Peter Hart*

Relationship between fitness performance and a newly developed continuous body composition score in U.S. adolescent boys

<https://doi.org/10.1515/ijamh-2020-0198>

Received August 3, 2020; accepted August 30, 2020; published online September 23, 2020

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 ($\alpha=0.96$). The CRM analysis indicated all CBC scale variables fit a unidimensional construct with adequate discrimination ($as: 0.71-2.16$) and difficulty ($bs: -0.04-1.44$). CBC scores (Mean=0, $SD=1.00$) displayed strong reliability ($SEE.\theta=0.22$, $r.\theta=0.95$) 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

*Corresponding author: Peter Hart PhD, Health Promotion Research, 930 4th Ave, Havre, 59501, MT, USA, E-mail: pdhart@outlook.com

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/m^2 . 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 cardio-respiratory 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 (VO_2) during each of the middle two stages were used to estimate participant maximal oxygen consumption ($\text{VO}_{2\text{max}}$) in $\text{mL}/\text{kg}/\text{min}$. 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 1=\$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 polytomous [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

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

where

$$v = a\left(\theta - b - \frac{1}{\alpha} \ln \frac{x}{k-x}\right)$$

and a denotes the discrimination parameter, b the difficulty parameter, and α the scaling parameter linking the original response scale to the latent trait scale (θ). 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 polytomous 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 θ scale. Unique to the CRM,

however, is the α 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 α). Whereas, the GRM estimates a step parameter for each response category of a polytomous item. In summary, the above CRM represents the probability of a respondent with a specific θ 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 α .

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 $ps < 0.001$) [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.

Variable	Unweighted		Weighted			Weighted\Unweighted correlations							
	Mean	SE	Mean	SE	CV	BM	BMI	AC	WC	CC	CSF	TSF	SSF
BM (kg)	59.43	1.05	58.99	1.23	2.1	1	0.890	0.932	0.887	0.914	0.591	0.573	0.688
BMI (kg/m ²)	21.69	0.30	21.53	0.29	1.3	0.880	1	0.933	0.939	0.882	0.759	0.746	0.836
AC (cm)	27.09	0.29	27.00	0.29	1.1	0.924	0.927	1	0.890	0.905	0.693	0.691	0.758
WC (cm)	77.13	0.81	77.01	0.91	1.2	0.885	0.938	0.883	1	0.819	0.765	0.779	0.839
CC (cm)	35.06	0.26	34.95	0.24	0.7	0.908	0.879	0.900	0.819	1	0.643	0.601	0.653
CSF (mm)	14.29	0.51	14.41	0.47	3.2	0.545	0.731	0.663	0.735	0.617	1	0.912	0.818
TSF (mm)	13.52	0.44	13.70	0.38	2.8	0.532	0.723	0.662	0.747	0.582	0.909	1	0.856
SSF (mm)	11.92	0.47	11.84	0.49	4.1	0.662	0.824	0.741	0.825	0.640	0.815	0.849	1

Note. $n=212$. SE is standard error. All correlation coefficients are significant ($ps < 0.05$). 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. TSK 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.

Variable	Unweighted				Weighted			
	loading	r_{Total}	α_{del}	α	loading	r_{Total}	α_{del}	α
BM	0.898	0.861	0.965	0.968	0.870	0.846	0.962	0.965
BMI	0.967	0.954	0.959		0.925	0.949	0.956	
AC	0.939	0.920	0.961		0.923	0.912	0.958	
WC	0.955	0.941	0.960		0.941	0.936	0.956	
CC	0.881	0.851	0.965		0.858	0.847	0.961	
CSF	0.835	0.809	0.968		0.817	0.788	0.965	
TSF	0.837	0.805	0.968		0.831	0.787	0.965	
SSF	0.873	0.856	0.965		0.859	0.849	0.961	
Eigenvalue	6.47				6.36			
% explained	0.89				0.88			

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. α_{del} is standardized Cronbach alpha with respective variable deleted from the scale. α 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.

Variable	Discrimination		Difficulty		Scaling	
	a	SE	b	SE	α	SE
BM	2.059	0.101	0.309	0.034	1.178	0.024
BMI	1.419	0.071	0.777	0.055	1.060	0.033
AC	1.811	0.089	0.523	0.041	1.059	0.029
WC	2.162	0.105	0.521	0.035	1.230	0.022
CC	1.823	0.089	-0.042	0.037	0.988	0.030
CSF	0.855	0.048	0.944	0.092	1.059	0.046
TSF	0.705	0.043	1.437	0.123	0.847	0.063
SSF	1.163	0.060	1.370	0.081	1.087	0.037

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

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 ($ps < 0.05$) and positive and at least moderate strength ($rs > 0.53$). 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 ($r_{totals} > 0.78$ & $\alpha_{dels} \leq \alpha$). 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 low 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 ($rs > 0.96$) and significant ($ps < 0.05$). In particular, the strong correlation between CBC and SSS presents additional evidence for the use of simple SSS in BC

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.

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 – 4.63 . Additionally, score reliability was adequate with low error ($SEE.\theta=0.22$) 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

Table 4: Correlations between new CRM-derived CBC scores, CBC scale factor scores, and CBC scale sum scores.

Score	SFS	SSS	CBC
SFS	1	0.999	0.968
SSS	0.999	1	0.966
CBC	0.962	0.960	1

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 ($\beta = 0.54$, $p < 0.0001$, $R^2 = 0.30$) 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 saw a single fitness predictor fail to significantly contribute to the model. Therefore, in model 7, after removing LS from the

lowest CBC tertile vs. middle CBC tertile analysis, the remaining four fitness predictors significantly predicted CBC tertile placement with large explained variance ($p < 0.001$, $R^2_{RS} = 0.60$). Additionally, after removing PL from the lowest CBC tertile vs. highest CBC tertile analysis, the remaining four fitness predictors significantly 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.

	Min	Median	Mean	Max	SD	I	SEE. θ	r. θ
Theta	-5.20	-0.07	0.00	4.63	1.00	20.11	0.22	0.95

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.

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

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

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.

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 relationships 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]. Additionally, a recent study examined the associations between BMI as a measure of BC and muscular strength in adolescents 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 support 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 specifically the use of CRM, sets a precedent for the application

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

Predictor	Models 1 through 5				Model 6		Model 7	
	<i>N</i>	β	<i>p</i>	R^2	β	<i>p</i>	β	<i>p</i>
CRF (mL/kg/min)	190	−0.150	0.039	0.14	−0.071	0.127	−0.090	0.052
LS (lb)	204	0.387	0.005	0.24	0.176	0.055	0.188	0.052
MPU (#)	203	−0.288	<0.0001	0.19	−0.335	<0.0001	−0.395	<0.0001
GS (kg)	205	0.544	<0.0001	0.30	0.543	<0.0001	0.531	<0.0001
PL (sec)	205	−0.307	0.019	0.19	−0.116	0.247	–	–
R^2					0.46			0.45
<i>N</i>					189			189

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. R^2 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.

Predictor	1st CBC tertile vs. 2nd CBC tertile						1st CBC tertile vs. 3rd CBC tertile					
	Models 1 through 5			Model 6			Models 1 through 5			Model 6		
	<i>N</i>	<i>OR</i>	<i>95% CI</i>	<i>R²_{RS}</i>	<i>OR</i>	<i>95% CI</i>	<i>N</i>	<i>OR</i>	<i>95% CI</i>	<i>R²_{RS}</i>	<i>OR</i>	<i>95% CI</i>
CRF (mL/kg/min)	129	1.09	1.03–1.14	0.18	1.11	1.02–1.21	125	1.08	1.00–1.18	0.27	1.15	1.05–1.28
LS (lb)	134	0.98	0.97–0.99	0.13	0.99	0.96–1.01	136	0.97	0.95–0.99	0.35	0.98	0.97–0.99
MPU (#)	135	1.05	1.01–1.09	0.08	1.14	1.04–1.26	135	1.19	1.11–1.28	0.38	1.54	1.15–2.06
GS (kg)	135	0.82	0.77–0.87	0.32	0.69	0.58–0.82	136	0.80	0.73–0.87	0.47	0.54	0.42–0.69
PL (sec)	136	1.01	1.01–1.02	0.11	1.02	1.01–1.04	136	1.02	1.01–1.03	0.30	1.02	0.99–1.05
<i>R²_{RS}</i>					0.61						0.86	
<i>N</i>					126						123	

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

1. Wilmore JH, Costill DL, Kenny WL. *Physiology of sport and exercise*, 7th ed. Champaign, USA: Human Kinetics Publishers; 2019.
2. Lowry DW, Tomiyama AJ. Air displacement plethysmography versus dual-energy x-ray absorptiometry in underweight, normal-weight, and overweight/obese individuals. *PLoS One* 2015;10: e0115086.
3. McArdle WD, Katch FI, Katch VL. *Essentials of exercise physiology*. Philadelphia, USA: Lippincott Williams & Wilkins; 2006.
4. Orsso CE, Silva MIB, Gonzalez MC, Rubin DA, Heymsfield SB, Prado CM, et al. Assessment of body composition in pediatric overweight and obesity: a systematic review of the reliability and validity of common techniques. *Obes Rev* 2020;21:e13041.
5. Johnson AJ. Reliability, Cronbach's alpha. *The SAGE Encyclopedia of Communication Research Methods*; 2017, 1415–17 pp.
6. DeVellis RF. *Scale development: theory and applications*. Thousand Oaks: Sage Publications, Inc; 2003.
7. Zanon C, Hutz CS, Yoo HH, Hambleton RK. An application of item response theory to psychological test development. *Psicol Reflexão Crítica* 2016;29:1–10.
8. Hoepfner BB, Kelly JF, Urbanoski KA, Slaymaker V. Comparative utility of a single-item versus multiple-item measure of self-efficacy in predicting relapse among young adults. *J Subst Abuse Treat* 2011;41:305–12.
9. Sarstedt M, Wilczynski P. More for less? a comparison of single-item and multi-item measures. *Die Betriebswirtschaft* 2009;69:211.
10. National Center for Health Statistics. National health and nutrition examination survey: National youth fitness survey plan, operations, and analysis; 2012. Available from: http://www.cdc.gov/nchs/data/series/sr_02/sr02_163.pdf [Accessed 8 June 2020].
11. National Health and Nutrition, Examination Survey (NHANES). National youth fitness survey (NYFS) body measures procedures manual. Available from: http://www.cdc.gov/nchs/data/nyyfs/Body_Measures.pdf [Accessed 8 June 2020].
12. National Health and Nutrition, Examination Survey (NHANES). National youth fitness survey (NYFS) mobile center (MC) operations manual. Available from: https://www.cdc.gov/nchs/data/nyyfs/Operations_Manual.pdf [Accessed 8 June 2020].
13. Centers for Disease Control and Prevention. National youth fitness survey (NYFS) treadmill examination manual. Hyattsville, MD: National Center for Health Statistics; 2013. Available from: <http://www.cdc.gov/nchs/data/nyyfs/Treadmill.pdf> [Accessed 8 June 2020].
14. National Health and Nutrition, Examination Survey (NHANES). National youth fitness survey (NYFS) lower body muscle strength component procedures manual. Available from: https://www.cdc.gov/nchs/data/nyyfs/Lower_Body_Muscle_Strength.pdf [Accessed 8 June 2020].
15. National Health and Nutrition, Examination Survey (NHANES). National youth fitness survey (NYFS) modified pull-up exercise procedures manual. Available from: https://www.cdc.gov/nchs/data/nyyfs/Modified_Pullup.pdf [Accessed 8 June 2020].
16. National Health and Nutrition, Examination Survey (NHANES). National youth fitness survey (NYFS) muscle strength (grip) procedures manual. Available from: https://www.cdc.gov/nchs/data/nyyfs/Handgrip_Muscle_Strength.pdf [Accessed 8 June 2020].
17. National Health and Nutrition, Examination Survey (NHANES). National youth fitness survey (NYFS) plank exercise procedures manual. Available from: <https://www.cdc.gov/nchs/data/nyyfs/Plank.pdf> [Accessed 8 June 2020].
18. Edwards MC. An introduction to item response theory using the need for cognition scale. *Soc Personal Psychol Compass* 2009;3: 507–29.

19. Hart PD. Modern psychometric analysis of the muscle strengthening activity scale (MSAS) using item response theory. *Res Psychol Behav Sci* 2019;7:23–33.
20. Nguyen TH, Han HR, Kim MT, Chan KS. An introduction to item response theory for patient-reported outcome measurement. *Patient-Patient-Centered Outcome Res* 2014;7:23–35.
21. Baker FB, Kim SH. The basics of item response theory using R. New York: Springer; 2017, 55–67 pp.
22. Ostini R, Nering ML. Polytomous item response theory models. Cary, USA: Sage; 2006.
23. De Ayala RJ. The theory and practice of item response theory. New York, USA: Guilford Publications; 2013.
24. Samejima F. Homogeneous case of the continuous response model. *Psychometrika* 1973;38:203–19.
25. Samejima F. Graded response models. In *Handbook of item response theory*. Chapman and Hall/CRC; 2016, vol 1.
26. Wang T, Zeng L. Item parameter estimation for a continuous response model using an EM algorithm. *Appl Psychol Meas* 1998; 22:333–44.
27. Hambleton RK, Swaminathan H, Rogers HJ. Fundamentals of item response theory. Thousand Oaks: Sage; 1991.
28. IBM. Compute Mahalanobis Distance and flag multivariate outliers; 2020 April. Available from: <https://www.ibm.com/support/pages/compute-mahalanobis-distance-and-flag-multivariate-outliers> [Accessed 8 June 2020].
29. Tabachnick BG, Fidell LS, Ullman JB. Using multivariate statistics Boston, MA: Pearson; 2007, vol 5, 481–98 pp.
30. Kaiser HF. The application of electronic computers to factor analysis. *Educ Psychol Meas* 1960;20:141–51.
31. Hair JF, Black WC, Babin BJ, Anderson RE, Tatham RL. Multivariate data analysis. Upper Saddle River, NJ: Prentice-Hall; 2012.
32. Zopluoglu C. EstCRM: an R package for Samejima's continuous IRT model. *Appl Psychol Meas* 2012;36:149.
33. Zopluoglu C. EstCRM: Calibrating Parameters for the Samejima's Continuous IRT Model. R package version; 2011, vol 1.
34. Stokes ME, Davis CS, Koch GG. Categorical data analysis using SAS. Cary, USA: SAS institute; 2012.
35. Freund RJ, Littell RC. SAS system for regression. Cary, USA: SAS Publishing; 2000.
36. Johnson CL, Dohrmann SM, Van de Kerckhove W, Borrud LG, Chiappa M, Burt V, et al. National health and nutrition examination survey: National youth fitness survey estimation procedures, 2012. *Vital Health Stat* 2014;2.
37. Lewis TH. Complex survey data analysis with SAS. Boca Raton, USA: CRC Press; 2016.
38. SAS Institute. Base SAS 9.4 procedures guide. Cary, USA: SAS Institute; 2015.
39. IBM Corp. IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp; 2017.
40. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2018. Available from: <https://www.R-project.org/>.
41. Nunnally JC, Bernstein IH. Psychometric theory, 3rd ed. New York: McGraw-Hill; 1994.
42. Karchynskaya V, Kopcakova J, Klein D, Gába A, Madarasova-Geckova A, van Dijk JP, et al. Is BMI a valid indicator of overweight and obesity for adolescents?. *Int J Environ Res Publ Health* 2020; 17:4815.
43. Ripka WL, Orsso CE, Haqq AM, Prado CM, Ulbricht L, Leite N. Validity and accuracy of body fat prediction equations using anthropometrics measurements in adolescents. *Eat Weight Disord-Stud* 2020. <https://doi.org/10.1007/s40519-020-00918-3>. [Online ahead of print].
44. Telford RD, Telford RM, Welvaert M. BMI is a misleading proxy for adiposity in longitudinal studies with adolescent males: the Australian LOOK study. *J Sci Med Sport* 2019;22:307–10.
45. Gonçalves EM, Silva AM, Santos DA, Lemos-Marini SH, de Oliveira Santos A, Mendes-dos-Santos CT, et al. Accuracy of anthropometric measurements in estimating fat mass in individuals with 21-hydroxylase deficiency. *Nutrition* 2012;28: 984–90.
46. Deurenberg-Yap M, Niti M, Foo LL, Ng SA, Loke KY. Diagnostic accuracy of anthropometric indices for obesity screening among Asian adolescents. *Ann Acad Med Singapore* 2009;38:3–6.
47. Widhalm K, Schönegger K, Huemer C, Auterith A. Does the BMI reflect body fat in obese children and adolescents? A study using the TOBEC method. *Int J Obes* 2001;25:279–85.
48. Ramírez-Vélez R, García-Hermoso A, Alonso-Martínez AM, Agostinis-Sobrinho C, Correa-Bautista JE, Triana-Reina HR, et al. Cardiorespiratory fitness normative values in Latin-American adolescents: role of fatness parameters. *Int J Environ Res Publ Health* 2019;16:3889.
49. Valerio G, Licenziati MR, Tortorelli P, Calandriello LF, Alicante P, Scalfi L. Lower performance in the six-minute walk test in obese youth with cardiometabolic risk clustering. *Front Endocrinol* 2018;9:701.
50. Pérez-Bey A, Segura-Jiménez V, Fernández-Santos JD, Esteban-Cornejo I, Gómez-Martínez S, Veiga OL, et al. The influence of cardiorespiratory fitness on clustered cardiovascular disease risk factors and the mediator role of body mass index in youth: the UP & DOWN Study. *Pediatr Diabetes* 2019;20:32–40.
51. Tuan S, Su H, Chen Y, Li M, Tsai Y, Yang C, et al. Fat mass index and body mass index affect peak metabolic equivalent negatively during exercise test among children and adolescents in Taiwan. *Int J Environ Res Publ Health* 2018;15:263.
52. Barker AR, Gracia-Marco L, Ruiz JR, Castillo MJ, Aparicio-Ugarriza R, González-Gross M, et al. Physical activity, sedentary time, TV viewing, physical fitness and cardiovascular disease risk in adolescents: the HELENA study. *Int J Cardiol* 2018;254:303–9.
53. González-Gross M, Ruiz JR, Moreno LA, De Rufino-Rivas P, Garaulet M, Mesana MI, et al. Body composition and physical performance of Spanish adolescents: the AVENA pilot study. *Acta Diabetol* 2003;40:s299–301.
54. Kakon GA, Hadjiyannakis S, Sigal RJ, Doucette S, Goldfield GS, Kenny GP, et al. Edmonton obesity staging system for pediatrics, quality of life and fitness in adolescents with obesity. *Obes Sci Pract* 2019;5:449–58.
55. Ramírez-Vélez R, Izquierdo M, Correa-Bautista JE, Tordecilla-Sanders A, Correa-Rodríguez M, Schmidt Rio-Valle J, et al. Grip strength moderates the association between anthropometric and body composition indicators and liver fat in youth with an excess of adiposity. *J Clin Med* 2018;7:347.
56. He H, Pan L, Du J, Liu F, Jin Y, Ma J, et al. Muscle fitness and its association with body mass index in children and adolescents aged 7–18 years in China: a cross-sectional study. *BMC Pediatr* 2019;19:101.

57. Singla D, Hussain ME. Association between handgrip strength and back strength in adolescent and adult cricket players. *Int J Adolesc Med Health* 2018;32:20170177.
58. Dumith SC, Van Dusen D, Kohl HW. Physical fitness measures among children and adolescents: are they all necessary?. *J Sports Med Phys Fit* 2012;52:181–9.
59. Dumith SC, Ramires VV, Souza MA, Moraes DS, Petry FG, Oliveira ES, et al. Overweight/obesity and physical fitness among children and adolescents. *J Phys Activ Health* 2010;7:641–8.
60. Joensuu L, Syväoja H, Kallio J, Kulmala J, Kujala UM, Tammelin TH. Objectively measured physical activity, body composition and physical fitness: cross-sectional associations in 9- to 15-year-old children. *Eur J Sport Sci* 2018;18:882–92.
61. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, Young SL. Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Front Pub Health* 2018;6:149.
62. Jean-Pierre P, Cheng Y, Paskett E, Shao C, Fiscella K, Winters P. Item response theory analysis of the patient satisfaction with cancer-related care measure: a psychometric investigation in a multicultural sample of 1,296 participants. *Support Care Canc* 2014;22:2229–40.