Field Evaluation of the New FITNESSGRAM® Criterion-Referenced Standards

Gregory J. Welk, PhD, Pedro F. De Saint-Maurice Maduro, MS,
Kelly R. Laurson, PhD, Dale D. Brown, PhD

Background: The Cooper Institute established new criterion-referenced standards for the body composition and cardiovascular fitness standards for the FITNESSGRAM® program.

Purpose: The purpose of this study was to evaluate differences in the achievement of standards and classification agreement between the old and new standards.

Methods: Fitness data were collected during the 2009–2010 school year from 1686 youth (grades 3–12) from six schools (two districts) in a small Midwestern town. Data analyses were completed during the 2010 and 2011 school year. Descriptive statistics were computed to determine the percentage of students classified into the various fitness classifications using both the old and new standards. Classification agreement was examined for alternative tests of cardiovascular fitness (Progressive Aerobic Cardiovascular Endurance Run [PACER] and 1-mile run) and body composition (skinfold determination of body fat and BMI) using both the old and new standards.

Results: The descriptive results reveal important age and gender differences in reported levels of fitness levels in school-aged youth. For body composition, the percentage of youth achieving the healthy fitness zone (HFZ) was about 10% lower with the new standards than the old ones. For aerobic capacity, a larger percentage of young boys and a smaller percentage of older boys achieved the new HFZ for aerobic capacity. However, a smaller percentage of girls achieved the new HFZ at all ages. The use of test-equating in the revised standards led to significant improvements in the classification agreement between 1-mile run and the PACER.

Conclusions: The large data set provided a useful sample to examine the impact of the new fitness standards on the documentation of youth fitness levels in schools. The new standards address a number of measurement issues with the old standards and provide a more appropriate way to evaluate levels of health-related fitness in youth.
(Am J Prev Med 2011;41(4S2):S131–S142)

Introduction

The FITNESSGRAM® program has provided teachers with valid and reliable indicators of health-related fitness to use in school-based fitness assessments.[1,2] Although used predominantly for school-based assessments, the data can have important value for research and program evaluation. A unique advantage of the FITNESSGRAM® program is that the assessments are related to criterion health standards, based on how fit youth need to be to achieve health benefits. Research on the health benefits of physical activity has progressed considerably since the development of the original standards.[3] The advent of new statistical techniques and the availability of national databases also provide new opportunities to refine and improve the health-related standards. These factors contributed to the decision to develop new criterion-referenced standards for body composition and aerobic capacity in FITNESSGRAM.[4]

The sequential articles[5–9] in this supplement to the *American Journal of Preventive Medicine* provide documentation of the steps taken to establish and evaluate the new standards,[1] but a brief summary is appropriate. The standards were developed with the same nationally representative data set (National Health and Nutrition Examination Survey [NHANES]) and were based on the same health criterion (metabolic syndrome). They were

From the Department of Kinesiology, Iowa State University (Welk, De Saint-Maurice Maduro), Ames, Iowa; and the School of Kinesiology and Recreation, Illinois State University (Laurson, Brown), Normal, Illinois

Address correspondence to: Gregory J. Welk, PhD, Department of Kinesiology, Iowa State University, 257 Forker Building, Ames IA 50011. E-mail: gwelk@iastate.edu.

0749-3797/$17.00
doi: 10.1016/j.amepre.2011.07.011

also developed using the same empirical methodology: the linked usage of LMS (L=skewness, M=median, and S=coefficient of variation) curves and receiver operating characteristic (ROC) curves. The LMS procedures established age- and gender-specific centiles for body fatness[5] and cardiovascular fitness.[6] The use of ROC curves, in turn, provided an empirical way to establish health-related standards (based on risk for development of metabolic syndrome). A unique advantage of the methodology is that the fitness variable in the ROC analyses was a z-score derived using the LMS parameters. This made it possible to take normal growth and maturation into account when establishing the health-related standards for both body composition[7] and cardiovascular fitness.[8] Another advantage of this approach is that messages can accurately warn of potential health implications if levels of fitness or fatness continue to "track" at the same level. These types of warning statements are defensible because the standards (developed using LMS) values reflect the same centile value across different ages.

Historically, the FITNESSGRAM program has used the term *healthy fitness zone* (HFZ) to denote levels of fitness that meet the health-related standards. Students that achieve the HFZ receive feedback that their level of fitness is sufficient for good health whereas students that do not achieve the HFZ are classified into a needs improvement zone (NIZ). As described in this supplement,[7,8] the new standards were established with three distinct zones. The use of three zones makes it possible to provide more effective prescriptive messages to youth since more distinct messages can be delivered. The same basic terminology was used for the new standards, but the new standards include two different NIZs (NIZ–higher risk and NIZ–some risk). Children below the NIZ–higher risk now receive messages warning them of potential risk if they continue tracking at that level. Children in the NIZ–some risk would receive a message that they could reduce potential risks by striving to move into the HFZ.

The new body composition standards are similar to the previous standards, which have been largely intact since the development of the program. However, the new standards differ in some important ways.[1] The new body fat standards take gender- and age-related changes in body fat deposition into account, whereas the previous standards were static across ages. The associated BMI standards[9] track similarly to the commonly used percentile standards developed by the CDC since both use LMS procedures to account for growth and maturation (Figure 1). The HFZ and the NIZ are criterion referenced (based on risk for metabolic syndrome), but the values correspond with the 83rd percentile and 90th percentiles in the CDC charts, respectively. The percentage of youth classified into the NIZ–some risk would be slightly higher than the percentage classified as being overweight or obese by the CDC standards.

The new cardiovascular fitness standards follow the same basic age and gender pattern as the previous standards, but there are some important differences for specific age and gender combinations (Figure 2). The standards for young boys are slightly lower (easier), but the standards for young girls are slightly higher (harder) than the previous standards. The new standards are now equivalent for young children, but on the basis of the thresholds, fewer girls and more boys will achieve the HFZ. Importantly, the new standards now follow age- and gender-specific centiles that take normal growth and maturation into account. The standards for older boys and girls trend toward higher values than the previous standards, theoretically reducing the number of older youth that will meet the HFZ. These changes would result in more adolescents in the two NIZs than the previous standards. The studies in the supplement provide good justification for the creation of new standards, but research is needed to determine the impact of the new standards on reported fitness levels in schools.

Research is also needed to evaluate the new approaches established to improve classification agreement among alternative assessments of body composition (body fat and BMI) and aerobic capacity (Progressive Aerobic Cardiovascular Endurance Run [PACER] and 1-mile run). For body composition, ROC curves were used to create age- and gender-specific BMI standards from associated age- and gender-specific body fat standards.[9] The use of ROC curves help to maximize classification agreement between these two different indicators of body composition. For aerobic capacity, a test-equating approach was used to link PACER scores to estimated 1-mile run times.[10] This approach allowed data from the two tests to be scored from using the same prediction equation. These steps were taken to improve classification agreement but the effects of these approaches on classification agreement have not been directly evaluated in independent samples.

The primary purpose of this study is to evaluate age and gender differences in fitness achievement with the new FITNESSGRAM standards. Direct comparisons were made with the previous standards to highlight specific changes in the percentage of children achieving the HFZ. Comparisons were also made to examine the impact of employing three feedback zones instead of two. A secondary purpose was to evaluate classification agreement among alternative assessments of body composition and aerobic capacity. The study capitalized on the availability of a large data set of school-aged youth that had matching data from

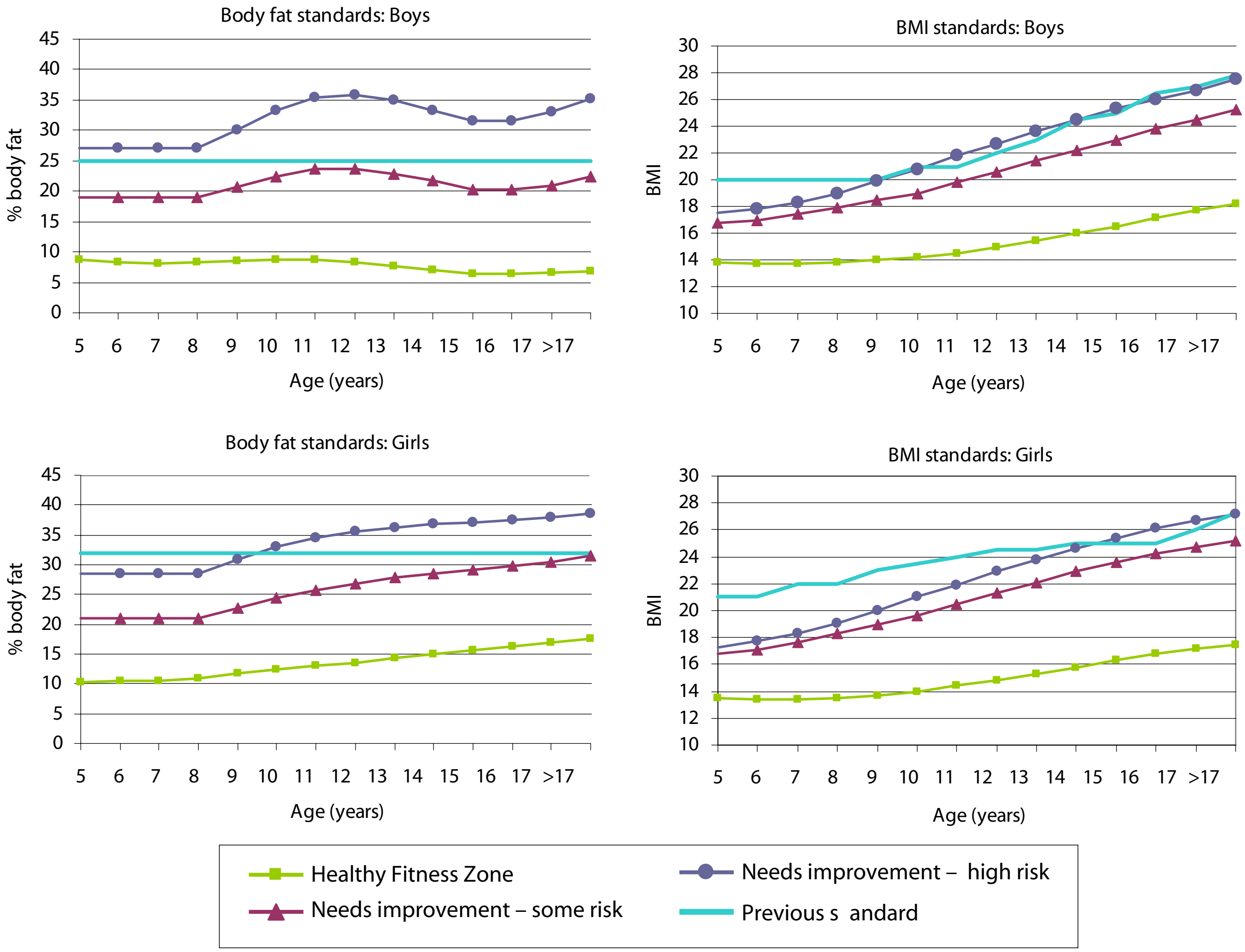

Figure 1. Age-related changes in body fat standards and BMI standards for boys and girls

several aerobic fitness tests (1-mile run and PACER) along with estimates of both BMI and body fatness.

Methods

Data Collection

The data for the present study were collected from 1711 participants in 3rd through 12th grade during the 2009–2010 school year from two school districts and two private schools in a small Midwestern town. The schools included one elementary school, two middle schools, and three high schools. Data were collected from these six schools during normal physical education sessions, with less than 2 weeks separating the various tests. The standard FITNESSGRAM protocol was followed to obtain data on health-related fitness in the participants. Emphasis was placed on obtaining indicators of aerobic fitness and BMI, but data from the complete battery were collected. Graduate research assistants went to the participating schools and offered assistance to the teachers in collecting and recording fitness data. The district has been involved in ongoing projects with the local university research team, and the teachers and students were familiar with the FITNESSGRAM fitness test battery. The use of the FITNESSGRAM fitness test is a standard part of the normal physical education curriculum for each school involved in this project. Research collaboration, and the university's involvement in the school's physical education curriculum, was approved by the school board and by the university's IRB.

Data Processing

Data from the various fitness assessments were compiled and examined to test for missing values and outliers. It was important to have an adequate sample size for each age group, so frequency distributions were examined by age and gender. The sample included only 10 participants that were aged 10 years, so analyses were restricted to youth aged 11–18 years. Further evaluation revealed six cases with missing demographic or BMI data and nine cases with outliers in one of the key fitness outcomes (z-scores ≥3.0 on BMI or aerobic fitness). These cases were removed leaving a total sample of 1666 participants.

The data on BMI and body fat are tracked and processed in their reported units, but the aerobic fitness data required conversion to estimates of aerobic capacity. The data from the PACER test were processed using both the Leger equation[11] and with the new test equating procedure.[10] The Leger equation includes terms for age

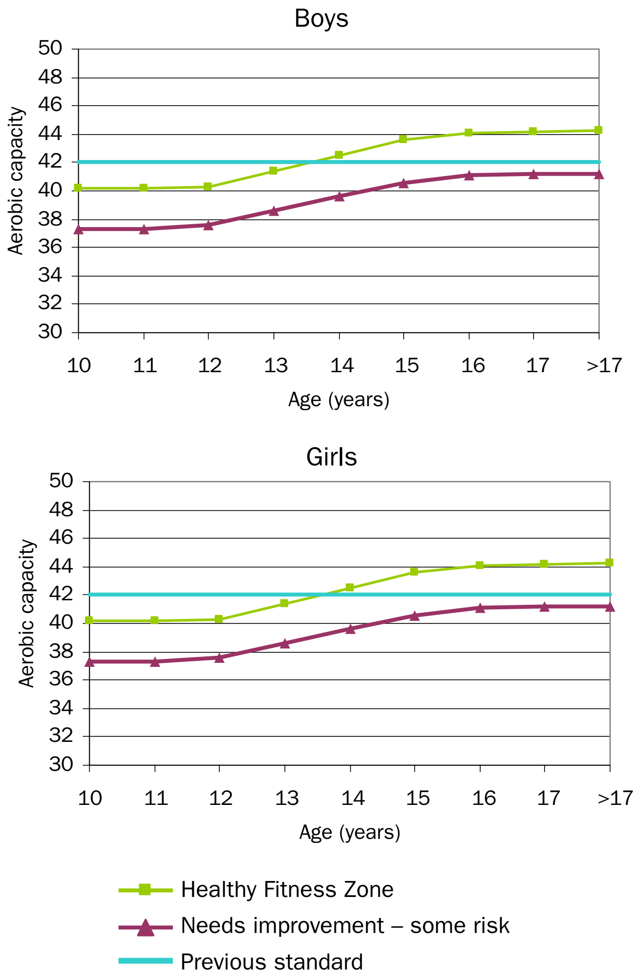

Figure 2. Age-related changes in aerobic capacity standards for boys and girls

but is not gender-specific [$VO_2max = 31.025 + (3.238 \times speed) - (3.248 \times age) + (0.1536 \times speed \times age)$]. The 1-mile run data and the PACER equated times were converted into estimates of maximal aerobic capacity using the established Cureton 1-mile run equation,[12] This equation includes terms for age (yr), gender (F=0; M=1), 1-mile run time, and BMI [VO_2max (mL/kg/min) = ($-8.41 \times min) + (0.34 \times min^2) + (0.21 \times (age \times gender) - (0.84 \times BMI) + 108.94$]. The individual fitness scores were then processed using both the old and new FITNESSGRAM standards. Individual data were coded into the specific fitness categories for each set of standards.

Data Analyses

Data analysis was conducted during the 2010 and 2011 school year. Descriptive analyses were performed to examine age and gender differences in body composition and aerobic capacity in the sample. Two-way (age X gender) ANOVAs were used to test for differences in fitness achievement in the sample. Post hoc analyses (with Tukey adjustments) were conducted to examine specific age X gender differences.

The impact of the new FITNESSGRAM standards on observed school-level fitness achievement was examined by comparing frequency distributions for fitness classifications from both the old and the new standards. Descriptive and graphical analyses were conducted to examine the variability in fitness classification by age and gender. These analyses were conducted separately for the two body composition measures (body fatness and BMI) and the two aerobic fitness measures (1-mile run and PACER). The changes in fitness allocation were further examined by comparing the relative distribution into two zones (old standards) and three zones (new standards). These were examined using stacked bar plots to show the relative distribution among the three fitness categories.

Classification agreement among alternative indicators of fitness for both body composition and aerobic fitness were examined quantitatively with percentage agreement and kappa statistics.[13,14] The two separate NIZs in the new standards were combined into one to allow for direct comparisons with the previous classifications of fitness. Kappa agreement was classified as poor if <0.20; fair if 0.20–0.40; moderate if 0.40–0.60; good if 0.60–0.80; and very good if 0.80–1.00.[15]

Results

The final sample included 1666 youth aged 11–18 years (911 boys and 755 girls). Descriptive statistics for body composition tests are provided in Table 1. The analyses indicated a trend for higher BMI levels with age in both boys and girls. The analyses across groups indicated there were no differences between boys and girls in measured BMI ($M_{boys-girls} = -0.25 \pm 0.28$, $F(1, 1670) = 0.78$, $p = 0.379$). Age- and gender-matched comparisons revealed a tendency for higher BMI values for girls, but there were no significant differences ($p > 0.05$). The other indicator of body composition used by FITNESSGRAM (percent body fat) exhibited a different trend. In boys, percent body fat scores decreased from age 11 years to age 13 years but then remained stable through age 17 years. In girls, body fat levels tended to increase with age. The analyses across groups indicated there were significant differences between boys and girls in percent body fat ($M_{boys-girls} = -10.79 \pm 0.61$, $F(1, 1670) = 314.28$, $p < 0.001$). Age- and gender-matched comparisons suggested that percent body fat values were significantly higher in girls at all ages ($p < 0.05$).

Descriptive statistics for aerobic fitness tests are provided in Table 2. Analyses revealed general increases in levels of fitness from age 11 years until age 15 years, but the changes were more dramatic in boys. Levels of fitness tended to stabilize or decline thereafter until age 18 years. The analyses across groups indicated that boys performed the 1-mile run test significantly faster than girls ($M_{boys-girls} = -1.64 \pm 0.16$ minutes, $F(1, 1670) = 105.55$, $p < 0.001$). Boys also completed a greater number of laps on the PACER test ($M_{boys-girls} = 18.80 \pm 1.16$, $F(1, 1670) = 263.28$, $p < 0.001$). Gender differences in absolute performance were consistent across age and were significant for ages 13, 15, 16, and 17 years in the 1-mile run test

Table 1. Gender differences in BMI and body fat by age and gender, *n* (M ± SD)

	Boys	Girls	Total
BMI			
Age (years)			
11	133 (19.63±4.36)	133 (20.52±5.05)	266 (20.07±4.73)
12	164 (20.20±4.85)	169 (21.25±4.87)	333 (20.73±4.88)
13	187 (20.51±4.12)	173 (21.24±4.68)	360 (20.86±4.40)
14	120 (21.40±4.77)	91 (21.37±3.70)	211 (21.39±4.33)
15	73 (22.05±4.22)	62 (21.51±3.91)	135 (21.81±4.07)
16	98 (22.95±4.16)	68 (23.47±5.55)	166 (23.16±4.77)
17	101 (23.84±4.72)	65 (22.72±4.25)	166 (23.40±4.56)
18	35 (24.56±6.79)	14 (25.05±6.23)	49 (24.70±6.57)
BODY FAT (%)			
Age (years)			
11*	133 (18.78±9.82)	133 (24.33±10.03)	266 (21.55±10.29)
12*	164 (16.97±9.23)	169 (24.94±9.76)	333 (21.02±10.30)
13*	185 (15.17±7.72)	171 (25.57±9.02)	356 (20.17±9.84)
14*	119 (14.90±9.30)	90 (25.63±8.04)	209 (19.52±10.25)
15*	65 (13.88±7.68)	50 (24.49±7.12)	115 (18.50±9.10)
16*	88 (14.32±6.90)	59 (27.83±8.50)	147 (19.74±10.06)
17*	93 (13.86±6.88)	55 (25.54±7.46)	148 (18.20±9.07)
18*	30 (16.2±10.83)	8 (32.06±8.22)	38 (19.54±12.15)

*Significant difference (M: boys – girls) with $p<0.05$.

and from ages 13 to 18 years with the PACER test ($p<0.05$).

The fitness outcome measures were processed using both the new and previous standards to determine the impact on fitness classification. Figure 3 provides a comparison of age-related patterns of fitness achievement (percent achieving the HFZ) for body composition, and Figure 4 provides a similar comparison for aerobic fitness. In both figures, results with the previous standards are shown on the left, whereas results for the new standards are shown on the right. The top panels compare the distributions in boys, and the bottom panels compare the distributions for girls.

For body composition (Figure 3), gender and age differences are noted for the percentage of youth achieving the HFZ age, with some differences noted between the two sets of standards. With the old standards, the percentage of boys achieving the HFZ for body fatness increased from around 75% to around 90% and stayed fairly stable at older ages. With BMI, the percentage achieving the HFZ ranged from 70% to 80%. The patterns were similar with the new standards, but the percentage of youth achieving the fitness standards was about 10% lower (70%–80% for body fatness and 60%–70% for BMI). Similar general patterns were evident for the girls. The percentage achieving the old body fatness standard ranged from 70% to 80%, whereas the percentages for the new standards ranged from 50% to 70%. For BMI, the percentage achieving the HFZ ranged from 70% to 90% with the old standard and from approximately 60% to 80% with the new standards.

For aerobic fitness (Figure 4), there were clear age and gender patterns, and the relationships were different when comparing results between the old and the new standards. With the old standards, the percentage of boys achieving the HFZ on the 1-mile run ranged from 70% to 80%, whereas values for girls ranged from 80% to 90%. There were no clear age-related patterns of achievement with the 1-mile run, but there were large decreases in the percentage of youth achieving the HFZ with the PACER. In boys, values dropped from approximately 80% to 30%, and in girls, values dropped from approximately 98% to 50%. Very different age and gender patterns were evident with the new

Table 2. Gender differences in 1-mile run and PACER performance by age and gender, *n* (M ± SD)

	Boys	Girls	Total
1-MILE RUN (in minutes)			
Age (years)			
11	133 (10.08±2.22)	133 (10.91±2.30)	266 (10.49±2.30)
12	164 (10.38±2.98)	169 (10.94±3.18)	333 (10.66±3.09)
13*	187 (9.21±2.71)	173 (10.48±3.00)	360 (9.82±2.92)
14	120 (9.41±2.65)	91 (10.71±2.51)	211 (9.97±2.66)
15*	73 (8.75±2.83)	62 (10.46±2.15)	135 (9.53±2.67)
16*	98 (8.73±2.29)	68 (11.18±2.30)	166 (9.74±2.59)
17*	101 (8.42±2.04)	65 (10.44±2.21)	166 (9.21±2.32)
18	35 (8.80±2.29)	14 (11.76±1.71)	49 (9.65±2.52)
PACER (number of laps)			
Age (years)			
11	133 (37.65±19.94)	133 (29.62±15.59)	266 (33.63±18.31)
12	164 (36.88±20.27)	169 (30.86±15.54)	333 (33.83±18.25)
13*	187 (44.05±21.80)	173 (34.17±16.11)	360 (39.31±19.87)
14*	120 (49.47±22.81)	91 (31.96±13.89)	211 (41.91±21.28)
15*	73 (61.95±26.97)	62 (33.53±14.47)	135 (48.90±26.23)
16*	98 (57.09±23.61)	68 (29.82±13.03)	166 (45.92±24.03)
17*	101 (57.35±21.53)	65 (34.88±13.96)	166 (48.55±21.85)
18*	35 (54.37±20.89)	14 (22.93±6.63)	49 (45.39±22.96)

*Significant difference (M: boys – girls) with $p<0.05$
PACER, Progressive Aerobic Cardiovascular Endurance Run

standards. The percentage of boys achieving the HFZ in the 1-mile run ranged from approximately 80% at younger ages to around 70% during late adolescence. The pattern reveals a clear age-related decline that was not clear with the old standards.

The percentage of girls achieving the new 1-mile run standard increased from 70% to 80% (those aged 15 years) and then down to approximately 60% for older girls. The percentage of girls achieving the HFZ was consistently higher than boys with the old standards, but this pattern is reversed with the new standards. The clearer age-related trends in the new standards suggest that the percentage of youth at risk for metabolic syndrome increases in late adolescence.

The differences in fitness classification were also visually examined using stacked bar plots to show the relative differences in the allocation of youth into three fitness zones (compared to two with the old standards). Figure 5 and Figure 6 show the aggregated fitness classification for the younger age sample (11–14 years) using both the old standards (left) and new standards (right). The top panels compare the distributions in boys, whereas the bottom panels compare the distributions for girls. With the old BMI standards, 73.3% of the boys were classified in the HFZ and 26.7% were in the NIZ. When the same sample was classified according to the new standards, the proportion of boys shifted to 63.8% in HFZ, to 12.2% in the NIZ–some risk, and to 23.9% in the NIZ– higher risk. The distribution of girls' BMI in the old standards resulted in 82.8% in the HFZ and 17.2% in the NIZ. The BMI classification based on the new standards shifted the pattern to 65.2% in the HFZ, 12.3% in the NIZ–some risk, and 22.5% in the NIZ– higher risk.

Body fat classification with the old standards resulted in 81.9% in the HFZ and 18.1% in the NIZ. The body fat classifications based on the new standards resulted in slightly higher percentage of boys in the HFZ (77.4%), 17.6% in the NIZ–some risk, and 5.1% in the NIZ– higher risk. For girls, the body fat classification resulted in 79.1% in the HFZ and 20.9% in the NIZ. After applying the new standards, there were fewer classified into the HFZ (56.9%) and higher percentages distributed into the NIZs

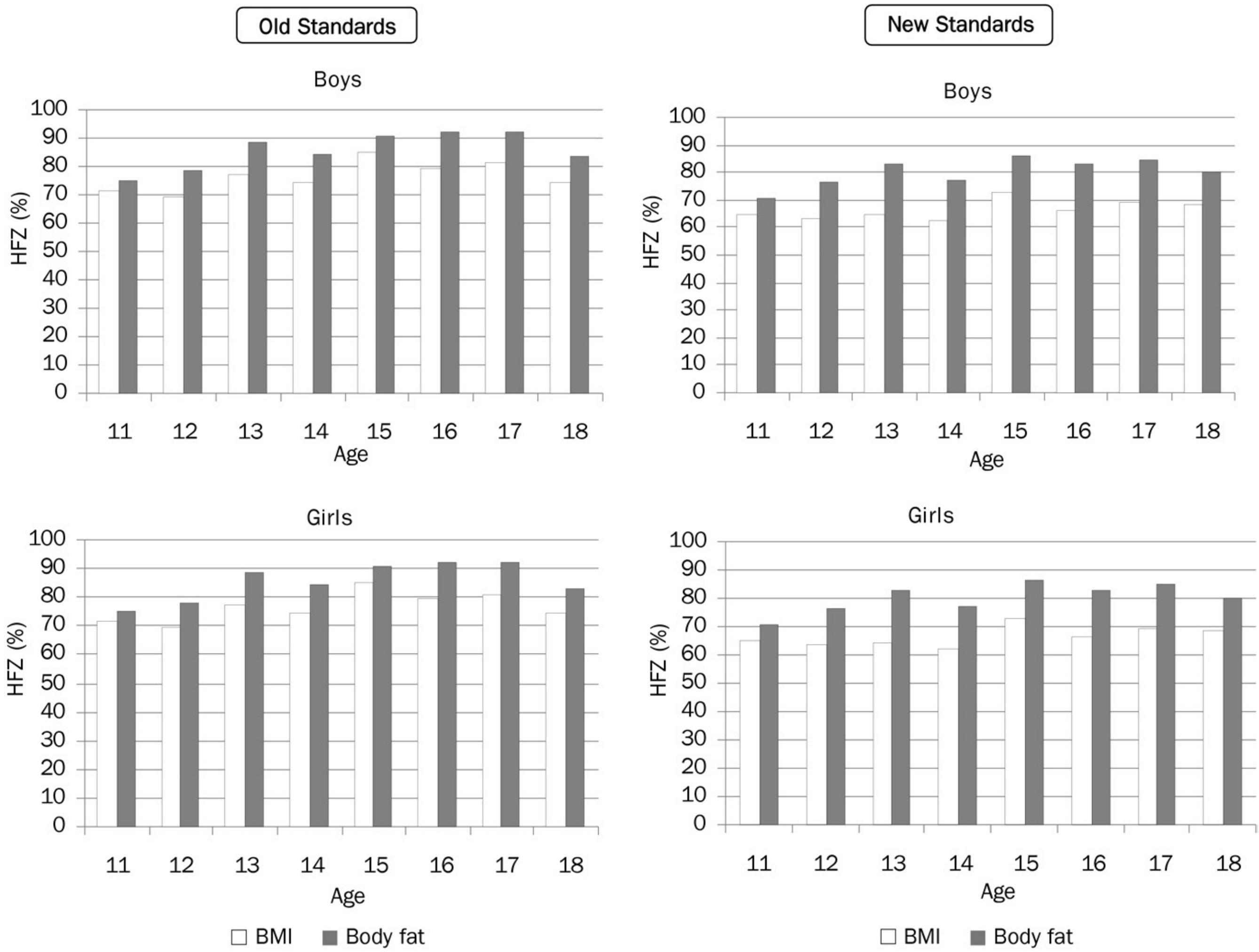

Figure 3. Gender- and age-related differences in the achievement of healthy fitness zone for body composition
HFZ, healthy fitness zone

(30.1% and 13.0% in the NIZ–some risk and NIZ–higher risk, respectively). Overall, Figure 5 demonstrates that the change to three zones essentially allocated the needs improvement category into two distinct needs improvement categories.

Similar results are presented in Figure 6 for aerobic fitness, but these plots also provide direct comparisons among the fitness classifications that would have resulted if the PACER data were processed with the Leger equation used in the previous standards,[11] rather than with the Zhu test-equating approach,[10] With the old 1-mile run standards, 79.0% of the boys were classified in the HFZ and 21.0% were in the NIZ. When the same sample was classified according to the new standards, the proportion of boys in the HFZ shifted to 82.5%, with 8.3% in the NIZ–some risk and 9.2% in the NIZ–higher risk. The Zhu test-equating approach employed with the new standards yielded very similar distributions (81.0%, 8.9%, and 10.1%, respectively). The respective distributions that would have resulted if the Leger equation was used instead were 73.8%, 13.4%, and 12.8%.

Similar classification agreement was evident with the plots for girls. With the old 1-mile run standards, 85% of girls were classified into the HFZ and 15% were classified into the NIZ. With the new standards, the distributions changed to 72.9% in the HFZ, 12.5% in the NIZ–some risk, and 14.7% into the NIZ–higher risk. The Zhu test-equating approach[10] yielded similar distributions (71.0%, 14.2%, and 14.8%). The values if the Leger equation was used were 77.3%, 17.3% NIZ–some risk, and 5.4% into the NIZ–higher risk. Overall, the classification agreement between the 1-mile run and the PACER are more similar when the PACER data are processed with the Zhu test-equating approach.

Classification agreement was empirically evaluated using kappa statistics and computed classification agreement values. The classification agreement among alternative tests for body composition varied by age and gender, but improve-

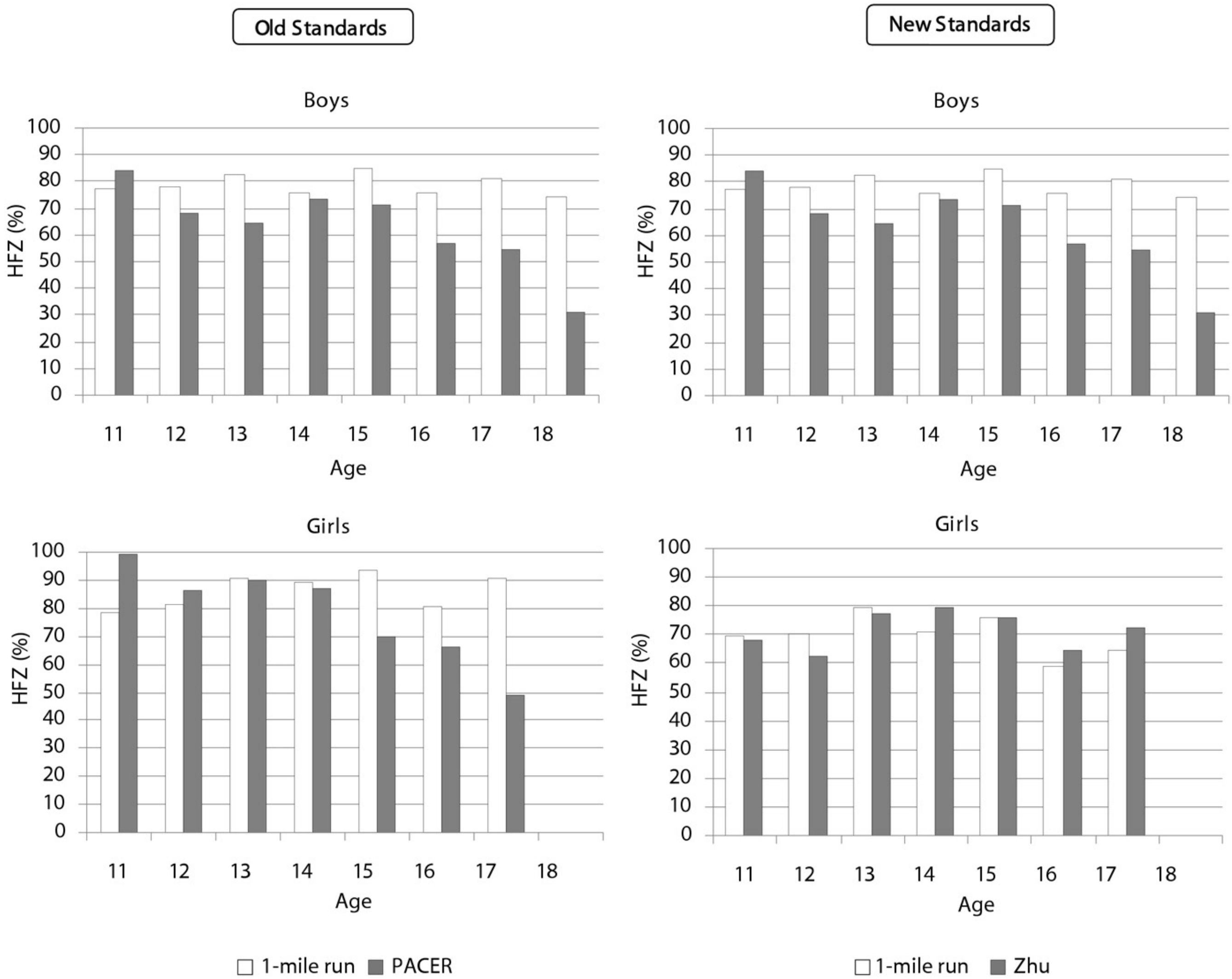

Figure 4. Gender- and age-related differences in the achievement of healthy fitness zone for aerobic fitness
HFZ, healthy fitness zone; PACER, Progressive Aerobic Cardiovascular Endurance Run

ments in agreement were not evident with the newer standards (Table 3). Agreement among alternative body composition tests using the old standards and the new ones was different for older youth (kappa=0.65 versus 0.45, for old and new standards, respectively), but results for the younger sample were similar (kappa=0.67 versus 0.64, for old and new standards, respectively). Agreement with the new standards was better for girls (kappa=0.62) than for boys (kappa=0.55). The results for aerobic fitness demonstrated significant improvements in classification agreement with the new standards. The kappa values with the old standards ranged from 0.45 to 0.51, and the percentage agreement values were approximately 75%. The kappa values with the new standards were consistently above 0.78, with classification agreement exceeding 90%. These results suggest the importance of the equating approach to improve the agreement among aerobic fitness tests.

Discussion

The FITNESSGRAM program recently released new age- and gender-specific, health-related fitness standards,[1] and these values were based on findings reported in previous papers in this supplement.[7,8] Because the program includes several options for each fitness component, special consideration was given to help ensure good classification agreement among the alternative assessments. For body composition, the standards were first established for health risks associated with excess body fatness. The values from these analyses were then equated or matched to corresponding BMI values using ROC analyses. For aerobic capacity, the PACER results are now equated to a 1-mile run time and these values were then processed using the same equation used to process the 1-mile run data.

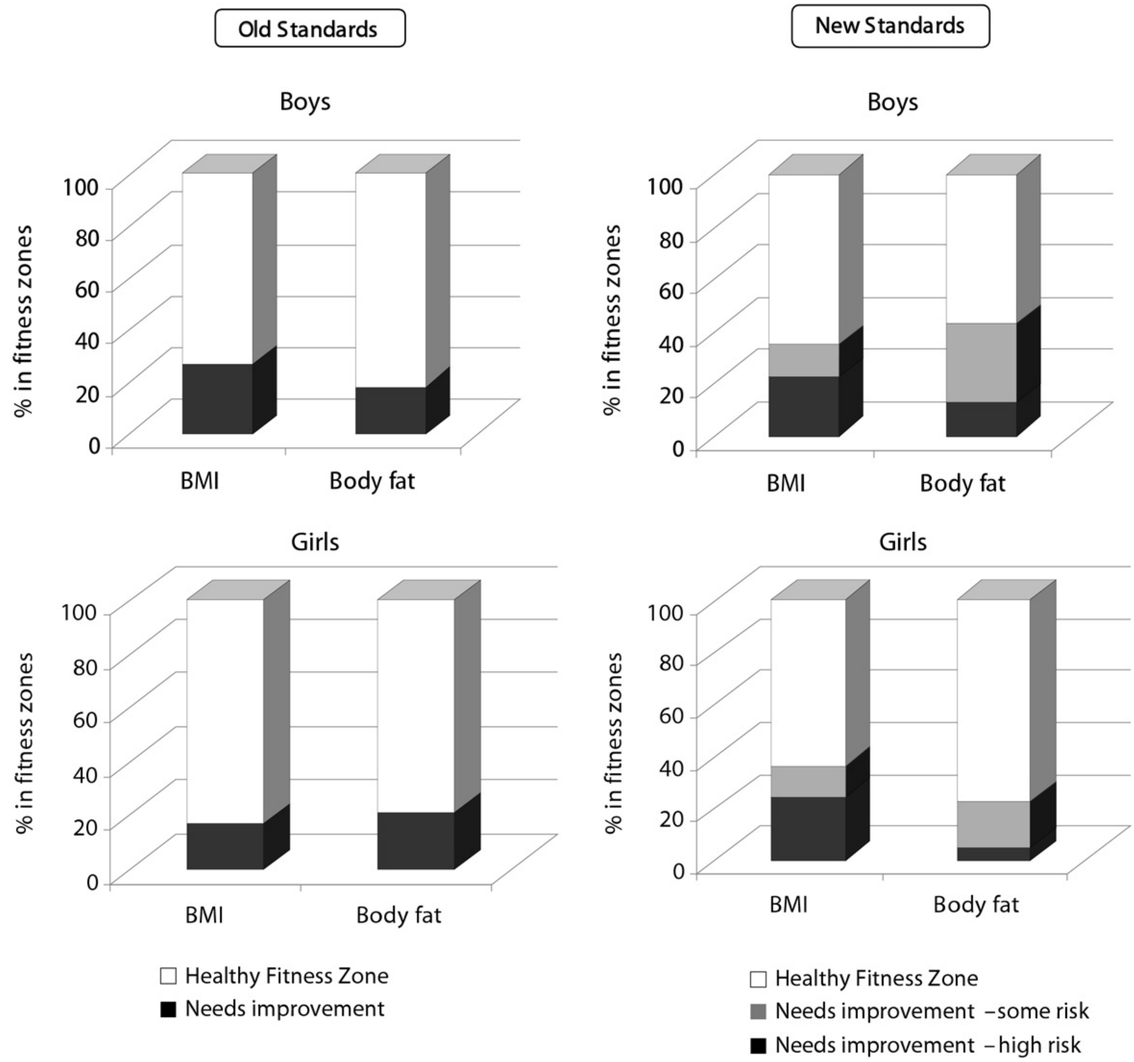

Figure 5. Gender-related differences in the achievement of healthy fitness zone for body composition

Because these reflect substantial departures from the previous standards, it is important to characterize how the new standards will influence reports of fitness achievement in youth.

The body composition results yielded similar overall patterns, but the percentage of youth classified in the HFZ was lower than with the previous standards. The previous standards led to less than 10% of youth classified into the NIZ, and this is inconsistent with other national data. Using the new standards, approximately 20%–30% of youth are classified into the NIZ, and this is more consistent with other data. The results indicated that the percentage of youth classified in the HFZ would be slightly higher for the body fat assessment than the percentage that would result from the BMI assessment. This pattern was evident in the older standards as well, so the revised standards and matching of standards did not seem to directly improve classification agreement.

The kappa values and computed classification agreement values tended to be more favorable for the old standards compared to the new ones. It is not clear why this would be the case, because specific efforts were made to help ensure good classification agreement. The previous standards were static across age for body fat but varied for BMI, so poor classification agreement might be expected with these standards. The new standards were developed using ROC curves (to correspond with the derived health-related fitness standards) and then linked with LMS parameters to take growth and maturation into account. The validation studies for these BMI standards[7,9] reported classification agreement values exceeding 90%, so it is hard to reconcile the findings in this sample.

The most likely explanation is that different methods were used to estimate percent body fat. The NHANES estimates used to develop the standards were

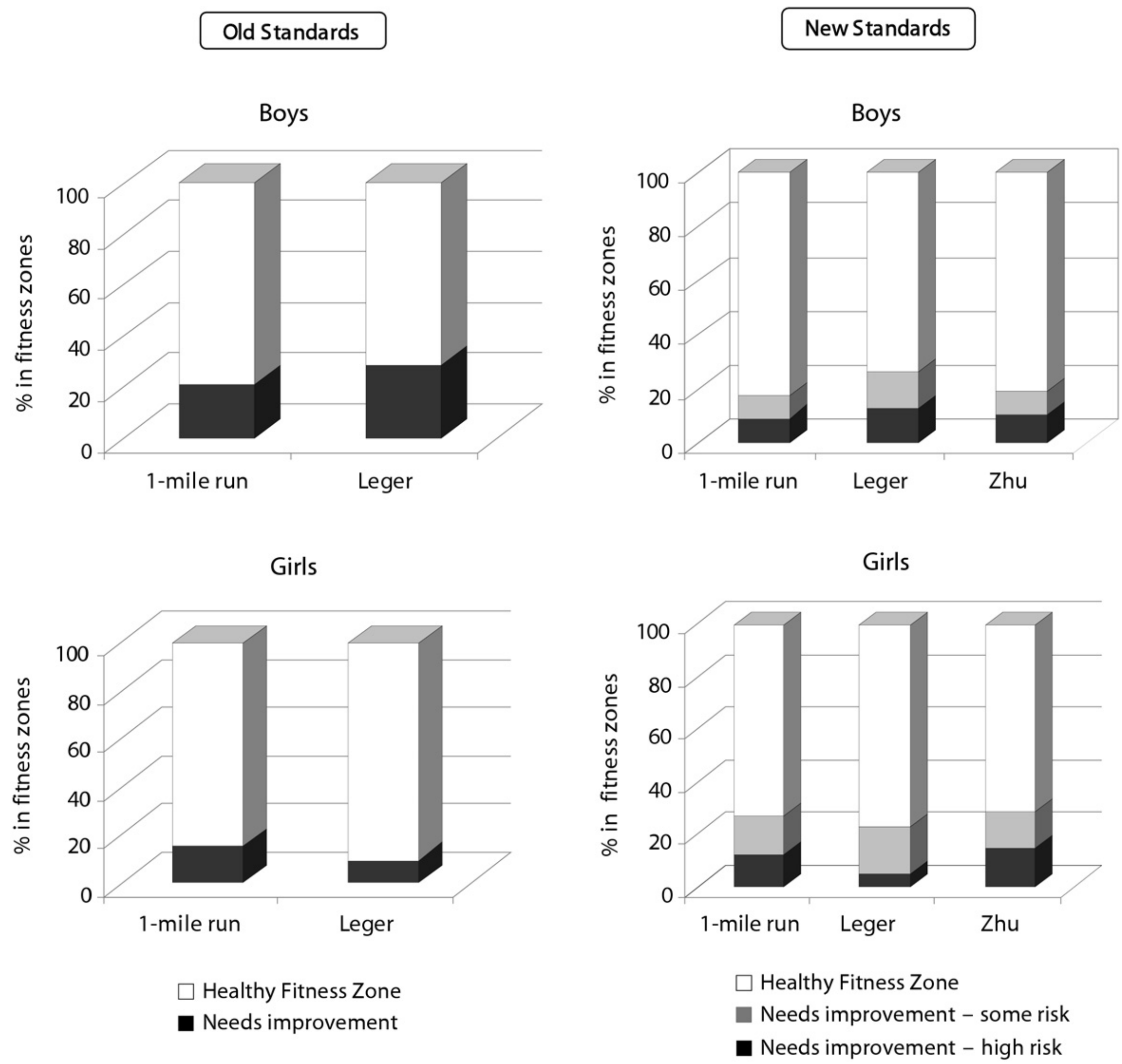

Figure 6. Gender-related differences in the achievement of healthy fitness zone for aerobic fitness

based on skinfolds, whereas the estimates in this study were obtained from bioelectric impedance analysis (BIA). The accuracy of some portable BIA devices has been questioned, so it is possible that the poor agreement is due to error or bias in the estimated body fat levels—agreement between BMI and body fat can be good only if the estimates are reasonably accurate. The present results for body composition should be interpreted with caution until additional research can be conducted to test these relationships.

The aerobic-capacity results yielded some interesting differences between the old and new standards. The old standards yielded significant differences in the percentage of youth achieving the HFZ with the 1-mile run compared with the PACER. Specifically, the results with the PACER showed large declines in the percentage of youth achieving the standards with age, but this pattern was not evident for the 1-mile run. This observation was consistent with previous findings from a large statewide study[16] but is difficult to explain since the data are processed by the same standards. The results also indicated that, on average, higher percentages of girls achieved the standards. This is also difficult to explain since the consensus in the literature is that boys tend to be more active than girls, particularly in later adolescence. The age and gender differences in the old standards could reflect actual differences in fitness and health risk, but they also may be due to the nature of the standards themselves. These underlying issues were part of the underlying justification for reexamining the FITNESSGRAM standards.[4] The articles in this supplement provided background on the research used to establish these new standards, and a key goal in this final paper was to evaluate the overall impact of the new standards on fitness classification, as well as on classification agreement.

Table 3. Classification agreement (kappa and % agreement) of alternative tests for body composition and aerobic fitness

	%BF – BMI		PACER – 1-mile run	
	Old	New	Old	New
Age (years)				
11–14	0.67	0.64	0.45	0.78
	89.2	83.7	76.1	92
15–18	0.65	0.45	0.51	0.79
	90.1	76.6	75.3	91.3
Gender				
Boys	0.63	0.55	0.51	0.80
	88.2	81.2	75.9	93.1
Girls	0.71	0.62	0.47	0.77
	90.9	82.1	75.7	90.3

BF, body fat; PACER, Progressive Aerobic Cardiovascular Endurance Run

The results with the new aerobic-capacity standards led to some noteworthy differences in the percentages of youth achieving the standards. In general, a larger percentage of young boys and a smaller percentage of older boys achieve the new HFZ. A smaller percentage of girls achieve the new aerobic-capacity HFZ at all ages. The absolute changes were relatively small, but when examined collectively, they lead to important differences in the pattern of aerobic fitness achievement. The results show higher percentages of boys achieving the HFZ, and they reveal a trend to decreasing numbers of children achieving the standards in late adolescence for both boys and girls. Additional research is needed to evaluate the validity of these new standards for detecting potential disease risk, but the results do seem to fit with prevailing views about levels of health-related fitness in youth.

An advantage of the present study is that it provided a unique perspective on the implications of the adoption of new FITNESSGRAM standards for school leaders, public health officials, and researchers. The previous FITNESSGRAM standards had been widely used, so the decision to develop and adopt new standards was undertaken only after careful deliberation about the relative advantages and disadvantages of change. Members of the FITNESSGRAM Scientific Advisory Board worked with the various research groups to consider both the scientific considerations as well as practical considerations about the transition to the new standards.

From a school perspective, it was important to consider the implications of the new standards on reported levels of fitness in schools. Schools that have had a history of tracking changes in the percentage of youth achieving the HFZ will need to understand the need for new standards and the implications of the new standards on their past trends. Children and parents would need to be educated about the differences in results if their child's level of fitness was coded differently in the old and the new standards. The ability to provide more prescriptive feedback (three zones rather than two) and the ability to provide more accurate indicators of potential health risk contributed to the decision that the benefits outweighed confusion regarding the changes.

The present study provides valuable descriptive information about the changes in fitness classification that would result from changes to the new standards and to evaluate whether classification agreement improved as a result of the specific efforts made in this area. The results show clearly different age and gender patterns in the percentage of youth using the old and the new standards. It is not possible to formally evaluate the predictive utility of the new standards on the present sample since there are no data on potential health risks. However, the descriptive age and gender patterns seem more consistent with the consensus findings about levels of fitness and fatness in youth. Thus, the new standards appear to have resolved some of the lingering issues that led to the need to revise the FITNESSGRAM health-related standards.[4]

The study also provides new information about classification agreement. This was an important consideration in the standards because schools typically have a choice about what assessments to use to complete their required or elective fitness testing. Direct comparisons were made between the old and the new standards to determine whether the changes in the standards and the processing of data led to improved classification agreement. The results demonstrate no significant improvement in classification agreement for the body composition measures but considerable improvements with the aerobic fitness indicators. Additional work is needed to examine these issues in other samples. The newly developed (empirically derived) fitness standards take growth and maturation into account and provide a sound method for evaluating levels of health-related fitness in youth. The transition to the use of three prescriptive zones for feedback will also improve the utility of these standards for school-based activity promotion and for promoting awareness to parents and clinicians. New research on these standards will continue to improve the scientific basis for youth fitness assessment and promotion.

Publication of this article was supported by The Cooper Institute through a philanthropic gift from Lyda Hill.

No financial disclosures were reported by the authors of this paper.

References

1. Meredith M, Welk GJ, eds. FITNESSGRAM & ACTIVITYGRAM: test administration manual. Developed by the Cooper Institute (Dallas, TX). Updated 4th edition. Champaign IL: Human Kinetics, 2010.
2. Welk GJ, Morrow JR Jr. FITNESSGRAM reference guide. Dallas TX: The Cooper Institute, 2008.
3. Physical Activity Guidelines Advisory Committee. Physical Activity Guidelines Advisory Committee Report, 2008. Washington, DC: DHHS, 2008.
4. Welk GJ, Going SB, Morrow JR Jr., Meredith MD. Development of new criterion-referenced fitness standards in the FITNESSGRAM® program: rationale and conceptual overview. Am J Prev Med 2011;41(4S2):S63–7.
5. Laurson KR, Eisenmann JC, Welk GJ. Body fat percentile curves for U.S. children and adolescents. Am J Prev Med 2011;41(4S2):S87–92.
6. Eisenmann JC, Laurson KR, Welk GJ. Aerobic fitness percentiles for U.S. adolescents. Am J Prev Med 2011;41(4S2):S106–10.
7. Laurson KR, Eisenmann JC, Welk GJ. Development of youth percent body fat standards using receiver operating characteristic curves. Am J Prev Med 2011;41(4S2):S93–9.
8. Welk GJ, Laurson KR, Eisenmann JC, Cureton KJ. Development of youth aerobic-capacity standards using receiver operating characteristic curves. Am J Prev Med 2011;41(4S2):S111–6.
9. Laurson KR, Eisenmann JC, Welk GJ. Body mass index standards based on agreement with health-related body fat. Am J Prev Med 2011;41(4S2):S100–5.
10. Zhu W, Plowman SA, Park Y. A primary field test centered equating method for cut-off score setting. Res Q Exerc Sport 2010;81(4):400–9.
11. Leger LA, Mercier D, Gadoury C, Lambert J. The multistage 20 metre shuttle run test for aerobic fitness. J Sports Sci 1988;6(2):93–101.
12. Cureton KJ, Sloniger MA, O'Bannon JP, Black DN, McCormack WP. A generalized equation for prediction of VO2 peak from one-mile run/walk performance in youth. Med Sci Sports Exerc 1994;27:445–51.
13. Cohen JA. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;(20):37–46.
14. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 1968;70(4): 213–20.
15. Altman DG. Practical statistics for medical research. Hall CA, ed. London, U.K.: Chapman and Hall, 1991.
16. Welk GJ, Meredith MD, Lhmels M, Seeger C. Distribution of health-related physical fitness in Texas youth: a demographic and geographic analysis. Res Q for Exerc Sport 2010;81(3):S6–15.