Abstract

Objectives: Body composition (BC) assessment typically requires the administration of a single test and can have different evaluation outcomes depending on the selected test and the specific population. The purpose of this study was twofold. Firstly, to develop and validate a novel continuous body composition (CBC) score using the continuous response model (CRM). Secondly, to examine the relationship between CBC scores and fitness performance.

Methods: Data from the 2012 NHANES National Youth Fitness Survey (NNYFS) were used and consisted of n=212 adolescent boys 12–15 years of age. CBC scale variables included body mass (BM), body mass index (BMI), arm circumference (AC), waist circumference (WC), calf circumference (CC), calf skinfold (CSF), triceps skinfold (TSF), and subscapular skinfold (SSF). Fitness performance variables included cardiorespiratory fitness (CRF, mL/kg/min), leg strength (LS, lb), modified pull-ups (MPU, #), grip strength (GS, kg), and plank (PL, sec). Samejima’s CRM, factor analysis, convergent validity coefficients and score reliability were used to validate the CBC scale. Multinomial logistic regression and multiple linear regression were used to examine the relationship between CBC scores and fitness performance variables.

Results: Factor analysis of the CBC scale variables retained a single factor (loadings > 0.81, 88% explained variance) with strong internal consistency (α=0.96). The CRM analysis indicated all CBC scale variables fit a unidimensional construct with adequate discrimination (α: 0.71–2.16) and difficulty (βs: −0.04–1.44). CBC scores (Mean=0, SD=1.00) displayed strong reliability (SEE.θ =0.22, r.θ =0.95) with lower values representing smaller-more-lean individuals and higher values representing larger-less-lean individuals. All fully adjusted regression models showed significant (ps<0.05) negative relationships between CBC scores and CRF, MPU, and PL and positive relationships between CBC scores and LS and GS.

Conclusion: The CRM-derived CBC score is a novel measure of BC and found to be positively associated with strength performance and negatively associated with endurance performance in U.S. adolescent boys.

Keywords: adolescent fitness; body composition; continuous response model (CRM); item response theory (IRT).

Introduction

Body composition (BC) concerns the chemical components of the body and is most commonly assessed under the two-compartment model of fat and fat-free mass [1]. Several laboratory-based techniques provide valid measures of percent body fat (PBF), such as hydrostatic weighing, air displacement plethysmography, and dual-energy X-ray absorptiometry (DXA) [2]. However, these said techniques are costly, require extensive clinician training, and can be burdensome to participants. BC assessment using field-based techniques can address these shortcomings and include a range of measures such as body mass index (BMI), body girth, as well as PBF via skinfold technique, body girth, and bioelectrical impedance analysis (BIA) [3]. Although field-based techniques add convenience to the assessment of BC, especially when large numbers of participants are considered, their ability to measure consistently is hindered by considerable error [4]. Therefore, an alternative BC assessment that can be conveniently administered to a relatively large number of individuals while providing adequate reliability would be valuable to both researchers and health practitioners.

One way to design such a BC assessment is to consider BC a discernible latent construct that can be measured using several observed tests or items. There are numerous psychometric advantages to measuring a construct using multiple items over a single item. Firstly, multi-item scales allow for the empirical assessment of internal consistency of the scales [5]. Secondly, multi-item scales allow for the canceling regression models showed significant (ps<0.05) negative relationships between CBC scores and CRF, MPU, and PL and positive relationships between CBC scores and LS and GS.
out (averaging) of random measurement error, providing increased reliability [6]. Thirdly, multi-item scales allow for greater ability to separate individuals across the construct spectrum, providing greater discrimination and measured information [7]. Fourthly, multi-item scales can include items that target a broader range of a trait, providing greater construct validity [8, 9]. Much of multi-item measurement theory has been investigated in the context of written aptitude tests and/or self-reported attitudinal scales with dichotomous and polytomous responses. However, physical traits, such as BC, can also be evaluated using multiple items but which are each measured on a continuous scale. The added advantage of multi-items in this context is the ability to create a BC score that provides for much richer psychometric properties than any one single BC test.

The primary purpose of this study was to develop and validate a novel continuous body composition (CBC) score using item response theory (IRT) for continuous response items, specifically, the continuous response model (CRM). The secondary purpose of this research was to examine the relationship between CRM-derived CBC scores and fitness performance in U.S. adolescent boys.

Materials and methods

Study procedures

Data for this research came from the 2012 National Health and Nutrition Examination Survey’s (NHANES) National Youth Fitness Survey (NNYFS). The purpose of the 2012 NNYFS was to assess both PA and physical fitness levels in U.S. youth aged 3–15 years [10]. The NNYFS design employed a four-stage probability sample of noninstitutionalized civilian U.S. residents and included 1,640 youth who were interviewed and 1,576 who were examined. NNYFS data are available to the public and organized into categories labeled: Demographics, Dietary, Examination, Questionnaire, and Limited Access. For the current study, Demographic and Examination data only were used. Due to the sex differences in many BC variables, this study was delimited to adolescent boys aged 12–15 years.

Body composition (BC) variables

Eight BC variables were used in this study and each assessed by trained medical personnel using standardized methods [11]. Detailed protocols are explained elsewhere but briefly described here. BC variables included body mass (BM), BMI, arm circumference (AC), waist circumference (WC), calf circumference (CC), calf skinfold (CSF), triceps skinfold (TSF), and subscapular skinfold (SSF). BM was measured in kilograms (kg) using a portable floor scale. BMI was assessed using both BM and standing height (by wall stadiometer) and computed as kg/m². AC was measured in centimeters (cm) and taken at the midpoint of the right upper arm in a relaxed state. WC was measured in cm at a horizontal plane (judged using a mirror) just above the iliac crest. CC was measured in cm at the maximal circumference with participant sitting down. CSF was measured in millimeters (mm) on the inside (medial side) of the lower right leg at the level of the CC point. TSF was measured in mm at the posterior surface midpoint mark of the upper arm. SSF was measured in mm at the inferior angle of the right scapula. All skinfold measurement protocols required a double thickness of skin with underlying adipose tissue.

Fitness performance variables

Five fitness performance variables were used in this study and similarly assessed by trained health professionals using standardized protocols [12]. Briefly, fitness performance variables included cardiorespiratory fitness (CRF), leg strength (LS), modified pull-ups (MPU), grip strength (GS), and plank (PL). CRF was measured using one of five submaximal exercise treadmill protocols varying in speed and grade [13]. Participants were assigned to a specific four-stage protocol based on their age, sex, BMI, and self-reported physical activity readiness (PAR) score. Submaximal heart rate and predicted submaximal oxygen consumption (VO₂) during each of the middle two stages were used to estimate participant maximal oxygen consumption (VO₂max) in mL/kg/min. LS was measured using an isometric knee extension procedure with a practitioner grasped hand-held dynamometer placed on the participant’s shin [14]. Participants were tested while seated and strapped to a chair so as to isolate the action of the quadriceps. Maximal knee extension force (in pounds) was measured thrice while alternating each leg and allowing a 1 min rest period between trials. Peak force across all six tests was used as a measure of LS in this study. MPU was assessed using a modifiable horizontal bar system positioned 2 inches above the participant’s extended reach while lying on their back [15]. During the MPU test, after grasping the bar using an overhand grip, the participant lifts their body upward off the surface while maintaining floor contact with their heels only. The participant flexes at the elbows until their chest reaches a predetermined height (8 inches below the bar) and then extends the elbows to the starting position, while maintain a straight-bodied form throughout the movement. The maximum number of correctly completed pull-ups was used as the measure of MPU in this study. GS was measured while the participant was standing and holding a handgrip dynamometer [16]. After completing a submaximal practice trial, the participant was instructed to squeeze the handgrip device as hard as they could. Each hand was tested three times, alternating hands with a 1 min rest between trials. The summed force (in kg) of each hand’s maximum score was used as a measure of GS in this study. Finally, PL was assessed using a front plank position with the participant lying face down with only their forearms and toes touching the floor while maintaining a straight back [17]. The isometric position of the front plank is held for as long as possible. The maximum number of seconds the plank was held was used as a measure of PL in this study.

Demographic variables

In order to control for possible demographic confounding, age, race, and income were used in this study. Age was used as a numeric covariate ranging from 12 to 15 years. Race was used as a categorical covariate and comprised the following four groups: (1) Non-Hispanic White, (2) Non-Hispanic Black, (3) Mexican/Hispanic, and (4) Other Races and Multi-racial. Finally, income was used as a numeric
covariate, collected as family income, and comprised 12 different income brackets ranging from $0–$4,999 to $12–$100,000 and over.

Continuous response model (CRM)

Item response theory (IRT) is a modern approach to scale validation that entails modeling the relationship between observed responses to items and the latent trait purported to be measured by those items [18]. There are many benefits to IRT over traditional methods and can be found elsewhere [19]. Briefly, with the use of IRT: (1) models can be tested for appropriate fit to data, (2) item parameters are considered invariant to changes in persons (i.e., regardless of sample characteristics), and (3) person parameters are considered invariant to changes in the test (i.e., easy vs. difficult tests). Additionally, an IRT model can be depicted as a monotonic plot of the probability of endorsing an item in relation to a person’s ability [20]. Most IRT models used by researchers are for either dichotomous [21] or polytomial [22] response items [23]. Despite being neglected by researchers, IRT may also be employed with continuous response items. Specifically, Samejima introduced the continuous response model (CRM) as a limiting form of the graded response model (GRM) [24, 25]. The three parameters CRM used in this study [26] is specified for a single person and single item as

\[ P(X \geq x|\theta, a, b, a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\theta} e^{-t^2/2} dt \]

where

\[ v = a^2 \theta - b - \frac{1}{a} \ln \left( \frac{x}{k-x} \right) \]

and \( a \) denotes the discrimination parameter, \( b \) the difficulty parameter, and \( a \) the scaling parameter linking the original response scale to the latent trait scale (\( \theta \)). Similar to the 2-parameter logistic (2 PL) IRT model [27], \( a \) parameter represents the steepness of the item’s probability curve. In the context of polytomial response items, these probability curves are called category characteristic curves (CCC) or operating characteristic curves (OCC) if category boundaries are modeled in cumulative fashion. As well, the \( b \) parameter represents the location of the OCC curves on the \( \theta \) scale. Unique to the CRM, however, is the \( a \) parameter which represents the distance between an item’s OCC curves. Also unique to the CRM is that regardless of the item’s scale range (0–\( k \)), only three item parameters are estimated (i.e., \( a, b, \) and \( a \)). Whereas, the GRM estimates a step parameter for each response category of a polytomial item. In summary, the above CRM represents the probability of a respondent with a specific \( \theta \) obtaining a score of \( x \) or higher on a particular item with a continuous measurement scale ranging from 0 to \( k \), given the parameters \( a, b, \) and \( a \).

Statistical analyses

The statistical analysis plan was separated into two phases. Phase I concerned the development and validation of the continuous body composition (CBC) score. Phase II concerned examining the relationship between CBC scores and fitness performance variables. For phase I, exploratory data analysis was conducted and descriptive statistics with bivariate correlation coefficients computed for the eight BC scale variables. During this step, three cases were identified as multivariate outliers, with unusually large Mahalanobis Distance values (chi-square \( p<0.001 \)) [28] and were removed from the analysis. Factor analysis was also performed to ensure the BC variables measured a unidimensional trait. The eigenvalue greater than 1.00 criteria was used to retain factors [29, 30]. Additionally, item-test correlations, Cronbach alpha, and alpha with item deleted were used to examine validity of scale items [31]. The final step in phase I involved fitting the CRM to the data where item parameters, test information, standard error of estimate, and score reliability were examined along with three-dimensional CCCs [32, 33]. Scores were outputted from the descriptive statistics (standardized sum scores), factor analysis (factor scores), and CRM (CBC scores) steps. For phase II, exploratory data analysis was performed and descriptive statistics computed for the five physical fitness variables. Multinomial logistic regression was run to estimate the fitness-related odds of being in the lowest CBC tertile relative to middle and then highest CBC tertiles [34]. Finally, multiple linear regression was run to examine the relationship between fitness performance variables and continuous CBC scores [35]. Analyses were weighted to produce generalizations representative of noninstitutionalized U.S. boys aged 12–15 years [36, 37]. SAS version 9.4, SPSS version 26, and R version 3.5 were used for all analyses [38–40].

Table 1: Continuous body composition (CBC) scale variable statistics and bivariate correlations, 2012 NNYFS boys 12–15 years of age.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unweighted</th>
<th>Weighted</th>
<th>Weighted correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SE</td>
<td>Mean</td>
</tr>
<tr>
<td>BM (kg)</td>
<td>59.43</td>
<td>1.05</td>
<td>58.99</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>21.69</td>
<td>0.30</td>
<td>21.53</td>
</tr>
<tr>
<td>AC (cm)</td>
<td>27.09</td>
<td>0.29</td>
<td>27.00</td>
</tr>
<tr>
<td>WC (cm)</td>
<td>77.13</td>
<td>0.81</td>
<td>77.01</td>
</tr>
<tr>
<td>CC (cm)</td>
<td>35.06</td>
<td>0.26</td>
<td>34.95</td>
</tr>
<tr>
<td>CSF (mm)</td>
<td>14.29</td>
<td>0.51</td>
<td>14.41</td>
</tr>
<tr>
<td>TSF (mm)</td>
<td>13.52</td>
<td>0.44</td>
<td>13.70</td>
</tr>
<tr>
<td>SSF (mm)</td>
<td>11.92</td>
<td>0.47</td>
<td>11.84</td>
</tr>
</tbody>
</table>

Note. \( n=212. \) SE is standard error. All correlation coefficients are significant \((p<0.05)\). Upper diagonal are unweighted correlations. Lower diagonal are weighted correlations. BM is body mass in kg. BMI is body mass index in kg/m². AC is arm circumference in cm. WC is waist circumference in cm. CC is maximal calf circumference in cm. CSF is calf skinfold in mm. TSF is triceps skinfold in mm. SSF is subscapular skinfold in mm.
Table 2: Factor loadings, eigenvalues and scale reliability coefficients for CBC scale items, 2012 NNYFS boys 12–15 years of age.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unweighted</th>
<th></th>
<th></th>
<th>Weighted</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>loading</td>
<td>$r_{total}$</td>
<td>$a_{del}$</td>
<td>$\alpha$</td>
<td>loading</td>
<td>$r_{total}$</td>
</tr>
<tr>
<td>BM</td>
<td>0.898</td>
<td>0.861</td>
<td>0.965</td>
<td>0.968</td>
<td>0.870</td>
<td>0.846</td>
</tr>
<tr>
<td>BMI</td>
<td>0.967</td>
<td>0.954</td>
<td>0.959</td>
<td></td>
<td>0.925</td>
<td>0.949</td>
</tr>
<tr>
<td>AC</td>
<td>0.939</td>
<td>0.920</td>
<td>0.961</td>
<td></td>
<td>0.923</td>
<td>0.912</td>
</tr>
<tr>
<td>WC</td>
<td>0.955</td>
<td>0.941</td>
<td>0.960</td>
<td></td>
<td>0.941</td>
<td>0.936</td>
</tr>
<tr>
<td>CC</td>
<td>0.881</td>
<td>0.851</td>
<td>0.965</td>
<td></td>
<td>0.858</td>
<td>0.847</td>
</tr>
<tr>
<td>CSF</td>
<td>0.835</td>
<td>0.809</td>
<td>0.968</td>
<td></td>
<td>0.817</td>
<td>0.788</td>
</tr>
<tr>
<td>TSF</td>
<td>0.837</td>
<td>0.805</td>
<td>0.968</td>
<td></td>
<td>0.831</td>
<td>0.787</td>
</tr>
<tr>
<td>SSF</td>
<td>0.873</td>
<td>0.856</td>
<td>0.965</td>
<td></td>
<td>0.859</td>
<td>0.849</td>
</tr>
<tr>
<td>Eigenvalue</td>
<td>6.47</td>
<td></td>
<td></td>
<td>6.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% explained</td>
<td>0.89</td>
<td></td>
<td></td>
<td>0.88</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. n=212. One factor was retained by the eigenvalue >1.0 criterion. $r_{total}$ is correlation coefficient representing correlation between respective variable and standardized scale total score. $a_{del}$ is standardized Cronbach alpha with respective variable deleted from the scale. $\alpha$ is overall standardized Cronbach alpha for all scale variables.

Table 3: Continuous response model (CRM) parameter estimates for the CBC scale, 2012 NNYFS boys 12–15 years of age.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Discrimination</th>
<th>Difficulty</th>
<th>Scaling</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$a$</td>
<td>SE</td>
<td>$b$</td>
</tr>
<tr>
<td>BM</td>
<td>2.059</td>
<td>0.101</td>
<td>0.309</td>
</tr>
<tr>
<td>BMI</td>
<td>1.419</td>
<td>0.071</td>
<td>0.777</td>
</tr>
<tr>
<td>AC</td>
<td>1.811</td>
<td>0.089</td>
<td>0.523</td>
</tr>
<tr>
<td>WC</td>
<td>2.162</td>
<td>0.105</td>
<td>0.521</td>
</tr>
<tr>
<td>CC</td>
<td>1.823</td>
<td>0.089</td>
<td>-0.042</td>
</tr>
<tr>
<td>CSF</td>
<td>0.855</td>
<td>0.048</td>
<td>0.944</td>
</tr>
<tr>
<td>TSF</td>
<td>0.705</td>
<td>0.043</td>
<td>1.437</td>
</tr>
<tr>
<td>SSF</td>
<td>1.163</td>
<td>0.060</td>
<td>1.370</td>
</tr>
</tbody>
</table>

Note. $a$ is discrimination parameter. $b$ is difficulty parameter. $\alpha$ is the scaling parameter. SE is standard error. All parameter estimates are significant ($p<0.05$).

Results

A total of 212 adolescent boys (Mean=13.4, $SE=0.08$ years of age) had complete BC data and were used in phase I of the study. Table 1 contains descriptive statistics and bivariate correlations for the CBC scale variables. Interestingly, SSF (Mean=11.8, $SE=0.49$ kg, $CV=4.1\%$), CSF (Mean=14.4, $SE=0.47$ kg, $CV=3.2\%$), and TSF (Mean=13.7, $SE=0.38$ kg, $CV=2.8\%$) showed the greatest amount of variability in adolescent boys with all other scale variables showing relatively less variation. Additionally, all correlation coefficients were significant ($p<0.05$) and positive and at least moderate strength ($r>0.53$). Table 2 contains construct validity and internal consistency evidence for the CBC scale items. A single BC factor was retained using the Kaiser criterion with 88% explained variance and all loadings greater than 0.81. Additionally, overall alpha ($\alpha=0.96$) indicated strong reliability with all items contributing to internal consistency ($r_{totals}>0.78$ & $a_{deles}=\alpha$).

Table 3 contains parameter estimates resulting from the CRM analysis. All CBC scale variables fit a unidimensional construct as noted by significant parameter estimates showing adequate discrimination ($as: 0.71–2.16$) and moderate difficulty coverage ($bs: -0.04–1.44$). WC and TSF were the most and least discriminating CBC scale items, respectively. Additionally, TSF was the most difficult CBC scale item, indicating a higher BC trait required in order to receive higher TSF values. Figure 1 displays three-dimensional CCCs for each CBC scale item. For any CCC, proper item functioning would be characterized as category curves peaking (indicating greatest probability) only as moving from left to right on the ability (BC trait) axis. Therefore, a proper three-dimensional CCC would see the peaks of each curve moving along a diagonal line from the upper left corner (low ability and high response scale) to the lower right corner (high ability and high response scale). Additionally, an item’s difficulty ($b$) and discrimination ($a$) can be inspected in the CCCs with the center of the curves representing the item’s $b$ and the spread of the curves relative to the response scale representing its $a$. Each CCC comprising the CBC scale indicated proper item functioning.

Table 4 contains additional CBC scale validity evidence with bivariate correlations between the new CRM-derived CBC scores (CBC), CBC scale item factor scores (SFS), and standardized scale sum scores (SSS) [40]. All correlation coefficients were strong ($r>0.96$) and significant ($p<0.05$). In particular, the strong correlation between CBC and SSS presents additional evidence for the use of simple SSS in BC
assessment. Table 5 contains descriptive statistics and reliability estimates for the new CBC score. CBC score values resemble $Z$-scores (Mean=0, SD=1.00) ranging from $-5.2$ to $4.63$. Additionally, score reliability was adequate with low error ($\text{SEE} \theta = 0.22$) and strong marginal reliability ($r \theta = 0.95$). CBC score values were conceptualized to represent an adolescent's body size and leanness. Therefore, lower values represent smaller-more-lean individuals and higher values represent larger-less-lean individuals.

Table 6 contains descriptive statistics for the fitness-related predictor variables. MPU (Mean=10.7, $SE=0.89$ kg, $CV=8.3\%$) displayed the greatest amount of variability in adolescent boys. Table 7 contains results from the multiple regression modeling using fitness variables to predict the continuous form of CBC scores. Models 1 through 5 are

Figure 1: Continuous response model (CRM) three-dimensional CCCs for the CBC scale, 2012 NNYFS boys 12–15 years of age. Item 1 = BM, Item 2 = BMI, Item 3 = AC, Item 4 = WC, Item 5 = CC, Item 6 = CSF, Item 7 = TSF, Item 8 = SSF.
Table 4: Correlations between new CRM-derived CBC scores, CBC scale factor scores, and CBC scale sum scores.

<table>
<thead>
<tr>
<th>Score</th>
<th>SFS</th>
<th>SSS</th>
<th>CBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>SFS</td>
<td>1</td>
<td>0.999</td>
<td>0.968</td>
</tr>
<tr>
<td>SSS</td>
<td>0.999</td>
<td>1</td>
<td>0.966</td>
</tr>
<tr>
<td>CBC</td>
<td>0.962</td>
<td>0.960</td>
<td>1</td>
</tr>
</tbody>
</table>

Note. CBC are new CRM-derived CBC scores. SFS are scale factor scores. SSS are scale (standardized) sum scores. All correlation coefficients are significant (p<0.05). Upper diagonal are unweighted correlations. Lower diagonal are weighted correlations.

single fitness predictor models, adjusted for age, race, and income. These initial models each saw fitness values significantly (p<0.05) predict CBC scores, with GS (β=0.54, p<0.0001, $R^2=0.30$) showing the largest explained variance. Additionally, CRF, MPU, and PL each displayed a negative relationship with CBC. Whereas, LS and GS displayed a positive relationship with CBC. Model 6 is a fully adjusted model that included all five fitness predictors. Model 6 saw two fitness predictors (CRF and PL) drop below the level of significance (p>0.05). In model 7, after removing PL from the previous analysis, the remaining four fitness predictors significantly (p<0.053) predicted CBC scores with large explained variance (p<0.0001, $R^2=0.45$).

Table 8 contains results from the multinomial logistic regression analyses. Tabled values represent the fitness-related odds of being in the lowest CBC tertile relative to middle tertile and then relative to the highest CBC tertile. Models 1 through 5 are again single fitness predictor models, adjusted for age, race, and income. These initial models for both sets of logistic regressions each saw fitness values significantly (p<0.05) predict CBC tertile membership, with GS showing the greatest predictive power for the lowest CBC tertile vs. the middle CBC tertile (OR=0.82, CI: 0.77–0.87, $R^2_{RS}=0.32$) and the lowest CBC tertile vs. the highest CBC tertile (OR=0.80, CI: 0.73–0.87, $R^2_{RS}=0.47$). Similar to the regression results, CRF, MPU, and PL each displayed a negative relationship and LS and GS each displayed a positive relationship with CBC tertile placement. Model 6 for both sets of logistic regressions each saw significant predicted CBC tertile placement with large explained variance (p<0.001, $R^2_{RS}=0.85$).

Discussion

The primary purpose of this study was to develop and validate a CBC score using the IRT-based CRM. Results clearly showed that eight individual anthropometric items represent a unidimensional BC trait. With the eight-item BC scale displaying adequate construct validity and internal consistency using classical test theory methods. Furthermore, scale items fit the CRM, indicating adequate item functioning and reliable BC measurement. These results are the first to present the development and validation of a multi-item BC assessment in adolescent boys. As previously mentioned, there are many advantages in using a multi-item assessment to measure a relatively complex trait. In context with this study’s results, these advantages include (1) an improved association between the CBC scores and the latent BC trait, (2) an increased ability to measure individuals across a wider BC trait spectrum, (3) a reduced measurement error associated with BC assessment, and (4) an increased amount of measured information that allows for greater categorization of individuals [41]. The need for these improved measurement properties can be underscored by examining the literature for consistency among the commonly used BC assessments. For example, a recent study using over 700 adolescents, compared the use of BMI and percent body fat (criterion measure) in classifying individuals into overweight and obesity categories [42]. Results of this study showed that the BMI-derived overweight and obesity classification underestimated the prevalence in both categories, as compared to the PBF-derived classification estimates. These results highlight that two of the more common forms of BC assessment do not agree in terms of evaluating

Table 5: Summary of CBC scores (theta) and score reliability from CRM.

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Median</th>
<th>Mean</th>
<th>Max</th>
<th>SD</th>
<th>I</th>
<th>SEE.θ</th>
<th>r.θ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theta</td>
<td>−5.20</td>
<td>−0.07</td>
<td>0.00</td>
<td>4.63</td>
<td>1.00</td>
<td>20.11</td>
<td>0.22</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Note. Theta values resemble Z-scores. I is total test information. SEE.θ is standard error of estimated theta scores. r.θ is average theta reliability.
adolescents. Several other studies support the same
divergence between BC assessments in adolescents [43–
47]. Consequently, the development and validation of
the CBC scale provides a more psychometrically sound
assessment of overall BC in adolescent boys.

The secondary purpose of this research was to examine
the relationship between CRM-derived CBC scores and
fitness performance in U.S. adolescent boys. Results
showed unequivocally that CBC is independently related
to cardiorespiratory, muscular endurance, and muscular
strength fitness components. Specifically, negative re-
lationships were found between CBC and cardiorespiratory
performance and muscular endurance performance.
Conversely, positive relationships were seen between CBC
and muscular strength performance. These findings are
consistent with findings using single-item BC measures.
For example, a large population-based fitness study of
Latin-American adolescents examined the relationship
between cardiorespiratory performance scores and BC [48].
Results from this research showed that negative
associations were found between peak oxygen uptake and
measures of BMI, WC, and weight-to-height ratio in
adolescent boys. Several other studies support this indirect
CRF and BC relationship in adolescents [49–53]. Addition-
ally, a recent study examined the associations between
BMI as a measure of BC and muscular strength in adoles-
cents with obesity [54]. Results from this research indicated
that BMI in adolescents was positively associated with
muscular strength, as measured by eight-repetition
maximum bench press and leg press tests. Other studies
converge around these positive associations of muscular
strength and BC results [55–57]. Several studies also sup-
port the negative association between muscular endurance
and BC [58–60]. Altogether, the findings from the second
part of this study serve as additional validity evidence for
the CBC scale, given its ability to detect known fitness
performance relationships in adolescent boys.

There are some future research suggestions for the CBC
scale worth mentioning. Although internal consistency
and score reliability were both established in this study,
future studies should examine the stability of the CBC scale
over time. Also, a more in-depth construct validity study is
recommended to determine the extent to which CBC scores
can detect differences in certain athletes with known BC
profiles (e.g., wrestlers, cross-country runners, gymnasts,
etc.). Finally, future studies should evaluate a similar CBC
scale in adolescent girls.

One strength regarding this current study is its use of
IRT to validate a novel CBC scale. IRT is increasingly being
applied in the health sciences to assess the measurement
properties of latent behavioral, attitudinal, and patient-
reported outcome scales [61, 62]. The use of this modern
psychometric analysis to validate the CBC scale, and spec-
ifically the use of CRM, sets a precedent for the application

Table 6: Descriptive statistics for study fitness variables, 2012
NNYFS boys 12–15 years of age.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Min</th>
<th>Max</th>
<th>Median</th>
<th>Mean</th>
<th>SE</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRF (mL/kg/min)</td>
<td>194</td>
<td>28.89</td>
<td>92.60</td>
<td>42.93</td>
<td>44.17</td>
<td>0.94</td>
<td>2.1</td>
</tr>
<tr>
<td>LS (lb)</td>
<td>208</td>
<td>19.30</td>
<td>238.50</td>
<td>87.11</td>
<td>88.90</td>
<td>6.15</td>
<td>6.9</td>
</tr>
<tr>
<td>MPU (#)</td>
<td>207</td>
<td>0.00</td>
<td>30.00</td>
<td>9.42</td>
<td>10.73</td>
<td>0.89</td>
<td>8.3</td>
</tr>
<tr>
<td>GS (kg)</td>
<td>208</td>
<td>24.30</td>
<td>108.70</td>
<td>63.87</td>
<td>65.55</td>
<td>0.87</td>
<td>1.3</td>
</tr>
<tr>
<td>PL (sec)</td>
<td>209</td>
<td>5.00</td>
<td>450.00</td>
<td>86.07</td>
<td>93.24</td>
<td>5.58</td>
<td>6.0</td>
</tr>
</tbody>
</table>

Note. Statistics are weighted. CRF is cardiorespiratory fitness in mL/
kg/min. LS is leg strength in pounds. MPU is maximum number of pull-
ups. GS is grip strength in kg. PL is plank in seconds. N is sample size.
Min is minimum value. SE is standard error. Max is maximum value. CV is
coefficient of variation.

Table 7: Regression estimates for new CBC scores (theta) regressed on fitness variables, 2012 NNYFS boys 12–15 years of age.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Models 1 through 5</th>
<th>Model 6</th>
<th>Model 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRF (mL/kg/min)</td>
<td>β</td>
<td>p</td>
<td>R²</td>
</tr>
<tr>
<td>190</td>
<td>−0.150</td>
<td>0.039</td>
<td>0.14</td>
</tr>
<tr>
<td>LS (lb)</td>
<td>0.23</td>
<td>0.005</td>
<td>0.24</td>
</tr>
<tr>
<td>MPU (#)</td>
<td>−0.0288</td>
<td>&lt;0.0001</td>
<td>0.19</td>
</tr>
<tr>
<td>GS (kg)</td>
<td>0.544</td>
<td>&lt;0.0001</td>
<td>0.30</td>
</tr>
<tr>
<td>PL (sec)</td>
<td>−0.307</td>
<td>0.019</td>
<td>0.19</td>
</tr>
<tr>
<td>R²</td>
<td>0.46</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>189</td>
<td>189</td>
<td></td>
</tr>
</tbody>
</table>

Note. Theta values resemble Z-scores where lower values represent smaller-more-lean individuals and higher values represent larger-less-lean
individuals. CRF is cardiorespiratory fitness in mL/kg/min. LS is leg strength in pounds. MPU is maximum number of pull-ups. GS is grip strength
in kg. PL is plank in seconds. All models are adjusted for age, race, and income. β is standardized slope coefficient. R² is model coefficient of
determination. N is sample size. Models 1 through 5 are single fitness predictor models. Model 6 includes all five fitness predictor variables.
Model 7 includes only contributing fitness predictors.
Table 8: Multinomial logistic regression fitness-related odds of being in lowest CBC tertile relative to middle and highest CBC tertiles, 2012 NNYFS boys 12–15 years of age.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>1st CBC tertile vs. 2nd CBC tertile</th>
<th>1st CBC tertile vs. 3rd CBC tertile</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Models 1 through 5</td>
<td>Model 6</td>
</tr>
<tr>
<td>N</td>
<td>OR</td>
<td>95% CI</td>
</tr>
<tr>
<td>CRF (mL/kg/min)</td>
<td>129</td>
<td>1.09</td>
</tr>
<tr>
<td>LS (lb)</td>
<td>134</td>
<td>0.98</td>
</tr>
<tr>
<td>MPU (#)</td>
<td>135</td>
<td>1.05</td>
</tr>
<tr>
<td>GS (kg)</td>
<td>135</td>
<td>0.82</td>
</tr>
<tr>
<td>PL (sec)</td>
<td>136</td>
<td>1.01</td>
</tr>
<tr>
<td>R²_RS</td>
<td>0.61</td>
<td>0.60</td>
</tr>
<tr>
<td>N</td>
<td>126</td>
<td>127</td>
</tr>
</tbody>
</table>

Note. CRF is cardiorespiratory fitness in mL/kg/min. LS is leg strength in pounds. MPU is maximum number of pull-ups. GS is grip strength in kg. PL is plank in seconds. All models are adjusted for age, race, and income. Models 1 through 5 are single fitness predictor models. Model 6 includes all five fitness predictor variables. Model 7 includes only contributing fitness predictors. Lowest (1st) CBC tertile represents smaller-more-lean individuals. Highest (3rd) CBC tertile represents larger-less-lean individuals. R²_RS is a rescaled version for logistic regression and useful for comparing models.
of multi-item scales to assess a unidimensional BC trait. Another strength of this study was its use of a nationally representative sample of U.S. boys ages 12–15 years, increasing this study’s external validity. A final strength worth mentioning is the use of objectively measured BC and objectively measured physical fitness measures assessed by trained medical professionals, distinguishing these findings from other studies utilizing self-reported measures and non-standardized methods. Despite these strengths, there are some limitations worth declaring. The NNYFS is not of continuous or longitudinal nature and therefore the results of this cross-sectional study should only be considered as correlational. Another limitation of this study was its inability to control for puberty. It is possible that some adolescents of the same BC profile (response pattern) differed in fitness characteristics due to differences in natural hormones. Therefore, owing to these limitations, study results should be considered with caution.

Conclusions

This study presents development and validation evidence for a multi-item BC scale, resulting in a novel CBC score for adolescent boys. The psychometric evidence supports the use of simple standardized sum scores, across individual BC assessments, as a sufficient measure of CBC. Moreover, CBC scores were found to be positively associated with strength performance and negatively associated with endurance performance in U.S. adolescent boys. Health promotion specialists should be aware of the advantages of using multi-item scales to assess BC for evaluation.

Research funding: None declared.

Author contribution: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Competing interests: The authors declare no conflicts of interest regarding this article.

References


