBTough could be expanded to include mental toughness items relating to participation in a broad range of physical activities or more specialized populations such as athletic training rehabilitation.

The study results also revealed significant DIF for the two gender groups at the item level. It may be that items 2 and 17 had DIF because they revealed gender-schematic processing (i.e., learning what is conventionally appropriate for each gender; Bem, 1981) in addition to mental toughness. Perhaps item 2 favors males because it contains fairly masculine language ("I can take a punch emotionally and recover quickly."), which was perceived as more conventionally appropriate for males than females. Conversely, item 17 ("I have the ability to assess powerful positive emotions during competition.") may favor females because it is more culturally acceptable for females to be in touch with and able to access their emotions than for males. Therefore, items 2 and 17 may have functioned differently because there was a difference in schematic processing between males and females who have the same mental toughness ability. Based on the judgmental review, gender-appropriate characteristics are not relevant to the ability (i.e., mental toughness) being measured by the MeBTough, therefore, items 2 and 17 are considered to be biased in this study. Follow-up analysis showed significant DIF favoring males still existed when the two items were bundled for DIF analysis. As assessed by the MeBTough, males tended to have higher total mental toughness scores than females even in the situation that respondents from the two groups actually have the same mental toughness ability. Thus, conclusions about mental toughness for the male and female groups might be incorrect if the DIF were not accounted for.

**Proposed Solution to Addressing Bias**

To address the bias, three different and viable options as recommended by Myers et al. (2006) are presented: (a) Eliminate the two items (2 & 17) because of the gender bias, (b) Reword the two items and perform additional

This is an electronic version of an article published in *Measurement in Physical Education and Exercise Science*, Volume 16, Issue 3, 2012.
*Measurement in Physical Education and Exercise Science* is available online at: http://www.tandfonline.com.
DOI: 10.1080/1091367X.2012.693349

testing, and (c) Establish different norms for males and females (i.e., can't make comparisons between males and females) using all 43 items. A discussion of each option follows.

*Eliminate the two items (2 & 17) because of the gender bias.* The original intent of the MeBTough was to create a single measure of mental toughness that could be used for a wide population of competitive athletes. Eliminating the two questions (Items 2 "I can take a punch emotionally and recover quickly." and 17 "I have the ability to access powerful positive emotions during competition.") would make the MeBTough a DIF free instrument for assessing mental toughness, and reduce the overall length of the test without causing any of the nine constructs to have less than four questions. However, the two items were from different content domains and thus, while eliminating the DIF problem, whether removing the items could potentially hurt the overall MeBTough discrimination needs further investigation.

*Reword the items and perform additional testing.* It is somewhat surprising that there are so few gender DIF items in the MeBTough. Previous research suggests that the successful female athlete tends to exhibit personality traits (i.e., assertive, aggressive, dominant) much more like the normative male and male athlete than the normative female (Cox, 2007). Unfortunately, the relatively small number of female athletes (n = 44) in this sample limits the ability to do additional DIF analyses focusing on the interaction between gender and athletic membership (e.g., compare the responses to the two items between female athletes and females who are not an athlete). Thus, additional research increasing the number of female athletes is warranted regarding these two items. This option would consist of rewording the two items and administering it to additional subjects to see if this addresses the problem. One of the strengths of the Rasch analysis model used previously to psychometrically examine the MeBTough is that both the items and participants are placed on a common metric so that additional items could be included on the same metric at a later time (Zhu, Timm, & Ainsworth, 2001), which would allow for future DIF analyses of samples including more female athletes.

*Establish different norms for males and females using all 43 items.* An examination of the mean scores found that males had significantly higher mental toughness scores (M = 138.7, SD = 16.0) than did females (M = 132.1, SD = 17.6). In addition, while not statistically significant, the mean male scores listed in Table 2 are higher on 39 of the 43 items. Thus, there may be real gender differences in mental toughness abilities that are revealed by the present MeBTough. Additional research could examine possible cultural or psychosocial influences on mental toughness. By limiting the comparisons within the same gender, the DIF differences between genders would be mute and the overall MeBTough integrity would remain the same.

**Limitations**

It should be noted that the current study is not without limitations. This study used a relatively homogeneous sample of participants (i.e., college students and/or collegiate athletes who are at similar ages and education level). Therefore, the generalization of the study's results and conclusions to other populations should be made with caution. In addition, DIF analysis in this study was conducted based only on two grouping variables (i.e., gender and athletic membership). It is possible that other demographic variables such as race/ethnicity and cultural preference may play a role in participants' responses to a particular item in the MeBTough. Considering the large number of minority athletes in many of today's sports, it is important to examine whether any race/ethnicity related DIF items exist in the MeBTough, and if so, how the DIF items influence the aggregated MeBTough scores between different race/ethnicity groups. Such investigations will further advance our understanding of the underlying factors contributing to mental toughness discrepancies between groups.

## Conclusions

In summary, this study indicates there are gender and athletic membership DIF items in the MeBTough. However, only gender DIF still exists at the scale level. Thus, when using cumulative MeBTough scores from the current version of 43 items, conclusions regarding potential mental toughness differences between males and females should be made with caution. The current study also highlights the importance of conducting DIF analysis for measures used to investigate between group differences and provides further validity evidence for the MeBTough.

## References

Agresti, A. (1990). *Categorical data analysis*. New York: John Wiley & Sons.

Bem, S. L. (1981). Gender schema theory: A cognitive accounts of sex typing. *Psychological Review, 88,* 354-364.

Camilli, G., & Shepard, L. A. (1994). Methods for identifying biased test items. Thousand Oaks, CA: Sage.

Chang, H. H., Mazzeo, J., & Roussos, L. (1996). Detecting DIF for polytomously scored items: An adaptation of the SIBTEST procedure. *Journal of Educational Measurement, 33*(3), 333-353.

Cohen, A. S. (2006). Item bias and differential item functioning. In T. M. Wood & W. Zhu (Eds.), *Measurement issues and practice in physical activity.* (pp. 113-126.). Champaign, IL: Human Kinetics.

Cox, R. H. (2007). *Sport psychology: Concepts and applications* (6th ed.). New York: McGraw-Hill.

Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P.W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 35-66). Hillsdale, NJ: Lawrence Erlbaum.

Douglas, J., Roussos, L., & Stout, W. (1996). Item bundle DIF hypothesis testing: Identifying suspect bundles and assessing their differential item functioning. *Journal of Educational Measurement, 33,* 465-484.

Gao, Y., & Zhu, W. (2011a). Identifying group sensitive physical activities: A Differential item functioning analysis of NHANES data. *Medicine & Science in Sport & Exercise, 43*(5): 922-929.

Gao, Y., & Zhu, W. (2011b). Differential item functioning analysis of the 2003-04 NHANES physical activity questionnaire. *Research Quarterly for Exercise and Sport, 82*(3): 381-390.

Gould, D., Dieffenbach, K., & Moffatt, A. (2002) Psychological characteristics and their development in Olympic champions. *Journal of Applied Sport Psychologist, 14,* 172-204.

Hambleton, R. K. (2006). Good practices for identifying differential item functioning. *Medical Care, 44*(11 Suppl 3):S182-188.

Holland, P. W., & Thayer, D. T. (1988). Differential item functioning and the Mantel-Haenszel procedures. In H. Wainer & H. Brain (Eds.), *Test validity*. Hillsdale, NJ: Lawrence Erlbaum.

Jones, G., Hanton, S., & Connaughton, D. (2007). A framework of mental toughness in the world's best performers. *The Sport Psychologist, 21,* 243-264.

Loehr, J. E. (1994) *The new toughness training for sports: mental, emotional, and physical conditioning from one of the world's premier sports psychologists*. New York: Penguin Putnam.

Lord, F. M. (1980). *Applications of item response theory*. Hillsdale, NJ: Erlbaum.

Looney, M.A., Spray, J.A., & Castelli, D. (1996). The task difficulty of free throw shooting for males and females. *Research Quarterly for Exercise and Sport, 67*(3), 265-271.

Mack, M. G., & Ragan, B. G. (2008) Development of the Mental, Emotional, and Bodily Toughness Inventory in collegiate athletes and nonathletes. *Journal of Athletic Training, 43*, 125-132.

Mantel, N. (1963). Chi-square tests with one degree of freedom: Extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association, 58*, 690-700.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22*, 719-748.

Myers, N. D., Wolfe, E. W., Feltz, D. L., & Penfield, R. D. (2006). Identifying differential item functioning of rating scale items with the Rasch model: An introduction and an application. *Measurement in Physical Education and Exercise Science, 10*(4), 215-240.

Rasch G. (1980). *Probabilistic models for some intelligence and achievement tests*. 2nd ed. Chicago, IL: University of Chicago Press.

Roussos, L., & Stout, W. (1996). A multimensionality-Based DIF analysis paradigm. *Applied Psychological Measurement, 20*(4), 355-371.

Roznowski, M. (1988). Review of test validity. *Journal of Educational Measurement, 25*, 357-361.

SAS Institute Inc. 2008. SAS 9.1 for Windows. SAS Institute Inc., Cary, NC.

Shealy, R. T., & Stout, W. F. (1993a). An item response theory model for test bias and differential test functioning. In P. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 197-240). Hillsdale, NJ: Erlbaum.

Shealy, R., & Stout, W. (1993b). A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DTF as well as item bias/DIF. *Psychometrika, 58*, 159-194.

William Stout Institute for Measurement. (2007). DIFPACK. Assessment Systems Corporation, St. Paul, MN.

Zhu, W., & Kurz, K. A. (1996). Graphical DIF analysis for assessing biased motor items/tasks. *Research Quarterly for Exercise and Sport, 67*(Suppl. 1), A-63 - A-64.

This is an electronic version of an article published in *Measurement in Physical Education and Exercise Science*, Volume 16, Issue 3, 2012.
*Measurement in Physical Education and Exercise Science* is available online at: http://www.tandfonline.com.
DOI: 10.1080/1091367X.2012.693349

Zhu, W., Timm, G., & Ainsworth, B. (2001). Rasch calibration and optimal categorization of an instrument measuring women's exercise perseverance and barriers. *Research Quarterly for Exercise and Sport, 72,* 104-116.

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores*. Ottawa, ON, Canada: Directorate of Human Resources Research and Evaluation, Department of National Defense.

**Authors' Note**

The Mental, Emotional, and Bodily Toughness Inventory will be selling at Mick Mack, Moira Ragan and Brian Ragan's company -- Measuremental LLC.

.

This is an electronic version of an article published in *Measurement in Physical Education and Exercise Science*, Volume 16, Issue 3, 2012.
*Measurement in Physical Education and Exercise Science* is available online at: http://www.tandfonline.com.
DOI: 10.1080/1091367X.2012.693349

## Figure Captions

Figure 1: Athletic Membership DIF: **A. Item 21:** "I sometimes allow my negative emotions and feelings to lead me into negative thinking." **B. Item 40**: "I can sustain a powerful fighting spirit against almost impossible odds."

Figure 2: Gender DIF: **A. Item 2:** "I can take a punch emotionally and recover quickly." **B. Item 17**: "I have the ability to assess powerful positive emotions during competition."

This is an electronic version of an article published in *Measurement in Physical Education and Exercise Science*, Volume 16, Issue 3, 2012.
*Measurement in Physical Education and Exercise Science* is available online at: http://www.tandfonline.com.
DOI: 10.1080/1091367X.2012.693349

Table 1: Descriptive Statistics for MeBTough Items by Athlete Membership

| | Non-Athlete (Alpha = 0.95) | | | | Athlete (Alpha = 0.95) | | | |
|---|---|---|---|---|---|---|---|---|
| Item | Mean | SD | Item-Test | Alpha | Mean | SD | Item-Test | Alpha |
| q1 | 3.14 | 0.71 | 0.47 | 0.94 | 3.39 | 0.73 | 0.57 | 0.95 |
| q2 | 2.95 | 0.76 | 0.52 | 0.94 | 3.17 | 0.77 | 0.51 | 0.95 |
| q3 | 3.06 | 0.64 | 0.36 | 0.95 | 3.19 | 0.66 | 0.56 | 0.95 |
| q4 | 2.60 | 0.74 | 0.37 | 0.95 | 2.80 | 0.71 | 0.26 | 0.95 |
| q5 | 2.82 | 0.74 | 0.34 | 0.95 | 2.81 | 0.81 | 0.26 | 0.95 |
| q6 | 3.14 | 0.67 | 0.52 | 0.94 | 3.30 | 0.68 | 0.47 | 0.95 |
| q7 | 3.27 | 0.67 | 0.46 | 0.94 | 3.36 | 0.71 | 0.62 | 0.95 |
| q8 | 3.21 | 0.68 | 0.56 | 0.94 | 3.28 | 0.72 | 0.66 | 0.95 |
| q9 | 3.01 | 0.68 | 0.53 | 0.94 | 3.20 | 0.67 | 0.50 | 0.95 |
| q10 | 2.92 | 0.78 | 0.31 | 0.95 | 2.97 | 0.78 | 0.39 | 0.95 |
| q11 | 3.16 | 0.70 | 0.45 | 0.94 | 3.26 | 0.74 | 0.57 | 0.95 |
| q12 | 2.93 | 0.68 | 0.58 | 0.94 | 3.15 | 0.65 | 0.58 | 0.95 |
| q13 | 3.20 | 0.64 | 0.56 | 0.94 | 3.20 | 0.69 | 0.66 | 0.95 |
| q14 | 3.08 | 0.73 | 0.49 | 0.94 | 3.09 | 0.81 | 0.32 | 0.95 |
| q15 | 3.27 | 0.67 | 0.61 | 0.94 | 3.24 | 0.78 | 0.55 | 0.95 |
| q16 | 3.42 | 0.69 | 0.49 | 0.94 | 3.58 | 0.62 | 0.49 | 0.95 |
| q17 | 3.30 | 0.64 | 0.54 | 0.94 | 3.38 | 0.74 | 0.65 | 0.95 |
| q18 | 3.03 | 0.78 | 0.38 | 0.95 | 3.02 | 0.82 | 0.41 | 0.95 |
| q19 | 3.07 | 0.69 | 0.65 | 0.94 | 3.26 | 0.71 | 0.52 | 0.95 |
| q20 | 3.15 | 0.67 | 0.62 | 0.94 | 3.28 | 0.67 | 0.66 | 0.95 |
| q21 | 2.77 | 0.75 | 0.60 | 0.94 | 2.74 | 0.89 | 0.49 | 0.95 |
| q22 | 2.89 | 0.74 | 0.52 | 0.94 | 3.05 | 0.79 | 0.53 | 0.95 |
| q23 | 3.30 | 0.65 | 0.64 | 0.94 | 3.40 | 0.69 | 0.67 | 0.95 |
| q24 | 3.05 | 0.66 | 0.48 | 0.94 | 3.30 | 0.69 | 0.31 | 0.95 |
| q25 | 3.15 | 0.62 | 0.59 | 0.94 | 3.33 | 0.73 | 0.68 | 0.95 |
| q26 | 3.03 | 0.79 | 0.49 | 0.94 | 3.18 | 0.81 | 0.50 | 0.95 |
| q27 | 3.22 | 0.66 | 0.45 | 0.94 | 3.36 | 0.69 | 0.42 | 0.95 |
| q28 | 3.09 | 0.74 | 0.71 | 0.94 | 3.23 | 0.76 | 0.66 | 0.95 |
| q29 | 3.13 | 0.75 | 0.45 | 0.94 | 3.11 | 0.89 | 0.47 | 0.95 |
| q30 | 2.97 | 0.68 | 0.55 | 0.94 | 3.28 | 0.67 | 0.46 | 0.95 |
| q31 | 3.09 | 0.71 | 0.60 | 0.94 | 3.28 | 0.77 | 0.72 | 0.95 |
| q32 | 3.23 | 0.66 | 0.63 | 0.94 | 3.39 | 0.71 | 0.66 | 0.95 |
| q33 | 3.26 | 0.83 | 0.49 | 0.94 | 3.48 | 0.74 | 0.56 | 0.95 |
| q34 | 3.64 | 0.62 | 0.46 | 0.94 | 3.73 | 0.59 | 0.56 | 0.95 |
| q35 | 3.12 | 0.69 | 0.56 | 0.94 | 3.26 | 0.69 | 0.57 | 0.95 |
| q36 | 3.19 | 0.64 | 0.65 | 0.94 | 3.31 | 0.72 | 0.68 | 0.95 |
| q37 | 3.17 | 0.65 | 0.53 | 0.94 | 3.31 | 0.69 | 0.61 | 0.95 |
| q38 | 2.97 | 0.59 | 0.66 | 0.94 | 3.11 | 0.69 | 0.66 | 0.95 |
| q39 | 3.00 | 0.67 | 0.55 | 0.94 | 3.04 | 0.81 | 0.58 | 0.95 |
| q40 | 3.02 | 0.66 | 0.59 | 0.94 | 3.34 | 0.64 | 0.49 | 0.95 |
| q41 | 3.28 | 0.64 | 0.58 | 0.94 | 3.52 | 0.62 | 0.54 | 0.95 |
| q42 | 3.26 | 0.77 | 0.44 | 0.95 | 3.23 | 0.89 | 0.56 | 0.95 |
| q43 | 3.53 | 0.61 | 0.60 | 0.94 | 3.63 | 0.66 | 0.58 | 0.95 |

This is an electronic version of an article published in *Measurement in Physical Education and Exercise Science*, Volume 16, Issue 3, 2012.
*Measurement in Physical Education and Exercise Science* is available online at: http://www.tandfonline.com.
DOI: 10.1080/1091367X.2012.693349

Table 2: Descriptive Statistics for MeBTough Items by Gender

| | Female (Alpha = 0.95) | | | | Male (Alpha = 0.94) | | | |
|---|---|---|---|---|---|---|---|---|
| Item | Mean | SD | Item- | Alpha | Mean | SD | Item- | Alpha |
| q1 | 3.05 | 0.77 | 0.50 | 0.95 | 3.35 | 0.65 | 0.48 | 0.94 |
| q2 | 2.72 | 0.79 | 0.50 | 0.95 | 3.25 | 0.67 | 0.50 | 0.94 |
| q3 | 3.03 | 0.66 | 0.42 | 0.95 | 3.16 | 0.63 | 0.42 | 0.94 |
| q4 | 2.52 | 0.71 | 0.43 | 0.95 | 2.78 | 0.74 | 0.24 | 0.94 |
| q5 | 2.77 | 0.71 | 0.30 | 0.95 | 2.86 | 0.80 | 0.31 | 0.94 |
| q6 | 3.12 | 0.66 | 0.55 | 0.95 | 3.24 | 0.69 | 0.48 | 0.94 |
| q7 | 3.16 | 0.72 | 0.55 | 0.95 | 3.40 | 0.63 | 0.45 | 0.94 |
| q8 | 3.14 | 0.72 | 0.60 | 0.95 | 3.30 | 0.66 | 0.57 | 0.94 |
| q9 | 3.03 | 0.66 | 0.54 | 0.95 | 3.10 | 0.69 | 0.52 | 0.94 |
| q10 | 2.84 | 0.71 | 0.25 | 0.95 | 3.01 | 0.81 | 0.38 | 0.94 |
| q11 | 3.16 | 0.73 | 0.50 | 0.95 | 3.22 | 0.70 | 0.50 | 0.94 |
| q12 | 2.79 | 0.75 | 0.66 | 0.95 | 3.16 | 0.57 | 0.48 | 0.94 |
| q13 | 3.03 | 0.68 | 0.55 | 0.95 | 3.33 | 0.60 | 0.60 | 0.94 |
| q14 | 3.09 | 0.69 | 0.49 | 0.95 | 3.07 | 0.80 | 0.41 | 0.94 |
| q15 | 3.19 | 0.73 | 0.57 | 0.95 | 3.32 | 0.68 | 0.58 | 0.94 |
| q16 | 3.40 | 0.69 | 0.50 | 0.95 | 3.52 | 0.65 | 0.49 | 0.94 |
| q17 | 3.34 | 0.68 | 0.57 | 0.95 | 3.31 | 0.67 | 0.62 | 0.94 |
| q18 | 2.99 | 0.75 | 0.39 | 0.95 | 3.05 | 0.82 | 0.39 | 0.94 |
| q19 | 2.98 | 0.73 | 0.65 | 0.95 | 3.23 | 0.65 | 0.54 | 0.94 |
| q20 | 3.12 | 0.66 | 0.63 | 0.95 | 3.24 | 0.68 | 0.63 | 0.94 |
| q21 | 2.61 | 0.79 | 0.59 | 0.95 | 2.88 | 0.78 | 0.49 | 0.94 |
| q22 | 2.89 | 0.79 | 0.59 | 0.95 | 2.97 | 0.73 | 0.48 | 0.94 |
| q23 | 3.23 | 0.70 | 0.65 | 0.95 | 3.42 | 0.62 | 0.64 | 0.94 |
| q24 | 3.13 | 0.69 | 0.55 | 0.95 | 3.13 | 0.67 | 0.36 | 0.94 |
| q25 | 3.13 | 0.68 | 0.66 | 0.95 | 3.26 | 0.64 | 0.59 | 0.94 |
| q26 | 3.04 | 0.78 | 0.47 | 0.95 | 3.11 | 0.81 | 0.53 | 0.94 |
| q27 | 3.26 | 0.69 | 0.49 | 0.95 | 3.27 | 0.66 | 0.42 | 0.94 |
| q28 | 2.98 | 0.77 | 0.69 | 0.95 | 3.25 | 0.70 | 0.67 | 0.94 |
| q29 | 3.10 | 0.75 | 0.42 | 0.95 | 3.14 | 0.83 | 0.49 | 0.94 |
| q30 | 2.95 | 0.73 | 0.59 | 0.95 | 3.16 | 0.65 | 0.46 | 0.94 |
| q31 | 3.07 | 0.74 | 0.68 | 0.95 | 3.21 | 0.73 | 0.62 | 0.94 |
| q32 | 3.19 | 0.65 | 0.71 | 0.95 | 3.36 | 0.70 | 0.58 | 0.94 |
| q33 | 3.19 | 0.87 | 0.52 | 0.95 | 3.44 | 0.74 | 0.49 | 0.94 |
| q34 | 3.59 | 0.70 | 0.47 | 0.95 | 3.74 | 0.52 | 0.49 | 0.94 |
| q35 | 2.99 | 0.74 | 0.58 | 0.95 | 3.29 | 0.62 | 0.53 | 0.94 |
| q36 | 3.11 | 0.68 | 0.66 | 0.95 | 3.32 | 0.65 | 0.65 | 0.94 |
| q37 | 3.22 | 0.67 | 0.58 | 0.95 | 3.21 | 0.66 | 0.58 | 0.94 |
| q38 | 2.87 | 0.64 | 0.68 | 0.95 | 3.13 | 0.59 | 0.62 | 0.94 |
| q39 | 2.95 | 0.69 | 0.48 | 0.95 | 3.06 | 0.73 | 0.63 | 0.94 |
| q40 | 3.04 | 0.66 | 0.60 | 0.95 | 3.18 | 0.68 | 0.53 | 0.94 |
| q41 | 3.29 | 0.67 | 0.58 | 0.95 | 3.40 | 0.62 | 0.57 | 0.94 |
| q42 | 3.23 | 0.76 | 0.49 | 0.95 | 3.27 | 0.84 | 0.47 | 0.94 |
| q43 | 3.54 | 0.63 | 0.59 | 0.95 | 3.58 | 0.62 | 0.60 | 0.94 |

This is an electronic version of an article published in *Measurement in Physical Education and Exercise Science*, Volume 16, Issue 3, 2012.
*Measurement in Physical Education and Exercise Science* is available online at: http://www.tandfonline.com.
DOI: 10.1080/1091367X.2012.693349

Table 3: DIF Analysis Results by Athlete Membership

| | SIBTEST | | MH | | |
|---|---|---|---|---|---|
| Item | Beta | p-value | Chi-Square | p-value | DIF Evaluation |
| 1 | 0.133 | 0.090 | 6.457 | 0.011 | |
| 2 | -0.070 | 0.385 | 0.349 | 0.555 | |
| 3 | 0.086 | 0.210 | 1.404 | 0.236 | |
| 4 | 0.139 | 0.102 | 1.852 | 0.174 | |
| 5 | -0.273 | 0.011 | 0.726 | 0.394 | |
| 6 | -0.020 | 0.777 | 0.170 | 0.680 | |
| 7 | -0.040 | 0.527 | 0.000 | 0.986 | |
| 8 | -0.115 | 0.110 | 0.847 | 0.357 | |
| 9 | 0.013 | 0.861 | 0.617 | 0.432 | |
| 10 | -0.090 | 0.342 | 0.213 | 0.644 | |
| 11 | -0.064 | 0.431 | 0.008 | 0.929 | |
| 12 | -0.009 | 0.893 | 0.132 | 0.716 | |
| 13 | -0.176 | 0.014 | 4.639 | 0.031 | |
| 14 | -0.128 | 0.179 | 4.967 | 0.026 | |
| 15 | -0.142 | 0.057 | 3.002 | 0.083 | |
| 16 | 0.029 | 0.699 | 1.051 | 0.305 | |
| 17 | -0.032 | 0.640 | 0.034 | 0.853 | |
| 18 | 0.044 | 0.681 | 3.091 | 0.079 | |
| 19 | 0.046 | 0.497 | 0.477 | 0.490 | |
| 20 | -0.028 | 0.646 | 0.007 | 0.933 | |
| 21 | -0.227 | 0.001 | 10.287 | 0.001 | DIF |
| 22 | 0.032 | 0.663 | 0.155 | 0.694 | |
| 23 | 0.006 | 0.921 | 0.129 | 0.720 | |
| 24 | 0.187 | 0.011 | 2.534 | 0.111 | |
| 25 | 0.028 | 0.669 | 0.906 | 0.341 | |
| 26 | 0.030 | 0.722 | 1.481 | 0.224 | |
| 27 | 0.012 | 0.883 | 0.001 | 0.977 | |
| 28 | -0.104 | 0.143 | 1.554 | 0.213 | |
| 29 | -0.097 | 0.269 | 6.620 | 0.010 | |
| 30 | 0.112 | 0.128 | 5.141 | 0.023 | |
| 31 | 0.026 | 0.707 | 1.676 | 0.195 | |
| 32 | -0.019 | 0.760 | 0.042 | 0.837 | |
| 33 | 0.157 | 0.027 | 3.488 | 0.062 | |
| 34 | 0.043 | 0.524 | 0.919 | 0.338 | |
| 35 | -0.013 | 0.862 | 0.034 | 0.853 | |
| 36 | -0.028 | 0.681 | 0.123 | 0.726 | |
| 37 | 0.011 | 0.867 | 0.000 | 0.988 | |
| 38 | -0.083 | 0.194 | 0.636 | 0.425 | |
| 39 | -0.136 | 0.084 | 3.598 | 0.058 | |
| 40 | 0.215 | 0.000 | 17.471 | 0.000 | DIF |
| 41 | 0.124 | 0.051 | 3.692 | 0.055 | |
| 42 | -0.063 | 0.522 | 3.799 | 0.051 | |
| 43 | 0.004 | 0.947 | 0.048 | 0.827 | |

This is an electronic version of an article published in *Measurement in Physical Education and Exercise Science*, Volume 16, Issue 3, 2012.
*Measurement in Physical Education and Exercise Science* is available online at: http://www.tandfonline.com.
DOI: 10.1080/1091367X.2012.693349

Note. Significant level has been set at 0.001 to account for potential inflation of α from multiple comparisons; Positive Beta indicates DIF favoring the reference group and negative Beta value indicates DIF favoring the focal group.

This is an electronic version of an article published in *Measurement in Physical Education and Exercise Science*, Volume 16, Issue 3, 2012.
*Measurement in Physical Education and Exercise Science* is available online at: http://www.tandfonline.com.
DOI: 10.1080/1091367X.2012.693349

Table 4: DIF Analysis Results by Gender

| | SIBTEST | | MH | | |
|---|---|---|---|---|---|
| Item | Beta | p-value | Chi-Square | p-value | DIF Evaluation |
| 1 | 0.248 | 0.001 | 3.272 | 0.071 | |
| 2 | 0.354 | 0.000 | 26.675 | 0.000 | DIF |
| 3 | -0.057 | 0.342 | 0.694 | 0.405 | |
| 4 | 0.155 | 0.093 | 3.900 | 0.048 | |
| 5 | -0.140 | 0.096 | 0.764 | 0.382 | |
| 6 | -0.021 | 0.767 | 0.073 | 0.788 | |
| 7 | 0.157 | 0.018 | 2.105 | 0.147 | |
| 8 | 0.065 | 0.340 | 0.796 | 0.372 | |
| 9 | -0.114 | 0.084 | 1.193 | 0.275 | |
| 10 | 0.149 | 0.050 | 1.167 | 0.280 | |
| 11 | -0.080 | 0.264 | 0.989 | 0.320 | |
| 12 | 0.188 | 0.001 | 8.626 | 0.003 | |
| 13 | 0.219 | 0.000 | 7.827 | 0.005 | |
| 14 | -0.154 | 0.034 | 5.968 | 0.015 | |
| 15 | -0.001 | 0.991 | 0.704 | 0.401 | |
| 16 | -0.029 | 0.655 | 0.656 | 0.418 | |
| 17 | -0.199 | 0.001 | 10.216 | 0.001 | DIF |
| 18 | -0.098 | 0.248 | 3.046 | 0.081 | |
| 19 | 0.022 | 0.734 | 1.161 | 0.281 | |
| 20 | -0.056 | 0.380 | 0.041 | 0.839 | |
| 21 | 0.076 | 0.316 | 0.609 | 0.435 | |
| 22 | -0.134 | 0.063 | 0.901 | 0.342 | |
| 23 | 0.062 | 0.289 | 0.549 | 0.459 | |
| 24 | -0.137 | 0.053 | 6.667 | 0.010 | |
| 25 | -0.076 | 0.192 | 2.372 | 0.124 | |
| 26 | -0.073 | 0.328 | 2.539 | 0.111 | |
| 27 | -0.058 | 0.399 | 5.633 | 0.018 | |
| 28 | 0.046 | 0.473 | 0.973 | 0.324 | |
| 29 | -0.049 | 0.534 | 0.404 | 0.525 | |
| 30 | 0.073 | 0.269 | 0.857 | 0.355 | |
| 31 | -0.073 | 0.251 | 0.753 | 0.386 | |
| 32 | -0.020 | 0.748 | 0.014 | 0.906 | |
| 33 | 0.098 | 0.247 | 1.204 | 0.273 | |
| 34 | -0.042 | 0.473 | 0.018 | 0.893 | |
| 35 | 0.171 | 0.016 | 7.535 | 0.006 | |
| 36 | 0.037 | 0.547 | 1.079 | 0.299 | |
| 37 | -0.180 | 0.004 | 7.984 | 0.005 | |
| 38 | 0.111 | 0.055 | 4.917 | 0.027 | |
| 39 | -0.083 | 0.204 | 0.200 | 0.655 | |
| 40 | -0.041 | 0.530 | 0.153 | 0.696 | |
| 41 | -0.067 | 0.228 | 1.388 | 0.239 | |
| 42 | -0.062 | 0.462 | 3.709 | 0.054 | |
| 43 | -0.139 | 0.010 | 5.709 | 0.017 | |

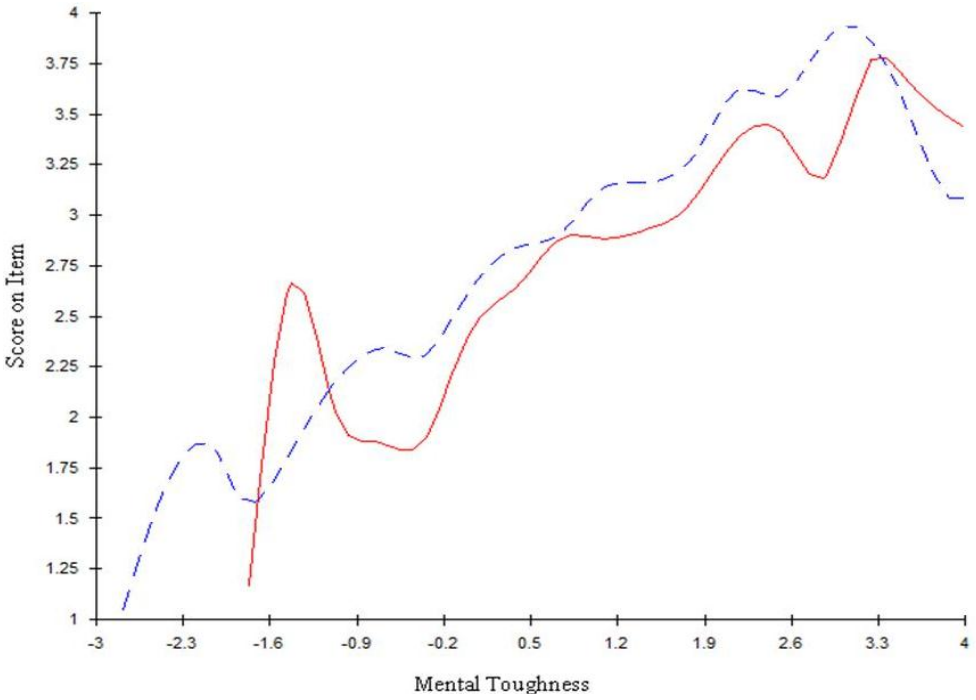
This is an electronic version of an article published in *Measurement in Physical Education and Exercise Science*, Volume 16, Issue 3, 2012.
*Measurement in Physical Education and Exercise Science* is available online at: http://www.tandfonline.com.
DOI: 10.1080/1091367X.2012.693349

Note. Significant level has been set at 0.001 to account for potential inflation of α from multiple comparisons; Positive Beta indicates DIF favoring the reference group and negative Beta value indicates DIF favoring the focal group.

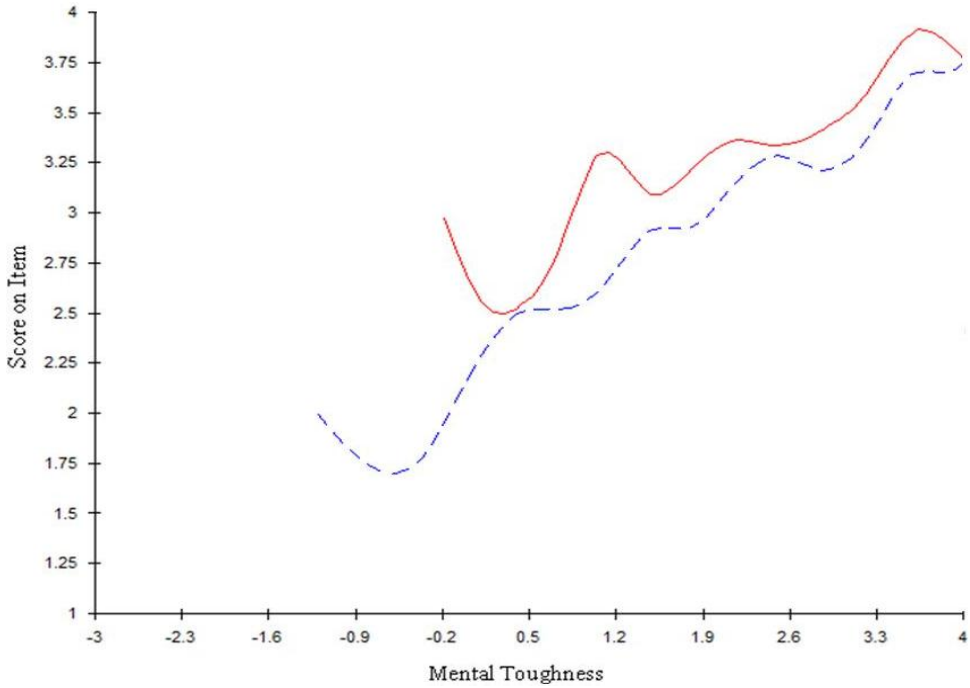
This is an electronic version of an article published in *Measurement in Physical Education and Exercise Science*, Volume 16, Issue 3, 2012.
*Measurement in Physical Education and Exercise Science* is available online at: http://www.tandfonline.com.
DOI: 10.1080/1091367X.2012.693349

Figure 1

a. Item 21




b. Item 40

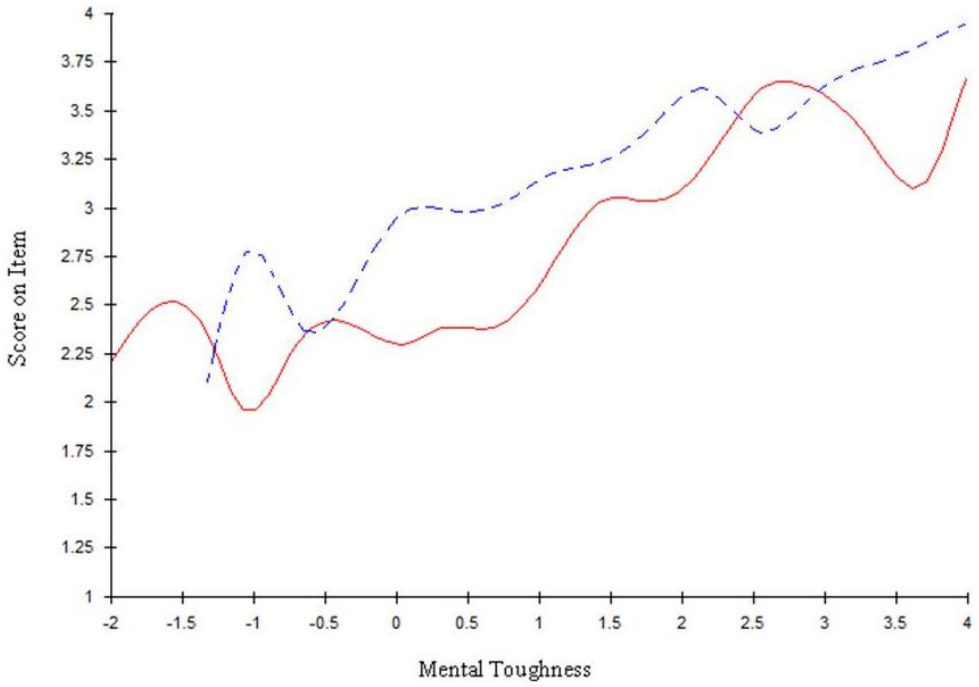



Note: Solid line: Athlete group; Dash line: Non-Athlete group.

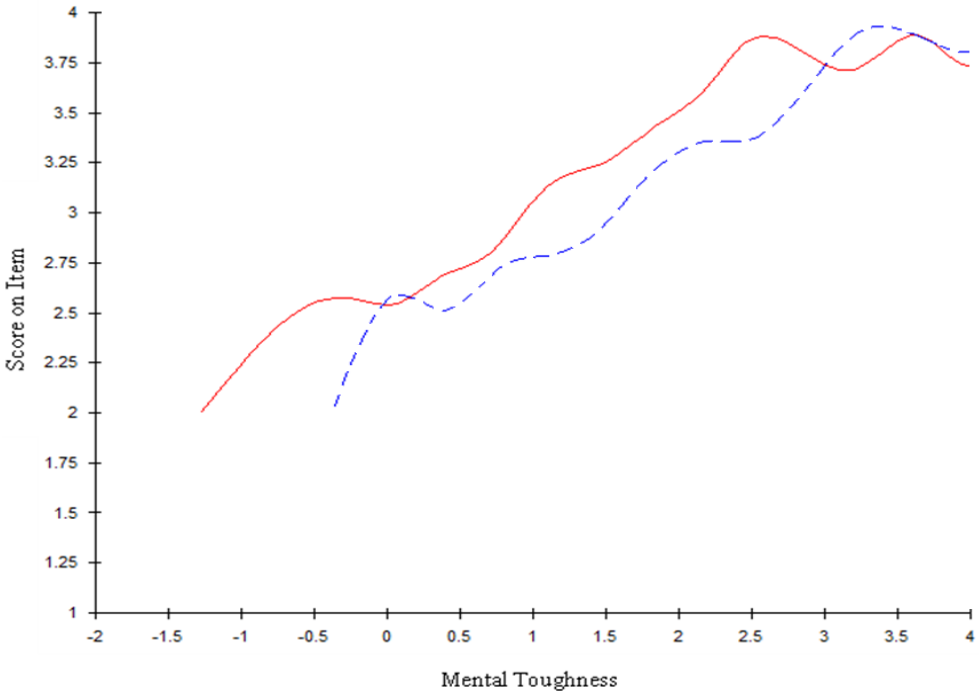
This is an electronic version of an article published in *Measurement in Physical Education and Exercise Science*, Volume 16, Issue 3, 2012.
*Measurement in Physical Education and Exercise Science* is available online at: http://www.tandfonline.com.
DOI: 10.1080/1091367X.2012.693349

Figure 2

a. Item 2




b. Item 17




Note: Solid line: Female group; Dash line: Male group.