Fitness Testing for Children: Let’s Mount the Zebra!

Daniel D. Cohen, Christine Voss, and Gavin R.H. Sandercock

Childhood obesity has increased greatly over the past few decades, while children’s fitness levels have been declining worldwide. Physical fitness, as a multidimensional construct that includes skill and health related components of which cardiorespiratory fitness (CRF) and muscular fitness in particular are powerful determinants of health in youth. The association between fitness and health is well documented for CRF, whereby good CRF is protective against cardio-metabolic risk factors across BMI/fatness categories. Muscular fitness is also inversely associated with metabolic risk and is a valuable part of health monitoring in children. Specifically, poor muscular fitness is associated with elevated cardio-metabolic risk factors in adolescence and an increased risk of developing obesity, cardiovascular disease and with cardiovascular and total mortality. This effect appears independent of the associations between metabolic health and low CRF.

Given such strong, independent associations between fitness and health, declines in pediatric fitness are worrying from a public health perspective and underpin the former UK Chief Medical Officer’s (CMO) recommendation for routine fitness testing in schools. The CMO stated that “The introduction of a standardised school-based fitness assessment in England may have multiple benefits that extend beyond the benefits for the individual.” These benefits include: lowering the lifetime risk of 6 diseases, building a lifelong habit of participation in physical activity, higher educational achievement, maintaining a healthy weight, as well as improving social and mental wellbeing. Field-tests of fitness, such as Leger’s 20-m shuttle run test or “bleep test,” “beep test,” “PACER,” “multistage fitness test”) used to estimate CRF, as well as hand grip strength, jump performance, and/or trunk muscular endurance to assess muscular fitness, particularly lend themselves to school-based assessments due to relatively low space and equipment requirements. More importantly, because they are often already, or can be easily integrated into Physical Education (PE) lessons. In part, the UK CMO’s recommendation for school-based fitness testing builds upon a variety of well-reported school-based fitness projects, and the wealth of information that these have yielded. For example, the FITNESSGRAM fitness testing battery is implemented annually in a number of US states and in New York City public schools, while the ALPHA fitness testing battery based on the pan-European HELENA and Spanish AVENA studies has been successfully piloted in Spanish schools. In the UK, in addition to being implemented sporadically in schools, youth fitness testing is implemented regularly by fitness professionals in conjunction with academics from Liverpool John Moores University.

The study allowed for the development of various fitness test norms and revealed declining trends in fitness and a number of mediating factors. Furthermore, the Texas Youth Study based on fitness assessment of over 2.5 million children revealed positive associations between fitness and academic achievement, school attendance, several psychosocial measures and negative associations with indicators of delinquency. While much of these data are not new, one should not ignore the potential value of media coverage of these physical and cognitive or behavioral correlates of fitness and highlighting regional differences as leverage to promote governmental action on physical activity (PA).

While few would argue against the value of monitoring trends in children’s health, routine school-based fitness testing is not greeted with approval by all. The Association for Physical Education and parent groups criticized such plans for the UK, echoing concerns about fitness testing in children aired in the academic literature. One specific concern is the potential for fitness testing to be an unpleasant and embarrassing experience for overweight children, one that might reinforce poor physical self-concept. While there is little empirical research examining children’s experience of fitness testing, there is indeed some evidence to suggest that there may be negative psychosocial consequences, such as embarrassment or teasing by other children. Since physical self-concept is a determinant of PA, this is indeed an important concern for school-based fitness testing. In addition, fitness testing has also been criticized for not promoting PA per se, yet even its strongest advocates do not regard this as an expected outcome of the process. For example, the National Child Measurement Program, which assesses BMI in UK schoolchildren, does not cite weight loss as an objective of the measurement process but rather “to inform local planning and delivery of services for children.” Another important point for consideration, and one which has previously been alluded to in attempts to advance the debate around fitness testing for children, is that while assessment of academic progress is almost universally accepted, there remains strong resistance to school-based health and fitness assessments.

Fitness Testing as a Zebra

Based on the aforementioned concerns, as well as additional concerns regarding fitness tests validity, use of criterion standards or inappropriate use of data, a number of authors have challenged the value of youth fitness testing and likened it to a dead horse, from which “it is time to dismount.” Other authors, in keeping with the metaphor of a horse, have defended youth fitness testing, asserting that its value depends on the skill and training of the riders.
or the understanding that is not a horse, but a zebra that is being ridden. The zebra's stripes symbolize the potential for fitness testing to have multiple applications and outcomes, including a role in the development of "physical literacy, a multidimensional and interactive construct."

We argue that the principal aim of youth fitness testing, which is the mapping and tracking of trends in fitness, is of substantial value and provides data beyond that of BMI alone. This contributes to the surveillance of youth health and to the evaluation of the effectiveness of PA promotion strategies and can inform school curricula or policies. While fitness testing itself may not promote increased PA and fitness, improvements in fitness measures could detect increases in the quantity and quality of school PE or PA overall, and provide an noninvasive objective assessment of the physiological impact of interventions. Both cross-sectional and longitudinal fitness data may also provide convincing empirical data and thus impetus for legislative action to support programs aimed at PA promotion and obesity prevention.

Further, we believe that the delivery and contextualization of youth fitness testing and provision of feedback, or "the skill of the rider," could transform the horse into the zebra which can achieve improved health, as well as psychosocial and educational outcomes. We argue that youth fitness testing can be a vehicle both for the promotion of PA and the development of physical literacy and it can do so without undermining physical self-concept in overweight children. However, these outcomes are not an inevitable consequence of fitness testing and we urge for diligence when selecting fitness test batteries, as well as when providing feedback.

We now highlight issues which we believe are important for the design and delivery of youth fitness testing, with the aim of achieving objectives beyond surveillance while addressing concerns about negative psychosocial outcomes and lack of contextualization.

Fitness Testing and Physical Self-Concept

We acknowledge that fitness testing has the potential to negatively impact on the physical self-concept of the overweight (or unfit) child. Although there is little empirical evidence to support this notion, there are limited reports of negative psychosocial consequences of fitness testing, such as embarrassment or teasing by other children. More specifically, a survey of more than 2500 PE teachers, who administered the FITNESSGRAM in Texas, revealed that approximately 25% reported observing some negative consequences of fitness testing, while in another report, few PE teachers reported such incidents. Since physical self-concept is a determinant of PA, we emphasize that it is essential to minimize such negative experiences and ensure that for less active or overweight children—those for whom increased levels of PA and fitness are most important—fitness testing does not undermine physical self-concept. We further believe that this outcome is not an unavoidable consequence of the process and we argue that fitness testing could, on the contrary, be a means to enhance the physical self-concept in such children.

Fitness testing can also be a vehicle used to challenge pervasive and potentially harmful assumptions about the relationship between body size, health, and fitness. In particular, it challenges the notion that being overweight is incompatible with being fit and healthy, while thinness equates to good health and fitness. For example, Bacon and Aphramor recently reported that 51% of overweight adults had 'normal' cardio-metabolic health, whereas 24% of normal weight adults had 'abnormal' cardio-metabolic health, illustrating the limitations associated with using BMI alone as a proxy for health. Similarly, evidence from the Aerobics Longitudinal Study long ago introduced the 'fat-fit' phenomenon, whereby overweight but fit individuals had a lower risk of cardiovascular mortality than normal weight but unfit individuals. This phenomenon is also observed in youth, where higher fitness is associated with lower cardio-metabolic risk factors in children classified as normal and overweight by BMI. Furthermore, Parrett et al found that in children with body composition assessed by DEXA, higher % fat children with high aerobic fitness had significantly lower metabolic risk score than their low fitness, higher % fat counterparts.

Indeed, based on the work of Blair and others it is now widely agreed that differences in fitness and PA contribute greatly to the variability in cardio-metabolic health observed at any given BMI. To further illustrate this point, we cross-tabulated the prevalence of children with varying levels of CRF according to BMI weight status from our East of England Healthy Hearts Study dataset (Figure 1, Panel A). As expected, low CRF was typically higher in overweight and obese children compared with normal-weight children; yet, our data also revealed that ~50% of overweight and a smaller proportion of obese children were in the middle and top tertile of CRF. By not testing them, we deny 'fat-fit' children an opportunity to demonstrate their adequate levels of CRF and thus positively reinforce the healthy PA behaviors that presumably underlie this trait. In addition, the notion that fitness testing does not provide additional information to BMI reinforces the false assumption that normal weight or thinness equates to good fitness, or indeed health. According to our findings, the proportion of unfit normal weight children is lower than the proportion of unfit overweight children (Figure 1, Panel A), but overall, there was a greater number of normal weight unfit than overweight unfit children (n = 855 vs. n = 615). This further highlights the importance of communicating to normal weight children, and perhaps more importantly their parents, that their levels of PA or fitness are inadequate. To add to this, we previously reported significant secular declines in CRF and muscular fitness in English 10-year-olds over a 10-year period, even though BMI remained stable; these data further call into question the usefulness of measuring BMI alone to monitor youth health.

The Importance of Muscular Fitness Tests

It appears that in the debate in the UK over youth fitness and fitness testing, 'fitness' has become synonymous with cardio-respiratory fitness (CRF), with critics referring to the ad hoc way in which the 20-m shuttle-run is implemented in British schools. In this context, the multidimensional nature of fitness cannot be emphasized enough and an opportunity to give overweight children fitness-related positive feedback is potentially missed if only CRF and weight-bearing tests of muscular fitness are administered. Overweight children tend to outperform normal weight children in nonweight bearing tests of strength and power, such as hand grip strength (Figure 1, Panel B), medicine ball throw (Figure 1, Panel C), and medicine ball throw (Figure 1, Panel C). Morano et al reported a bolstering of physical self-esteem in overweight children when tested with the medicine ball throw; a measure of upper body power. We can anecdotally confirm observing a similar effect in the East of England Healthy Hearts Study for hand grip strength testing. Indeed, muscular strength is one of the only subcomponents of physical self-concept scales on which overweight children rate themselves highly. Highlighting these pupils' capabilities in this specific aspect of fitness through testing could be a way to positively reinforce a specific "athletic identity." This construct...
predicts sports participation and PA levels independently of gender, body mass, ethnicity, and perceived task competence. Therefore, the inclusion of nonweight bearing strength and power tests increases the likelihood of good physical performances and is critical to ensuring that children across the range of BMI values receive positive feedback on physical performance. This approach also more accurately reflects the multidimensional nature of fitness and athletic performance, enshrined in such sporting events as the Olympic Games where the athlete comes in a range of body shapes and sizes.

To conceptualize this point we have created an arbitrary 'athletic index' score; a composite score of CRF (20-m shuttle-run test z-score), strength (absolute hand grip z-score), vertical jump height (countermovement jump z-score) and peak power (z-score, derived from vertical jump height and body mass) for participants in the East of England Healthy Hearts Study (see Online Appendix for methodology of study). There was no significant difference in the prevalence of upper tertile athletic index between normal and overweight children (Figure 1, Panel D), although the contribution of each athletic index component differed by IOTF weight category (Figure 1, Panel A–C). On this basis, the inclusion of muscular fitness tests not only improves the social and educational experience for all children—and especially those who are overweight—but importantly, it also provides a more complete evaluation of health-related fitness, given the accumulating evidence associating muscular fitness with cardio-metabolic health.

Feedback and Expression of Data

In addition to the importance of carefully selecting the fitness testing battery, we also argue that careful consideration must be given to how the results are communicated to the child. Domangue and Solmon highlighted this when examining the association between 5th grade children's motivational constructs and attaining awards based on achieving norm-referenced fitness standards. Achievement goals, intrinsic motivation and future intention to participate were significantly lower in those who did not receive the awards. Domangue and Solmon suggested that this would serve to accentuate gaps in fitness between high- and low-performing children by influencing...
motivation to participate in sports. Implicitly, these findings also demonstrate the potential for positive fitness test feedback as a tool to enhance intrinsic motivation for PA. Positive feedback and fitness awards based on personal improvement rather than attainment of criterion referenced standards, might provide an alternative and more achievable source of PA related positive feedback. This may be particularly important in promoting physical self-competence in those children with negative experiences in team sports, which has in turn been identified as a predictor of reduced PA levels.55

While it is evident that fewer overweight children are high performers in weight-bearing fitness tests,49,56 results can be expressed in a number of ways. For example, in the assessment of CRF using the 20-m shuttle-run test, performance may be reported as levels or shuttles achieved, or as relative VO2peak predicted based on the number of shuttles completed (or corresponding running speed).17,49 Figure 2 compares a variety of performance values of 1 normal weight and 1 overweight 13-year-old boy who participated in the East of England Healthy Hearts Study. When their CRF test results were expressed as shuttle count, the normal weight boy completed more than twice as many shuttle laps as the overweight boy (74 shuttles (or level 8.9) vs. 34 shuttles (or level 5.2), respectively). Unsurprisingly, predicted relative VO2peak7 was also higher in the normal weight boy (45 ml·kg⁻¹·min⁻¹) compared with the overweight boy (39 ml·kg⁻¹·min⁻¹); however, when relative VO2peak values were converted to absolute VO2peak (l·min⁻¹) based on the individuals' body mass (kg), the normal and overweight boys achieved VO2peak values of 2.91 l·min⁻¹ and 2.82 l·min⁻¹, respectively. These comparably small differences between the normal weight and overweight child are in agreement with previous work showing that after adjusting for fat mass, there were nonsignificant differences in 20-m shuttle-run test performance between overweight and normal weight girls and higher performances in overweight boys.49 It also indicates that the overweight child is not necessarily characterized by a lower capacity for aerobic energy production, but their greater body mass limits their ability to express this in weight-bearing tests such as the 20-m shuttle-run test. Similarly, Figure 2 also compares jump height, as well as peak power, of the same normal weight and overweight 13-year-old boys who participated in the East of England Healthy Hearts Study. Peak power is calculated from vertical jump height, but is relative to body mass;26 although the overweight boy's vertical jump height was lower than that of the normal weight boy (38 cm vs. 33 cm, respectively), his absolute peak power was greater (2540 W vs. 2615 W, respectively). Offering feedback in absolute as well as relative values, thus offers another potential avenue for providing more positive feedback.

Moreover, the underlying calculations to determine relative as well as total performance outcomes could be performed by the children themselves within PE, math, science, and/or personal social education class and would enhance the educational value of fitness testing. Involving pupils in this process and developing their understanding of the interaction between different fitness measures, body size and physical capacity, as well as the underlying biology, serves several purposes. Apart from addressing the criticism of lack of context associated with the administration of fitness testing in schools,21 increased scientific content may also increase pupils' engagement in PE. Trudeau and Shephard55 suggested that a declining interest in PE in adolescence, particularly in girls, is associated more with the low academic prestige of the subject than with a lack of interest in PE per se. Developing the scientific content around fitness testing to develop pupils' understanding of how these measures relate to underlying biology, health and physical performance could enhance the status and interest in PE and be used to increase engagement in science. Indeed, under their remit for engagement with science, the Wellcome Trust58 launched the "in the zone" initiative; coinciding with the 2012 Olympic and Paralympic games, this initiative offers age-appropriate curricula materials for schools that combine physical performance and physiological measures. Similarly, the Active Science Scheme,59 which has been piloted in some US schools, also involves pupils measuring, calculating and understanding their PA levels by downloading and interpreting their own pedometer data.

Figure 2 — Performance portrait of a normal weight (left) and overweight (right) 13-year-old boy with identical athletic index scores from the East of England Healthy Hearts Study. Individual performance scores (text), and corresponding age-sex specific percentile scores (bars) based on East of England Healthy Hearts cross-sectional cohort. AI—athletic index; composite z-scores of cardiorespiratory fitness, hand grip, vertical jump and peak power. VO2peak—peak oxygen uptake; relative VO2peak predicted;57 total VO2peak derived from relative VO2peak and body mass; Hand Grip—dominant arm maximum grip strength; Vertical Power—calculated based on Taylor et al;26 PAQ-score—Physical Activity Questionnaire for Adolescents,69 scored between 1 (inactive) and 5 (very active).
Why not Simply Measure Physical Activity?

It has previously been argued that instead of testing fitness, efforts should be redirected to the measurement of PA. Yet the methodological difficulties of measuring PA in youth are well documented; self-reported PA is subject to questionable recall ability in youth, whereas the objective assessment of PA via accelerometry is still relatively expensive and time consuming, and equally not free from measurement limitations (ie, cycling). Anecdotally, we have observed that adherence to the accelerometer wear-time protocol in school-based studies is comparably poor, thereby not only potentially introducing bias, but also calling into question the usefulness of accelerometry as a tool to monitor youth health at the population level. More problematic still is the ongoing scientific debate over various technical issues related to accelerometry, such as appropriate choice of manufacturer and model, epoch length, inclusion criteria for what constitutes a valid file, and ultimately what accelerometry count cut-points offer meaningful interpretation and information on PA. In contrast, the feasibility and reliability of standardized large scale and school-based fitness assessment by teachers or other trained individuals is well documented in the UK and internationally. Importantly, while fitness and PA are independently associated with health, studies in adults and children suggest that fitness is more strongly associated with cardio-metabolic risk than is PA. Given these health-associations, we argue that any cross-sectional, longitudinal or secular assessments of youth health should include a measure of CRF as well as at least 1 component of muscular fitness.

Conclusion

In the UK and many other countries, youth health monitoring, if carried out at all, relies on assessments of BMI. Yet, secular trends in BMI may mask a concurrent loss of lean mass and gain in fat mass, and furthermore, may not accurately describe population health or secular trends thereof. Irrespective, an exclusive emphasis on weight (or BMI) in children is of concern for various obvious reasons. We argue that the troublesome decline in health-related fitness is a major public health concern, and urge to give the UK's former CMO's recommendation for routine fitness testing in schools due consideration.

All the same, we agree that valid concerns have been raised for the vast potential for negative experiences of fitness testing, particularly for the overweight and/or less fit individuals; poor performance, in the wrong context, can be detrimental to physical self-concept, and reinforce low PA. However, we believe that this is not an inevitable consequence of fitness testing and depends on the skills and knowledge of the test administrators, as well as the context in which fitness testing is carried out. Indeed, it is thanks to the ongoing debate—academic and public—that we can suggest ways by which to overcome concerns, and furthermore, recognize the opportunities for effective and engaging health education, as well as health- and PA promotion that comes with school-based fitness testing.

For effective school-based testing, we suggest a multidisciplinary approach involving educators, health professionals and academics from sport science and psychology, all of which maintain an up to date understanding of the association between various components of fitness, body composition and health. The large scale school-based testing conducted by UK Universities of Essex and Liverpool John Moores, which involve sports science (or similar) undergraduate and graduate students, as well as fitness industry trainees, as part of the vocational/research elements of their training, provide an economical model of large-scale implementation, that also minimizes additional demands on schools and PE teachers. More importantly, perhaps, is that establishing such university-school relationships then provides a framework for the development of health- and PA-promotion programs, for which schools are an ideal setting. There needs to be effective communication of the documented associations between health, CRF and muscular fitness to physical educators and health professionals working with children as part of initial or continuing training, an understanding which in turn would be passed on to parents and children. Such knowledge is essential for successful implementation of fitness testing and health-promotion through physical activity. There is clear need for a greater emphasis on routes to health that are not directed solely by measurements of body weight, where an over-emphasis on weight must be considered alongside the risk of promoting the development of eating disorders.

In summary, we strongly believe that a carefully chosen fitness test battery, which is delivered in an educational context, can transform the horse into the zebra, and it is time for the skillful rider to mount and ride it.

Acknowledgments

We wish to thank all of the pupils and PE teachers who participated in the East of England Healthy Heart Study.

References


28. Sandercock GRH, Voss C, Dye L. Associations between habitual school-day breakfast consumption, body mass index, physical activity and


69. Kowalski KC, Crocker RE, Donen RM. The Physical Activity Questionnaire for Older Children (PAC-C) and Adolescents (PAQ-A) manual. Saskatoon, Canada: University of Saskatchewan; 2004.