FITNESSGRAM® /ACTIVITYGRAM® **Reference Guide (4th Edition)**

Editors

Sharon A. Plowman and Marilu D. Meredith

Suggested citations from this source:

Reference to a chapter (sample) in the book:
Plowman, S.A. (2013). Muscular Strength, Endurance, and Flexibility Assessments. In S. A. Plowman & M.D. Meredith (Eds.), *Fitnessgram/Activitygram Reference Guide (4th Edition)* (pp. Internet Resource). Dallas, TX: The Cooper Institute, 8-1 - 8-55.

Reference to the whole book:
Plowman, S.A. & Meredith, M.D. (Eds.). (2013). *Fitnessgram/Activitygram Reference Guide (4th Edition)*. Dallas, TX: The Cooper Institute.

Document Navigation:
At the bottom of each page two links will assist in navigating this document:
⇦TOC - links back to the Table of Contents
⇦Chapter - links back to the beginning of the current Chapter

Chapter 1

The History of FITNESSGRAM®

Sharon A. Plowman, Marilu D. Meredith, Charles L. Sterling, Charles B. Corbin, Gregory J. Welk, James R. Morrow, Jr.

(Note: This manuscript is an update of a previously published article: Plowman, S.A., Sterling, C.L., Corbin, C.B., Meredith, M.D., Welk, G.J., and Morrow, J.R., Jr.: (2006). The History of FITNESSGRAM®. *Journal of Physical Activity & Health*, 3 (Supplement 2), S5-S20. Additions/modifications made with permission of the publisher, Human Kinetics, Champaign, IL

⇦TOC ⇦Chapter

Abstract

Initially designed by Charles L. Sterling as a physical fitness "report card," FITNESSGRAM® is now the educational assessment and reporting software portion of the Presidential Youth Fitness Program. Based on physiological/epidemiological, behavioral, and pedagogical research, FITNESSGRAM® is committed to health-related physical fitness, criterion-referenced standards, an emphasis on physical activity, and the latest in technology. The evolution of these major concepts including the inclusion of ACTIVITYGRAM® and NUTRIGRAM® is described in this history of FITNESSGRAM®.

Introduction

The concept for FITNESSGRAM® had its beginning in 1977 when Charles L. Sterling, Ed.D., the Director of Health and Physical Education of the Richardson, Texas school system, recognized school administrators' and parents' interest in a physical fitness "report card" similar to those used in other educational areas. He also recognized the potential for using computers to print reports and keep student records. Sterling and teachers Marilu Meredith, Nancy Voith, Cindy Raymond, and Don Rainey administered the Texas Physical Fitness—Motor Ability Test (Governor's Commission on Physical Fitness, 1972) in their schools. Personalized fitness report cards were then generated for all students using customized software developed for the school district's mainframe computer.

In 1981, Dr. Sterling joined the staff of the Cooper Institute for Aerobics Research (CIAR/IAR/CI) in Dallas. The Institute had a mainframe computer that allowed batch processing of the physical fitness reports. This created the opportunity to take the concept to a wider audience, but a name was needed. FITNESSGRAM® was chosen through a contest in the local school district—Nancy Voith is credited with the winning entry. This name played off the concept of a telegram and suited the intended purpose of the report—namely, to communicate important fitness information to children and parents. The Campbell Soup Company's Institute for Health and Fitness signed on as a national sponsor to support the promotion and dissemination of the tool. Dr. Marilu Meredith was hired as National Project Director in 1982, a position she held until 2012. At that time Dr. Meredith became the Director of the Perot International Youth Data Repository with part-time FITNESSGRAM® duties. In 2013 Catherine Vowell assumed the directorship of FITNESSGRAM®.

FITNESSGRAM® (FG) was implemented in phases with the first pilot conducted in 30 schools in the Tulsa, Oklahoma School District (1982-83) using the AAHPERD Youth Fitness Test (YFT) (American Alliance for Health, Physical Education, and Recreation [AAHPER], 1976; Lacy & Marshall, 1984; Razor, 1984). In the second year (1983-84), approximately 125 schools throughout Oklahoma participated and were able to select either the AAHPERD YFT or the AAHPERD Health Related Fitness Test (HRPFT) (American Alliance for Health, Physical Education, Recreation and Dance [AAHPERD], 1980). After these successes, FG was implemented on a national basis first as a pilot, one district per state in addition to OK (1984-85), and then unrestricted (1985-86).

Now, in 2013, FITNESSGRAM®/ACTIVITYGRAM® (FG) is an educational assessment and reporting software system that has been used by thousands of teachers with millions of youth in schools worldwide to help teachers track health-related fitness and physical activity information over time and produce personalized reports for children, parents, and school administrators. In conjunction with a variety of partners, FG has pushed the evolution of physical

⇦TOC ⇦Chapter

fitness and physical activity philosophy, research, evaluation, education, and promotion. This evolution has occurred in four major areas:

1. A commitment to the concept of health-related physical fitness,
2. A concentration on criterion-referenced evaluation in place of percentile norm-referenced evaluation,
3. A consistent emphasis on fitness behavior/physical activity, and
4. Systematic updating and sophistication of the computerized reporting system.

A Commitment to Health-Related Physical Fitness

At its inception FG was not a test battery. The route to embracing the current health related battery reflects the basic changes that were made in the last half of the twentieth century in the concept of physical fitness.

The history of youth physical fitness testing from approximately the 1860s to 1988 is described in a government document prepared by Roberta Park (1988) and much of it need not be repeated here. Suffice it to say that since its inception in 1885, the organization now known as the American Alliance for Health, Physical Education, Recreation and Dance (AAHPER/AAHPERD) was deeply concerned with the physical fitness of American youth. Formal governmental involvement began in 1956 when President Dwight David Eisenhower established the President's Council on Youth Fitness (PCYF) [name later changed to President's Council on Physical Fitness (PCPF), then the President's Council on Physical Fitness and Sports (PCPFS) and now the President's Council on Fitness, Sports and Nutrition (PCFSN)] in response to published data that American children were less fit than European children (Kraus & Hirschland, 1953; 1954). Shortly thereafter the Research Council of AAHPER agreed on a battery of tests and the AAHPER Youth Fitness Project, a nationwide pilot study of the fitness levels of 5-12 year old boys and girls headed by Dr. Paul Hunsicker, was completed. As a result the AAHPER YFT Manual with national norms was published in 1958. The test items included pull-ups for boys/ modified pull-ups for girls, straight leg sit-ups, shuttle run, standing broad (long) jump, 50-yd dash, softball throw for distance, 600-yd run/walk, and three aquatic tests that were rarely used. In 1966 the then President's Council on Physical Fitness, at the direction of President Lyndon Johnson, established a Presidential Award Program based on AAHPER's YFT. Criteria for this award for youngsters between 10 and 17years included being in good academic standing, a recommendation from the school principal, and scoring in the 85th percentile on all seven items. Between 1958 and 1975 minor changes were made in the test items and norms (AAHPER, 1965). However, by the early 1970s there was mounting dissatisfaction with the actual test items and philosophy behind the test and award program from both practitioners and researchers. In 1972-73, the Texas Physical Fitness-Motor Ability Test (Governor's Commission on Physical Fitness, 1972) was developed. At the same time a joint committee from the Measurement and Evaluation Council (MEC) and Physical Fitness Council (PFC) of AAHPER, chaired by Dale Mood and then Mike Reuter, was put in place to "recommend…appropriate activities concerning testing of components of physical fitness" (D. Mood, personal communication, November 14, 1972). The committee was usurped when a small group of AAHPER and PCYF people, apparently at the initiation of the AAHPER staff, changed the YFT. The Texas distance run items and norms were incorporated as options, the California version of a one minute flexed knee sit-up replaced the straight leg sit-up, and the softball throw

⇦TOC ⇦Chapter

for distance was deleted for the 1975 AAHPER Youth fitness survey (and 1976 published manual) (AAHPER, 1976).

In 1975, a joint committee was established to systematically study whether the AAHPER YFT needed major revision. Don Franks, Frank Katch, Vic Katch, Sharon Plowman, Margaret J. Safrit, and Andrew Jackson (chairperson), representing the MEC, PFC, and Research Council (RC, later renamed Research Consortium) of AAHPER, comprised the committee. Dr. Ray Ciszek attended the meetings as the AAHPER staff liaison and Dr. Ash Hayes was invited to represent the PCPFS (AAHPER, 1977; Plowman & Falls, 1978; Plowman & Falls, 1979). After extensive review of the literature, much discussion, open hearings at the 1976 national convention, and solicitation of opinions from colleagues, A Position Paper on Physical Fitness was submitted to AAHPER. This position paper called for a revision of the AAHPER YFT and set as a basic goal the relating of physical fitness to functional health and not sport performance. A six member Task Force on Youth Fitness was appointed in 1977 to implement the proposals made in the position paper. Members of this task force included Steven Blair, Charles Corbin (who resigned after the initial meeting and contributions; replaced by Don Franks), Andrew (Tony) Jackson, Michael Pollock, Margaret (Jo) Safrit, and Harold Falls (chairperson). Ray Ciszek served as the AAHPER staff consultant (AAHPER, 1977; Plowman & Falls, 1978; Plowman & Falls, 1979; AAHPER, 1978). Throughout 1978 the task force established goals and gathered information. Consultants, who joined the task force members for the 1979 meetings to finalize the test items, identify normative sources and establish norms, and write the manual, included Charles Dotson, Dennis Humphrey, Tim Lohman, Russ Pate, Sharon Plowman, and Glen Swengros (PCPFS) (Plowman & Falls, 1979; Falls, 1979). Additional input was obtained from Gary Krahenbuhl, William Stone, Kirk Cureton, Robert Serfass, Ed Burke, Frank Katch, Vic Katch, and Ash Hayes (AAHPERD, 1980). The components and items agreed upon were cardiorespiratory function (1 mi/9 min or 1.5mi/12 min), body composition (triceps or sum of triceps and subscapular skinfolds), and abdominal and low back-hamstring musculoskeletal function (bent knee, timed sit-ups; sit-and-reach). Although the task force recommended to AAHPER that the YFT items be relegated to an optional appendix in the new manual (and a study be undertaken for the performance related motor fitness items), that the new test be called the AAHPERD Physical Fitness test, and that the current award system be eliminated (Plowman & Falls, 1979), the Board of Governors did not concur. The result was the 1980 publication of the AAHPERD Health Related Physical Fitness Test Manual (AAHPERD, 1980) and continuance of the AAHPER YFT and Presidential Award system (theoretically, both for a period of two years). FG continued to support both tests. Thus, during this time, AAHPERD, PCPFS, and the CIAR worked together.

In 1984, a Technical Manual (AAHPERD, 1984) for the Health Related Physical Fitness test (HRPFT) was published. Also in 1984, a report of the AAHPERD RC Committee to Evaluate the Two-Test System, chaired by Ed Burke, reiterated the recommendation that the HRPFT be made the primary test with the non-overlapping YFT motor fitness items combined into a second part of the testing manual. A five year transition phase ending in 1989-1990 was suggested (E. Burke, July 11, 1984).

As a result of the Burke committee report, yet another AAHPERD task force was appointed in 1985. This task force (Manual Task Force), chaired by Harold Falls and made up of members of the RC, MEC and PFC, was charged with developing a single AAHPERD fitness test battery, establishing criterion-referenced standards, examining the existing awards schemes, and writing the appropriate manual (E. Haymes, personal communication, June 26, 1985). At the same time, the PCPFS was conducting the 1985 School Population Survey and developing its own youth fitness test and awards program. In late 1985 to early 1986, the AAHPERD Executive Committee was approached and given the “opportunity” to approve, publish, and promote the new PCPFS fitness test as well as continue to administer the Presidential Award (B.D. Franks, personal communication, June 1, 1986; June 24, 1986b). The Manual Task Force was asked to “advise the Alliance on the data and test items included in the population fitness survey for updating and inclusion of the AAHPERD Youth Fitness Test” (A. Annarino, personal communication, March 4, 1986). A series of phone discussions failed to produce an agreement from January to April 1986 (B.D. Franks, personal communication, June 24, 1986a). The Manual Task Force had scheduled its first meeting to take place at the AAHPERD national convention on April 10, 1986 and invited representatives from the PCPFS and CIAR to attend for discussions in an attempt to reach a compromise. However, on April 9, the PCPFS distributed its new fitness test and awards flyer: Fitness Testing and the Presidential Physical Fitness Award (The President’s Council on Physical Fitness and Sports [PCPFS}, 1986; B.D. Franks, personal communication, April 6, 1987). Selected test items included pull-ups for boys and flexed arm hang (FAH) for girls, sit-ups, one-mile run, shuttle run, and sit-and-reach. The creation of a test and awards system by the PCPFS represented a change in policy. Prior to 1984, the PCPFS had left the decision on test composition and the selling of awards (a major source of revenue) to AAHPERD, although the PCPFS had determined the criteria for the Presidential award (B. Orr, personal communication, March 14, 1986; J. Razor, personal communication, May 22, 1984).

Despite the unilateral and unexpected action of the presentation of a new test by the PCPFS, representatives from the Manual Task Force met with representatives of the PCPFS as previously scheduled. Complete agreement could not be reached. Following the meeting the Manual Task Force recommended to the AAHPERD Executive Committee that AAHPERD support the HRPFT and not the PCPFS test and award system. This decision and the reasons for it were communicated directly from Harold Falls to Ash Hayes (H. Falls, personal communication, April 23, 1986). Specifically, the four major concerns were the omission of any item to measure body composition, the continued use of the 85^{th} percentile for the Presidential award, the inclusion of the shuttle run, and the choice of the items used to measure upper arm and shoulder girdle strength and endurance. FG remained committed to whatever AAHPERD decided (B.D. Franks, personal communication, April 21, 1986). Negotiations among the Manual Task Force (and the councils the members represented), the AAHPERD Executive Board, and Board of Governors, as well as among AAHPERD, PCPFS, and CIAR continued throughout the spring of 1986 (B.D. Franks, personal communication, June 24, 1986b). The perceived lack of commitment from the AAHPERD leadership to the health-related physical fitness concept, the awkward and time consuming decision making process utilized by the AAHPERD structure, and the overwhelming financial considerations linked to the awards led concerned members of AAHPERD to hold several meetings at the annual American College of Sports Medicine Meetings in Indianapolis. At one of these meetings Steve Blair, Harold Falls, Patty Freedson, Don Franks, Dennis Humphrey, Tim Lohman, Pat McSwegin, Jim Morrow, Russ Pate, Sharon Plowman, and Jack Wilmore decided that they would work to provide the best physical fitness test to this nation whether through AAHPERD or other avenues (B.D. Franks, personal communication, June 24, 1986a). The Manual Task Force (on which many of these individuals served) continued to work.

⇦TOC ⇦Chapter

In July the AAHPERD Executive Committee and PCPFS agreed to form a Joint Task Force made up of the respective presidents and chairs of the RC (Don Franks), MEC (Jim Morrow) and PFC (Sharon Plowman), with Ash Hayes (Executive Director), Guy Reiff and Bill Savage representing the PCPFS, and Barbara Lockhart (President), and Hal Haywood (Acting Executive Vice President) representing AAHPERD ex officio. The Task Force was charged with finding a compromise solution for a fitness test and award system that would be endorsed by AAHPERD, PCPFS, and CIAR (B. Lockhart, personal communication, June 14, 1986). This Joint Task Force called for the presentation of position statements from any interested professionals at hearings that were held in Chicago, IL October 3-4, 1986 (B. Lockhart, personal communication, August 22, 1986). Immediately after the hearings the Joint Task Force met and devised a plan that appeared to be agreeable to all parties (Hayes, 1986). However, the plan was never ratified. Part of the difficulty was that CIAR/Campbell Soup could not support the plan, and FG was integral to the AAHPERD Manual Task Forces' recommendations (B.D. Franks, personal communication, April 6, 1987). The "compromise" required computer programming of two tests (albeit with some overlapping items, but with a total of nine different ones) with norms that were to be criterion-referenced for AAHPERD awards and percentile-referenced for the Presidential Award (Hayes, 1986). In a letter (C. Sterling, personal communication, October 30, 1986), Charles Sterling informed AAHPERD that the Campbell Soup Company had informed CIAR that "…reprogramming more than one test is not an economic reality." Thus, a decision had been made "in house" (Charles Sterling, Lee Dukes, Marilu Meredith, Steve Blair) (M. Meredith, personal communication, July 11, 2005) to utilize a single test consisting of five items: one mile run, modified sit-up, sit-and-reach, pull-up/flexed arm hang (either sex), body composition (grades 4-12) assessed by triceps and calf skinfolds (default to body mass index [BMI] if no skinfolds taken), and an optional shuttle run for K-3. Thus, the first FG test battery was established. AAHPERD and the PCPFS were invited to adopt the new test and program. Dialogue continued among the three parties until time simply ran out. The Manual Task Force ceased writing in December, 1986 (H. Falls, personal communication, December 8, 1986), material was returned to the members in February (B.D. Franks, personal communication, February 26, 1987), and the committee was formally disbanded by AAHPERD in March, 1987 (B. Lockhart, personal communication, March 19, 1987).

On February 23, 1987 Charles Sterling informed Hal Haywood (C. Sterling, personal communication, February 23, 1987) that "The institute must now move forward with the finalization of test methodology, manual writing, and refinement of the awards program. We will, as of today, begin contacting content experts to form an advisory committee to contribute to this effort." The initial meeting of the FITNESSGRAM® "advisory committee," later changed to Advisory Council, and currently called the Scientific Advisory Board was held in Atlanta, GA, March 9-10, 1987. Persons attending were Steve Blair, Lee Dukes (Campbell Soup), Chuck Corbin, Harold Falls, Tim Lohman, Marilu Meredith, Jim Morrow, Russ Pate, Sharon Plowman, Charles Sterling, and Katie Stone (Campbell Soup) (FITNESSGRAM® minutes, March 9, 1987). Kirk Cureton, also a founding member of the Advisory Council, did not attend. The CIAR and individual members of the FG scientific advisory board were committed to health-related fitness based on research evidence that would dictate the test items and program. The unity of purpose and ability to move quickly on decisions was instantly apparent. Material prepared by the former AAHPERD Manual Task Force members and others were discussed at great length and used as the basis for developing the new FITNESSGRAM® Test Administration Manual (Cooper Institute for Aerobics Research [CIAR], 1987). This Advisory Council provided the scientific core from 1987 to the present while adding specialists to enhance the group as the need arose and replacing individuals due to retirement and individual circumstances. A complete list of the advisors and their years of service is included in Table 1. Table 2 summarizes the various committees that ultimately lead to the formation of the FG Advisory Council.

Table 1. FITNESSGRAM® Scientific Advisory Board

Advisory Council Member	Dates Served
Steven N. Blair	1987-2007
Dave B. Buller	2003-2004
Chuck B. Corbin	1987-2011
Kirk J. Cureton	1987-
Don L. Disney	2012
Joey C. Eisenmann	2012-
Harold B. Falls	1987-2005
Scott B. Going	2005-
Baker C. Harrell	2010-
Harold W. Kohl	1993-1996
Dolly D. Lambdin	2008
Timothy G. Lohman	1987-2005
Matthew T. Mahar	2008-
Marilu D. Meredith	1982-
James R. Morrow, Jr.	1987-
Robert P. Pangrazi	1993-2005; 2007-2011
Russell R. Pate	1987-2011
Sharon A. Plowman	1987-
Stephen J. Pont	2013
Judith J. Prochaska	2005-
Sara Jane Quinn	2006-2009
Georgianne Roberts	2010-
Margaret J. Safrit	1989-1995
James F. Sallis	1989-2004
Charles L. Sterling	1977-2002; 2008-2012
Catherine L. Vowell	2013-
Gregory J. Welk	1996-
Weimo Zhu	2003-

Table 2. The Road to the FITNESSGRAM® Scientific Advisory Board

Date	Name of Committee	Charge	Outcome
1975	Joint Task Force to Study Revision of the AAHPERD Youth Fitness Test (YFT)	To determine if the YFT needed revising	Position Paper on Physical Fitness recommending switch to health-related physical fitness
1977	Task Force on Youth Fitness	To implement recommendations for	1980 AAHPERD Health Related Physical Fitness Manual

		revision to health-related physical fitness	
1985	Physical Fitness Test Manual Task Force	To develop a single AAHPERD fitness battery, establish criterion-referenced standards, examine awards, write manual	Disbanded
1986	Joint AAHPERD-PCPFS Task Force	To find a test and award system that represented a compromise and could be endorsed by AAHPERD, PCPFS, and CIAR	Unsuccessful
1987	FG Advisory Council	To devise a health-related fitness test, criterion-referenced standards, physical activity promotion and reporting system	FITNESSGRAM®/ACTIVITYGRAM®

The goal of a unified test was not achieved in the 1980s due to philosophical differences between the PCPFS, AAHPERD and the CIAR. The net result was that, for the first time, FG had a test battery and was developing standards and awards; the PCPFS had its own President's Challenge Test (Association for Research, Administration, Professional Councils and Societies [ARAPCS], 1987) and awards program; and AAHPERD continued with both the HRPFT and the YFT until publication of Physical Best (PB) test in 1988 (AAHPERD, 1988). The PB test battery included the one-mile run, sum of triceps and calf skinfolds, sit-and-reach, modified sit-ups, and pull-up/modified pull-up. Several years later AAHPERD developed its own fitness reporting system as part of the PB program. Obviously, there was considerable overlap between the tests and the philosophies of FG and PB. In 1991 Prudential Insurance began its six-year sponsorship of FITNESSGRAM®. In December 1993, a strategic partnership was formed between AAHPERD (represented by Mike Davis, president) and CIAR/Prudential (represented by Charles Sterling, Bill Kohl [FG Scientific Director], Marilu Meredith, and Don Southwell [Prudential president]). FG was designated the fitness and activity assessment and reporting program and AAHPERD's PB became the education program (C. Sterling, personal communication, December, 20. 1993). In 1999, the cooperative work of Charles Sterling, Marilu Meredith, Greg Welk (FG Scientific Director) and Steve Blair from CIAR, Mike Davis and Gayle Claman of AAHPERD, and Rainer Martens and Scott Wikgren of Human Kinetics Publishers resulted in an ongoing agreement for Human Kinetics to publish, market, and distribute all FG materials. Human Kinetics was already publishing the PB materials so this agreement brought together an array of educational resources to support youth fitness and activity programming. At one point, the PCPFS included an optional health-related fitness test as part of the President's Challenge (PCPFS, 2005), but as of 2005 their test and FG continued to operate independently.

⇦TOC ⇦Chapter

FITNESSGRAM® was adopted by more and more individual school districts. As concerns deepened about rising obesity levels major metropolitan areas and whole states began mandating fitness testing for children and adolescents. Among those who mandated FG were California (1996), New York City (2006), Texas (2007), Delaware (2008), and Georgia (2010). In December of 2009, NFL Charities announced a three-year (subsequently renewed through at least 2015) grant to fund FITNESSGRAM® assessment in more than 1,120 schools nationwide (35 schools for each of the 32 teams). Play60 is the NFL's national youth health and fitness campaign, focused on making the next generation of kids active and healthy by encouraging them to be active for at least 60 minutes per day. Each selected school receives a FITNESSGRAM® license, access to a website promoting physical activity, best practices, etc., and is part of a longitudinal study (E. Snyder, personal communication, December 15, 2009; C. Sterling & M. Meredith, personal communication, December 17, 2009).

Sporadic and nonproductive talks occurred between the President's Council and the Cooper Institute from 2006 through 2010. Finally in September, 2011, Cindy Sessions (former president of the HopSports Corporation, a professional colleague of the new Director of Youth Initiatives at the Cooper Institute, Don Disney, and friend and colleague of Shellie Pfohl, now Director of the PCFSN) facilitated a dialog between the CI and the PCFSN. At an expanded meeting, in January 2012, the official idea of a partnership/alliance was formalized by leaders from selected nonprofits, government agencies, educational associations, educational vendors, and award products administrators (D. Disney, Report to FG Scientific Advisory Board, June 21, 2012). After many meetings and much negotiation this idea ultimately became the President's Youth Fitness Program (PYFP)—launched officially in September, 2012.

The PYFP is a partnership of five non-profit and government organizations, each of which brings unique strengths to the relationship:

1. The American Alliance for Health, Physical Education, Recreation and Dance provides the lead in staff development and teacher training;
2. The Amateur Athletic Union (AAU) administers the youth fitness recognition (awards) system structure;
3. The Centers for Disease Control and Prevention (CDC) is responsible for developing a plan of national surveillance to track and evaluate the PYFP;
4. The Cooper Institute provides FITNESSGRAM® as the official fitness assessment tool and has designed a series of educational vignettes and,
5. The President's Council on Fitness, Sports, & Nutrition brings its brand recognition and promotes, facilitates, and motivates individuals to participate in the programs. Human Kinetics continues as the official publisher. Individuals interested in more information about the PYFP are referred to the website: www.presidentialyouthfitnessprogram.org.

Plans for FITNESSGRAM® International, renamed the Cooper International Youth Fitness Test to avoid any confusion with the use of the term "gram"—a unit of mass in the metric system), began in 2009. In 2010, the test kicked off in China. In 2013, a partnership agreement was signed with the Hungarian School Sport Federation to establish a national platform for children's fitness assessment in that country (http://cooperinstitute.org/pub/news.cfm?id=146, accessed 7/15/13).

It was fitting that as FG celebrated its 30th anniversary in 2012, Charles L Sterling, Founder of FITNESSGRAM®, received President's Council Presidential Lifetime Achievement

Award. Don Disney was appointed Director of Youth Initiatives, which oversees FG for CI, in 2012.

A Concentration on Criterion-Referenced Standards (CRS)

Normative referenced standards (NRS) rank an individual's performance relative to the performance of all other individuals in the group used for reference. The make-up of the reference group is critical, especially in terms of physical activity and health history in relation to physical fitness standards, and part of the question has always been whether these norms should be based on what the population can currently do, or whether only trained individuals should be tested to represent goals. The AAHPER(D) Youth Fitness Test was scored and the Presidential Award recipients determined on the basis of percentile normative standards. The 85th percentile award standard of the latter was a source of controversy throughout the 1970s and 1980s and remained so until introduction of the PYFP in 2012, although lower percentile awards and criterion-referenced health-related awards were at one point available from the PCPFS (PCPFS, 2005). Despite the 1977 Task Force recommendation, the AAHPERD Health Related Physical Fitness Test (AAHPERD, 1980) also utilized percentile rather than criterion-referenced standards (CRS). Teachers were encouraged, however, to interpret the test results following guidelines that functioned as CRS. The first true CRS were developed in 1978 for the South Carolina Physical Fitness Test (Pate, 1983). Currently, as part of the Presidential Youth Fitness Program, all fitness test results are evaluated against CRS and students who score in the CRS healthy fitness zone in five out of six fitness tests are eligible to be recognized with the Presidential Youth Fitness Award.

In 1987, the FG Scientific Advisory Board established CRS for the mile run, %BF/BMI, sit-and-reach, sit-ups, pull-ups, and FAH (CIAR, 1987; Sterling, 1988). These standards set one cut-off point. Scores above the cut-off were classified as acceptable; no label was associated with scores below the cut-off. The cut-off points were based on empirical data, normative data, and the professional judgment of the advisory council members (Cureton & Warren, 1990). They were intended to set a specific minimal level of performance on each test item that was consistent with acceptable good health (minimal disease risk) and adequate function (the ability to carry on with tasks of daily life) independent of the population tested, or the proportion of the population that meets the standard (CIAR, 1987). The FG CRS were the first for youth fitness that were put into widespread national and international use. In the 1992 Test Administration Manual (CIAR, 1992), healthy fitness zones (HFZ) replaced the single cut-off score. Results of each test could then be evaluated as "Needs improvement" (NI), in the "Healthy Fitness Zone" (HFZ), or above the Healthy Fitness Zone.

In 2010, new aerobic capacity and body composition standards, more closely linked with health outcomes (Morrow, Going, & Welk, 2010), were introduced with three categories of evaluation: "Healthy Fitness Zone," "Needs Improvement-Some Risk," and "Needs Improvement-High Risk" (Meredith, & Welk, 2010). In 2013, these new standards were refined and the terminology modified to "Healthy Fitness Zone," "Needs Improvement," and "Needs Improvement-Health Risk". The goal remains achievement of the HFZ for all students, but it is still recognized that scores higher than the upper limit of the HFZ are both attainable and healthy (with the possible exception of excessive leanness).

CRS for health-related physical fitness require both criterion and field test items that relate to health status and function. They also require scores that are responsive to health status and physical activity. This means that the reliability and validity not only for the field test items,

⇦TOC ⇦Chapter

but also for the criterion referenced standards must be established. Additional criteria for field tests include objectivity and ease of administration. Research is the basis for providing this information and to this end the Cooper Institute has funded a number of studies. A listing is presented in Table 3. In addition, three supplements to research journals have been published on strengthening the scientific basis of FG (Blair & Morrow, 2006), the Texas Youth Fitness Study (Martin & Morrow, 2010) and the development of criterion-referenced standards for aerobic capacity and body composition (Morrow, Going, & Welk, 2011).

Table 3. FITNESSGRAM® Research Studies Funded by the Cooper Institute and Supplements Published

2006	• Validation of FITNESSGRAM® Aerobic Fitness Components and ACTIVITYGRAM® in 10- to 11-year-old Children; Matthew Mahar • ROC Analysis of fatness measures: Application to FITNESSGRAM®; Joe Eisenmann • Primer-Test Centered Equating Method for Cut-off Score Setting; Weimo Zhu • Blair, S.N, & Morrow, J.R. ,Jr. (editors) (2006). *The Journal of Physical Activity and Health*, 2, Supplement 2.
2007	• Assessment of FITNESSGRAM® Body Composition Standards in 12-14 year old Youth; Matt Mahar
2008	• Derivation of aerobic fitness cutpoints using LMS and ROC analyses: application to FITNESSGRAM®; Greg Welk • The Cooper Back Extension Study; Weimo Zhu
2009	• Development of a Model to Predict Aerobic Fitness from PACER Performance; Matt Mahar • Establishment of Longitudinal Database FITNESSGRAM® and Other Measurements; Kelly Laurson
2010	• FITNESSGRAM® Trunk Test Research; James Hannon • Martin, S.B., & Morrow, J.R., Jr. (editors) (2010): *Texas Youth Fitness Study. Research Quarterly for Exercise and Sport*, 81(3): Supplement.
2011	• Setting Equivalent Cut-off Scores for FITNESSGRAM's One-Mile Walk Test; Weimo Zhu • Morrow, J.R., Jr., Going, S. B. & Welk, G.J. (2011): FITNESSGRAM®: Development of Criterion-Referenced Standards for aerobic capacity and body composition. *American Journal of Preventive Medicine*. 41(4): Supplement 2.
2012	• Development of a Walking Test to Predict Fitness in Children,; Matt Mahar • Evaluating the Impact of Body Composition on Aerobic Capacity Standards; James Hannon • FITNESSGRAM® BMI Diagnostic Comparison; Kelly Laurson

Available physiologic and psychometric research from all sources on each individual item in the FG battery was presented first in 1994 (Morrow, Falls, & Kohl), published online in 2001, and updated in online versions of the FITNESSGRAM® Reference Guide in 2003, 2008 (Welk & Meredith), and 2013. Areas of needed research are constantly being explored and, as information is available, test items and standards are changed. Table 4 presents a listing of the deletions and additions of test items that have occurred from 1987 to 2005. Cognate chapters which follow in this Reference Guide explain the development of the criterion reference standards for aerobic capacity, body composition, and the various musculoskeletal test items.

Table 4. Additions and Deletions to the FITNESSGRAM® Health-Related Fitness Battery 1987-2005

Fitness Component	Test Item	Year Included	Year Deleted
Aerobic Capacity	One Mile Run/Walk	1987	
	PACER (20 meters)	1992	
	PACER (15 meters)	2007	
	One Mile Walk Test	1999	
Body Composition	Skinfold Measure of Percent Body Fat	1987	
	Body Mass Index (Height and Weight)	1987	
	Portable Bioelectric Impedance Analyzers	2004	
Muscular Strength and Endurance	Modified Sit-up Test	1987	1992
	Curl-up Test	1992	
	Pull-up	1987	2005
	Flexed Arm Hang	1987	
	90^0 Push-up	1992	
	Modified Pull-up	1992	
	Trunk Lift	1992	

A Consistent Emphasis on Fitness Behavior and Physical Activity

The emphasis on fitness behavior and physical activity is seen in three major areas of the FG program: the "award" structure, the development of ACTIVITYGRAM®, and the development of the Youth Activity Profile. In 1987, FG offered awards that were labeled as such (CIAR, 1987). By 1992 (CIAR, 1992), however, the decision had been made to not use an "award" system, but instead to institute a "recognition" system. Because maintaining good fitness depends upon establishing patterns of regular physical activity, activity participation should be reinforced. In the 2004 manual (Meredith & Welk, 2004) these ideas were formalized into the HELP philosophy. The essence of the HELP philosophy is that Health comes from regular physical activity and the development of health-related physical fitness is for Everyone for a Lifetime and it should be designed to meet Personal needs (Meredith & Welk, 2004; Corbin & Lindsey, 2005).

FG allows recognition both for fitness attainment (the product), for all individuals who attain the HFZ, but emphasizes rewarding fitness behavior (the process). Teachers are encouraged not to use test performance recognition to the exclusion of activity participation recognition. Available recognitions have changed through the years. "I'm Fit" was designed to recognize either achievement of the HFZ on five of six (or four of five) test items or improvement in performance on at least two test items. "Get Fit", "Fit for Life", the FITNESSGRAM® Honor Award, and "SMARTCHOICE" programs were intended to recognize participants for completion of exercise logs, achievement of specific goals, fulfillment of contractual agreements, and completion of the test assessment plus physical activities at home, in school, or the community (CIAR, 1987; CIAR, 1992). "It's Your Move" was introduced in 1994 and consisted of a series of activity booklets for K-6 graders that incorporated a recognition

⮐TOC ⮐Chapter

system. “You Stay Active”, a joint AAHPERD/CIAR project (FITNESSGRAM® minutes, 1995), was introduced in 1995 and consisted of comprehensive activity programs, assessment activities, cognitive activities, activity promoting events, goal setting performance recognition, and a model school and teacher recognition program for conducting programs that focused attention on and encouraged regular physical activity (CIAR, 1999). Now, in 2013, the “Get Fit” or a Fitness Contract Recognition (determined by the teacher) are still available in the FITNESSGRAM® & ACTIVITYGRAM® Test Administration Manual (Meredith & Welk, 2010). In addition, students can earn the Presidential Active Lifestyle Award (PALA) sponsored by the PCFSN. The FITNESSGRAM® software can track student eligibility for the PALA through data entry in the Activity Log module (Meredith & Welk, 2004).

The culmination of the emphasis on physical activity occurred with the development of ACTIVITYGRAM® in 1999 (CIAR, 1999; Meredith & Welk, 2004) predominantly targeted for grades 5 and higher. The ACTIVITYGRAM® module was based conceptually on a previously validated instrument known as the Previous Day Physical Activity Recall (Weston, Petosa, & Pate, 1997), but includes a number of enhancements that take advantage of the computer interface and other features to help promote interest and involvement in physical activity. The student is asked to report his/her activity for each 30 min block on two schooldays and one weekend day, selecting from a list of activities categorized according to the Physical Activity Pyramid (Corbin & Lindsey, 1995; Meredith & Welk, 2004). Duration and intensity are quantified. Students completing the assessment receive personalized reports similar to the existing FG reports, but evaluating their minutes of activity, times during the day when they are active, and the types of activity in which they are currently engaged.

Because ACTIVITYGRAM® is cumbersome for some schools to use, an easier way to collect and compile activity data was developed in 2011 for students in grades 5-12. The Youth Activity Profile (YAP) includes 15 questions: 5 for school physical activity; 5 for home physical activity; and 5 for sedentary habits. The YAP has been built into the evaluation plan for the NFL Play 60 FITNESSGRAM® Partnership project. The Youth Activity Profile is used as the basis for the new ACTIVITYGRAM® Profile introduced in 2013.

Cognitive tests became part of the FG Student App in 2013 to help students understand the necessity for activity, what the test items were intended to measure, and evaluate this understanding. The cognitive tests are available in six levels.

Systematic Updating and Sophistication of the Computerized Reporting System

As stated previously, FG was originally conceived and produced as a mainframe computerized reporting system for physical fitness test results. Table 5 presents the evolution of the FITNESSGRAM® software (as well as important partners and sponsorships enabling and facilitating this development) from version 1.0 to the current Version 10.0. Note that (for reasons that have been forgotten) there never was a version 4.0, and the missing version 7.0 was an attempted web based system that never was widely available. As can be seen from Table 5, the first two versions supported the AAHPER(D) YFT and HRPFT. Since then the FITNESSGRAM® test battery has been the only test battery supported. ACTIVITYGRAM® first appeared in version 6.0.

⇦TOC ⇦Chapter

Table 5. Key Highlights in the Evolution of FITNESSGRAM® Software and Reports

1982–1984 (Version 1.0)	
Sponsor/Partnerships	• Campbell Soup Company's Institute for Health and Fitness
Technology	• Mini-mainframe computer at Cooper Institute for Aerobics Research, Dallas, Texas • Service bureau approach—all data sent to Dallas where reports were produced and returned to districts • Data entered via "bubble" cards completed by teacher then scanned into database • Graphical presentation of current test results
Notable Features	• Schools could administer the AAHPERD Health Related Fitness Test or the AAHPERD Youth Fitness Test • Results were presented using percentile norms • Basic reports included: FITNESSGRAM® (single sheet report for student and parents), Summary Report
1985–1987 (Version 2.0)	
Sponsor/Partnerships	Campbell Soup Company's Institute for Health and Fitness
Technology	• Apple IIe, dual disk • Menu driven application • Easy to use software • FITNESSGRAM® reports were printed on pre-printed forms with a line printer. • Teachers entered data by typing
Notable Features	• Software was available for both the AAHPERD Health Related Fitness Test and the AAHPERD Youth Fitness Test • Results were presented using percentile norms
1987–1993 (Version 3.0)	
Sponsor/Partnerships	• Campbell Soup Company continued as sponsor through 1989-90 school year • Prudential Insurance Company began sponsorship in 1991-92 school year
Technology	• Versions for Apple IIe, Apple IIgs, and DOS • Dual disk version, added a DOS hard disk version in 1989 • Easy to use software • FITNESSGRAM® reports produced on pre-printed forms with a line printer. • Basic group statistical report included in software
Notable Features	• FITNESSGRAM® health related test battery • Results were presented using criterion referenced standards indicating minimal levels of fitness for health

1994–1998 (Version 5.0)	
Sponsor/Partnerships	• Prudential Insurance Company continued sponsorship through 1996-97 school year • No sponsor beyond 1997 • Implemented partnership with AAHPERD and their Physical Best curriculum program
Technology	• Versions for DOS, Macintosh and later Windows • Hard drive data storage • Relational database using multiple related tables of data • Printing available on laser printers • Introduced importing and exporting of data • Included new utilities to facilitate management of data such as promoting students and moving them from class to class • Added the Achievement of Standards Report • Teachers enter data by typing or scanning with Scantron forms
Notable Features	• Introduction of Healthy Fitness Zone rather than a single standard • Introduction of the PACER aerobic capacity assessment and new musculoskeletal fitness tests (curl-up, trunk lift, 900 push-up, back saver sit and reach, and shoulder stretch) • Reported calculated VO2max to allow for comparison between aerobic capacity assessments from one test date to another • Included a Spanish translation of the FITNESSGRAM® report • Introduced handwritten version of the Long Term Tracking Report
1999–2004 (Version 6.0)	
Sponsor/Partnerships	• Continued partnership with AAHPERD and their Physical Best curriculum program • Finalized partnership with Human Kinetics to publish and distribute all FITNESSGRAM® materials in 1999 • AAHPERD, CIAR, and HK formed the American Fitness Alliance • Entered into cooperative arrangement with PCPFS and AAHPERD to promote the Presidential Active Lifestyle Award (PALA)
Technology	• Versions for Windows and Macintosh • Network version of software to allow use in school computer labs • Student interface available to allow students to do own

	data entry • Graphical presentation of both current and past test results
Notable Features	• Introduction of ACTIVITYGRAM® module that included a three-day physical activity recall and report of results • Introduction of questions regarding activity levels for integration with fitness output
2005-2008(Version 8.0)	
Sponsor/Partnerships	• Continued partnership with AAHPERD and their Physical Best curriculum program • Continued partnership with Human Kinetics to publish and distribute all FITNESSGRAM® materials Continued cooperative arrangement with PCPFS and AAHPERD to promote the Presidential Active Lifestyle Award (PALA)
Technology	• Versions for Windows and Macintosh • Use of SQL database engine • Standalone version, Local Area Network version (school building) and Wide Area Network version (district server) • Centralized database for network versions • Greatly enhanced import and export capabilities • Improved security features • Data entry via a pocket PC module • Teachers can select from the test items and order them on the input screen and on the score sheet for recording data
Notable Features	• Introduction of Activity Log module for entering pedometer steps per day or minutes of activity per day • Activity Log includes feature to allow teachers to develop customized incentive challenges for students and classes • New report specifically for parents explaining FITNESSGRAM® test results • Modification in presentation of body composition information • New preprinted output forms for ACTIVITYGRAM® • Access to ACTIVITYGRAM® through the teacher application • Computerized long term tracking report for FITNESSGRAM® data • Free online software training videos developed by CI • Free online course on philosophy and test items developed and hosted by HK

2009-2013 (Version 9)	
Sponsor/Partnerships	• Continued partnership with AAHPERD and their Physical Best curriculum program • Contract renewed with Human Kinetics to publish and distribute all FITNESSGRAM® materials • Texas Department of Agriculture awarded grant funding for the development of NUTRIGRAM® • Partnership with NFL Play60 was obtained in 2009 and launched in 2010 • Centers of Excellence project implemented in local Dallas area schools and funded by private contributions to The Cooper Institute • United Way of Metropolitan Dallas funds the Healthy Zone School project • Perot International Youth Data Center began operations Brockport Assessment adopts “Healthy fitness zone” terminology for special populations
Technology	• School and District versions of FG 9 released for sale in 2009 • Web-based application • State, district, and school versions • Greatly expanded import/export utility • All users require username and password • Security based on user level • Numerous locally configurable factors • Ability to batch email student and parent reports • App for smartphones released in 2011 for teachers to do mobile input of test scores
Notable Features	• Extensive aggregate group reporting capabilities—scalable from individual class through the state level • In 2010, new aerobic capacity and body composition standards introduced with three categories: Healthy Fitness Zone, Needs Improvement-Some Risk, Needs Improvement-High Risk • Included use of test equating procedure to obtain classification agreement between PACER and One-mile Run • Made extensive use of online training webinars, hosted by Human Kinetics • Began development of ACTIVITYGRAM® Profile based on the Youth Activity Profile developed by Gregory J. Welk, Ph.D., Iowa State University.
2013 (Version 10.0)	
Technology	• Planned dashboard specific to user level and permission

	• Planned ability for schools to automatically update student information from student management systems • Planned ability for schools to deliver FITNESSGRAM® and ACTIVITYGRAM® reports to school district portals • Mobile apps available for students
Notable Features	• New messaging for aerobic capacity and body composition, renamed three levels to be Healthy Fitness Zone, Needs Improvement, and Needs Improvement-Health Risk • ACTIVITYGRAM® Profile and FITNESSGRAM® Cognitive Tests released in mobile platforms for students • Adoption of new PACER algorithms that do not require the input of height and weight.

The importance of the sponsorship of Campbell Soup (1982-1990) and then Prudential Insurance (1991-1997) cannot be stressed enough. This support allowed initially for the data entry to be done in Dallas and ultimately for personal computer (PC) software to be programmed and distributed to schools first without charge and then essentially for pennies per student involved. Prudential life insurance agents were given kits and encouraged to become involved with their local schools. By 1997, sponsorship ended and FG was supported directly from product sales. The publication agreement with Human Kinetics has allowed FG to continue to grow the user base, provide training, and to enhance and upgrade the software and reports.

As alluded to above, data entry was initially conducted at CIAR in Dallas. Teachers completed “bubble” cards that were then scanned into the mini-mainframe computer that printed the FG report cards that, in turn, were sent back to the teachers. By version 2.0 teachers could type in the data that was then printed out via an inline printer on forms that were purchased. Scantron forms for scanning data into the computer and the ability to utilize laser printers were the next innovations. Student input of data became an option with version 6.0.

Initially the only output that was available was the individual student report—the FG card. Gradually group statistical information became part of the output starting with version 3.0. Reports can now be generated for individual classes, schools, or districts. With version 8.0 and later it is possible to longitudinally track students graphically on each of the items with available data throughout their school career.

The sophistication of the PC software has always depended in large part on the sophistication of the available computers in the schools. At first, that meant Apple technology in the form of IIe, IIgs, and Macintosh. By version 3.0 (1989) a DOS version was available and this evolved into Windows. Networking versions that allowed use in school computer labs first became available in version 6.0. Version 8.0 was designed to enhance the use of assessments in school and district computer networks. Previous versions allowed multiple teachers to be linked in the same school. Version 8.0 allowed multiple schools to be linked within a larger school district. The use of unique individual, teacher, school, district, and state ID numbers facilitated the use of this data for large-scale tracking and surveillance projects. Enhancements in the software also allowed for more personalized monitoring of physical activity. An Activity Log provided a calendar-type interface that allowed youth to monitor and track their personal activity levels (using minutes or pedometer steps). Youth that met the requirements for the PCPFS PALA award are automatically flagged within the software so that teachers can send requests to PCPFS for awards.

The software for 9.0 is an entirely web based application with district, state, and school versions that incorporates the characteristics of Version 8.0 but in addition allows numerous locally configurable factors and has the ability to batch e-mail student and parent reports. As of 2011 teachers were able to input test scores via Smartphone apps.

The development of the Perot International Youth Data Center provides a platform for coordinating and tracking results from the web based versions of FG. The Perot Data Center allows schools districts that do not have the capability to host their own FG software the opportunity to use the web application. Web hosting customers include states, large school districts, very small school districts, and numerous international schools.

Version 10.0, introduced in the Fall of 2013, is a Software as a Service subscription based web application. It is hosted at the Perot International Youth Data Center. This version allows data collection for 1) FITNESSGRAM® (the Health Related fitness test items; state, district, school, teacher, and individual reporting; student cognitive tests; and instructional resources); 2) ACTIVITYGRAM® (the 3-day activity recall; the Activity Log, including the PALA and Challenges; and the gaming survey Activity Profile; 3) NUTRIGRAM® (grade based knowledge and behavior surveys, Quest to Lava Mountain game based learning, student and teacher reports, and instructional resources). FG 10.0 expands the usefulness of the software with mobile apps available for students and allows direct coordination with school wide student management systems.

As technology continues to improve, so will the sophistication of the FG software, with the objective always being to make the results easier for students, teachers, administrators, and parents to utilize, interpret, and act on to encourage physical activity and a comprehensive healthy lifestyle.

Conclusion

FITNESSGRAM®/ACTIVITYGRAM® represents an important innovation in the field of physical education/youth fitness. It is dedicated to providing the best possible physical fitness assessment, activity promotion, and feedback system for students, teachers, and parents to encourage lifelong physical activity and lifetime health-related physical fitness. Materials included in the Administration Manual and Reference Guide are constantly updated based on physiological/epidemiological, behavioral, and pedagogical research to support these unchanging goals utilizing the latest technology. Thus, FITNESSGRAM®/ACTIVITYGRAM® and NUTRIGRAM® will always be evolving-a work in progress.

⇦TOC ⇦Chapter

Bibliography

(Note: strict APA formatting would not include personal communications in the reference list. However, because it is important to see who the recipient was as well as the sender, personal communications have been listed here).

American Alliance for Health, Physical Education, and Recreation. (1976). *AAHPER Youth Fitness Test Manual*. Washington, DC: Author.

American Alliance for Health, Physical Education, Recreation and Dance.(1980). *AAHPERD Health Related Physical Fitness Test Manual*. Reston, VA: Author.

American Alliance for Health, Physical Education, Recreation and Dance. (1984). *AAHPERD Technical Manual Health Related Physical Fitness*. Reston, VA: Author.

American Alliance for Health, Physical Education, Recreation, and Dance. (1988). *Physical Best: The American Alliance Physical Fitness Education & Assessment Program*. Reston, VA: Author.

American Alliance for Health, Physical Education, and Recreation. (1978). *Task force on youth fitness. Research Consortium Newsletter*, March, 3(1), 1.

American Alliance of Health, Physical Education and Recreation. (1977). *Youth fitness task force appointed by AAHPER. Research Consortium Newsletter*, October, 2 (2), 1.

American Association for Health, Physical Education, and Recreation. (1965). *AAHPER Youth Fitness Test Manual.* Washington, DC: Author.

Annarino A. (3/04/1986). Memorandum to Physical Fitness Test Manual Revision Committee: Harold Falls, Chairperson, Patty Freedson, Dennis Humphrey, Wendell Leimohm, Tim Lohman, James Morrow, Russ Pate, Sharon Plowman and Terry Wood. Re: Alliance Executive Committee Action—1985 PCPFS School Population Survey.

Association for Research, Administration, Professional Councils, and Societies. (1987). *AAHPERD Fitness Test Update*. ARAPCS Physical Fitness Council Newsletter, Winter, 3(2), 4.

Blair, S.N., & Morrow, J.R., Jr. (Eds.) (2006). *Journal of Physical Activity & Health*, 3 (Supplement 2).

Burke E.J. (7/11/1984). Letter to Kroll W.

Cooper Institute for Aerobics Research. (1987). *FITNESSGRAM® Test Administration Manual*. Dallas, TX: Author.

Cooper Institute for Aerobics Research. (1999). *FITNESSGRAM® Test Administration Manual*. (2nd ed.). Champaign, IL: Human Kinetics.

Cooper Institute for Aerobics Research. (1992). *The Prudential FITNESSGRAM® Test Administration Manual*. Dallas, TX: The Cooper Institute for Aerobics Research.

Cooper Institute (2013). Report available at www.cooperinstitute.org/pub/news.cfm?id=146. Accessed 7/15/2013.

Corbin C.B., Lindsey R. (1995). *Fitness for Life*. (4th ed.) Champaign, IL: Human Kinetics.

Corbin C.B., Lindsey R. (2005). *Fitness for Life*. (5th ed.) Champaign, IL: Human Kinetics.

Cureton, K.J., Warren G.L. (1990). Criterion-referenced standards for youth health-related fitness tests: A tutorial. *Research Quarterly for Exercise and Sport*, 61, 7-19.

Disney, D. (6/21/2012). Report to the FITNESSGRAM® Scientific Advisory Board.

Falls H.B. (1979). *Task force on youth fitness. Research Consortium Newsletter*. , 4(1), 1-3.

Falls H.B. (4/23/1986). Letter to Hayes A.

⇦TOC ⇦Chapter

Falls H. (12/08/1986).Letter to Individuals involved in writing the new AAHPERD Youth Fitness Manual.
Franks B.D. (6/24/2986a). Letter to AAHPERD Executive Committee.
Franks B.D. (2/26/1987). Letter to Falls H.
Franks B.D. (4/06/1987). Letter to Hayes A.E.
Franks B.D. (6/01/1986). Letter to Measurement and Evaluation Council, Physical Fitness Council, and Research Consortium Executive Committees.
Franks B.D. (4/21/1986). Letter to Physical Fitness Manual Revision Committee.
Franks B.D. (6/24/1986b). Significant events in development of physical fitness test in AAHPERD.
Governor's Commission on Physical Fitness. (1972). *Texas Physical Fitness-Motor Ability Test*. Austin, TX: Author.
Hayes A. (10/16/1986). Corrected Task Force Report.
Haymes E. (6/26/1985). Letter to Falls H.
Kraus H, Hirschland R.P. (1953). Muscular fitness and health. *JOHPER*, 24, (10), 17-19.
Kraus H, Hirschland R.P. (1954). Minimum muscular fitness tests in school children. *Research Quarterly*, 25, 178-188.
Lacy E, Marshall B. (1984). FITNESSGRAM®: An answer to physical fitness improvement for school children. *JOPERD*, 55 (1), 18-19.
Lockhart B.D. (3/19/1987). Letter to Franks B.D.
Lockhart B.D. (6/14/1986). Letter to Franks B.D. representing Research Consortium, ARAPCS Measurement & Evaluation and Physical Fitness Councils. Re: Fitness Test—AAHPERD and PCPFS.
Lockhart B.D. (8/22/1986). Letter to "Persons interested in AAHPERD Youth Fitness Testing Program".
Meredith M.D. (7/11/2005). Personal communication (e-mail) to Plowman S.A.
Meredith M.D., Welk G.J. (Eds.). (2004). *FITNESSGRAM®/ACTIVITYGRAM® Test Administration Manual*. Champaign, IL: Human Kinetics.
Meredith, M.D., & Welk, G.J. (Eds.) (2010). *FITNESSGRAM® & ACTIVITYGRAM® Test Administration Manual* (updated 4th edition). Champaign, IL: Human Kinetics.
Minutes of FITNESSGRAM® Meeting. (3/9-10/1987). Atlanta, GA.
Minutes of the Prudential FITNESSGRAM® Advisory Council. (6/15/1995). Jackson Hole, WY.
Mood, D. (11/14/1972). Letter to Baumgartner T, Nelson J, & Reuter M.
Morrow J.R. Jr., Falls H.B., Kohl H.W., III. (Eds.) (1994). *The Prudential FITNESSGRAM® Technical Reference Manual*. Dallas, TX: Cooper Institute for Aerobics Research.
Morrow, J. R., Jr., Going, S.B., & Welk, G.J. (Eds.) (2011). FITNESSGRAM®: Development of criterion-referenced standards for aerobic capacity and body composition. *American Journal of Preventive Medicine*, 41 (4), Supplement 2.
Orr B. (3/14/1986). Memorandum to AAHPERD Board of Governors and Alliance Assembly Delegates. Preliminary 1986-87 budget.
Park, R.J. (1988). Measurement of physical fitness: A historical perspective. *ODPHP Monograph Series*. Washington, DC: US Department of Health and Human Services; Public Health Service.
Pate, R. (Ed.). (1983). *South Carolina Physical Fitness Test Manual*. (2nd ed.) Columbia: South Carolina Association for Health, Physical Education, Recreation and Dance.

Plowman S.A., Falls H.B. (1978). How fit? And for what? AAHPER Youth Fitness Test revision. *JOPER*, 49 (9), 22-24.

Plowman S.A., Falls H.B. (1979). Fitness testing in the schools: Revision of the AAHPER Youth Fitness Test. In: Cundiff, D.E. (Ed.) *Implementation of Aerobic Programs*. Washington, DC: AAHPER, 70-77.

Razor J.E. (1984). AAHPERD and a fit America. *JOPERD*, 55 (6), 54-60.

Razor J.E. (5/22/1984). Memorandum to Ciszek R.

Sterling C.L. (10/30/1986). Letter to Haywood H.

Sterling C.L. (2/23/1987). Letter to Haywood H.

Sterling C.L. (12/20/1993). Letter to Plowman S.A.

Sterling C.L. (10/14/1988). Testimony presented at President's Council on Physical Fitness and Sport Hearing on Youth Fitness. Dallas, TX.

Sterling, C., & Meredith, M. (12/17/2009). E-mail letter to FITNESSGRAM® Advisors.

Snyder, E. (12/15/2009). E-mail letter to FITNESSGRAM® Advisors.

The President's Council on Physical Fitness and Sports. (4/1986). *Fitness testing and the Presidential Physical Fitness Award*.

The President's Council on Physical Fitness and Sports [homepage on the Internet]. [cited 2005 July 27] Washington, DC: *The President's Challenge; Active Lifestyle Program*. Available from: http://www.presidentschallenge.org.

Welk G.J, Meredith, M. (Eds.). (2008). *FITNESSGRAM® Reference Guide*. Dallas, TX: The Cooper Institute. Available at: http://www.cooperinst.org/reference-guide Accessed 7/15/2013.

Weston A.T, Petosa R, Pate R.R. (1997). Validation of an instrument for measurement of physical activity in youth. *Medicine & Science in Sports & Exercise*, 29,138-143.

Chapter 2

Why Test? Effective Use of Fitness and Activity Assessments

**Charles B. Corbin, Dolly D. Lambdin, Matthew T. Mahar,
Georgi Roberts, Robert P. Pangrazi**

In the previous chapter the mission, philosophy, and history of the FITNESSGRAM® and ACTIVITYGRAM® programs are described. In this chapter guidelines for effective, efficient, and safe use of the assessments are provided. The first part of the chapter focuses on physical fitness assessment and the second part of the chapter focuses on physical activity assessment. The Scientific Advisory Board periodically prepares position statements for the information of users regarding use of the assessment programs. Information from the various position statements is included in this chapter.

⇦TOC ⇦Chapter

Why Use Physical Fitness Tests?

Facilitating Fitness Education: The Primary Use of FITNESSGRAM®

The mission statement, as well as the program philosophy, clearly outlines education with a focus on lifelong physical activity as the primary goal of FITNESSGRAM®. For this reason the educational value of the program should be central in using FITNESSGRAM® tests and related program materials. Going through the FITNESSGRAM® assessment procedures helps students to understand the various components of fitness. Reviewing personal fitness scores, included on the FITNESSGRAM® report, helps students determine if they meet health standards and have adequate fitness to meet their own personal needs. They can answer the question, "Am I where I need to be with my fitness?" Once these needs are determined, students can concentrate on fitness components that need attention by setting goals and creating and enacting plans to help achieve the goals. Self-monitoring (using reports) helps students determine if they are maintaining or improving their fitness and if they are meeting personal fitness goals.

A report of the Institute of Medicine (IOM, 2012, p. 9-2) concluded that, "Along with improving the fitness performance of individuals, fitness tests in educational settings can yield other benefits when appropriately conducted and interpreted. One benefit is that, when integrated into physical education programs in school settings, fitness testing can provide clear technical performance expectations . . ." "Fitness test results can also be used for assessing learning outcomes and physical education content standards. Given the connection between physical activity/fitness and cognitive performance (Castelli, Hillman, Buck, & Erwin, 2007; Hillman, Buck, Themanson, Pontifex, & Castelli, 2009; Kamijo et al., 2011; Welk et al., 2010), moreover, it becomes important for knowledge, attributes, and awareness of fitness to be promoted in educational settings as part of fostering healthy lifestyle choices across the life span." The IOM report also notes (IOM, 2012, page 9-2) that, "When the primary objectives of physical education or physical activity programming are achieved as intended, such programming can lead to the development of habitual healthy behaviors. The inclusion of fitness testing in physical education provides a forum for supporting and measuring the attainment of learning standards associated with physical fitness (Tremblay and Lloyd, 2010)."

Providing Feedback

The availability of individual and group fitness reports is an important benefit of using FITNESSGRAM®. Reports provide ratings of fitness based on health criteria, feedback to help interpret results, and information that is useful in planning programs for improvement of fitness through regular physical activity. Teachers should include student reports as part of student physical education portfolios along with other information related to important physical education objectives. Individual reports can be used to aid students in achieving the benefits described in Table 1. Reports should also be provided to parents. If this is done, it is recommended that school personnel (i.e., physical education teachers and/or nurses) meet with parents to help them interpret test results. A physical educator or nurse going over the FITNESSGRAM® report with parents provides a powerful message that the FITNESSGRAM® results are important and that it is possible for individuals to change their level of fitness by assessing where they are, setting appropriate goals, and creating and enacting plans that work toward the chosen goals. Parents should be encouraged to use the messages on the FITNESSGRAM® report to help students plan personal physical activity programs that are suited to each child's individual needs. The American Academy of Pediatrics (2006) recommends that physicians track the fitness and exercise patterns of youth and parents. Also

⇦TOC ⇦Chapter

research (Chomitz, et al., 2003, p. 771) has "…demonstrated that a health report card approach may be an important tool for schools interested in informing and motivating parents …."

Table 1. Summary of Recommended Appropriate Use of *FITNESSGRAM*® Test Items

Uses	Types of Testing			
	Self-Testing	Individual	Institutional	Personal Best
Student Benefits				
learning to assess personal fitness levels	✓	✓		✓
learning about health-related fitness	✓	✓	✓	✓
learning the health-benefits of activity	✓	✓	✓	✓
learning about criterion-references health standards	✓	✓	✓	
learning to interpret fitness test results	✓	✓	✓	✓
keeping personal fitness records (self-monitoring and tracking)	✓	✓	✓	✓
developing personal fitness goals	✓	✓	✓	✓
motivating students to be active	✓	✓	✓	✓
planning activity for fitness	✓	✓	✓	✓
learning about confidentiality of results	✓	✓	✓	✓
preparing student portfolios	✓	✓	✓	✓
verifying of accuracy of self-assessment information		✓	✓	
providing evidence of capabilities to perform in special settings (e.g., sports)				✓
Teacher Benefits				
providing fitness and activity education	✓	✓	✓	
providing individual data for student guidance	✓	✓	✓	✓
group data for curriculum development	✓		✓	
using fitness information, activity goals, and plans as part of student portfolios	✓	✓	✓	✓
tracking students over time—identifying health and fitness problems	✓	✓	✓	✓
providing verification of accuracy of self-assessment information		✓	✓	✓
Parent Benefits				
reporting results to parents	✓	✓	✓	✓
gaining information about children	✓	✓	✓	✓
involving parents in fitness and activity education of the child	✓	✓	✓	✓
providing information that leads to remedial help when necessary	✓	✓	✓	
developing family activity plans	✓	✓	✓	
Other Benefits				

⇦TOC ⇦Chapter

evidence of fitness education in schools	✓	✓	✓	
documenting use of fitness testing	✓	✓	✓	✓
planning curriculum	✓	✓	✓	
conducting research	✓	✓	✓	✓
centralized record keeping	✓	✓	✓	

Personal Tracking

Personal Tracking is another way of using FITNESSGRAM®. Student test results are plotted on a regular basis to see if youth retain their fitness status over time. The goal is to help all youth to score in the Healthy Fitness Zone (HFZ) (consistent with personal goals) on all parts of fitness over time. Improvement can be tracked and celebrated for each component. When dramatic changes in personal performance occur, tracking helps the student, teacher, and parent identify reasons for changes. Self-testing results and/or institutional testing results can be used for tracking changes over time (Meredith & Welk, 2010).

Meeting National Physical Education Standards and Guidelines

One of the principal national physical education standards is the achievement of health enhancing physical fitness (National Association for Sport and Physical Education [NASPE], 2004). A report of the IOM (2012) indicates that, "as of June 2011, all 50 states had learning standards centered on health-related fitness (Centeio and Keating, 2011); 14 states mandated direct measurement of physical fitness (NASPE, 2010a)." Physical education and the implementation of models such as Coordinated School Health and Comprehensive School Physical Activity Programs have outcomes concentrated on both the achievement and maintenance of health-enhancing levels of fitness and regular engagement in physical activity, as these variables are independent risk factors associated with health (Plowman, 2005). FITNESSGRAM® and associated educational programs help students meet national physical education standards and guidelines.

Why Use FITNESSGRAM®?

Scientific Basis

FITNESSGRAM® was developed by a board of scientists (advisors) with extensive expertise in youth fitness. The advisors meet regularly to consider the scientific basis for including test items and health-related fitness standards. The board publishes this Reference Guide to provide information about development of tests and standards as well as guidelines for use of FITNESSGRAM®. Such a Reference Guide explaining the scientific basis for the program is unique to FITNESSGRAM®.

Criterion-Referenced Health Standards

FITNESSGRAM® uses health standards for helping users interpret their personal fitness. Rather than providing comparisons to group norms, indicating how students compare to others their age, which doesn't provide much information about health risk, these criterion-referenced standards are based on the best evidence available of a score's relationship to current and future health. Health-related fitness and health-related fitness standards have been widely endorsed over the years and recently reaffirmed in the Institute for Medicine report, Fitness Measures and Health Outcomes in Youth (2012).

FITNESSGRAM® Reports

FITNESSGRAM® provides a unique personal fitness report card for each student. This report provides information to students and parents, not only about what the student's scores from the tests on different components of fitness are, but also feedback about how those scores relate to health. In addition advice is provided as to how to proceed in order to realize health benefits related to fitness. The use of report cards has been shown to be motivating to parents (Chomitz et al., 2003) and can be used by physicians consistent with policy of the American Academy of Pediatrics (2006).

Basis for Fitness Education

As noted earlier, a principal purpose of fitness testing is fitness education. FITNESSGRAM® reports and accumulated data files provide information that students can use to determine personal needs, set goals, plan programs, and self-monitor behavior that promotes fitness.

Partnerships

Because of its scientific basis and history of effectiveness, FITNESSGRAM® and The Cooper Institute now partner with the President's Council on Fitness, Sports, and Nutrition (PCFSN), the American Alliance for Health, Physical Education, Recreation and Dance (AAHPERD), the Centers for Disease Control and Prevention (CDC), the Amateur Athletic Union (AAU), and the National Association for Sport and Physical Education's (NASPE) Physical Best to deliver a fitness assessment program as well as fitness education programs collectively known as the Presidential Youth Fitness Program. In addition, the American Academy of Pediatrics (2003) has endorsed the use of body composition screening (BMI) for children and adolescents.

Widely Used

FITNESSGRAM® is used in all 50 states and 14 different countries. It is estimated that FITNESSGRAM® is in 67,000 schools and over 22 million students got tested in 2012. In Texas alone nearly 2.8 million students are assessed annually. As of fall 2012 the following states mandate the use of FITNESSGRAM®: California, Delaware, Georgia, Texas, and Kansas. In addition many large school districts require FITNESSGRAM® assessment including New York City, Miami-Dade County, District of Columbia, Indianapolis, Milwaukee, Tulsa, Prince Georges County (MD), Anne Arundel County (MD), Memphis City (TN), and Columbus (OH) City Schools. The NFL Play 60 supports its use in 1,120 schools they work with in all 32 NFL markets.

Tracking and Data Management

In addition to individual reports, FITNESSGRAM® offers a host of individual and group summary reports (see Chapter 9 Interpreting FITNESSGRAM® and ACTIVITYGRAM® Reports) that enable individuals to go from just seeing test results to tracking the test results over time and using those test results to impact future planning and achievement. For the individual there are graphs that show scores plotted over time for each component. For teachers there are group summary reports that support analysis for curriculum planning as well as program evaluation.

What Are Different Ways in Which FITNESSGRAM® Testing Can Be Conducted?

There are four different ways in which fitness tests are typically conducted. These types of organization are described in the section that follows.

Self-Testing and Assessment

Personal fitness self-testing is considered to be the principal use for FITNESSGRAM® test items. Students are taught to evaluate themselves and interpret their test results. If this objective is met, students can test themselves and plan personal fitness programs throughout life. It takes a considerable amount of practice to self-test effectively so multiple opportunities to practice are necessary. It is also important to help students interpret the results. Students who fail to reach the healthy fitness zone (HFZ) should be assisted in developing and enacting a program for improvement. Students who reach the HFZ should be taught how to determine goals for fitness within the zone and how to maintain that level of fitness.

In this type of testing students evaluate themselves so special teams of fitness testers are not necessary. Test results for beginning self-testers may not be totally accurate but, with practice, self-testing skills improve and become more useful in program planning. Students self-testing on a regular basis begin to really understand that the testing helps them know where they are and see that working on their fitness can lead to improvement. Self-testing results, as all testing results, are considered personal and should generally be kept private if a student so desires. An exception is when FITNESSGRAM® reports are printed and used to report results to parents and teachers. If self-testing results are reported to parents, especially by beginning self-testers, parents should be aware that the results might be less accurate than results of more formal testing. Repeated self-testing allows students to be responsible for their own data, and soon it becomes apparent that working to ensure accurate data is in their own best interest. Over time students learn to accurately assess their own fitness and enjoy the process.

Individualized Testing

Individualized Testing refers to testing done with the principal goal of providing personal information to individual students much as a personal trainer would do. Self-testing as described in the previous paragraph is a form of individualized testing. Individualized testing could, however, be done with the assistance of others such as a partner, parent, or teacher. Test results are used for personal feedback and to provide feedback to parents. The results of individualized testing can be used to help students and parents plan personal activity programs and track progress over time.

Personal Best Testing

Personal Best Testing is for students who want to see how well they can perform on each fitness test item as opposed to seeing if they are in the healthy zone. Because such testing takes considerable time and because all children and youth may not be interested in this type of testing, it is recommended that this type of testing be done before or after school on a voluntary basis. The FITNESSGRAM® philosophy focuses on good health and high levels of fitness are not necessary for good health. Some youth, however, may be interested in achieving high levels of fitness to achieve performance goals, and teachers may wish to provide the opportunity for personal best testing.

⇦TOC ⇦Chapter

Institutional Testing/National Surveillance

Institutional Testing is done to help teachers and other educators determine the fitness level of groups of students and should be used to provide direction for curriculum planning. This type of testing takes teams of people trained to correctly administer the test to large groups of students and takes a considerable amount of class time. Reports to students and parents may also be prepared using institutional testing data. The FITNESSGRAM® advisors suggest that this type of testing need only be done periodically, for example every third year. If periodical institutional testing is to be done it is recommended that it always be done at the same time of the year (beginning or end). Care should be taken when interpreting data obtained from this type of testing since any individual can have a "bad" day where he/she is not feeling up to par and so the person does not perform up to ability on that specific day. As noted later in this chapter, the FITNESSGRAM® advisors discourage the use of FITNESSGRAM® for determining student grades, long-term student achievement, and/or teacher success. (See section on Inappropriate Uses of Fitness tests). As noted elsewhere in this Reference Guide, there are too many factors other than personal effort and physical activity that influence fitness to use fitness tests as major indicators of student achievement (see Chapter 3 on Health Benefits).

In addition to school program tracking, institutional testing may be done for statewide or national surveillance. Many states currently test students statewide and national surveys have been used in the past to assess the fitness of American youth. FITNESSGRAM® is the most frequently used test for statewide educational testing. Recently a committee of the Institute for Medicine (2012) prepared a report that provides direction for testing youth fitness in national surveys.

Recommended appropriate uses of the FITNESSGRAM® physical fitness test items are also listed in the Test Administration Manual, Chapter 2 (Meredith & Welk, 2010). The appropriate uses are listed in Table 1. Some additional comments related to other fitness testing uses are provided in the following section.

Guidelines for Institutional Testing

Institutional testing, one type of testing, was described in the previous section. When conducting institutional testing these guidelines should be considered:

- **Take care in interpreting results**. Group score differences among classes and among schools are often due to factors other than the quality of teaching and the level of student learning in a class or in a school. Motivational levels associated with a variety of factors (often beyond the control of the teacher), play a role in determining fitness test scores. Interclass and interschool comparisons should be made with great caution.
- **Take care in generalizing from pre-test to post-test data**. Fitness scores will typically be higher at the end of the school year than they are at the beginning of the school year because youth are ¾ of a year older. Older students do better on fitness tests than younger students (Pangrazi & Corbin, 1990). Pre- and post-testing takes considerable time and may not be warranted if too much time is taken from the regular curriculum. Additional comments about pre- and post-testing are provided later in this chapter.
- **Consider nutrition and other factors when generalizing about body composition results** (Lee et al., 2006). The number of youth who are overweight or obese has increased in recent years. This is, no doubt, because lifestyles (eating and activity patterns) have a major impact on body composition even in youth (Lee et al., 2006). Overweight and obesity are associated with many behaviors and solving problems

associated with youth overweight and obesity is complex. Physical education can help with the problem. However, physical activity and other learning in physical education are only two of many factors that should be considered when preparing a comprehensive plan to solve the problem (Lee et al., 2006; Lohman, Going, & Metcalfe, 2004).

- **Reports indicating the proportion of youth who meet health standards are more meaningful than reports containing mean scores for individual tests or percentile scores for students, classes, or schools**. Knowing the proportion of students that fail to meet minimum health standards may help in curriculum development and guide teachers and parents in helping more students achieve health standards. Meeting minimum standards is a reasonable goal when adequate time is provided for change and when multiple entities (teachers, schools, parents, communities, physicians) have the opportunity to work together. Factors other than physical activity (as described previously) should be considered when making curriculum plans. Percentile scores are of little value in this effort. (Percentile scores are not used in FITNESSGRAM®.)
- **Care should be taken to accurately report institutional test results**. Evidence exists to indicate instances of incorrect reporting of data in other areas of the curriculum when "high stakes" are associated with test results (Harrington- Leuker, 2000; Kohn, 2001, Sloane & Kelly, 2003). Care should be taken to assure teachers and students that scores on tests are for personal use by students, their families, and by teachers in curriculum development rather than less appropriate reasons (see inappropriate uses section). If "high stakes" are not associated with the testing, results are more likely to be reported accurately. It is also important that teachers receive in-service education in correct test administration if results are to be accurate.
- **Care should be taken to avoid overgeneralizations concerning the meaning of test scores**. Schools should rightly be concerned about issues related to youth fitness, physical activity, and health (including youth overweight and obesity) and it is true that physical education can be helpful. If large numbers of students are not reaching the healthy fitness zone, a quality physical education program with adequate instructional time can have impact on group and individual scores. However, as Ernst, Corbin, Beigle, and Pangrazi (2006) note, care should be taken not to overstate the potential of physical education in improving fitness. Many factors influence the fitness of youth and physical education, while important, is only one. Ernst et al. (2006, p. S97) note: "*It is reasonable to assume that if schools, homes, and communities worked together that physical activity levels of youth could be increased. Such increases in activity would, no doubt, help many adolescent youth to meet minimal criterion level fitness standards. Certainly younger children would also benefit from the increased activity but not all will have success in meeting fitness standards. However, until cooperative efforts among those in schools, homes, and communities are implemented, changes in year-to-year fitness scores are unlikely, especially among younger youth.*"

What Are the Appropriate Practice Guidelines for Using Fitness Tests?

The 2012 IOM report entitled, *Fitness Measures and Health Outcomes in Youth* stated: "If physical fitness tests are to be used effectively in schools and other educational settings, appropriate practices must be employed in their administration. Appropriate practice varies by maturation stage; thus what may be suitable for elementary school students may be inappropriate for adolescents (IOM, 2012, page 9-7). The National Association for Sport and Physical Education (2010b) position statement entitled Appropriate Uses of Fitness Measurement also reviews many of the items listed below. Factors such as insuring safety and confidentiality, incorporating education as well as fitness testing in the curriculum, preparing students, preparing teachers and providing ongoing professional development, factors that affect test performance, testing students with disabilities, use of fitness testing for research, and assuring appropriate scheduling of testing are discussed in the section that follows.

Insuring Safety and Confidentiality

Physical and emotional safety of participants should always be paramount. The Institute of Medicine report states that "it is vital as well for administrators to ensure the safety of fitness test participants by being sensitive to such variables as participants' pre-existing disease(s), body composition, and maturation stage" (IOM, 2012, page 9-4). Most schools have a policy for approving students for participation in physical education, active play during recess, and before and after school activities. Children who have limitations in activity are typically identified. When administered as part of an organized school testing or physical education testing program, fitness testing is typically covered under normal school policies. The Institute of Medicine further indicates that, "the articles selected for this review do not report any injuries during testing. One recent manuscript (Ruiz et al., 2011) does address the safety of the 20-meter shuttle run, finding that no complications occurred during the testing, with only one report of a lower-body muscle cramp. The authors note that they have experienced no safety issues in more than 10,000 children they have tested (2012, page 5-26)."

The following excerpt from Ernst et al. (2006, p. S97) provides information related to emotional safety and confidentiality:

> "One advantage of paper and pencil tests is that the results can easily be kept confidential. No one other than the person being tested, parents, teachers and other appropriate school officials know the results of a student's tests unless the student or parent chooses to reveal results or unless school officials reveal results inappropriately. With physical fitness testing the actual testing process is often quite public. Setting positive expectations leading to the creation of a supportive testing environment is extremely important. Appropriate protocol can be used to assure as much privacy as possible (e.g., separation of testing stations, screens to avoid observation of measurements—especially body composition measures) and to educate students concerning the confidentiality of the results of others. When partners or groups are used in testing it should be understood that test results revealed to a partner or observed by others in the group (e.g., PACER) are confidential. A major advantage of self-testing is that it can be done in privacy or relative privacy". Individually identifiable fitness test results should not be posted in public places.

Incorporating Fitness Testing into the Curriculum

Shortly after the first health-related physical fitness test was introduced in 1980 (American Alliance for Health, Physical Education, Recreation, and Dance, 1980) Pate and Corbin (1981, p. 37) noted that, "the Health Related Fitness Test should be fully integrated into the physical education curriculum. In the fitness domain, the ultimate objective of physical education should be to aid the student in acquiring the skills, knowledge, and attitudes needed to

⇦TOC ⇦Chapter

become a lifetime exerciser and to maintain a good level of health-related fitness." They further noted that the test "…should be used only to help accomplish predetermined educational objectives" (Pate & Corbin, 1981, p. 38).

More recently the IOM indicated that,

> When fitness testing is integrated into educational programs or curricula, it provides a mechanism for longitudinally tracking and monitoring physical fitness trends and risk for disease among individuals and groups. In an educational setting, individual tracking is most relevant as school is one of the few places where feedback can be provided to both participants and their parents. However, group tracking over time also can be useful for physical education teachers, enabling them to utilize trends to inform instruction by identifying the needs of the current student body. It has been suggested that, regardless of developmental stage, the benefits of being able to monitor progress, set goals, provide feedback, give incentives, and design and implement a personalized physical activity plan outweigh the risks of participation in physical fitness testing (Safrit, 1995). Clearly communicating to participants the meaning of each test item and discussing the training principle of specificity (i.e., the activity's association with an identified joint or muscle group) is important. Participants then can set personalized goals and create an individualized plan for achieving those goals that purposefully links modes of physical activity to health-related fitness components. Learning experiences that apply knowledge to authentic situations increase the likelihood that conceptual learning will lead to enhanced participation in physical activity. (IOM, 2012, p. 9-7)

Preparing Students for Testing

Many authors have noted the importance of student preparation prior to conducting fitness testing (Corbin, 2009; Corbin & Pangrazi, 2008; Ernst et al., 2006; National Association for Sport and Physical Education, 2009a, 2009b, 2009c; Pate & Corbin, 1981). NASPE (2009a, 2009b, 2009c) indicates that a basic tenant of fitness testing is that youth should be physically prepared to participate in fitness testing. The IOM (2012) also notes the importance of student preparation. Some important factors to consider are listed below.

- **Explain to students why they are taking a health related physical fitness test.** Part of preparation for fitness testing should be providing educational information of how the tests are done, why the tests are done, how the test results will be used, and how the test results are beneficial to the student" (Corbin, 2009, p. 25).
- **Provide instruction as to how to properly perform the test and provide practice**. "Allowing students to practice tests helps students understand concepts such as pacing, and helps students use good technique. Practice also helps to eliminate mistakes that may result in loss of repetitions (repetitions not counted toward score) when doing institutional testing" (Corbin, 2009, p. 25).
- **Train before testing**. "Taking a fitness test can be a daunting experience. Students who are unprepared will experience soreness, fatigue, and for many anxiety. Performing regular activity including practicing test items (see above) helps to prepare students for a more satisfying experience" (Corbin, 2009, p. 25).
- **Assure proper dress and testing conditions**. Students should have appropriate clothing and shoes for the testing experience. They should also have the time and

⇦TOC ⇦Chapter

appropriate facilities to get dressed before and after testing with teacher supervision. Appropriate testing equipment and facilities should be provided to ensure a satisfying and safe testing experience.

- **An appropriate testing climate should be provided**. Pate and Corbin (1981, pp. 37-38) suggest that "fitness testing be done in an enthusiastic, positive, nonpunitive manner," that testers "provide copious positive reinforcement for students who make good effort," and that "the test should never be used to embarrass learners."

The IOM noted that:

> An extensive body of literature expands on components of effective and sustainable professional development, a topic that is beyond the scope of this report. In general, however, professional development enables physical education teachers to administer physical fitness tests accurately and with minimal bias (Morrow et al., 2010) while providing physical activity opportunities that enhance fitness (Kibbe et al., 2011). A recent meta-analysis suggests that in general, students are motivated to participate and to learn in physical education (Chen et al., 2012). Yet student motivation is influenced by the school climate, specifically the task or ego orientation of the activities offered during physical education (Parish and Treasure, 2003; Standage et al., 2003). Teachers who develop a positive and mastery-oriented climate are more likely to have students who perform better on assessments such as fitness testing. When introducing students to fitness testing, for example, the use of instructions that provide personal relevance and meaning for a student can lead to enhanced performance (Simons et al., 2003). Accordingly, it is important for teachers to be consistent in the delivery of content related to fitness testing, as well as to be equally supportive to learners of all ability levels, or the test may be biased. (IOM, 2012, p. 9-5)

Assure students about confidentiality. Confidentiality was addressed in a previous section. It is important that students be informed about confidentiality of their information and about the need to keep confidential information of other students.

Preparing Teachers and Providing Ongoing Professional Development

The 2012 IOM report on Fitness measures and health outcomes in youth emphasized the importance of preparing teachers and providing ongoing professional development to enhance youth fitness testing in the schools. The report stated that the following factors should be considered to effectively conduct fitness testing in schools:

- "School-based professional development that is applicable to the daily routine of teachers and includes instruction in how to integrate fitness testing into the curriculum should be provided.
- Professional development should include training in the administration of protocols and interpretation and communication of test results, with emphasis on educating participants about the importance of fitness, supporting the achievement of fitness goals, and developing healthy living habits. Those interpreting and communicating test results should ensure confidentiality, consider each individual's demographic characteristics, provide for the involvement of parents, and offer positive feedback and recommendations to students and parents" (IOM, 2012, p. 9-1).

The IOM report (2012, p. 9-5) further stated: "Professional development aimed at preparing physical education teachers to administer a battery of fitness tests can include a combination of the following components:

- how to integrate fitness testing into the curriculum;
- protocols and use of proper equipment for fitness test items;
- how to familiarize participants with the test, together with specifications regarding the amount and type of practice;
- how to communicate consistently with the students in ways that create a positive and encouraging environment for learners of all ability levels;
- teacher burden;
- participant burden;
- the validity and reliability of test items;
- class management during test periods; and
- how to interpret and communicate test results."

Considering Factors that Affect Test Performance

In addition to genetic and body composition factors that affect performance there are aspects of the environment that have a considerable effect on test performance. These include physical factors such as the temperature and weather conditions when the testing is being performed, as well as time of day (especially whether it is just before or after lunch or at the very end of the day). In addition social-emotional environmental factors such as the self-efficacy and feelings of support are important. Finally, familiarity with the testing procedures is important in getting accurate results. Helping students to understand how these and other factors can affect scores on specific days provides a realistic view of their performance and progress.

Testing Students with Disabilities

The FITNESSGRAM®/ACTIVITYGRAM® philosophy highlights the need for inclusion of "everyone" and the need to "personalize" (the E and P of the HELP philosophy). Students who are physically or mentally challenged can benefit from FITNESSGRAM® and ACTIVITYGRAM® but special adaptations may be necessary. The IOM (2012) notes that

> …it is important for students with disabilities to be included in fitness testing whenever possible and for the interpretation of test results to be modified accordingly. Specifically, those students with personal fitness goals should be encouraged to participate in fitness testing as a means of tracking progress toward their goals. The Brockport Fitness Test is an example of how specific fitness tests can be modified for students with disabilities, and the Brockport Physical Fitness Technical Manual provides criterion-referenced cut-points (cutoff scores) for a variety of disabilities (Winnick and Short, 1999). While the relationship between health outcomes and physical activity in people with disabilities is not the focus of this report, other reviews, such as the Physical Activity Guidelines Advisory Committee Report (Physical Activity Guidelines Advisory Committee, 2008), specifically examine this issue. (IOM, 2012, p. 9-3)

Using Fitness Testing for Research

One of the appropriate uses of fitness test results is research. As is the case in other uses of fitness test results, care should be taken to protect the integrity of confidential personal data when conducting research. Also, care should be taken in interpreting research results (see Ernst et al., 2006). Procedures for preparation of test administrators (teachers or others) and preparations for students outlined in this chapter should be followed.

Assuring Appropriate Scheduling of Tests

FITNESSGRAM® and ACTIVITYGRAM® offer options in test selection and in test scheduling. The following guidelines should be considered when scheduling testing.

- **Determining which tests to use**. It is not necessary to give every FITNESSGRAM® test every time an assessment is done. It is very reasonable to test youth, or have them test themselves, on a component of fitness that is being studied at any particular time. In addition, FITNESSGRAM® offers options for using test items. Test administrators can select from a variety of test items for each testing period.
- **Pre-testing and post-testing**. Some teachers feel that tests at the beginning of the year and again at the end of the year are good indicators of student achievement. While this type of testing may be used, the results must be interpreted with caution. First, students will improve whether they are doing activity or not, just because they are getting older. For this reason incorrect messages may be conveyed. Second, students learn over a period of time to “be bad” on initial tests and “get good” on later tests if grades are based on improvement. Third, giving pre-tests and post-tests can lead to the feeling that somehow fitness “is done” after the post-test rather than that it is an ongoing part of life. The FITNESSGRAM® Scientific Advisory Board recommends that students be given many opportunities to learn to self-test accurately. Also, keeping logs of fitness test results helps students set fitness and activity goals and plan personal programs. Once students become accomplished in self-testing they can repeat testing periodically to assess personal improvement. Using pre-post fitness tests as a primary method of grading students is strongly discouraged because some individuals do not respond to training as easily as others and because of the difficulty of determining appropriate levels of improvement for different levels of fitness. If a student starts out with extremely good fitness it may be difficult to show improvement. Having students learn to self-test and keep records for goal setting and program planning are encouraged. Like FITNESSGRAM®, ACTIVITYGRAM® can be used for institutional evaluation. ACTIVITYGRAM® can also be used as a means of assessing activity patterns for research purposes. More information is available in the chapter devoted to ACTIVITYGRAM®.

What Are Inappropriate Uses of FITNESSGRAM®?

Appropriate uses of FITNESSGRAM® consistent with the HELP philosophy are encouraged. (It’s about Health, for Everyone, for a Lifetime, and it’s Personal.) Appropriate uses also require knowledge of the goals of physical education. The National Outcomes Project by NASPE outlined five key characteristics of a physically educated person (National Association for Sport and Physical Education, 1992). Principal objectives have also been outlined by NASPE in the most recent version of a book describing National Standards for Physical Education (National Association for Sport and Physical Education, 2004). Virtually all states have now

⇦TOC ⇦Chapter

outlined standards and objectives for physical education based on NASPE recommendations. Table 2 describes the five characteristics of a physically educated person and the six major national standards for physical education because they will be referenced in the paragraphs that follow.

Table 2. The Multiple Objectives of Physical Education

A physically educated person is one who (NASPE, 1992):

- has learned skills necessary to perform a variety of physical activities
- knows the implications of and the health benefits from involvement in physical activity
- does participate regularly in physical activity
- is physically fit
- values physical activity and its contribution to a healthful lifestyle

A quality physical education program produces a student who (NASPE, 2004):

- demonstrates competency in motor skills and movement patterns needed to perform a variety of physical activities (Standard 1)
- demonstrates understanding of movement concepts, principles, strategies, and tactics as they apply to the learning and performance of physical activities (Standard 2)
- participates regularly in physical activity (Standard 3)
- achieves and maintains a health-enhancing level of physical fitness (Standard 4)
- exhibits responsible personal and social behavior that respects self and others in physical activity settings (Standard 5)
- values physical activity for health, enjoyment, challenge, self-expression, and/or social interaction (Standard 6)

In addition to the comments that follow concerning inappropriate practices, the reader is referred to Chapter 2 of the Test Administration Manual, and an article listed in the reference section by Ernst et al. (2006).

Sometimes methods of using fitness tests and/or assessments of physical activity levels violate the HELP philosophy or are inconsistent with the goals of physical education programs. Such uses are considered to be "inappropriate practice" and are discouraged. Specific inappropriate uses include the following.

Inappropriate Use of Test Results to Grade Students in Physical Education

Using fitness test results as a primary method of grading students in physical education is strongly discouraged. There are many good reasons why fitness test scores should not be used as a primary method of grading students in physical education. Several of the most important reasons include:

- **Physical fitness is only one of many goals of physical education** (see Table 2). Because physical fitness is one important objective of physical education programs, assessment of physical fitness related objectives can be considered when grading. However, the fitness objective should be considered with other important objectives as well. All fitness related objectives (not just fitness scores) should be considered and the limitations of using fitness scores should be considered when developing grading plans.

- **Scores in the healthy zone are the goal**. It would be inappropriate for students to be graded down if their scores are in the Healthy Fitness Zone, but not equal to others in a class.
- **Fitness test results are only one indicator of accomplishment of the fitness goals and objectives described in Table 2**. Examples of other important fitness related objectives are learning to do self-assessments, interpreting test results, and planning activity programs to promote fitness (see Table 1 for other student benefits related to fitness testing). These and other fitness related objectives should be considered. Differences in developmental level should be considered when grading, especially when considering fitness related objectives. Elementary students are less likely to respond to training than older students. Also, young children are concrete thinkers rather than abstract thinkers, therefore, fitness performance objectives and objectives that relate to learning fitness concepts should be adapted to meet the developmental level of learners.
- **Fitness does not correlate well with time spent in activity especially among preadolescent and young teens** (Morrow & Freedson, 1994; Morrow, Jackson, & Payne, 1999). Students can be "turned off" to physical activity when they make little progress despite regularly participating in appropriate physical activity and their grade is affected. Reasons for the lack of correlation are described in the next bullet point.
- **Many factors other than effort and physical activity influence fitness test results**. For example: heredity, maturation, gender, chronological age, and other factors beyond the control of the student and teacher affect fitness test scores, especially among preadolescent youth.
- **Physical education time is often limited in current school programs**. For elementary school students, physical education classes may be conducted only a few days a week for periods of 20-30 minutes. Physical education can help promote activity for students so that they will get the total daily activity needed to promote optimal fitness, but it is unrealistic to assume that fitness scores can be impacted in programs offering only limited time for participation.

The FITNESSGRAM® Scientific Advisory Board encourages teachers to evaluate students on all important physical education objectives.

Inappropriate Use of Test Results as Indicators of Student Achievement in Physical Education (e.g., District or State Tests)

The use of test results as indicators of student achievement in physical education (e.g., district and state tests) is considered an inappropriate use. Federal and state mandates have led to regular standardized testing in areas such as math, science, and language arts. The tests are intended to determine if students are meeting state standards in specific areas of study. Because well designed fitness tests such as FITNESSGRAM® are based on sound science and educational principles, some schools have proposed that fitness tests be used as a local or state test to assess student achievement of local or state standards. The appropriate uses and benefits of FITNESSGRAM® are clearly described earlier in this document. FITNESSGRAM® was not developed as a standardized test of comprehensive physical education standards.

The reasons why the use of fitness test results are discouraged as a primary method of assessing student achievement in physical education are similar to those described in the previous section for grading (see previous section). It is not within the purview of the

FITNESSGRAM® Scientific Advisory Board to determine local or state standards for physical education or to determine assessment procedures for assessing these standards. However, the FITNESSGRAM® Scientific Advisory Board would like to go on record indicating that any program to assess achievement of local or state standards should be based on a comprehensive body of evidence that a student has met ALL important physical education program standards (objectives).

Inappropriate Use of Test Results to Exempt Students from Physical Education Classes

Exemptions from physical education that are inconsistent with the practice used for exemptions in other areas of the curriculum are discouraged. Using student fitness scores as a primary method of exempting students from physical education classes is considered to be inappropriate practice as fitness is only one of the physical education standards and is something that must be continually maintained.

As noted in the statement above, exemptions from any class in the school should be consistent with a comprehensive policy that applies to all areas of the curriculum. For many of the same reasons described in the previous two sections (concerning grading and student achievement), exemptions are discouraged, especially those based exclusively on fitness test scores. The following summary characterizes the position of the Board.

The Scientific Advisory Board has taken great pains to provide quality fitness and activity assessment programs with scientifically based standards. Developing policy related to exemptions for high school physical education is beyond the purview of the FITNESSGRAM®/ACTIVITYGRAM® Scientific Advisory Board. However, the Board would like to go on record indicating that any exemption from high school physical education should be based on a comprehensive body of evidence that a student has met ALL of the important standards (objectives) of physical education programs. Further, the Board encourages "appropriate" uses and discourages "inappropriate" uses of the program.

Inappropriate Use of Fitness Test Results as a Measure of Teacher Success

In some instances the use of physical fitness scores have been used as a primary indicator of teacher success in physical education and student fitness scores have been used in assigning raises for teachers. As indicated in previous sections of this chapter, there are many reasons why fitness scores are not a good overall indicator of student success. Accordingly, student performances on fitness tests are not a good overall indicator of teacher success. In addition when teacher success is based solely on student fitness performance, class sessions can resemble "fitness training" rather than physical education, resulting in dislike rather than enjoyment of activity. When teacher success is based on student fitness performance, cheating on fitness tests becomes a problem as it has in academic areas where cheating has been documented (Harrington-Leuker, 2000; Kohn, 2001; Sloane & Kelly, 2003).

Developing policy related to teacher evaluation in physical education is beyond the purview of the FITNESSGRAM®/ACTIVITYGRAM® Scientific Advisory Board. However, the Board would like to go on record indicating that teacher evaluations in physical education should be based on comprehensive teacher effectiveness criteria and student achievement on ALL important standards (objectives) of physical education programs.

As the Institute of Medicine report states:

> Although physical fitness can be increased through engagement in specific types of physical activity, factors other than physical activity affect a student's fitness that are beyond the control of the student and physical education teacher. Examples include heredity, caloric consumption, access to opportunities to be physically active both within and beyond the school day, and possibly socioeconomic status. For similar reasons, physical fitness testing for the purpose of teacher and school accountability is also inappropriate. (Institute of Medicine, 2012, p. 9-8)

Why Use Activity Assessments?

Facilitating Fitness Education: The Primary Use of ACTIVITYGRAM®

As indicted earlier the primary goal of FITNESSGRAM®/ACTIVITYGRAM® is to promote lifelong physical activity. Physical activity, separate from physical fitness, has been shown to be an important factor related to health. Standard 3 of the NASPE standards described earlier is a physically educated student "participates regularly in physical activity." National guidelines (United States Department of Health and Human Services, 2008) recommend 60 minutes of daily physical activity including vigorous activity, muscle strengthening activity, and flexibility activities. Fitness testing allows individuals to see "whether I am where I need to be in terms of fitness" whereas monitoring physical activity allows the individual to determine "whether I am where I need to be in terms of daily physical activity." Both are important but only activity assessment can determine if students are meeting national physical activity guidelines.

Providing Feedback

Just as the availability of individual and group fitness reports is an important benefit of using FITNESSGRAM®, ACTIVITYGRAM® provides similar information about physical activity. Reports provide ratings of physical activity based on health criteria, feedback to help interpret results, and information that is useful in planning programs for improvement of fitness through regular physical activity. Teachers should include ACTIVITYGRAM® student reports as part of student physical education portfolios along with other information related to important physical education objectives. Reports should also be provided to parents to help them understand daily physical activity needs. ACTIVITYGRAM® sends a powerful message that by assessing where they are, setting appropriate goals, and creating and enacting plans that work toward the chosen goals, students can achieve healthy levels of physical activity. Parents should be encouraged to use the messages on the ACTIVITYGRAM® report to help students plan personal physical activity programs that are suited to each child's individual needs.

Personal Tracking

Personal Tracking can also be used with ACTIVITYGRAM®. Student activity results are plotted on a regular basis to see if youth maintain active lifestyles over time. The goal is to help all youth to achieve 60 minutes of daily physical activity over time. Improvement can be tracked and celebrated for physical activity just as it is for the various fitness components.

Meeting National Physical Education Standards and Guidelines

One of the principal national physical education standards is the achievement of regular physical activity (National Association for Sport and Physical Education, 2004). Physical

⇦TOC ⇦Chapter

education and the implementation of models such as Coordinated School Health and Comprehensive School Physical Activity Programs have outcomes concentrated on both the achievement and maintenance of health-enhancing levels of fitness and regular engagement in physical activity, as these variables are independent risk factors associated with health (Plowman, 2005). ACTIVITYGRAM® and associated educational programs help students meet national physical education standards and guidelines.

Why Use ACTIVITYGRAM®?

Basis for Fitness Education

ACTIVITYGRAM® was designed to help youth learn to self-monitor their personal physical activity patterns. It is a self-assessment program that helps students determine current activity levels in a variety of different activities. Like FITNESSGRAM®, the primary purpose of ACTIVITYGRAM® is to facilitate the promotion of physical activity for a lifetime. Learning to self-assess and regularly monitor physical activity helps students see "how active they really are" and helps them set goals for planning lifetime physical activity programs. Self-monitoring, goal-setting, and program planning are considered to be "self-management skills" and learning self-management skills is considered essential to lifetime physical activity adherence (Dale & Corbin, 2000; Dale, Corbin, & Cuddihy, 1998). ACTIVITYGRAM® is designed as a tool to aid in effective learning of self-management skills. The ACTIVITYGRAM® produces a report that summarizes the results of the individual's activity and provides feedback to the individual.

ACTIVITYGRAM® Reports

As noted earlier, the FITNESSGRAM® software offers many reporting options for students, parents, teachers, and schools (local, district, and state) including a fitness summary report. The software also provides an ACTIVITYGRAM® report that summarizes physical activity patterns of the student. Activity amounts in each of four areas are reported: moderate (lifestyle activity), vigorous (active aerobics, active sports, and active recreation), muscle fitness exercise, and flexibility exercise. In addition advice is provided as to how to proceed in order to become more active. As also noted earlier, the use of health report cards has been shown to be motivating to parents (Chomitz et al., 2003) and is consistent with the policy of the American Academy of Pediatrics (2006) that recommends tracking of activity patterns of youth and parents.

Partnerships

Because of its scientific basis and history of effectiveness, FITNESSGRAM®/ACTIVITYGRAM® and The Cooper Institute now partner with the President's Council on Fitness, Sports, and Nutrition (PCFSN), the American Alliance for Health, Physical Education, Recreation and Dance (AAHPERD), the Centers for Disease Control and Prevention (CDC), the Amateur Athletic Union (AAU), and the National Association for Sport and Physical Education's (NASPE) Physical Best to deliver a fitness assessment program as well as fitness education programs known as the Presidential Youth Fitness Program. The ACTIVITYGRAM® is a part of this comprehensive fitness education program providing an important indicator of student success in achieving a health-enhancing level of physical activity.

⇦TOC ⇦Chapter

Tracking and Data Management

In addition to individual reports, ACTIVITYGRAM® offers individual and group summary reports (see Chapter 9 Interpreting FITNESSGRAM® and ACTIVITYGRAM® Reports) that enable individuals to go from just seeing results to tracking the results over time and using those results to impact future planning and achievement.

What Are the Recommended or Appropriate Uses of ACTIVITYGRAM®?

Using the ACTIVITYGRAM® periodically provides students with an assessment of their activity levels, a time profile for when they are most active, and a profile of the types of activity in which they participate. Students, teachers, and parents can use reports of student activity levels to achieve benefits similar to those outlined for FITNESSGRAM® in the preceding table. Group data based on activity scores from ACTIVITYGRAM® can be used for curriculum development, research, and other appropriate institutional purposes. More information is available in the chapter devoted to physical activity assessments.

Bibliography[1]

[1]The advisors would like to acknowledge the Institute of Medicine. (2012). *Fitness measures and health outcomes in youth*. Washington, DC: The National Academies Press, as the source of many of the bibliographical listings included in this chapter.

American Academy of Pediatrics Committee on Nutrition. (2003). Policy Statement: Prevention of Pediatric Overweight and Obesity. *Pediatrics*, *112*, 424-430.

American Academy of Pediatrics Council on Sports Medicine and Fitness and Council on School Health. (2006). Policy Statement: Active Healthy Living—Prevention of Childhood Obesity Through Increased Physical Activity. *Pediatrics*, *125,(2),* 1156.

American Alliance for Health, Physical Education, Recreation, and Dance. (1980). *Health-related physical fitness test manual*. Reston, VA: Author.

Armour, K. M., & Yelling, M. (2007). Effective professional development for physical education teachers: The role of informal, collaborative learning. *Journal of Teaching in Physical Education*, *26*, 177-200.

Artero, E. G., Espana-Romero, V., Castro-Pinero, J., Ortega, F. B., Suni, J., Castillo-Garzon, M. J., & Ruiz, J. R. (2011). Reliability of field-based fitness tests in youth. *International Journal of Sports Medicine, 32,* 159-169.

Baranowski, T., Mendlein, J., Resnicow, K., Frank, E., Cullen, K. W., & Baranowski, J. (2000). Physical activity and nutrition in children and youth: An overview of obesity prevention. *Preventive Medicine*, *31*, S1-S10.

Bouchard, C., & Shephard, R. J. (1994). Physical activity, fitness and health: The model and key concepts: International proceedings and consensus statement. In C. Bouchard, R. J. Shephard, and T. Stephens (eds.), *Physical activity, fitness and health* (pp. 11-22). Champaign, IL: Human Kinetics.

Cale, L., & Harris, J. (2009). Fitness testing in physical education: A misdirected effort in promoting healthy lifestyles and physical activity? *Physical Education and Sport Pedagogy*, *14*(1), 89-108.

Castelli, D. M., Hillman, C. H., Buck, S. M., & Erwin, H. E. (2007). Physical fitness and academic achievement in third- and fifth-grade students. *Journal of Sport and Exercise Psychology*, *29*, 239-252.

Castro-Pinero, J., Artero, E. G., Espana-Romero, V., Ortega, F. B., Sjostrom, M., Suni, J., & Ruiz, J. R. (2010). Criterion-related validity of field-based fitness tests in youth: A systematic review. *British Journal of Sports Medicine*, *44*, 934-943.

Centers for Disease Control and Prevention. (2010). *The association between school-based physical activity, including physical education, and academic performance.* Atlanta, GA: U.S. Department of Health and Human Services.

Centeio, E. F., & Keating, X. D. (2011). How phat is your fitness knowledge? *Research Quarterly for Exercise and Sport, 82,* A3.

Chen, S., Chen, A., & Zhu, X. (2012). Are K-12 learners motivated in physical education? A meta-analysis. *Research Quarterly for Exercise and Sport, 83,* 36-48.

Chomitz, V. R. (2003). Promoting healthy weight among elementary school children via a Health Report Card approach. *Archives of Pediatrics and Adolescent Medicine*. 157 (8), 765-772.

Corbin, C. B. (2009). Appropriate use of physical fitness tests. *The CAHPERD Journal*, *72*(1),

20-25, 40.
Corbin, C. B. (2010). Texas Youth Fitness Study: A commentary. *Research Quarterly for Exercise and Sport, 81*(Suppl. 3), S75-S78.
Corbin, C. B., Lovejoy, P. Y., Steingard, P., & Emerson, R. (1990). Fitness awards: Do they accomplish their intended objectives? *American Journal of Health Promotion, 4,* 345-351.
Corbin, C. B., & Pangrazi, R. P. (1992). Are American children and youth fit? *Research Quarterly for Exercise and Sport, 63*, 96-106.
Corbin, C. B., & Pangrazi, R. P. (2008). Appropriate uses of *FITNESSGRAM*®. In G. J. Welk and M. D. Meredith (Eds.). *FITNESSGRAM®/ACTIVITYGRAM®* reference guide (3rd ed.). Dallas, TX: The Cooper Institute.
Corbin, C. B., Pangrazi, R. P., & LeMasurier, G. C. (2004). Physical activity for children: Current patterns and guidelines. *President's Council on Physical Fitness and Sports Research Digest*, *5*(2), 1-8.
Corbin, C. B., Whitehead, J. R., & Lovejoy, P. Y. (1988). Youth physical fitness awards. *Quest, 40*, 200-218.
Dale, D., & Corbin, C. B. (2000). Physical activity participation of high school graduates following exposure to conceptual or traditional physical education. *Research Quarterly for Exercise and Sport, 71,* 61-68.
Dale, D., Corbin, C. B., & Cuddihy, T. F. (1998). Can conceptual physical education promote physically active lifestyles? *Pediatric Exercise Science*, *10*, 97-109.
Donnelly, J. E., & Lambourne, K. (2011). Classroom-based physical activity, cognition, and academic achievement. *Preventive Medicine, 52,* S36-S42.
Engelman, M. E., & Morrow, J. R., Jr. (1991). Reliability and skinfold correlates for traditional and modified pull-ups in children grades 3-5. *Research Quarterly for Exercise and Sport, 62,* 88-91.
Ennis, C. D. (2007). Defining learning as conceptual change in physical education and physical activity settings. *Research Quarterly for Exercise and Sport, 78,* 138-150.
Erbaugh, S. J. (1990). Reliability of physical fitness tests administered to young children. *Perceptual and Motor Skills, 71*(3, Pt. 2), 1123-1128.
Ernst, M. P., Corbin, C. B., Beighle, A., & Pangrazi, R. P. (2006). Appropriate and inappropriate uses of *FITNESSGRAM*®: A commentary. *Journal of Physical Activity and Health, 3(*Suppl. 2), S90-S100.
Espana-Romero, V., Artero, E. G., Jimenez-Pavon, D., Cuenca-Garcia, M., Ortega, F. B., Castro-Pinero, J., Flegal, K. M., Carroll, M. D., Kuczmarski, R. J., & Johnson, C. L. (1998). Overweight and obesity in the United States: Prevalence and trends, 1960-1994. *International Journal of Obesity, 22,* 39-47.
Espana-Romero, V., Artero, E. G., Jimenez-Pavon, D., Cuenca-Garcia, M., Ortega, F. B., Castro-Pinero, J., Sjostrom, M., Castillo-Garzon, M. J., & Ruiz, J. R. (2010). Assessing health-related fitness tests in the school setting: Reliability, feasibility and safety; the ALPHA Study. *International Journal of Sports Medicine, 31,* 490-497.
Fox, K. R. (1988). The self-esteem complex and youth fitness. *Quest, 40*, 230-246.
Freedson, P. S., Cureton, K. J., & Heath, G. W. (2000). Status of field-based fitness testing in children and youth. *Preventive Medicine, 31,* S77-S85.
Gordon-Larsen, P., McMurray, R. G., & Popkin, B. M. (2000). Determinants of adolescent physical activity and inactivity patterns. *Pediatrics*, *105*(6):E83.

Harrington-Leuker, D. (2000). When educators cheat. *School Administrator, 11,* 32-39.
Harris, J., & Cale, L. (2007). Children's fitness testing: A feasibility study. *Health Education Journal, 66*, 153-172.
Hasselstrom, H., Hansen, S. E., Froberg, K., & Andersen, L. B. (2002). Physical fitness and physical activity during adolescence as predictors of cardiovascular disease risk in young adulthood. Danish Youth and Sports Study. An eight-year follow-up study. *International Journal of Sports Medicine, 23,* S27-S31.
Hillman, C. H., Buck, S. M., Themanson, J. R., Pontifex, M. B., & Castelli, D. M. (2009). Aerobic fitness and cognitive development: Event-related brain potential and task performance indices of executive control in preadolescent children. *Developmental Psychology, 45,* 114-129.
Institute of Medicine. (2012). *Fitness measures and health outcomes in youth*. Washington, DC: The National Academies Press.
Kamijo, K., Pontifex, M. B., O'Leary, K. C., Scudder, M. R., Wu, C. T., Castelli, D. M., & Hillman, C. H. (2011). The effects of an afterschool physical activity program on working memory in preadolescent children. *Developmental Science, 14,* 1046-1058.
Keating, X. D. (2003). The current often implemented fitness tests in physical in physical education programs: Problems and future directions. *Quest*, *55*, 141-160.
Kibbe, D. L., J. Hackett, J., Hurley, M., McFarland, A., Schubert, K. G., Schultz, A., & Harris, S. (2011). Ten years of take 10!®: Integrating physical activity with academic concepts in elementary school classrooms. *Preventive Medicine*, *52*, S43-S50.
Kohn, A. (2001). Fighting the tests: A practical guide to rescuing our schools. *Phi Delta Kappan, 82,* 349-357.
Kollath, J. A., Safrit, M. J., Zhu, W., & Gao, L. G. (1991). Measurement errors in modified pull-ups testing. *Research Quarterly for Exercise and Sport*, *62*, 432-435.
Kulinna, P. H., & Silverman, S. (2000). Teachers' attitudes toward teaching physical activity and fitness. *Research Quarterly for Exercise and Sport, 71,* 80-84.
Lee, S. M., Wechsler, H., & Balling, A. (2006). The role of schools in preventing childhood obesity. *President's Council on Physical Fitness and Sports Research Digest*, *7*(3), 1-8.
Lohman, T., Going, S. B., & Metcalfe, M. S. (2004). Seeing ourselves through the obesity epidemic. *President's Council on Physical Fitness and Sports Research Digest*, *5*(3), 1-8.
Martin, S. B., Ede, A., Morrow, J. R., Jr., & Jackson, A. W. (2010). Statewide physical fitness testing: Perspectives from the gym. *Research Quarterly for Exercise and Sport*, *81*(Suppl. 3), S31-S41.
Meredith, M. D., & Welk, G. J. (Eds.) (2010). *FITNESSGRAM® & ACTIVITYGRAM® test administration manual* (updated 4th ed). Champaign, IL: Human Kinetics.
Morrow, J. R., Jr., & Freedson, P. S. (1994). Relationship between habitual physical activity and aerobic fitness in adolescents. *Pediatric Exercise Science, 6,* 315-329.
Morrow, J. R., Jr., Jackson, A. W., & Payne, V. G. (1999). Physical activity promotion and school physical education. *President's Council on Physical Fitness and Sports Research Digest, 3*(7), 1-8.
Morrow, J. R., Jr., Martin, S. B., & Jackson, A. W. (2010). Reliability and validity of the *FITNESSGRAM®*: Quality of teacher-collected health-related fitness surveillance data. *Research Quarterly for Exercise and Sport*, *81*(Suppl. 3), S24-S30.

National Association for Sport and Physical Education. (1992). *Outcomes of a quality physical education program*. Reston, VA: NASPE.

National Association for Sport and Physical Education. (2004). *Moving into the future: National standards for physical education* (2nd ed.). Reston, VA: NASPE.

National Association for Sport and Physical Education. (2009a). *Appropriate instructional practice guidelines for elementary school physical education: A position statement* (3rd ed.). http://www.cahperd.org/cms-assets/documents/toolkit/naspe_apprroprac/5287-207931.elementaryappropr ac.pdf

National Association for Sport and Physical Education. (2009b). *Appropriate instructional practice guidelines for high school physical education: A position statement* (3rd ed.). http://www.cahperd.org/cms-assets/documents/toolkit/naspe_apprroprac/5288-573262.hsappropr ac.pdf

National Association for Sport and Physical Education. (2009c). *Appropriate instructional practice guidelines for middle school physical education: A position statement* (3rd ed.). http://www.cahperd.org/cms-assets/documents/toolkit/naspe_apprroprac/5289-666992.msappropr ac.pdf

National Association for Sport and Physical Education. (2010a). *Shape of the Nation Report: Status of physical education in the USA.* http://www.aahperd.org/naspe/publications/Shapeofthenation.cfm

National Association for Sport and Physical Education (2010b). *Appropriate uses of fitness measurement: Position statement* http://www.aahperd.org/naspe/standards/PEPS.cfm

National Association for Sport and Physical Education & American Heart Association. (2012). *2012 Shape of the Nation Report: Status of physical education in the USA*. Reston, VA: American Alliance for Health, Physical Education, Recreation and Dance.

Ogden, C. L., Carroll, M. D., Curtin, L. R., Lamb, M. M., & Flegal, K. M. (2010). Prevalence of high body mass index in US children and adolescents, 2007-2008. *Journal of the American Medical Association, 303,* 242-249.

O'Sullivan, M., & Deglau, D. (2006). Chapter 7: Principles of professional development. *Journal of Teaching in Physical Education, 25,* 441-449.

Pangrazi, R. P., & Corbin, C. B. (1990). Age as a factor relating to physical fitness test performance. *Research Quarterly for Exercise and Sport, 61,* 410-414.

Pangrazi, R. P., & Corbin, C. B. (2004). Factors that influence physical fitness in children and adolescents. In G. J. Welk, J. R. Morrow, and H. Falls. (Eds.). *FITNESSGRAM® reference guide (*pp. 28-36). Dallas, TX: The Cooper Institute.

Parish, L. E., & Treasure, D. C. (2003). Physical activity and situational motivation in physical education: Influence of the motivational climate and perceived ability. *Research Quarterly for Exercise and Sport, 74,* 173-182.

Pate, R., & Corbin, C. (1981). Implications for curriculum. *Journal of Physical Education and Recreation, 52*(1), 36-38.

Pate, R. R., Dowda, M., & Ross, J. G. (1990). Associations between physical activity and physical fitness in American children. *American Journal of Diseases in Children, 144,* 1123-1129.

Pate, R. R. Pratt, M., Blair, S. N., Haskell, W. L., Macera, C. A., Bouchard, C., et al.1995). Physical activity and public health. A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. *Journal of the American Medical Association, 273,* 402-407.

⇦TOC ⇦Chapter

Pate, R. R., Wang, C. Y., Dowda, M., Farrell, S. W., & O'Neill, J. R. (2006). Cardiorespiratory fitness levels among US youth 12 to 19 years of age: Findings from the 1999-2002 National Health and Nutrition Examination Survey. *Archives of Pediatrics and Adolescent Medicine, 160,* 1005-1012.

Pate, R. R., Ward, D. S., O'Neill, J. R., & Dowda, M. (2007). Enrollment in physical education is associated with overall physical activity in adolescent girls. *Research Quarterly for Exercise and Sport, 78*, 265-270.

Pate, R. R., Saunders, R. P., O'Neill, J. R., & Dowda, M. (2011). Overcoming barriers to physical activity: Helping youth be more active. *ACSM Health and Fitness Journal, 15*(1), 7-12.

Payne, V. G., & Morrow, J. R., Jr. (1993). Exercise and VO_2max in children: A meta-analysis. *Research Quarterly for Exercise and Sport, 64*, 305-313.

Physical Activity Guidelines Advisory Committee. (2008). *Physical activity guidelines advisory committee report, 2008.* Washington, DC: U.S. Department of Health and Human Services.

Placek, J. H., Griffin, L. L., Dodds, P., Raymond, C., Tremino, F., & James, A. (2001). Middle school students' conceptions of fitness: The long road to a healthy lifestyle. *Journal of Teaching in Physical Education, 20,* 314-323.

Plowman, S. A. (2005). Physical activity and physical fitness: Weighing the relative importance of each. *Journal of Physical Activity and Health, 2,* 143-158.

Plowman, S. A. (2008). Muscular strength, endurance, and flexibility assessments. In G. J. Welk and M. D. Meredith (Eds.), *FITNESSGRAM®/ACTIVITYGRAM®* reference guide (pp. 11.11-11.40). Dallas, TX: The Cooper Institute.

Plowman, S. A., Sterling, C. L., Corbin, C. B., Meredith, M. D., Welk, G. J., & Morrow, J. R., Jr. (2006). The history of *FITNESSGRAM®*. *Journal of Physical Activity and Health, 3*(Suppl. 2), S5-S20.

Ruiz, J. R., Castro-Pinero, J., Espana-Romero, V., Artero, E. G., Ortega, F. B., Cuenca, M. M., . . . Castillo, M. J. (2011). *British Journal of Sports Medicine, 45,* 518-524.

Safrit, M. J. (1990). The validity and reliability of fitness tests for children: A review. *Pediatric Exercise Science, 2,* 9-28.

Safrit, M. J. (1995). *Complete guide to youth fitness testing*. Champaign, IL: Human Kinetics.

Saint Romain, B., & Mahar, M. T. (2001). Norm-referenced and criterion-referenced reliability of the push-up and modified pull-up. *Measurement in Physical Education and Exercise Science, 5,* 67-80.

Simons, J., Dewitte, S., & Lens, W. (2003). "Don't do it for me. Do it for yourself!" Stressing the personal relevance enhances motivation in physical education. *Journal of Sport and Exercise Psychology, 25,* 145-160.

Sjostrom, M., Castillo-Garzon, M. J., & Ruiz, J. R. (2010). Assessing health-related fitness tests in the school setting: Reliability, feasibility and safety; the ALPHA Study. *International Journal of Sports Medicine, 31,* 490-497.

Sloane, F. C., & Kelly, A. E. (2003). Issues in high-stakes testing programs. *Theory into Practice*, *42*(1), 12-17.

Standage, M., Duda, J. L., & Ntoumanis, N. (2003). Predicting motivational regulations in physical education: The interplay between dispositional goal orientations, motivational climate and perceived competence. *Journal of Sports Sciences, 21*, 631-647.

Stroot, S. A., Collier, C., O'Sullivan, M., & England, K. (1994). Contextual hoops and hurdles:

⇦TOC ⇦Chapter

Workplace conditions in secondary physical-education. *Journal of Teaching in Physical Education, 13*, 342-360.

Tremblay, M., & Lloyd, M. (2010). Physical literacy measurement: The missing piece. *Physical and Health Education Journal, 76*(1), 26-30.

United States Department of Health and Human Services. (2008). *2008 Physical Activity Guidelines for Americans*. Retrieved from http://www.health.gov/paguidelines/

Welk, G. J., Jackson, A. W., Morrow, J. R. Jr., Haskell, W. H., Meredith, M. D., & Cooper, K. H. (2010). The association of health-related fitness with indicators of academic performance in Texas schools. *Research Quarterly for Exercise and Sport, 81*(Suppl. 3), S16-S23.

Whitehead, J. R., & Corbin, C. B. (1991). Youth fitness testing: The effect of percentile-based evaluative feedback on intrinsic motivation. *Research Quarterly for Exercise and Sport, 62*, 225-231.

Winnick, J. P., & Short, F. X. (1999). *The Brockport Physical Fitness Test Manual*. Champaign, IL: Human Kinetics Publishers.

Chapter 3

Health Benefits of Physical Activity and Fitness in Youth

Joey C. Eisenmann, Gregory J. Welk,
James R. Morrow, Charles B. Corbin

The FITNESSGRAM® Reference Guide is intended to provide answers to some common questions associated with use and interpretation of FITNESSGRAM® assessments. This chapter, devoted to the Health Benefits of Physical Activity and Fitness in Youth, describes the concepts of physical activity, fitness and health, and the links between them, in children and adolescents. Conceptually, this chapter provides a foundation for the health-related focus used in the FITNESSGRAM® and in the interpretation of the assessments. It is important that teachers, and those who administer FITNESSGRAM®/ACTIVITYGRAM®, understand the theoretical constructs that underlie the program.

What Is the Difference Between Physical Activity and Physical Fitness

Many people think that physical activity and physical fitness are the same thing and often use the terms physical activity and physical fitness interchangeably, assuming that they are directly related. Physical activity and physical fitness are related but are actually two very different concepts. Physical activity is a behavior (something that you do) while physical fitness is a biological or physical trait or characteristic (something that you have). Below are definitions of physical activity and physical fitness.

- *Physical activity:* Physical Activity is defined as "any bodily movement produced by skeletal muscles that results in energy expenditure" (Casperson et al, 1985). When educating youth about physical activity, the focus is on health-enhancing energy expenditure resulting from the large muscles (primarily moderate to vigorous activity). It should be noted, however, that activities of less intensity (light activity) can have some health benefits.
- *Physical fitness:* Physical Fitness can be defined as "a set of attributes that people have or achieve that relates to the ability to perform physical activity." This is the definition used in both the Surgeon General's Report on Physical Activity and Health (HHS, 1996, p. 21) and the Institute of Medicine (IOM) report. Furthermore, physical fitness has been divided into health-related and skill- or performance-related fitness. Given the focus on health, health-related physical fitness is defined as "a state of being that reflects a person's ability to perform specific forms of physical activity/exercise or functions, and is related to present and future health outcomes."

What Are the Dimensions or Components of Health-Related Fitness?

Health-related fitness has been defined by individuals and groups in different ways over the years and the definition greatly influences how it is further characterized. Despite different definitions, there is consensus that health-related fitness is a multi-dimensional construct, meaning that there are several attributes or components, not just one test, that define one's health-related physical fitness. With that said, some components also have been referred to in many ways. For example and most notably, the ability of a person to perform sustained endurance activity as a function of the heart, lungs, blood vessels, and the muscular system has been called maximal aerobic capacity, maximal aerobic power, aerobic fitness, cardiorespiratory fitness, cardiorespiratory endurance, etc. Therefore, the reader should understand that these terms will be used throughout this chapter based on the organization referring to this concept. Historically, FITNESSGRAM® has referred to this trait as aerobic capacity.

The Classic Definition (1985)

A classic definition by Caspersen and colleagues (Caspersen, Powell, and Christenson, 1985) has served as the basis for FITNESSGRAM ®and many youth fitness programs, curricula, and textbooks. This model suggests that there are five main components of physical fitness: 1) body composition, 2) cardiorespiratory endurance, 3) muscular strength, 4) muscular endurance, and 5) flexibility. This multi-dimensional model of health-related fitness is well accepted; however, variations of this model have been proposed and are described below. In addition, the view of health-related fitness by FITNESSGRAM® is also provided.

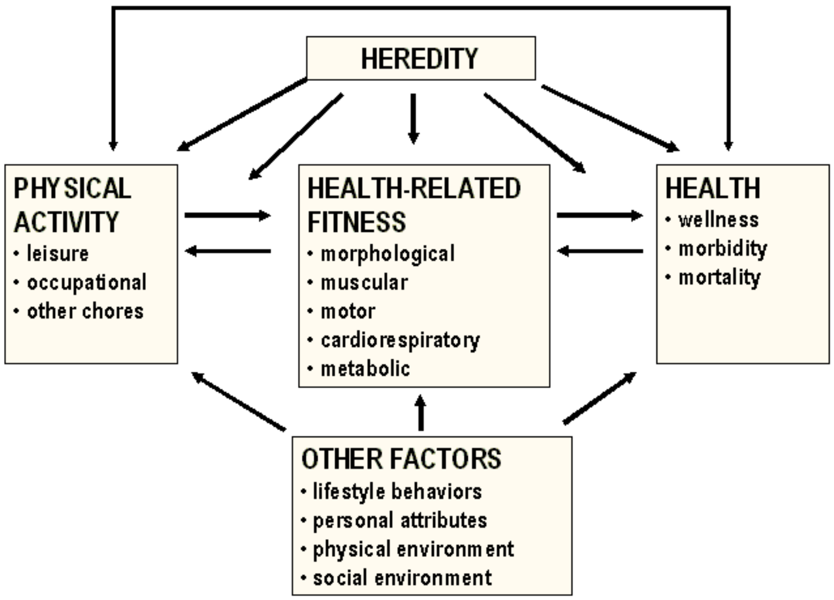
Consensus Guidelines Model (1994)

A landmark consensus conference on physical activity, fitness, and health led to the creation of a detailed conceptual framework for health-related fitness (Bouchard & Shephard,

⇦TOC ⇦Chapter

1994). The model is conceptually the same as the one shown in Figure 3 but considers the impact of heredity and proposes different dimensions or components of physical fitness (morphological, muscular, motor, cardiorespiratory, and metabolic). The model includes different dimensions than proposed by Caspersen et al. (1985), but this is because it takes a broader approach with regard to the concept of health-related fitness. The model, for example, illustrates a number of various sub-dimensions within each major dimension to capture the diverse impacts of fitness on health. Descriptions of each of the major dimensions (and sub-dimensions) are provided below.

Figure 1. Framework for Health-Related Fitness—Consensus Guidelines (1994).



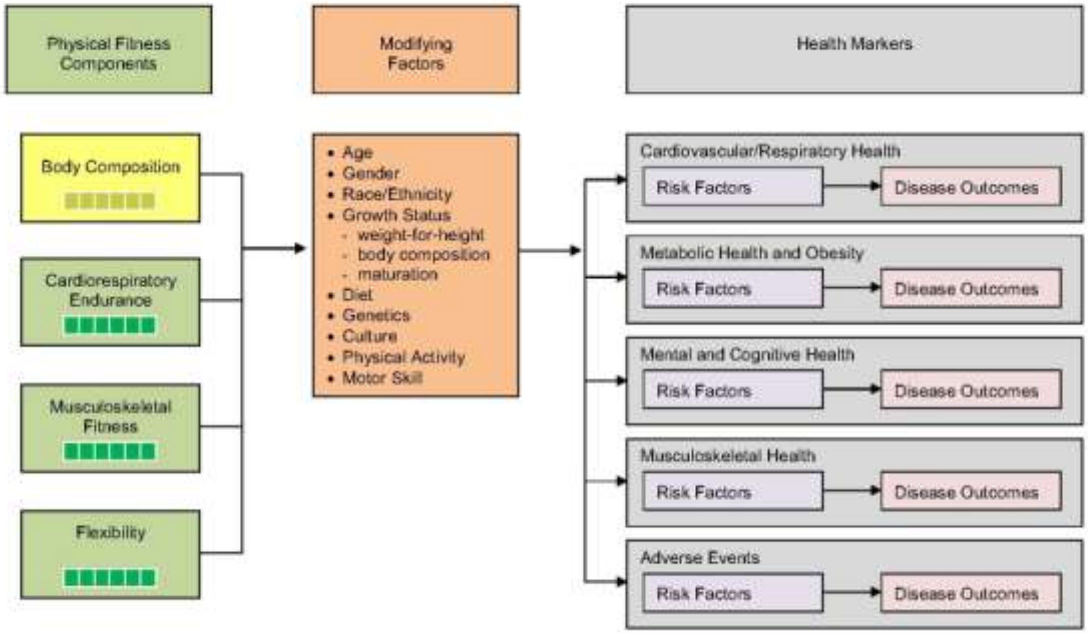
- *Morphological fitness*: Morphological fitness includes a variety of indicators that reflect the structure and composition of the body (e.g., subcutaneous and visceral adipose tissue, body fat distribution, and bone mineral density). This component has often been referred to as body composition but the term “morphological fitness” is broader and more inclusive. Body fatness is the most common indicator of morphological fitness and it is known to have important influences on health. Bone density is another important indicator of morphological fitness in youth since it is most effective to build bone density during adolescence and early adulthood.
- *Muscular fitness*: Muscular fitness captures a diverse range of muscular fitness constructs including dimensions of power, muscular strength, and muscular endurance. Aside from the importance of muscular fitness to athletic ability, recent studies indicate a relationship between muscular fitness and cardio-metabolic health in youth (Magnussen et al., 2012) and muscular fitness in adolescence and cardiovascular disease and mental health in young adulthood (Ortega et al. 2012).
- *Motor fitness*: Motor fitness refers to components that are thought to improve the ability to learn and perform motor skills and include balance, coordination, agility, and speed. Many refer to this component as skill-related or performance-related fitness. The acquisition and improvement of motor skills is often a primary focus of physical education and youth sports programs since it is assumed that this foundation provides a basis for the maintenance of active lifestyles over time.

- *Cardiorespiratory fitness*: Cardiorespiratory fitness refers to function of the heart, lungs, blood vessels, and the muscular system involved in movement. It is typically evaluated or quantified with measurements of aerobic power which reflect the maximal amount of oxygen that can be taken in and consumed during maximal exercise (called "VO_2max"). However, it is important to point out that the category of cardiorespiratory fitness also includes other indicators of heart function (e.g., blood pressure) and lung function, and also the ability to perform sustained submaximal exercise.
- *Metabolic fitness*: Metabolic fitness refers to biochemical indicators related to cardiovascular disease and type II diabetes (e.g., blood lipids and glucose metabolism). While these indicators have not been included in traditional conceptions of fitness they are known to be influenced by physical activity and cardiorespiratory fitness and to relate directly to health. Indicators of metabolic fitness are also known to cluster together as part of an overall 'metabolic syndrome' that is known to predispose individuals to cardiovascular disease and diabetes. Negative indicators of poor metabolic fitness were previously thought to occur only in adulthood but it is now well established that the risks can occur early in life. Physical inactivity has been associated with the onset of early cardiovascular disease in children but it has proven difficult to determine the independent contributions of physical activity/physical inactivity and obesity on metabolic fitness. For example, physical activity is known to moderate the impact of obesity on health outcomes. The current FITNESSGRAM® health-related standards for aerobic capacity and body composition were based on associations with metabolic syndrome (See chapters 6 and 7)

Institute of Medicine Framework (2012)

The U.S. government sanctioned Institute of Medicine (IOM) group released a report in 2012 entitled "Fitness Measures and Health Outcomes in Youth" (IOM, 2012) that provided a comprehensive overview of health-related physical fitness in youth. The IOM report included a new framework for understanding how dimensions of health-related fitness influence health (IOM 2012, section 3-2). In the IOM report, the focus was on examining the associations between fitness and health. Therefore, some sub-dimensions of health-related fitness proposed in the Consensus Guidelines (e.g., bone density, metabolic fitness) are viewed as outcomes rather than as separate components of health-related fitness. The term "health marker" was used to denote variables often referred to as "risk factors" in adults (e.g., elevated blood pressure, blood lipids, blood sugar, and body fat). This is because these markers do not typically have impacts on specific health outcomes such as cardiovascular disease or musculoskeletal problems until later in life. Similar to the models described above, the IOM model depicts a number of "modifying factors" that influence the relationship between physical fitness and health markers. The model is shown below in Figure 2 followed by definitions for each of the specific dimensions of health-related physical fitness (IOM, 2012, page 3-2).

Figure 2: Framework for Health-Related Fitness—Institute of Medicine.



The IOM report uses a simpler classification of health-related dimensions compared to the Consensus Guidelines described above with terms that are more consistent with the original categories proposed by Caspersen et al. (1985). The IOM report identified four general components of physical fitness:

- *Body Composition*: "The components that make up body weight, including fat, muscle, and bone content (IOM, 2012, p. 1-2)."
- *Cardiorespiratory Endurance*: "The ability to perform large-muscle, whole-body exercise at moderate to high intensities for extended periods of time (IOM, 2012, p. 1-2 adapted from Saltin, 1973)."
- *Musculoskeletal Fitness*:
  - *Muscle Strength*. The ability to use the muscles to lift a heavy weight or exert considerable force. Technically, muscle strength is defined as "the ability of a single muscle or group of muscles to produce force, torque, or movement about a single or multiple joints, typically during a single maximal voluntary contraction and under a defined set of controlled conditions, which include specificity of movement pattern, muscle contraction type (concentric, isometric, or eccentric), and contraction velocity (IOM, 2012, page 6-3 adapted from Farpour-Lambert and Blimkie, 2008; Kell et al., 2001; Sale and Norman, 1982)."
  - *Muscle Endurance*. The ability to repeatedly use muscles over time without tiring. The technical definition of muscle endurance is "the ability of a muscle or group of muscles to perform repeated contractions against a constant external load for an extended period of time. The constant load can be either an absolute external resistance, which provides a measure of absolute endurance, or a relative load based on an individual's maximal strength, which provides a measure of relative endurance (IOM, 2012, page 6-3 adapted from Kell et al., 2001)."

- *Muscle Power*. The ability to generate force quickly. "Muscle power is a physiological construct reflecting the rate at which work is performed" (IOM, 2012, page 6-3, adapted from Knuttgen and Kraemer, 1987). It is derived from the product of the force production of a muscle or group of muscles and the velocity of the muscle contraction during a single- or multi-joint action.

• Flexibility. "The ability to move the joints through a range of motion. Flexibility reflects the intrinsic property of body tissues (e.g., muscles, tendons, bones) that determines the range of motion achievable without injury at a joint or group of joints (IOM, 2012, page 1-2, adapted from Holt et al., 1996, p. 172)."

Dimensions of Health-Related Fitness in FITNESSGRAM®

The descriptions and conceptual models above are provided to show the different ways in which health-related physical fitness has been described. The FITNESSGRAM® program has utilized a hybrid depiction of health-related physical fitness that is consistent with the general IOM framework.

FITNESSGRAM® categorizes the dimensions of health-related fitness into 3 main categories: 1) Aerobic capacity; 2) Musculoskeletal fitness including muscle strength, muscular endurance, and flexibility; and 3) Body Composition.

These same categories are used in the IOM report; however, there are several important differences. One difference is that the IOM report includes muscle power as a specific dimension of musculoskeletal fitness. This is justified by recent evidence that shows associations between muscle power and various health markers, mainly in adults (see IOM report p 148). Another difference is that flexibility in the IOM report was viewed as an independent dimension of fitness rather than being linked to the dimension of musculoskeletal fitness. This is partly because there was limited evidence linking flexibility to specific health-related outcomes (i.e., low back pain, etc.) among youth.

It is important to note that the IOM recommendations were based on the fitness items that would be most important in a national survey of fitness while the focus of FITNESSGRAM® is on education and promotion of physical activity and health-related fitness. The inclusion of assessments of muscle power may be justified in national surveillance studies because they provide useful predictors of health outcomes but has not been adapted as a health-related fitness at this time. Further study is necessary. Similarly, the exclusion of flexibility may make sense for health-related surveillance, but it is important within FITNESSGRAM® for youth to learn about flexibility and its importance to lifelong health. The IOM report specifically acknowledged the distinction between surveillance and education applications as well as the unique needs for more practical assessments for school-based fitness evaluation.

The multi-dimensional nature of health-related fitness is an important consideration for planning and evaluating youth fitness programming. The distinction explains why youth may achieve a good level of health related fitness in one dimension but not others.

Other Considerations Related to Fitness

Health-Related Fitness and Performance. Typically, we think of the skill-related components of fitness when discussing performance. Skill-related fitness is associated with successful performance of sports and other skill-based activities. It is important that skill-related fitness be included in fitness education programs so that youth can understand the different components of fitness and the benefits of each. It should be noted, however, that health-related fitness components are also important to performance. As Corbin & Le Masurier (in press) note, cardiovascular endurance is important to success in most sports (e.g., cross country, track and field, soccer). Muscle fitness including strength, muscular endurance, and power are important to most sports (e.g., volleyball, football, wrestling) and many jobs (e.g., military, law enforcement, firefighting). While skill-related fitness is one type of fitness, it is not the focus of the FITNESSGRAM®/ACTIVITYGRAM® programs.

Functional Fitness. Functional fitness refers to the ability to function effectively when performing normal daily tasks (Corbin & LeMassurier, in press). While the first priority of the assessment program is health-related fitness, attention to functional fitness is also considered important. As noted by Corbin & LeMasurier (in press) functional fitness helps you "do your school work, get to and from school and participate in leisure time activities without fatigue, respond to emergency situations, and perform other daily tasks safely and without fatigue (e.g., drive a car, do house and yard work)." The tests in FITNESSGRAM® help youth to be healthy and to function effectively in daily living.

How Are Physical Activity, Physical Fitness, and Health Related?

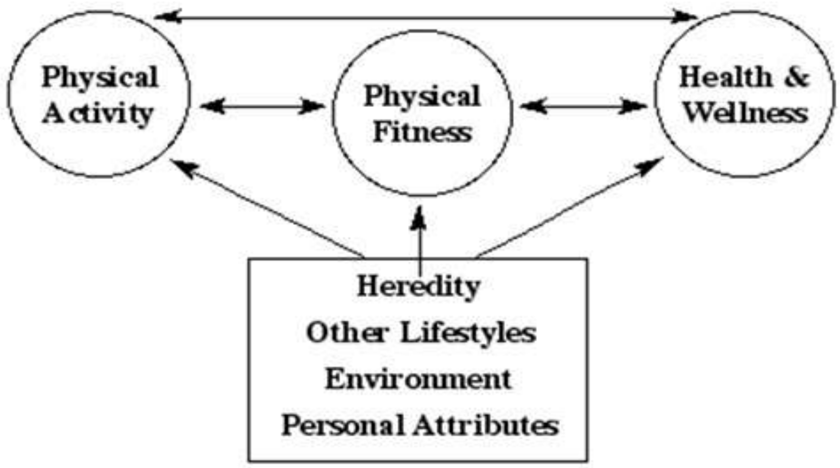
Although they have distinct meanings (see above), physical activity clearly contributes to physical fitness (and vice versa) but the relations between them are not as strong as many might expect. There are a several additional factors influencing levels of physical fitness, and many are outside a person's control.

There is also considerable debate about whether physical activity or physical fitness is more important to health. Research (Blair et al 2001; Paffenbarger et al 1996.) has consistently demonstrated that physical activity and physical fitness both influence health and that the effects are largely independent. This implies that a person needs to be physically active even if they have reasonable levels of fitness. Individuals with low levels of fitness can also obtain health benefits by remaining physically active.

A conceptual model highlighting the reciprocal relationships between physical activity, physical fitness, and health is shown below in Figure 3 (Bouchard (1990). The model also indicates that a number of other factors influence physical activity, physical fitness, and health status. Some of these factors are out of a person's control (e.g., genetics and rate of maturation) but lifestyle behaviors such as sleep, nutrition, and stress management can be modified to positively influence activity, fitness, and health.

The information in Figure 3 has important implications for fitness and activity assessment (particularly with regard to youth programming in FITNESSGRAM®). First, it points out the importance of assessing both physical activity and physical fitness. This is philosophically why the FITNESSGRAM® program embraces and emphasizes both the collection of fitness data (FITNESSGRAM®) and activity data (ACTIVITYGRAM®). Second, it points out the need to emphasize the promotion of physical activity since it is essential for developing physical fitness. The FITNESSGRAM® program emphasizes physical activity as the modifiable variable in the FITNESSGRAM® reports to help encourage youth to be more active or to maintain their physical activity level (See Chapter 5). Finally, it points out the need to individualize fitness expectations for youth based on multiple factors such as heredity, other lifestyles, environment, and, other personal attributes. The philosophy of FITNESSGRAM®—"Health is Available to Everyone for a Lifetime, and it's Personal"—reflects the individual and personal nature of the programming. Morrow et al. (2013) have illustrated that if students are physically active sufficient to meet the 2008 Physical Activity Guidelines, they are more likely to achieve the FITNESSGRAM® HFZ, illustrating the relation between physical activity and physical fitness.

Figure 3. Conception of Health Related Fitness by Bouchard (1990).



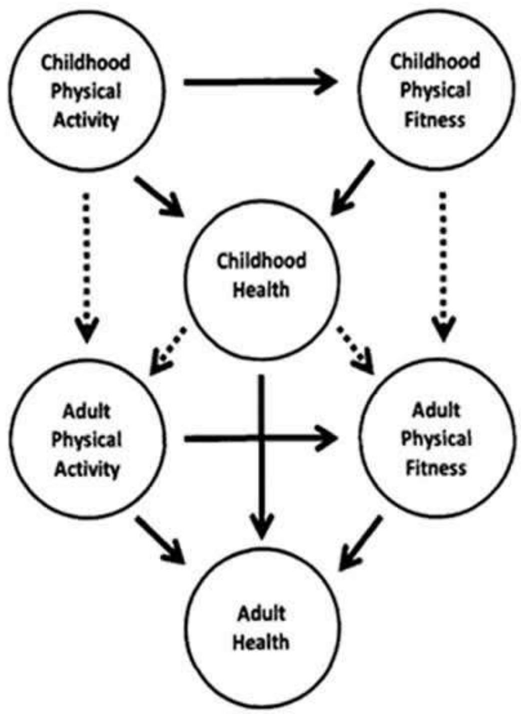
How Do Physical Activity and Physical Fitness Influence Health Across the Lifespan?

The generalized links between activity, fitness, and health have been described in a previous section of this chapter (see above). While many factors affect fitness and health, there is considerable evidence that physical activity contributes to both. No doubt, some people (young or old) respond differently to exercise training, but physical activity has benefits (beyond physical aspects—cognitive, emotional, social, or overall quality of life) for all persons. The Surgeon General's report on Physical Activity and Health released in 1996 provided strong documentation supporting this evidence (Physical Activity and Health Executive Summary). The release of official U.S. Physical Activity Guidelines in 2008 emphasized the important benefits of physical activity for all segments of the population (http://www.health.gov/paguidelines). Continued efforts by many public health organizations, foundations and agencies have sought to emphasize the importance of physical activity for optimal health and well-being (http://www.ncppa.org). Systematic efforts have been made to promote physical activity in different segments of society via the National Physical Activity Plan (http://www.physicalactivityplan.org), but one of the most significant advances has been the focused effort by the American College of Sports Medicine (ACSM) to legitimize and advance the promotion of physical activity by the medical community. The ACSM Exercise is Medicine campaign has launched efforts through a number of channels to advance the promotion of physical activity within the clinical setting and community (http://exerciseismedicine.org).

The FITNESSGRAM® program provides a key role in public health by promoting awareness and education about physical activity and physical fitness in youth. The programming and feedback is planned to promote health in children but also to increase the likelihood that children will grow up to be active and healthy adults. Considerable research has been done to understand how physical activity and physical fitness influence health across the lifespan. A Conceptual model showing the key associations is provided in Figure 4. This model developed by Morrow and Ede (2009) is based on a conceptual model proposed originally by Blair et al.

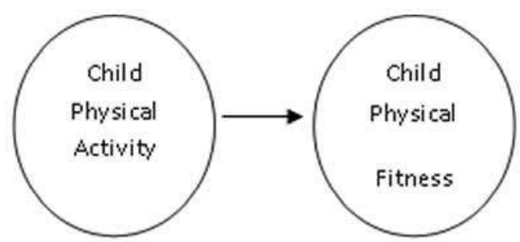
(1989). An advantage of the present model is that it depicts the strengths of the various associations based on the scientific literature. Solid lines depict associations that are more established while dotted lines depict associations that are less well supported. Part of the reason for weaker established relationships is that few studies have spanned childhood into adulthood as this requires tracking these individuals over long periods of time. Brief summaries of some aspects of the model are described below. Emphasis in the review is placed on how these findings impact programming and feedback provided through the FITNESSGRAM® program.

Figure 4. Relationship between Physical Activity, Physical Fitness, and Health Across the Lifespan.



Links Between Physical Activity and Physical Fitness

Based on extensive review of literature (NASPE, 1998, 2004; Strong et al., 2005), evidence-based data are strong for beneficial effects of physical activity on musculoskeletal health, several components of cardiovascular health, and adiposity in overweight youth. Evidence is adequate on the beneficial effects of physical activity on adiposity in normal weight children and adolescents.

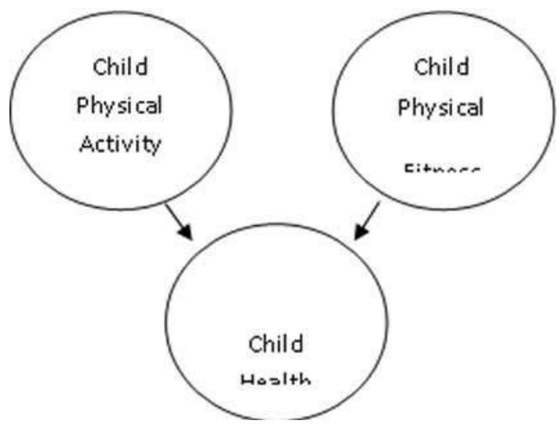


Much of the evidence used to draw these conclusions is based upon cross-sectional studies; however, exercise training studies in youth also were considered in the evaluation of evidence. Training studies are dependent on several factors, including the frequency, intensity, and duration of the training sessions and the training program. In addition, the baseline value is also important to consider. For example, one review of the literature for VO_2max indicates an average net increase of 8.6% and a range in the mean response from 1% to 29% (Pfeiffer et al., 2008). Thus, some youth respond to training whereas others do not respond. This may be due to several factors, including genetics, the baseline value, and the training program itself. Some youth will not respond to training because their baseline value is close to their “genetic ceiling” and they already participate in high levels of moderate-to-vigorous physical activity. Morrow et al. (2013) reported that middle-school students failing to meet the national physical activity guidelines were more likely to be in the FITNESSGRAM® Needs Improvement Zone.

The somewhat limited impact of physical activity on fitness indicates that these parameters, while related, should be viewed somewhat independently. This is one reason why physical activity is promoted as an independent construct within the FITNESSGRAM® program. The limited trainability of children also has implications for understanding and interpreting fitness data. It is possible for a child to be physically active but not physically fit. It is also possible for youth to be fit but not active. The feedback algorithms on the FITNESSGRAM® report incorporate information from the 3 self-report measures of typical physical activity in order to provide appropriate feedback to children. By acknowledging that a child is active (even if they are not fit) it helps to reward them for pursuing appropriate behaviors.

Links Between Child Physical Fitness and Child Health

The IOM report provides the most comprehensive review of health-related physical fitness in youth. The IOM report categorized the specific health outcomes into five different categories and findings are briefly summarized below. Emphasis in the report was focused on studies that showed links between fitness and health, but as shown in the model, it is well accepted that physical activity has independent influences on child health. It is difficult to separate out the independent paths from physical activity and physical fitness but both are clearly important. The IOM report focused on links between fitness and health partially because it is easier to measure fitness objectively. Brief summaries from the IOM report regarding the associations between dimensions of health-related fitness and child health are provided below. Readers interested in additional detail are referred to the complete report available at http://www.iom.edu/Activities/Nutrition/FitnessMeasuresYouth.aspx.



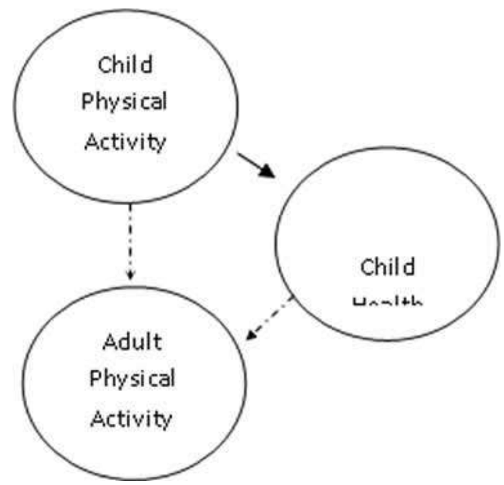
The IOM committee relied on a systematic review conducted by the Centers for Disease Control and Prevention as the primary evidence for the report. Separate analyses were conducted for body composition, cardiorespiratory endurance, musculoskeletal fitness, and flexibility. The committee concluded that there was a "substantial body of evidence" to support specific test items that are related to health for body composition and cardiorespiratory endurance. There was "adequate evidence" supporting hypothesized relationships between musculoskeletal fitness and health but less evidence linking specific musculoskeletal test items to health. The committee found little evidence linking flexibility and health in youth. Based on these conclusions, the committee recommends that national surveys of health-related fitness in youth include selected measures of body composition, cardiorespiratory endurance, and musculoskeletal fitness. These conclusions support the inclusion of these categories of assessment in the FITNESSGRAM® program.

Although flexibility was not recommended for fitness surveillance in the IOM report due to a lack of current evidence, it is the philosophy of FITNESSGRAM® that it is important that it be included as a component of health-related fitness in school physical education. Likewise, the IOM noted that "although the committee does not recommend a flexibility measure as a core component of a fitness test battery (for a national survey), administrators in schools and other educational settings may wish to include the sit-and-reach test or its alternatives (e.g., backsaver sit-and-reach) to measure flexibility" (IOM, 2012, page 9-12).

The documented associations between child fitness and child health provide a compelling case for the continued emphasis on youth fitness programming. However, as depicted in Figure 4, there are weak associations between childhood fitness and adult fitness as well as weak links between child health and adult fitness. This is because fitness cannot be maintained without regular involvement in physical activity. To ensure continued health benefits it is critical to maintain an active lifestyle over time. The potential for physical activity habits to track across the lifespan is discussed in the next section.

Links Between Childhood Activity and Adult Activity

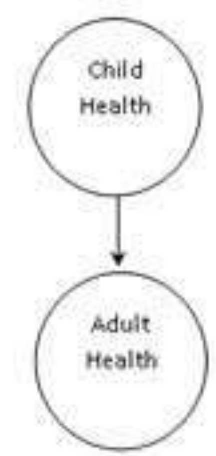
It is often assumed that a physically active child will become a physically active adult, thus influencing adult health outcomes. This assumption prompts the promotion of a physically active lifestyle during childhood. Indeed, a key goal of youth physical activity promotion programming is to increase the likelihood that youth will grow up to become active adults. One approach to examining this question is tracking. In general, the results of several studies indicate that physical activity tracks at low to moderate levels across all ages. This is largely due to the difficulty of measuring physical activity behavior over time but also due to other social and environmental factors that influence behavior.



As shown in the figure, there are clear links between physical activity and health (depicted with solid line) but for health to be maintained over the lifespan it is essential to emphasize the promotion of lifetime physical activity as stated throughout this chapter.

Links Between Childhood Health and Adult Health

The progressive nature of chronic conditions strongly suggests that the presence of risk factors (referred to as health markers in IOM report) during childhood increases the likelihood of health problems during adulthood. A number of studies indicate that risk factors/health parameters (cardiovascular risk factors, adiposity, and aerobic capacity) track fairly well throughout the lifespan.



Some studies have examined the association between body fatness and CVD risk factors during the growing years and their association with adult health. This literature is a bit limited due to the short length of follow-up only into the third decade of life in some studies which impacts the results given that the clinical manifestations of cardiovascular disease, type II diabetes, and the metabolic syndrome occur during mid-adulthood (i.e., 40-50 years). It is apparent that excess body fatness during childhood or

⇦TOC ⇦Chapter

adolescence can negatively impact CVD risk factors, type 2 diabetes, orthopedic complications, and all-cause and CVD mortality in adulthood (Maffeis and Tatò , 2001; Morrison et al., 2008).

Conclusions: The Health-Related Focus of FITNESSGRAM®

The focus of FITNESSGRAM® is on promoting a healthy lifestyle and overall health of youth that will carry forward throughout life. The first priority is given to physical fitness components for which a health association has been shown among youth. The second priority is given to physical fitness items for which a health association has been established later in life. Although not included in the discussion above, there is strong evidence linking sedentary living to health problems (Owen et al 2010). As noted in the NASPE physical activity guidelines for children (1998, 2004), extended periods of inactivity (two hours or more) are discouraged. So an additional priority is reducing inactivity among youth. Because the focus of FITNESSGRAM® is on “the promotion of lifelong physical fitness, physical activity, and other health-related behaviors,” and because FITNESSGRAM® and its partner programs are educationally based, learning about physical activity and fitness for application later in life is deemed important. Below is a list of key concepts for educators related to FITNESSGRAM®:

- Physical activity and physical fitness are independent, but related, and both are related to health.
- The amount of physical activity necessary to produce health-related fitness and health benefits varies with age and other factors.
- Fitness is generally defined as what your body can do. Body composition is an exception to the rule and is defined by the make-up (muscle, bone, fat, other tissue) of your body.
- Fitness is related to three key characteristics (health-fitness, functional fitness, skill-related or performance fitness).
- The focus of FITNESSGRAM®/ACTIVITYGRAM® is on health-related fitness with consideration of functional fitness.
- All parts of fitness, including health-related components, are related to performance.

Bibliography

Blair, S. N., Clark D. G., Cureton K. J., & Powell K. E. (1989). Exercise and fitness in childhood: implications for a lifetime of health. In C. V. Gisolfi & D. R. Lamb (Eds.), *Perspectives in exercise science and sports medicine* (pp. 401-431). Indianapolis, IN: Benchmark.

Blair, S. N., Cheng, Y., & Holder, J. S. (2001). Is physical activity or physical fitness more important in defining health benefits? *Medicine and Science in Sports and Exercise*, 33(6 Suppl), S379-99.

Bouchard, C. & Shephard, R. J. (1994). Physical activity, fitness, and health: The model and key concepts. In C. Bouchard, R.J. Shephard, & T. Stephens (Eds.) *Physical activity, fitness, and health: international proceedings and consensus statement*. Champaign IL: Human Kinetics.

Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public Health Report*, 100(2), 126-131.

Corbin, C. B. & Le Masurier, G. C. (in press). *Fitness for Life* (6th ed.). Champaign, IL: Human Kinetics.

Farpour-Lambert, N. J., & Blimkie. C. J. R. (2008). Muscle strength. In N. Armstrong N. & W. van Mechelen. *Paediatric Exercise Science and Medicine* (pp. 37-53).Oxford, UK: Oxford University Press.

Holt, J., L. E. Holt, & Pelham, T. W. (1996). Flexibility redefined. In T. Bauer (Ed.), *Biomechanics in Sports XIII* (pp. 170-174). Thunder Bay, Ontario: Lakehead University.

IOM (Institute of Medicine) (2012). *Fitness Measures and Health Outcomes in Youth*. Washington DC: The National Academies Press.

Kell, R. T., Bell, G. & Quinney, A. (2001). Musculoskeletal fitness, health outcomes and quality of life. *Sports Medicine*, 31, 863-873.

Knuttgen, H. G., & Kraemer, W. J. (1987). Terminology and measurement in exercise performance. *Journal of Strength and Conditioning Research*, 1:1-10.

Maffeis, C. & Tatò, L. (2001). Long-term effects of childhood obesity on morbidity and mortality. *Hormone Research*, 55 Suppl 1, 42-45.

Magnussen, C. G., Schmidt, M. D., Dwyer, T., &Venn A. (2012). Muscular fitness and clustered cardiovascular disease risk in Australian youth. *European Journal of Applied Physiology*, 112, 3167-3171.

Morrow, J. R. Jr., Tucker, J. S., Jackson, A. W., Martin, S. B., Greenleaf, C. A., & Petrie, T. A. (in press). Meeting physical activity guidelines and health-related fitness in youth. *American Journal of Preventive Medicine*.

Morrison, J. A., Friedman, L. A., Wang, P., & Glueck, C. J. (2008). Metabolic syndrome in childhood predicts adult metabolic syndrome and type 2 diabetes mellitus 25 to 30 years later. *Journal of Pediatrics*, 152, 201-206.

Morrow, J. R. Jr, & Ede, A. (2009). Statewide physical fitness testing: a big waist or a big waste? *Research Quarterly for Exercise and Sport*, 80, 696-701.

National Association for Sport and Physical Education. (2004). *Physical Activity Guidelines for Children: A Statement of Guidelines Ages 5-12* (2nd. Ed.). Reston, VA: National Association for Physical Education and Sports.

National Association for Sport and Physical Education. (1998). *Physical Activity Guidelines for Children: A Statement of Guidelines*. Reston, VA: National Association for Physical Education and Sports.

Ortega, F. B., Silventoinen, K., Tynelius, P., & Rasmussen, F. (2012) Muscular strength in male adolescents and premature death: cohort study of one million participants. *British Medical Journal*. 345, e7279.

Owen, N., Healy, G. N., Matthews, C. E., & Dunstan, D. W. (2010). Too much sitting: the population health science of sedentary behavior. *Exercise and Sports Science Reviews*, 38, 105-13.

Paffenbarger, R. S. & Lee, I-M. (1996) Physical activity and fitness for health and longevity. *Research Quarterly for Exercise and Sport*, 67 (Suppl. 3), 11-28.

Pfeiffer K. A, Lobelo, F, Ward, D., & Pate, R. R. (2008). Endurance trainability of children and youth. In H Hebestreit & O Bar-Or (Eds.) *The Young Athlete* (pp. 84-95) London: Blackwell Science.

Sale, D. G. & Norman, R. W. (1982). Testing strength and power. In J. D. Macdougall, H. A. Wenger & H. J. Green *Physiological Testing of the Elite Athlete* (pp. 7-37). Toronto: Mutual Press.

Saltin, B. (1973). Oxygen transport by the circulatory system during exercise in man. In J. Keul (Ed.) *Limiting Factors of Physical Performance* (pp. 235-252). Stuttgart, Germany: Thieme Publishers.

Strong, W. B., Malina, R. M., Blimkie, C. J., Daniels, S. R., Dishman, R. K., Gutin, B……. Trudeau, F. (2005). Evidence based physical activity for school-age youth. *Journal of Pediatrics*, 146, 732-737.

US Department of Health and Human Services. (1996). *Physical Activity and Health: A Report of the Surgeon General*. Atlanta, Georgia: US Department of Health and Human Services, Public Health Service, CDC, National Center for Chronic Disease Prevention and Health Promotion.

Web resources:

U.S. Surgeon General's report on Physical Activity and Health:
http://www.cdc.gov/nccdphp/sgr/index.htm
U.S. Physical Activity Guidelines: http://www.health.gov/paguidelines
National Coalition for Promoting Physical Activity: http://www.ncppa.org
National Physical Activity Plan: http://www.physicalactivityplan.org
American College of Sports Medicine Exercise is Medicine: http://exerciseismedicine.org
Institute of Medicine report on Fitness Measures and Health Outcomes in Youth:
http://www.iom.edu/Activities/Nutrition/FitnessMeasuresYouth.aspx

Chapter 4
Physical Fitness Standards for Children

James R. Morrow, Jr., Weimo Zhu, Matthew T. Mahar

This chapter, devoted to Physical Fitness Standards, describes the scientific rationale and procedures used when setting fitness standards (cut-offs) for the FITNESSGRAM assessments. The following questions are specifically addressed:

How Can Physical Fitness Scores Be Evaluated?

Scores on a physical fitness test can be meaningfully interpreted in several ways. Three ways of interest to people are norm-referenced evaluation, criterion-referenced evaluation, and improvement in performance. Each of these is presented below.

Norm-Referenced Evaluation

For many years, national fitness test data were used to develop percentile tables for boys and girls of all ages. A percentile represents the percentage of people who score at or below a performance or score value (e.g., a 10-year-old girl is at 90 percentile if she can finish the 1-mile run within 9 minutes and 9 seconds, according the percentile table generated from the National Children and Youth Fitness Study I [NCYFS, 1985]). The comparison is typically made to a specific, well-defined reference group (e.g., 10-year-old girls; 11-year-old boys; senior males over the age of 60; women between the ages of 40 and 49; etc.). Using these specific groups, test developers identified norms, that is, specific percentiles, as standards for students to achieve. The standard might be quite high (e.g., 85th percentile), achievable by only a small portion of the population of school-aged children. Or, the standard might represent the middle of the percentile table (i.e., 50th percentile, or an average performance). In the latter case, many more students could reach the standard.

There are advantages and disadvantages to norm-referenced (percentile) standards. The advantages are that students can learn how they compare with other children and youth in the well-defined group (e.g., their age, gender, school, etc.). Percentiles are also easy to interpret as they are used in most national standardized tests. Norm-referenced standards are relatively easy to develop as long as a representative sample is available. The primary disadvantage is that the standards are based on the children and adolescents' current levels of performance rather than the level they ought to achieve. Consider whether it is "good" for one to achieve "average" fatness if the average person has a level of fat that is unhealthy or puts the individual at risk. A shift in the performance distribution over time impacts interpretation. If the normative data change over time (either increase or decrease), yet an individual's performance does not change, this results in a relative change in the judgment/evaluation of the performance. Another disadvantage is that percentiles, particularly ones set at a high level, might discourage students whose fitness levels are moderate or low, as measured by the test, even though the fitness levels of those students may be adequate when viewed in another context such as health or some specific sports performance. Importantly, a disadvantage of the norm-referenced approach in evaluating health-related fitness is that the student's health status is not considered when interpreting the results.

Criterion-Referenced Evaluation

A solution to the disadvantages and problems of the norm-referenced evaluation is to use criterion-referenced evaluation where health status is used as the criterion. With criterion-referenced evaluations, a standard on a field test is determined which is related to a specific health outcome (i.e., the criterion) such as heart disease, body fatness, low back pain, etc.

With criterion-referenced evaluation, the most important interpretation of a fitness test score is the information about the student's health status. Use the 1-mile run test as an example. If an adolescent girl runs the 1-mile run test in 9 minutes, what does this mean in terms of her health status? The 1-mile run test is used to measure aerobic capacity. Does her performance put her at a low, medium, or higher level of risk for cardiac disease? While the precise answer to this question is unknown, there is evidence from adult populations substantiating that people with higher levels of aerobic capacity have a lower risk of cardiac disease. This evidence is well-documented in "Physical Activity and Health: A Report of the Surgeon General" (U.S. Department of Health and Human Services, 1996), the U.S. Department of Health and Human Service's Physical Activity Guidelines for Americans (Physical Activity Guidelines Advisory Committee, 2008; U.S. Department of Health and Human Services., 2008), and many research reports. We also know that even young children can show signs of cardiac disease (e.g., atherosclerotic changes) (Moller, Taubert, Allen, Clark, & Lauer, 1994). Numerous studies, conducted around the world illustrate the relations between physical activity behaviors and physical fitness status and health outcomes (Andersen et al., 2006; Andersen, Riddoch, Kriemler, & Hills, 2011; Chen & Wu, 2008; Eisenmann, 2004; Haas, Liepold, & Schwandt, 2011; Lobelo, Pate, Dowda, Liese, & Daniels, 2010; Tanha et al., 2011). Based on this evidence, the FITNESSGRAM® developers have concluded that criterion-referenced standards should be used when interpreting the FITNESSGRAM® scores. These standards have also been referred to as health-related criterion-referenced standards because of their link to the child's health status.

Note that criterion-referenced standards suggest that there is a minimum level of performance that must be achieved before a student is said to be fit or healthy or at reduced risk. The score representing the minimum level is called "cut-off" score. For example, for 10-year-old girls the cut-off score for the FITNESSGRAM® standards for the Healthy Fitness Zone is 40.2 $mL \cdot kg^{-1} \cdot min^{-1}$. Setting appropriate criterion-referenced standards requires evidence of reliability and validity. These issues are addressed in separate sections below.

In 2011, the American Journal of Preventive Medicine (2011) published an entire supplement providing the rationale and complete description of setting criterion-referenced standards for aerobic capacity and body composition for the FITNESSGRAM®.

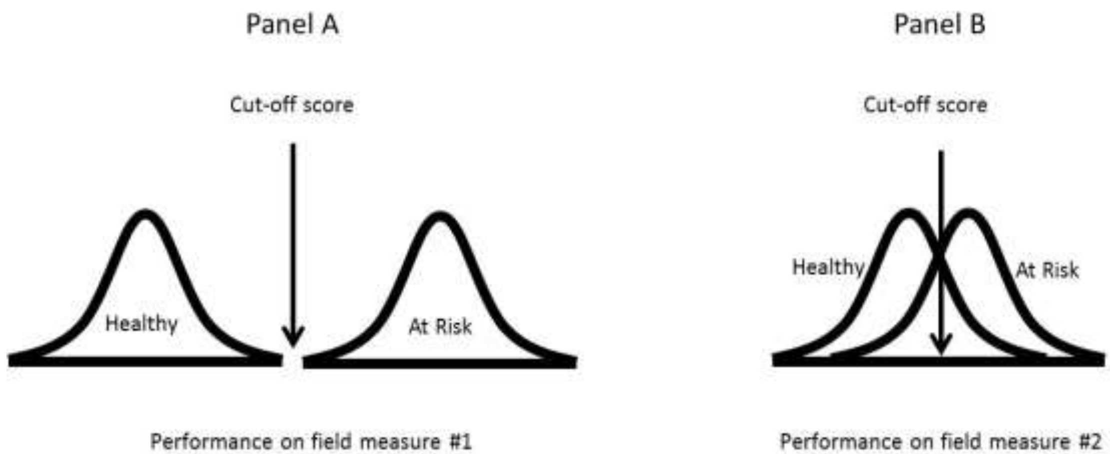
Improvement in Performance

A third way of interpreting scores is to look at the improvement in performance from one test administration to another. This is intuitively appealing but more difficult to conduct validly than it appears. If a student's score increases (or decreases) by a small amount, this change might be due to measurement error or, in case of improvement, might be due to practice or maturity. If the score increases more substantially, this increase should be interpreted in light of the initial score. If the initial score was low, a significant improvement is easier to attain than if the initial score was very good. If students are aware that the instructor looks for improvement, they might be tempted to perform poorly on the initial test so that their improvement looks much better at the second testing period. Equally important is the fact that improvement (change) scores tend to be very unreliable. Collectively, these issues make it difficult to validly assess students on improvement. Clearly, one wants to "improve" or change performance if it is unhealthy. However, it is quite difficult to accurately interpret a change score.

How Are Criterion-Referenced Standards (Cut-Off Scores) Set?

Setting criterion-referenced cut-off scores is a combination of art and science. Key to setting a cut-off score for a health-related fitness test is the identification of a single value that separates those at health risk from those who are at less risk (or higher risk). A criterion outcome must first be identified. Researchers then determine what field measure best "predicts" the risk category into which one would likely appear. Generically, one could be healthy or ill. This health status serves as the criterion outcome. Researchers then determine the cut-off score that most validly separates these groups on a field measure. For example, assume that the outcome measure is metabolic syndrome. Individuals are determined to either be positive (unhealthy) for metabolic syndrome or negative (healthy) for metabolic syndrome. A series of field tests are then identified that could be administered to children (or adults, of course) and values are determined to see which field test values best separate the unhealthy group from the healthy group. In practice, a series of cut-off score values are tested to identify the best cut-off score; that is, to identify the cut-score on the field test that best differentiates the healthy and at-risk groups. Figure 1 below illustrates two different field tests that are theoretically associated with metabolic syndrome. In panel A little overlap exists between the groups. Thus, the cut-off score might be easily identified. In panel B you can see a great deal of overlap on the field test performance for those with and without metabolic syndrome, so the cut-score would be more difficult to set and would ultimately result in a substantial number of misclassifications regarding risk for development of metabolic syndrome. A number of statistical procedures which are described later in this chapter can be used to identify the best cut-off score. Receiver Operating Characteristic (ROC) analysis is a procedure that has recently been applied to FITNESSGRAM® testing (American Journal of Preventive Medicine, 2011). Different values of cut-off scores are tested using criteria to determine which cut-off score is best. Important measures used to determine where to set the cut-off score are sensitivity and specificity. Sensitivity (or the true positive rate) is the ability of the measure to identify those who actually ARE positive on the criterion measure. Specificity (the true negative rate) is the ability of the measure to identify those who actually ARE negative on the criterion measure. The science is in using mathematics to adjust the cut-off score and see how sensitivity and specificity are changed. The art is in determining the final cut-off score realizing that there will not be perfect sensitivity or specificity. Scientists understand that there will always be true positives, false positives, true negatives, and false negatives. Setting the cut-off score impacts these values. The importance of each of these decisions helps one set the most optimal cut-off score.

Figure 1. Determining a Cut-Off Score.



Zhu et al. (2011) provide a review of approaches for developing criterion-referenced standards in health-related youth fitness tests. They suggest the following key steps in setting criterion-related fitness standards:

- Determine the components of health-related fitness (e.g., cardiorespiratory fitness, body composition, musculoskeletal health)

- Select a criterion measure and potential field tests
- Determine the relation between the criterion and the field tests
- Set the cut-off score
- Validate or cross-validate using additional measures and sample

How Is Reliability Determined for Criterion-Referenced Standards?

Classification Consistency

The concepts of reliability and validity are key to interpreting and trusting the results with criterion-referenced testing. With criterion-referenced measurement, reliability is often viewed as "classification consistency" because interest lies in the consistency with which individuals are classified into categories (e.g., Pass/Fail or Healthy/Unhealthy) on repeated administrations of a test. If a person is tested and then retested shortly thereafter on a test, he or she should be expected to be classified consistently across the administrations. Classification consistency is necessary but not sufficient to allow confidence in the criterion-referenced test results. Not only must the classification be consistent across test administrations, the classification must also truly represent the individual's level of achievement. That is, evidence of reliability and validity for criterion-referenced standards must exist for the test administrator, test taker, and important stakeholders to view the results with confidence.

The important comparison in criterion-referenced testing is whether or not the student has achieved the standard and not how well the student compares to one's peers. Obviously, training should not have occurred between the two testing occasions and nothing external should have occurred that would have changed the individual's true performance. Figure 2 illustrates the results you want to obtain when investigating the reliability of students tested on the same test on two occasions. It is important to realize that if test administration results in reliable positioning of test takers, those that fail to achieve the standard on the first administration would be expected to fail to achieve the standard on the second administration. Likewise, those that achieve the standard on the first administration would be expected to achieve the standard on the second administration. Consistency of measurement and consistency in decision-making criteria are keys to reliable testing. It is important to realize that no test is always reliable. That is, a test can result in reliable decisions for a given sample under given circumstances. Beets and Pitetti (2006) and Mahar et al. (1997) provide excellent examples of determining the criterion-referenced reliability of the FITNESSGRAM's PACER and 1-mile run items. Hartman and Looney (2003) provide similar procedures for the FITNESSGRAM's back-saver sit-and-reach test item. Saint-Romain and Mahar (2001) illustrate the criterion-referenced reliability of the push-up and modified pull-up. Ihmels et al. (2006) provide an illustration of reliability of tests with body composition measures from the FITNESSGRAM®.

Figure 2. Reliability for a Criterion-Referenced Test Administered on Two Days.

		Day 1	
		Pass	**Fail**
Day 2	**Pass**	*You want people to appear here on both days*	
	Fail		*You want people to appear here on both days*

Because an observer is used in scoring students' performance on several items of the FITNESSGRAM® (e.g., curl-up and push-up), attention is also needed in special cases of inter-observer and intra-observer reliability which are discussed below.

Field testing is often conducted with teachers or students as test administrators or raters (testers, observers, or scorers). Thus, agreement between raters should be considered. Two cases of rater agreement are important, inter-rater and intra-rater reliability.

Inter-Rater Reliability

Inter-rater reliability refers to the consistency (i.e., reliability) of two different testers administering the same test to the same students. Inter-rater reliability is also known as objectivity. You desire for students' abilities to achieve or not achieve the standards to be independent (i.e., unrelated) to who is administering the test. Note in Figure 3 that different raters are used to evaluate test results. This is key to inter-rater reliability. If raters use the same standards and observe the same behavior, it is expected that they will arrive at the same decision about whether the test taker has achieved the standard. Thus, good inter-rater reliability is illustrated when both raters agree on the interpretation of test results. Figure 3 below illustrates good inter-rater reliability.

Figure 3. Objectivity (Inter-rater Reliability).

		Rater 1	
		Pass	**Fail**
Rater 2	**Pass**	*You want people to appear here for both raters*	
	Fail		*You want people to appear here for both raters*

Intra-Rater Reliability

Intra-rater reliability refers to the ability for a single tester to observe the same performance by a student and place him or her in the same category each time. Figure 4 is similar to Figure 3 except in this case, it is the SAME rater who observes the SAME performance each time. That is, one is interested in the reliability (i.e., consistency) of decisions when there is a time interval between observing the same behavior. Figure 4 illustrates intra-rater reliability.

Figure 4. Reliability (Intra-rater Reliability).

		Rater 1–Occasion 1	
		Pass	**Fail**
Rater 1–Occasion 2	**Pass**	*You want people to appear here on both occasions*	
	Fail		*You want people to appear here on both occasions*

How Is the Validity of a Criterion-Referenced Standard Determined?

Relationship Between a Score and the Criterion

Any time you are discussing validity of a criterion-referenced test (i.e., the truthfulness of a score), you must have a criterion of some sort. With health standards, the criterion is typically the presence or absence of a disease, a disease risk factor, or some other health measure (even death can be an outcome measure in many epidemiologic studies). Setting standards that are criterion-referenced requires both scientific knowledge and measurement expertise. The standards are typically set through a combination of expert judgment, knowledge of the distribution of the field test, knowledge of the distribution of the criterion test, and the relationship between the field test and the criterion measure. The standard represents the level of risk for the aspect of health associated with each fitness component. The test score (or range of scores) associated with a defined level of risk is used as the criterion-referenced standard. In other words, the test score is referenced to the criterion. Examples can be found in the following published reports. Cureton and Warren (1990) provide an excellent example of setting valid standards in aerobic capacity while Going, Williams, and Lohman (1992) provide an excellent example of validation work in body fatness. Looney and Gilbert (2012) provide an example of setting standards for the sit-and-reach test.

Evaluating the Validity of a Criterion-Referenced Measure

Two specific requirements must be satisfied for a criterion-referenced measure to be valid. First, you want individuals who pass the criterion measure to successfully pass (i.e., meet or achieve) the criterion cut-score on the field test. Second, you want those who fail to meet the criterion score to also fail to achieve the criterion (i.e., cut-score) on the field test. If these two things occur, the test has resulted in your making a correct classification or decision.

Two possible errors (i.e., false positive or false negative) can result from these types of comparisons. In the context of health-related fitness testing, a false positive results when a participant fails to achieve the standard on the field test but can actually achieve the minimum level on a health-outcome criterion. A false negative occurs when the field test results indicate that everything is "OK" because the participant has achieved the minimum level needed on the field test, yet, the participant cannot actually achieve the minimum level necessary on the health-outcome criterion.

False positives and false negatives occur for a variety of reasons (e.g., unreliability of the test, participant motivation, recording errors, etc.). The foremost reason is that the field test is not perfectly valid. There will always be some measurement errors associated with testing.

Figure 5 illustrates the validity of a criterion-referenced measure. Note in Figure 5 that an individual who actually fails the criterion but DOES meet the standard on the field test is said to be a "false negative." One who fails to achieve the standard on the field test yet does achieve the standard on the criterion is said to be a "false positive."

Figure 5. Example of Types of Outcomes From Evaluations of Criterion-Referenced Standards in a Validity Study.

Field Test (e.g., skinfolds or BMI)		**Criterion (Health Outcome)**	
		Pass	**Fail**
	Pass	*Correct classification*	*False Negative*
	Fail	*False Positive*	*Correct classification*

⇦TOC ⇦Chapter

What Statistical Procedures Are Used to Estimate Reliability and Validity of Criterion-Referenced Measures?

The procedures often used to estimate criterion-referenced reliability and validity are proportion of agreement, kappa coefficient, the phi coefficient, and Chi Square.

Proportion of agreement (P) is simply the total of the correctly classified cells (depending on whether it is reliability or validity) divided by the total number of individuals tested. The kappa coefficient adjusts the proportion of agreement for agreements due to chance (Looney, 1989). While P is simple to interpret, its drawback is the effect chance can have on this statistic. Meaningful interpretable values of P range from .50 to 1.00 (a value of .50 could be obtained merely by chance). The kappa coefficient is interpreted as a correlation coefficient, except that negative values are considered un-interpretable. Thus meaningful interpretable values of kappa range from .00 to 1.00. Often both P and kappa are reported to give the user a more complete picture of the reliability or validity. While acceptable P and kappa values depend on the nature of the study, generally higher values are more acceptable. Moderate kappas are in the range of .41 to .60 and values above .61 are considered substantial (Viera & Garrett, 2005).

The phi coefficient is simply the Pearson product moment correlation coefficient between two variables that are scored dichotomously (i.e., 0 or 1). The Chi Square test of association is an inferential procedure used to determine if there is a non-chance relation between the two variables under investigation.

Each of these procedures can be used to estimate the reliability or validity of criterion-referenced measures. Whether it is reliability, objectivity, or validity that is being investigated depends on the variables that are used in the analysis. If the same two variables are related, it is some type of reliability. If a criterion measure is used, then validity is being investigated.

These analyses can be compared in assessing the reliability or validity of the standard. Setting the cut-score for the field test and the standard for the criterion is often a matter of adjusting each score until the ability to classify students is maximized. (However, it is most important that the criterion cut-score be truly related to the risk or health factor under investigation.) Then the scores are compared across analyses. If there is agreement on the most valid score, the evidence of reliability or validity is enhanced. The cut-score that was identified in this way, then, is used as the standard for that test. When there is no clear-cut agreement across the three methods, this suggests that the test or the criterion (or perhaps both) should be re-examined.

As mentioned earlier, Receiver Operating Characteristic (ROC) analysis is a procedure that has recently been applied to setting cut-scores in youth fitness testing. With ROC analyses, changes in the cut-off score are evaluated in terms of their sensitivity (the true positive rate) and specificity (the true negative rate). The cut-score is evaluated at several places to determine the desired sensitivity and specificity.

What Measures Must Have Their Reliability and Validity Estimated?

When a criterion-referenced field test is developed, the validity and reliability of the field test standard must be determined along with the reliability and validity of the criterion itself. The test developer must demonstrate that the field test is a valid measure of the attribute of interest. For example, the 1-mile run must be shown to be a valid estimate of aerobic capacity. In addition, the standards set for the 1-mile run test must also be shown to be reliable and valid. Few studies have been conducted in which standards have been adequately validated. This information is essential in the further development of health related physical fitness testing. See Mahar et al. (1997) for an example with aerobic fitness. Ihmels, Welk, McClain, and Schaben (2006) examined the reliability and validity of body composition measures from the FITNESSGRAM®. The aforementioned American Journal of Preventive Medicine supplement (2011) is an excellent resource on setting health-related criterion-referenced standards for aerobic capacity (cardiorespiratory fitness) and body composition in children and youth.

Are There Age, Gender, and Related Issues in Criterion-Referenced Evaluation?

Why Do Standards Differ for Different Ages?

Criterion-referenced standards may be different for individuals of different ages. For example, to achieve the aerobic capacity Healthy Fitness Zone a 10-year-old girl must have a VO_2max of 40.2 $mL \cdot kg^{-1} \cdot min^{-1}$. However, a 17-year-old girl must achieve only a VO_2max of 38.8·$mL \cdot kg^{-1} \cdot min^{-1}$. The values for boys of this age are 40.2 and 44.2, respectively. This is because evidence suggests that the best differentiation occurs between healthy and at-risk individuals at different VO_2max values for boys and girls and this is also a function of the individual's age. Thus, the cut-off score varies by both age and gender.

Why Do Standards Differ Among Different Tests of Physical Fitness?

Criterion-referenced standards may be different for the same tests in different test batteries. This will usually occur because the criteria used to set the standards differ. For example, let's assume that a test developer is setting standards for a 1-mile run test. Scores on the 1-mile run test will be compared with measured VO_2max to set the standards. One test developer might use a VO_2max of 32 $mL \cdot kg^{-1} \cdot min^{-1}$ to represent a minimally healthy person while another might use a VO_2max of 38 $mL \cdot kg^{-1} \cdot min^{-1}$ for the same age level and gender. To achieve the higher VO_2max a better run performance is necessary and, thus, a different health standard will result. This further illustrates the difficulty in setting criterion-referenced standards.

Why Do Some Standards for Boys and Girls Differ?

Two factors must be taken into account when determining criterion-referenced health standards: inherent physiologic differences between genders (performance) and differences in health risks between genders. Due to physiologic and anatomic differences between the genders, inherent performance differences may exist between boys and girls for a specific fitness component. For example, differences in cardiovascular function and body composition between adolescent boys and adolescent girls result in adolescent boys, as a general rule, having a higher aerobic capacity than adolescent girls. For example, if the minimum VO_2max for healthy girls is 36.0 $mL \cdot kg^{-1} \cdot min^{-1}$ and for healthy boys, 40.6 $mL \cdot kg^{-1} \cdot min^{-1}$, setting the same standard for both genders on the 1-mile run test would not be appropriate. In the case of aerobic capacity, the gender differences are taken into account, along with existing data on health risks in order to determine the standards. Likewise, should physiologic differences between genders occur, but existing data show health risks between genders occurring at the same absolute level, then the criterion standard should be the same for boys and girls, despite the performance differences. The key point is how differences in performance relate, in an absolute sense, to the criterion health standard. A difference may exist in the relation between the field test and the criterion for boys and girls. Thus, the standard for the boys and girls will differ because risk is elevated at different points. The accurate way to reflect this relation is to have different criterion-referenced standards for boys and girls.

Why Are Some Standards for Boys and Girls the Same?

In a few cases, the standards for boys and girls may not be different. When there is no valid reason for expecting a difference in the performance of boys and girls, the standards should be the same for both groups. For example, the trunk lift, a measure of trunk extension, has the same standard for boys and girls in the FITNESSGRAM® test. There are no known sex differences in trunk extension flexibility; thus, there is no valid rationale for different standards from a health-related perspective. Young children, particularly in grades 1-6, do not always possess the physical and physiological differences that appear as boys and girls approach puberty (Falls & Pate, 1993). When this is true, the same standards may be used for both groups. Some examples of this in the FITNESSGRAM® test are push-ups, curl-ups, modified pull-up, and flexed arm hang for some ages.

⇦TOC ⇦Chapter

Bibliography

American Journal of Preventive Medicine. (2011). *FITNESSGRAM* ®:Development of Criterion-Referenced Standards for Aerobic Capacity and Body Composition. *American Journal of Preventive Medicine 41*[4(Supplement)].

Andersen, L. B., Harro, M., Sardinha, L. B., Froberg, K., Ekelund, U., Brage, S. et al. (2006). Physical activity and clustered cardiovascular risk in children: a cross-sectional study (The European Youth Heart Study). *Lancet, 368*, 299-304.

Andersen, L. B., Riddoch, C., Kriemler, S., & Hills, A. P. (2011). Physical activity and cardiovascular risk factors in children. *British Journal of Sports Medicine, 45,* 871-876.

Beets, M. W., & Pitetti, K. H. (2006). Criterion-referenced reliability and equivalency between the PACER and 1-mile run/walk for high school students. *Journal of Physical Activity and Health, 3,* S21-S33.

Chen, J. L., & Wu, Y. (2008). Cardiovascular risk factors in Chinese American children: associations between overweight, acculturation, and physical activity. *Journal of Pediatric Health Care, 22,* 103-110.

Cureton, K. J., & Warren, G. L. (1990). Criterion-referenced standards for youth health-related fitness tests: a tutorial. *Research Quarterly for Exercise and Sport, 61,* 7-19.

Eisenmann, J. C. (2004). Physical activity and cardiovascular disease risk factors in children and adolescents: an overview. *Canadian Journal of Cardiiology, 20,* 295-301.

Falls, H. B., & Pate, R. R. (1993). Status of physical fitness in U.S. children. In M. J. Leppo & L. M. Summerfield (Eds.), *Healthy from the start: New perspectives on childhood fitness* (pp. 3-24). Washington, DC: Eric Clearinghouse on Teacher Education.

Going, S. B., Williams, D. P., & Lohman, T. G. (1992). Setting standards for health-related youth fitness tests: Determining critical body fat levels. *Journal of Physical Education and Recreation, 63(8),* 19-24.

Haas, G. M., Liepold, E., & Schwandt, P. (2011). Predicting cardiovascular risk factors by different body fat patterns in 3850 German children: The PEP Family Heart Study. *International Journal of Preventive Medicine, 2,* 15-19.

Hartman, J. G., & Looney, M. A. (2003). Norm-referenced and criterion-referenced reliability and validity of the back-saver sit-and-reach. *Measurement in Physical Education and Exercise Science, 7,* 71-87.

Ihmels, M., Welk, G. J., McClain, J. J., & Schaben, J. (2006). The reliability and convergent validity of field tests of body composition in young adolescents. *Journal of Physical Activity and Health, 3,* S67-S77.

Lobelo, F., Pate, R. R., Dowda, M., Liese, A. D., & Daniels, S. R. (2010). Cardiorespiratory fitness and clustered cardiovascular disease risk in U.S. adolescents. *Journal of Adolescent Health, 47,* 352-359.

Looney, M. A. (1989). Criterion-referenced measurement: Reliability. In M. J. Safrit & T. M. Wood (Eds.), *Meassurement concepts in physical education and exercise science* (pp. 137-152). Champaign, IL: Human Kinetics.

Looney, M. A., & Gilbert, J. (2012). Validity of alternative cut-off scores for the back-saver sit and reach test. *Measurement in Physical Education and Exercise Science, 16,* 268-283.

Mahar, M. T., Rowe, D. A., Parker, C. R., Mahar, F. J., Dawson, D. M., & Holt, J. E. (1997). Criterion-referenced and norm-referenced agreement between the mile run/walk and PACER. *Measurement in Physical Education and Exercise Science, 1,* 245-258.

⮌TOC ⮌Chapter

Moller, J. H., Taubert, K. A., Allen, H. D., Clark, E. B., & Lauer, R. M. (1994). Cardiovascular health and disease in children: current status. A Special Writing Group from the Task Force on Children and Youth, American Heart Association. *Circulation, 89*, 923-930.

NCYFS. (1985). National Children and Youth Fitness Study. *Journal of Physical Education, Recreation and Dance, 56(1)*, 43-90.

Physical Activity Guidelines Advisory Committee. (2008). *Physical activity guidelines advisory committee report, 2008*. Washington DC: U.S. Department of Health and Human Services.

Saint-Romain, B., & Mahar, M. T. (2001). Norm-referenced and criterion-referenced reliability of the push-up and modified pull-up. *Measurement in Physical Education and Exercise Science, 5*, 67-80.

Tanha, T., Wollmer, P., Thorsson, O., Karlsson, M. K., Linden, C., Andersen, L. B. et al. (2011). Lack of physical activity in young children is related to higher composite risk factor score for cardiovascular disease. *Acta Paediatrica*, *100*, 717-721.

U.S. Department of Health and Human Services. (1996). *Physical activity and health: A report of the Surgeon General.* Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion.

U.S. Department of Health and Human Services. (2008). *2008 Physical activity guidelines for Americans.* Washington DC: U.S. Department of Health and Human Services.

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine, 37,* 360-363.

Zhu, W., Mahar, M. T., Welk, G. J., Going, S. B., & Cureton, K. J. (2011). Approaches for development of criterion-referenced standards in health-related youth fitness tests. *American Jouranl of Preventive Medicine*, *41*, S68-S76.

Chapter 5
Physical Activity Assessment

Gregory J. Welk, Matthew T. Mahar , James R. Morrow, Jr.

The FITNESSGRAM® Reference Guide is intended to provide answers to some common questions associated with the use and interpretation of FITNESSGRAM® assessments. This chapter provides information about the importance of physical activity promotion in schools and how physical activity can be assessed in schools. Specific information is provided about the physical activity assessments that are available within the FITNESSGRAM® software.

Why Is Youth Physical Activity Behavior Important?

The promotion of physical activity (PA) in youth is an important public health priority. This is due in large part to concerns over the increasing prevalence of obesity but also to the growing consensus about the importance of physical activity for optimal health later in life. While children are the most active segment of the population, there are major concerns about the well documented declines in activity during adolescence (Caspersen, Pereira, & Curran, 2000; Jago, Anderson, Baranowski, & Watson, 2005; Wall et al., 2011) since this is a critical period for the development of lifestyle patterns. Studies have consistently shown that boys tend to be more active than girls at a given age; however, there is evidence to suggest that the gender differences in physical activity patterns may be due to differences in maturation rates since girls mature approximately two years earlier than boys. Gender differences in PA seem to be less evident when physical maturity is controlled (Cumming, 2008; Sherar, 2007; Thompson, 2003).

The decline in activity with age in both sexes is somewhat to be expected since this phenomena is evident in all species (e.g., frisky puppies also become less active as they become adult dogs). However, it is critical for youth to develop the behavioral and cognitive skills needed to establish healthy adult patterns of physical activity as they move from adolescence into adulthood. While studies are not conclusive, evidence suggests that physical activity patterns do track across the lifespan to at least a moderate degree (Malina, 1996). Details on the evidence linking youth physical activity to health are summarized in Chapter 3 (Health Benefits of Physical Activity and Fitness in Youth), and additional content is relevant in Chapter 6 (Aerobic Capacity Assessments), Chapter 7 (Body Composition Assessments), and Chapter 8 (Muscular Strength, Endurance, and Flexibility Assessments). Readers are also encouraged to consult other prominent public health documents (Strong et. al.). Some might assume that physical activity is only important for increasing physical fitness, but research suggests that physical activity provides health benefits that are independent of physical fitness (Blair, Cheng, & Holder, 2001). At this point, the evidence is sufficiently clear to warrant specific recommendations for the amount of physical activity needed for health. This section highlights youth physical activity guidelines and the unique roles that schools have for youth physical activity promotion strategies.

What Are the Guidelines for Youth Physical Activity?

The unique needs for children and adolescents warrant unique physical activity guidelines. The U.S. guidelines recommend that children and adolescents accumulate 60 minutes or more of physical activity daily (Physical Activity Guidelines Advisory Committee, 2008). The amount of physical activity recommended to youth is twice that of adults, not only because youth have more freedom and greater needs for physical activity, but also because forming a healthy lifestyle at an early age has an influence on lifestyle later on. Other countries passing similar guidelines include Australia, the United Kingdom, and Canada. Though there are minor discrepancies between them, all of the guidelines suggest that youth should engage in at least 60 minutes of moderate or vigorous intensity physical activity on a daily basis (Australia's physical activity recommendations for 5-12 year olds, 2004; Canadian Physical Activity Guidelines, 2011). The specific U.S. guidelines for youth physical activity are summarized below:

Key Components of the U.S. Youth Physical Activity Guidelines

- Children and adolescents should have 60 minutes (1 hour) or more of activity daily.

- ***Aerobic***: Most of the 60 or more minutes a day should be either moderate- or vigorous-intensity aerobic physical activity and should include vigorous-intensity physical activity at least 3 days a week.
- ***Muscle-strengthening***: As part of their 60 or more minutes of daily physical activity, children and adolescents should include muscle-strengthening physical activity on at least 3 days of the week.
- ***Bone-strengthening***: As part of their 60 or more minutes of daily physical activity, children and adolescents should include bone-strengthening physical activity on at least 3 days of the week.

- It is important to encourage young people to participate in physical activities that are appropriate for their age, that are enjoyable, and that offer variety.

Source: U.S. Department of Health and Human Services. Physical Activity Guidelines for Americans. Washington, DC: U.S. Department of Health and Human Services; 2008.

What Is the Role of Schools in Youth Activity Promotion?

Schools are not responsible for the declines in levels of physical activity in youth. However, they are clearly seen as part of the solution. Public health recommendations call for coordinated links between school, home, and community to promote physical activity in youth (Centers for Disease Control and Prevention, 2011), and specific efforts have been made to link physical education programming to school physical activity outcomes (Lee, Burgeson, Fulton, & Spain, 2007). The National Association for Sport and Physical Education (NASPE) also made formal recommendations for school physical activity to help schools commit to coordinated school activity promotion efforts. The NASPE guidelines state that “school-age children accumulate at least 60 minutes and up to several hours of physical activity per day while avoiding prolonged periods of inactivity” (See NASPE Guidelines, http://www.aahperd.org/naspe/standards/nationalGuidelines/PAguidelines.cfm). The NASPE also made formal recommendations for school physical activity. They recommend that “schools provide 150 minutes of instructional physical education for elementary school children, and 225 minutes for middle and high school students per week for the entire school year.” While not specifically indicated, the 150 minutes essentially represent half of the child’s recommended 300 minutes that would be captured over the five days at school (5 days × 60 minutes per day). Schools essentially need to accept responsibility for providing youth with opportunities to get at least 30 minutes of physical activity a day. Schools should also play a role in promoting physical activity at home by alerting parents that they are responsible for helping their children with the other half of their daily physical activity. The limits of class sessions and class time in physical education make it impossible for teachers to be personally responsible for the full activity guideline. However, physical education teachers can have a major impact by helping youth (and parents) become aware of how much activity is needed and how to obtain it. The subsequent sections highlight the importance of incorporating physical activity assessments into the school evaluation profile.

Why Should Physical Activity Be Assessed in Physical Education?

Promoting physical activity is a priority in physical education so it should also be a priority for program evaluation. The assessment of physical fitness has been a mainstay of most physical education programs (Morrow, 2005; Morrow & Ede, 2009); however, fitness achievement is influenced by a number of factors that are out of a child’s control (e.g., maturation, heredity, predisposition/trainability). Aerobic capacity estimates from aerobic fitness assessments are also directly related to body weight (i.e. body fat) and this may lead some youth to have a lower estimated aerobic capacity than would be expected. An advantage of incorporating physical activity assessments into a school evaluation is that it allows children to learn that they have control over their physical activity behavior and that it has independent effects on health (Welk, 2008). While it is possible to effectively use fitness testing to teach physical activity and fitness principles (Mahar & Rowe, 2008; Silverman, Keating, & Phillips, 2008; Wiersma & Sherman, 2008), a singular focus on physical fitness testing in physical education may lead to some unintended negative consequences on children's motivation for (and understanding of) physical activity and physical fitness. For example, some children may get discouraged in physical education if they score poorly on fitness tests despite being physically active. Alternately, children may incorrectly believe that they don't need to be physically active if their fitness levels are good. A child has more control over their physical activity behavior so feedback or goals based on this outcome may be more motivational.

The incorporation of physical activity assessments in physical education can provide a platform for reaching important educational goals. The inherent goal of physical education is to help children gain the skills (both physical and behavioral) needed to be active the rest of their lives. The NASPE standards for physical education describe the six characteristics of a "physically educated person" (NASPE, 2004); four of the six components specifically refer to physical activity. In addition to having good skills and reasonable levels of fitness, a "physically educated person" is someone who participates in regular activity, demonstrates understanding of principles related to performance of physical activities, knows the benefits of participation in physical activity, and values the contribution activity can make to a healthy lifestyle. While fitness testing provides considerable value in a well-planned physical education program, physical activity assessments can help address these other important curricular and educational goals. Because physical activity is a behavior, children need to specifically learn how much physical activity is needed for health as well as behavioral skills needed to plan and monitor their level of physical activity. These learning outcomes can be most effectively taught and evaluated using behaviorally-based physical activity assessments.

The FITNESSGRAM® Scientific Advisory Board believes that physical education programs should incorporate assessments of both physical fitness and physical activity to provide a more comprehensive and integrative view of physical development. Physical activity is a behavior and it is more amenable to change. Instruction based on physical activity provides a way to help children realize that they can take responsibility for their own health and well-being. Teachers can also more directly promote and influence physical activity behavior. Importantly, Morrow et al. (2013) report that adolescents who achieve the physical activity guidelines of 60 minutes of daily physical activity are more likely to achieve FITNESSGRAM® Healthy Fitness Zones. This validates the relation between physical activity (the behavior) and physical fitness (the health-related outcome).

What Techniques Are Available to Assess Physical Activity in Youth?

To advance understanding (and promotion) of physical activity behavior it is essential to have feasible, reliable and valid assessment techniques (Bauman, 2006). Considerable research has been done to improve the sophistication of current measurement methods and there is a large amount of literature on the utility of different methods. Several major research conferences have been held to help generate consensus and promote standardization in physical activity assessment

(Bowles, 2012; Freedson, Bowles, Troiano, & Haskell, 2012; Troiano, 2005) but it is still an imprecise science.

Assessing physical activity is challenging in all segments of the population, but it is more challenging in youth than adults because of inherent differences due to cognitions, growth/maturation, and physical activity patterns. A number of manuscripts have sought to summarize the key issues and challenges associated with assessing activity in this age group (Corder, Ekelund, Steel, Wareham, & Brage, 2008; Sirard & Pate, 2001; Welk, Corbin, & Dale 2000). Readers interested in understanding the detailed progression of work in youth physical activity are encouraged to consult these studies. The content here will describe some of the more practical methods that can be used in schools (Welk & Wood, 2000). Specific detail will also be provided about the physical activity assessments available within the FITNESSGRAM® program.

The key decision in selecting a physical activity assessment tool is the relative importance of feasibility and validity. In general, feasibility is inversely related to validity (i.e., more feasible instruments tend to be less valid and vice versa). However, other factors must also be considered including the ease of use, the goal for the assessment, the type of output measure, the burden on participant, and the cost. If the assessments are to be conducted primarily for educational and instructional purposes then the cost, ease of use, and utility should be emphasized. However, if assessments are needed for research or surveillance/evaluation purposes, the reliability and validity of the assessments may be more important factors (Welk, Corbin, & Dale, 2000).

In a review of methods, Sirard and Pate (2001) classified the various physical activity measurements into three categories: primary measures (e.g., direct observation, doubly labeled water, and indirect calorimetry), secondary measures (e.g., heart rate, pedometers, and accelerometers), and subjective measures (e.g., self-report, interviews, proxy-reports, and diaries). This categorization is consistent with the image presented above in that the primary measures are generally viewed as the most accurate and the secondary measures are considered somewhat less accurate. The categorization also highlights a key limitation of self-report measures, which is their subjectivity. The subjectivity of a self-report is viewed as a limitation since a person's perception or recollection of the information may contribute bias or error. However, subjectivity is also an advantage if a goal is to understand individual reactions to physical activity.

Welk and Wood (2000) conducted a review of tools that could be effectively used in school-based settings to evaluate activity in youth. The review emphasized that for use in physical education the most practical tools are heart rate monitors, pedometers, and self-report instruments. Accelerometers were not included in this list because they were expensive and primarily used for research. However, there has been a flurry of new developments with accelerometry-based devices in recent years and the costs have come down dramatically. Direct observation techniques were also not included as viable options in the original list, but newer techniques have been developed to facilitate use in school-based settings (e.g., SOPLAY, System for Observing Play and Leisure Activity in Youth). The basic advantages and disadvantages associated with each of the five primary techniques (direct observation, heart rate, accelerometer, pedometer, and self-report) are summarized below (see Table 1), followed by detailed reviews of each method. The devices are ordered from least practical to most practical for use within school physical education.

Table 1. Comparison of Different Types of Physical Activity Assessments

Type of Activity Measure	Advantage	Disadvantage
Direct observation	• Provides quantitative and qualitative information about physical activity	• Requires trained observers • Can only track several students at a time • Time consuming to collect and interpret
Heart rate monitor	• Accurate indicator of physical activity • Good educational potential to teach about the cardiovascular system	• High cost • Time-intensive to download • Difficult to assess large numbers of children • Relevant only to aerobic activity • Other factors affect heart rate (e.g., illness, anxiety, possible interference resulting in artifacts recorded, etc.)
Accelerometer (activity monitor)	• Accurate indicator of physical activity • Good educational potential to teach about "accumulating" activity over the whole day	• High cost • Time-intensive to download • Difficult to assess large numbers of children
Pedometer	• Inexpensive • Easy to use • Records distance	• Records "quantity" of movement but not "quality" (e.g., intensity) of movement
Self-report	• Low-cost • Easy to administer to large groups • Good educational potential for use in curriculum	• Potential problems with validity and reliability • The respondent must have the cognitive ability to self-report

Adapted from Welk and Wood (2000).

What Are Pros and Cons of Direct Observation Measures?

Direct observation techniques have been commonly used in physical education settings to assess activity behavior in children. In most systems, an observer codes the type and intensity of activity that is performed during a short periodic interval along with other details about the behavior or setting (McKenzie, 2002). The type of detail available through direct observation techniques offers some significant advantages for understanding youth activity behavior. Unfortunately, the time and cost of such assessments generally make this type of assessment only practical for research or instructional applications. A commonly used direct observation instrument called SOFIT (System for Observing Fitness Instruction Time) has been widely used in research to understand pedagogical and curricular strategies; however, it relies on individual observation and would have limited utility for use by teachers or school personnel. A more practical method for school-based evaluation is called SOPLAY (System for Observing Play and

⇦TOC ⇦Chapter

Leisure Activity in Youth, http://activelivingresearch.org/node/10642). Rather than monitoring an individual child, this system uses a scanning approach to capture the overall pattern of physical activity in a group of individuals (McKenzie, 2006; McKenzie, Marshall, & Sallis, 2000). An observer scans from left to right (once a minute) and records the number of youth that are currently sedentary, walking, or very active. This tool provides considerable value for evaluating youth activity behaviors (McKenzie, Crespo, & Baquero, 2010; Saint-Maurice, Welk, Silva, Siahpush, & Huberty, 2011). A recent calibration study of the SOPLAY (Saint-Maurice, Welk, Ihmels, & Krapfl, 2011) suggested that estimates of MVPA from SOPLAY were significantly higher than accelerometry-based PA estimates when codes of walking and very active were used (in combination) to reflect participation in moderate to vigorous PA. However, estimates were similar when only the SOPLAY code of very active was used to define MVPA. This alternative scoring method provides an empirically sound way to estimate participation in MVPA in school settings.

Can Heart Rate Monitors Be Used to Assess Physical Activity?

Heart rate monitors provide an accurate determination of exercise intensity and can record data over extended periods. They have been commonly used by endurance athletes to help monitor the intensity of their training, but they are also increasingly popular in many physical education programs to teach children about the cardiovascular system and to track activity within the class. If heart rate monitors are used in physical education, emphasis should be placed on the educational value rather than for evaluating children's performance or effort in physical education. Many teachers concerned about keeping students active have used heart monitors to ensure that the students are in the appropriate heart rate zone during their entire lesson. These efforts may be well-intentioned but they may impose a structure that makes exercise become more work than play. Children typically prefer intermittent activity and need opportunities for rest. Being forced to keep their heart rate elevated may make activity less enjoyable. Individual variability in heart rates may also make the use of specific target zones inappropriate for some children. If heart rate monitors are used in physical education, a low threshold should be used to define bouts of activity. The goal should also be to accumulate a certain number of minutes in the target zone rather than emphasizing continuous activity with elevated heart rates.

While heart rate monitors can provide a useful indicator during specific bouts of exercise (e.g., physical education), they are not particularly useful for tracking activity patterns under normal activities of daily living (Welk, Corbin, & Dale, 2000). For example, heart rate can be influenced by nervousness, dehydration, illness, or stress. There are also some transmission problems with the signal when heart rate monitors are worn over extended periods. Many children also find the transmission strap to be uncomfortable when worn over long periods. Therefore, heart rate monitors should be used primarily for educational purposes in school physical education and not for formalized individual or group assessments.

Are Accelerometers (Activity Monitors) Practical for School Assessments?

A variety of commercially available instruments can now be used to measure physical activity patterns under free-living conditions (Bassett, Rowlands, & Trost, 2012). The devices are typically about the size of a pager and clip to a belt or waistband. Most devices record body acceleration and store the raw movement counts collected in specific increments of time. These features allow them to assess the frequency, intensity, and duration of activity. They are widely used and accepted in research application and considerable work has been done to refine the validity and to improve the utility of these devices. Many review studies have summarized the key issues for using accelerometers to assess youth physical activity behavior (Butte, Ekelund, & Westerterp, 2012; De Vries, Van Hirtum, Bakker, Hopman-Rock et al., 2009; Freedson, Pober, & Janz, 2005; Rowlands, 2007; Trost, 2001).

The monitors are small, easy to use, and well suited to assessing physical activity in children; however, their cost and data management requirements make them impractical for use within the physical education curriculum. Newer lines of consumer based monitors have recently been released into the market to capitalize on the availability of low cost accelerometer technology, blue tooth data transfer capabilities, and social media communication channels. These devices are targeted primarily at adults for personalized weight loss and exercise training applications, but there are examples of technologies that have been developed. Over time, it is likely that the technology will enable more effective and cost effective activity monitoring for school-based activity assessments.

How Can Pedometers Be Used to Assess Physical Activity Behavior?

Pedometers are small, inexpensive, and easy to use devices that track the number of steps a person takes. They have become widely used by consumers and also for research applications. A key advantage of pedometers is that they provide immediate feedback using highly interpretable outcome measures (steps and/or distance).

The popularity of pedometers has led to an explosion of different devices and studies have confirmed that there is considerable variability in the quality of pedometers (Bassett & Crouter, 2003). In general, quality electronic pedometers have been shown to provide good indicators of daily steps. However, a limitation of pedometers is that they cannot measure non-locomotor activities (Welk et al., 2000). Most units also do not directly estimate minutes of physical activity or enable data to be stored internally for tracking and download. These characteristics (and the costs) limit the utility of pedometers for school evaluation. However, newer monitors now store data and many also now track and report step rate. This allows cadence (steps per minute) to be determined. Research has determined that step rates of 80-100 steps per minute can be used to reflect activity that is at least of moderate intensity. This makes it possible for pedometers to be used to estimate minutes of physical activity performed—a more useful outcome measure than steps for evaluation purposes.

As with activity monitors, there is considerable potential for pedometers to be used for large scale monitoring or surveillance, but cost is still a significant barrier. Despite this limitation, pedometers have considerable utility for education purposes and for promoting awareness about physical activity behavior. Children can clip them onto their belts or waistbands and record the number of steps taken during class. This provides a way to quantify activity levels during physical education (PE) class (Scruggs, Beveridge, Eisenmann, Watson, Schultz, & Ransdell, 2003). Pedometers also offer considerable promise for assessing physical activity outside of class if there are sufficient devices available for them to use at home. A final application is for school activity promotion efforts. Pedometers are widely used for activity challenges in worksites and they would have the same utility for use in schools. Readers interested in pedometers are referred to several articles in the literature (Bassett et al., 1996; Schneider, Crouter, & Bassett, 2004; Tudor-Locke, 2004; Tudor-Locke et al., 2011).

Can Self-Report Instruments Provide Useful Information?

Self-report instruments are the most commonly used format to collect information about physical activity. Depending on their scope, they can provide very detailed or very general information about physical activity. Advantages of self-reports are that they are inexpensive, easy to use, and can be administered to large groups in a cost-effective manner. Limitation of self-reports are that they usually require some form of recall and can be quite subjective (Matthews, 2002). The tendency for people to report socially desirable responses can be problematic, but this may be less of an issue with children. Despite these limitations, the low cost, ease of use, and education potential of self-report instruments make them well suited for use within the physical education curriculum, assuming students have the cognitive ability to complete the task in a valid manner.

Self-report measures vary considerably in the time frame and format used for the assessment. Some measures are designed to provide a general assessment of a child's normal level of physical activity. They often rely on a recall of activity completed over a representative period, such as one week. A limitation of this format is that it assumes that the recent week is representative of the child's activity in other weeks. Other instruments avoid this problem by using more general questions about "typical exercise behavior." These instruments, however, cannot provide the same detail as recall based measures. Another class of self-report measures utilizes detailed logs or activity records collected or recalled over several days. An advantage of this approach is that children have an easier time recalling specific activities from a previous day than generalizing over a longer period of time. Another advantage is that these instruments can provide considerable details regarding the type, intensity, and duration of activity. A limitation of these instruments is that the results may not generalize to a child's typical activity level. Readers interested in more specific information about the validity and reliability of various self-reports in children are referred to an excellent review (Chinapaw, Mokkink, van Poppel, van Mechelen, & Terwee, 2011). General information about self-report measures can be found in the following reviews (Ainsworth, Caspersen, Matthews, Masse, Baranowski, & Zhu, 2012; Troiano, Pettie-Gabriel, Welk, Owen, & Sternfeld, 2012). Importantly, self-report measures are widely used with adults in the CDC's Behavioral Risk Factor Surveillance System (BRFSS) and with children and youth in the Youth Risk Behavior Surveillance System (YRBSS).

Despite significant limitations, self-report tools still offer considerable potential for school applications (Welk, 2008; Welk & Wood, 2000). Self-report instruments provide a way to teach important principles about physical activity and help youth learn about the recommended types and amounts of physical activity. They also provide a way to evaluate group changes over time or to compare different schools to examine the relative effectiveness of different programs or environments. They provide the most effective way to evaluate school level activity promotion strategies so new methods are needed to overcome limitations highlighted in previous research. A detailed review of different self-report tools is provided in the next section.

What Are Some Practical Self-Report Instruments for Youth?

Self-report instruments provide the most practical and easy to use tool since they are time and cost-effective and easy to administer to large groups. Concerns about the reliability and validity of self-report measures in youth have contributed to the overall movement to objective monitoring methods in research applications. However, from an educational perspective, self-report measures provide a number of significant advantages (Welk & Woods, 2000).

A number of different self-report approaches are available for youth, but it is important to consider the relative advantages and limitations. Some instruments are based on recalling details of a previous day or series of days, while other instruments are focused on assessing typical or "general" activity profiles. Instruments also vary in how they collect data on activity. Some instruments are based on detailed lists of activities in which children are asked to indicate if they participate in a certain activity and how often. Other tools use time prompts that have children estimate activity levels during different times of the day. It is not possible to summarize all of the available instruments here, but detailed reviews are provided for two of the most commonly used instruments (Physical Activity Questionnaire for Children and Adolescents and the Previous Day Physical Activity Recall). The sections below summarize the psychometric properties of these two tools that have documented utility for school-based activity assessment.

Physical Activity Questionnaire (PAQ)

The PAQ is a simple self-report tool designed to assess activity over the past week. There are two versions of the Physical Activity Questionnaire: Physical Activity Questionnaire for Older Children (PAQ-C) and Physical Activity Questionnaire for Adolescents (PAQ-A). PAQ-C is designed for elementary school children ages 8 to 14 approximately (grades 4-8); PAQ-A is designed for high school students ages 14-20 approximately (grades 9-12). Both of the questionnaires are designed to measure general moderate to vigorous physical activity levels during a typical week in the school year (Crocker, 1997). The PAQ-C includes nine items (eight items for PAQ-A), each scored on a 5-point scale. The values are averaged to create a composite score with a higher value indicative of a higher activity level. The first question provides a physical activity checklist including over twenty kinds of sport and exercise activities asking the students how many times they did each in the past seven days. The next six questions examine their activity level in different school settings at certain periods in the last seven days (PE, recess, immediately after school, evening, weekends). The eighth question requires the students to summarize their general activity levels from among five different statements. The last question asks students to report their frequency in physical activities for each day of the previous week.

The PAQ-C has both limitations and strengths. A key limitation is that it does not provide a useful outcome measure such as energy expenditure or total minutes of physical activity. Additionally, the PAQ-C focuses on activity at school and is not appropriate for assessing physical activity during winter and summer breaks. Despite these limitations, the PAQ-C also has some advantages compared with other self-report instruments, including low cost, time efficient, large-scale usage, use of lunch and evening time periods to enhance recall ability, and short administration time to obtain a past week physical activity pattern.

The original validation studies (Crocker, Bailey, Faulkner, Kowalski, & McGrath, 1997; Kowalski, Crocker, & Faulkner, 1997; Kowalski, Crocker, & Kowalski, 1997) demonstrated that the PAQ-C has acceptable item-scale properties, reliability, internal consistency, and is sensitive to gender and seasonal differences. A more recent validation study (Janz, Lutuchy, Wenche, & Levy, 2008) demonstrated good concurrent validity when compared with an activity monitor (correlations ranging from $r = 0.56$ to $r = 0.63$). It has also been shown to have utility among different races (More, Hanes, Barbeau, Gutin, Trevino, & Yin, 2007).

The present review demonstrated that the PAQ meets established psychometric characteristics needed to provide validity evidence. It has been widely used in school-based research and provides an effective PA screening tool for school-based applications. The PAQ can be administered in short amounts of time (~5 minutes) and it provides useful insights into levels of activity at different times or in different settings. These attributes make it well suited for use in schools where education is the key goal. A major limitation of the PAQ is that it provides an outcome measure that is difficult to interpret. Research using innovative calibration methods demonstrated that equations can be used to adjust for measurement error and improve the utility of self-report measures (Saint-Maurice, Welk & Heelan, 2013). The calibration will make it possible to estimate minutes of time spent in physical activity from the self-report items, but additional research is needed to test the overall utility of the PAQ.

Previous Day Physical Activity Recall

The Previous Day Physical Activity Recall (PDPAR) is a self-report instrument intended to capture the previous day's physical activity patterns of children after school hours from 3:00 pm to 11:30 pm (Weston et al., 1997). It is a time-based recall approach and the time period is divided into 17 blocks, 30 minutes each. Children are asked to recall their specific activity from an activity checklist of 35 common activities which are grouped into the following categories: eating, sleep/bathing, transportation, work/school, spare timework, and physical activity. The children are also required to note the intensity of the activity by four levels (very light, light, moderate, or vigorous) per block of time. The PDPAR also provides some illustrations describing the characteristic of each intensity level to help children to rate their physical activity intensity. Each activity has its own corresponding MET values for all four intensity levels to facilitate the energy expenditure calculation. The PDPAR requires one day recall and uses a segmented day format to facilitate recall. Weston et al. (1997) tested the validity of PDPAR in youth with pedometers, Caltrac activity counts, and heart rate monitors as criteria. The correlation between the PDPAR and pedometer counts was high (r=0.88), as was the correlation between the PDPAR and Caltrac accelerometer (r=0.77), indicating good concurrent validity. Correlations between the PDPAR and heart rate were slightly lower, but still significant. In addition, results showed high test-retest reliability (R=0.98), and high interrater reliability (R=0.99) for PDPAR scoring. A general limitation of the PDPAR is that it only records physical activity pattern for one day, which is not long enough to capture the general habitual activity style. The authors recommended collecting data over several days (Weston et al., 1997), and this has become standard practice when the PDPAR has been used.

Trost et al. (1999) conducted a more comprehensive validation study of the PDPAR. The CSA 7164 accelerometer was used to evaluate the validity of PDPAR in 5th grade students. They found that the correlation between mean MET from PDPAR and CSA counts for each time block was 0.57, which is lower than the correlation in the Weston et al. (1997) study. Self-reported participation in vigorous activity (METS ≥ 6) had a higher correlation with the CSA MVPA (r=0.38) than corresponding correlations (r=0.19) for moderate activity (METS 3-6). This result indicated less favorable evidence to support the validity of PDPAR in young children, especially for estimating moderate physical activity. This study reached the same conclusion as other studies that the PDPAR is more valid among higher grade students than lower grade students. Despite these limitations, the PDPAR has been shown to provide good utility for school-based assessments.

One innovative approach is to develop calibration equations that can equate or link self-report data to objective estimates of physical activity. Tucker et al. (2011) developed and validated a prediction equation for the PDPAR that enabled the PDPAR blocks to be converted into estimates of time spent in physical activity. This study demonstrated the potential for developing and applying a calibration approach to improve the utility of self-report measures, but additional work is needed to enable use in practice.

What Physical Activity Assessments Are Available in FITNESSGRAM®

The FITNESSGRAM® program has been working to provide teachers with viable options to incorporate physical activity assessments into the curriculum. Several options are available for assessing physical activity through the FITNESSGRAM® program and teachers could also consider incorporating other options, such as the use of pedometers, activity monitors, or direct observation measures as described above. Emphasis in this section is on methods that are now available within the FITNESSGRAM® program (Meredith & Welk, 2010).

What Are the Physical Activity Questions in FITNESSGRAM?

A unique feature of the FITNESSGRAM® software is the inclusion of a simple set of physical activity questions WITHIN the FITNESSGRAM® software. This set of three individual items is selected from a tab within the student application of the FITNESSGRAM® software. The student clicks on separate items to provide information about his or her participation in aerobic, strength, and flexibility activity over the last seven days. Collectively, the items provide a general indicator of a child's activity patterns and are used within the software to improve the quality of the prescriptive feedback provided to the child.

Physical Activity Questions in the FITNESSGRAM® Fitness Battery

Aerobic Activity Question:

"On how many of the past seven days did you participate in physical activity for a total of 30-60 minutes, or more, over the course of a day? This includes moderate activities (walking, slow bicycling, or outdoor play) as well as vigorous activities (jogging, active games or active sports such as basketball, tennis, or soccer)." (0,1,2,3,4,5,6,7 days)

Strength Activity Question

"On how many of the past seven days did you do exercises to strengthen or tone your muscles? This includes exercises such as push-ups, sit-ups, or weight lifting." (0,1,2,3,4,5,6,7 days)

Flexibility Activity Question:

"On how many of the past seven days did you do stretching exercises to loosen up or relax your muscles? This includes exercises such as toe touches, knee bending, or leg stretching." (0,1,2,3,4,5,6,7 days)

How Can the Activity Items Be Used in Physical Education?

The physical activity items can help supplement and enhance the effectiveness of school fitness evaluations. If the child completes the additional items on physical activity, the software incorporates the responses to the physical activity items in the evaluative feedback that is provided on the FITNESSGRAM® report. For example, if a child has high fitness scores but low ratings on the physical activity items, the report would congratulate them about the fitness achievement but prompt them to be more active. In contrast, if a child scored poorly on the physical fitness items but reported being physically active, the messages would comment about the low fitness but congratulate the child for his or her healthy levels of physical activity. These messages provide an effective way to teach youth that it is important to be both active and fit.

⇦TOC ⇦Chapter

Teachers are strongly encouraged to have children complete these questions in order to activate the more integrated feedback messages.

What Is the Scientific Basis for the FITNESSGRAM® Physical Activity Items?

The items used in the FITNESSGRAM® module to assess physical activity are based on items from the Youth Risk Behavior Survey (YRBS) from the U.S. Centers for Disease Control and Prevention (Centers for Disease Control and Prevention, 2013). This survey is a comprehensive survey designed to collect information from adolescents about a variety of lifestyle behaviors. The inclusion of physical activity items to this battery provides researchers and professionals with some descriptive information about the normal activity patterns of children in the United States. Because of slight differences in wording and the use of self-report data, direct comparisons should be made with caution between FITNESSGRAM® and YRBS results.

The results of the Youth Risk Behavior Survey provide some comparison data on levels of activity among adolescents in the United States (grades 9-12). According to the most recent results from the YRBS (Centers for Disease Control and Prevention, 2012), approximately 50% of students reported getting at least 60 minutes of physical activity per day on at least 5 of the last 7 days. The prevalence rates for achieving 60 minutes of physical activity per day were higher among males (60%) than females (40%) and this pattern was consistent for whites (male: 62%, female: 43%), blacks (male: 57%, female: 32%), and Hispanics (male: 57%, female: 33%). The prevalence rates were higher among 9th-graders (53%) and 10th-graders (52%) compared with 11th -graders (47%) and 12th-graders (45%). The prevalence rates ranged from 38% to 55% across state surveys (median: 47%). Readers are encouraged to visit the YRBS website (http://www.cdc.gov/healthyyouth/data) to produce customized, state-specific queries and to examine trends over time.

How Does the ACTIVITYGRAM® Physical Activity Assessment Work?

ACTIVITYGRAM® is a separate module within the FITNESSGRAM® software that provides a detailed assessment of a child's level of physical activity. Children complete a recall of their previous days' activities and can print out a personalized report. Because the assessment requires detailed information from each child, this assessment is only available within the student application of the FITNESSGRAM® software. The assessment includes a time grid that allows a child to code the predominant activity he or she performed in each 30 minute block of the day. The activities are selected from the Activity Pyramid image above the grid. The child first selects the category (e.g., Rest, Aerobic Sports, Aerobic Activity, Muscular Activity, or Flexibility). Then the child is provided with five different options within each category. Once the child selects an activity, he or she is provided with an option to indicate the intensity (Rest, Light, Medium, or Hard) and then to indicate whether it was done "All of the Time," "Most of the Time," or "Some of the Time."

How Can ACTIVITYGRAM® Be Used in Physical Education?

The ACTIVITYGRAM® physical activity assessment module provides a powerful tool to help children learn about their physical activity patterns. The ACTIVITYGRAM® was designed to be appropriate for upper elementary, middle, and secondary students, however, the accuracy of the reports are likely to be better for middle school and high school youth. Younger students can still benefit from the experience of tracking their activity patterns, but emphasis should be placed on the educational value and not on the absolute data reported.

The ACTIVITYGRAM® assessment is accessed from within the student version of the FITNESSGRAM® software. Similar to FITNESSGRAM®, a teacher or district coordinator could set up the software to create an ACTIVITYGRAM® "event" so that the data are compiled together and enable group level estimates and aggregated reports by grade or by school. Details are provided in the FITNESSGRAM® manuals (Meredith & Welk, 2010).

What Is the Scientific Basis for the ACTIVITYGRAM® Assessment?

The ACTIVITYGRAM® assessment is based conceptually on a validated self-report instrument known as the Previous Day Physical Activity Recall (PDPAR). ACTIVITYGRAM® uses the same basic grid structure to help children record their activities from the previous day and uses a similar 4-point intensity classification. One major difference between the PDPAR and the ACTIVITYGRAM® assessment is that the PDPAR focuses on after school activity while the ACTIVITYGRAM® assessment includes activity during the whole day (7:00 am to 11:00 pm). The choices of activities and the way that they are selected are also different for the ACTIVITYGRAM®. Another distinction is that PDPAR uses whole 30 minute blocks of time to estimate duration while the ACTIVITYGRAM® assessment allows activities to be reported as "some of the time," "most of the time," or "all of the time." The three choices are operationalized as 10 minutes, 20 minutes, and 30 minutes respectively out of the 30-minute time block. While this is a rough approximation, it provides a better way to capture the total volume of physical activity reported by the child.

Welk et al. (2004) performed a convergent and criterion validation study of both the ACTIVITYGRAM® and the PDPAR. Data were collected on elementary students from two schools on three consecutive days using the ACTIVITYGRAM®, PDPAR and the Biotrainer monitor. The results revealed non-significant differences in the reported number of bouts between the two instruments, which provided evidence for the convergent validity of ACTIVITYGRAM®. The Biotrainer monitor provided a way to evaluate the relative validity of the two self-report formats (ACTIVITYGRAM® and PDPAR) in this study. The ACTIVITYGRAM® yielded average daily correlations of $r = .50$ against three days of objective data from the Biotrainer activity monitor. This study also made direct comparisons between the PDPAR and the ACTIVITYGRAM®. The correlations between the number of bouts on the ACTIVITYGRAM® and the number of bouts coded on the PDPAR were high ($r > .70$) across the three days of comparison. The classification agreement was also high for coding intensities of physical activity. These findings indicate that the instruments provide similar information about physical activity patterns. The correlations were highest for the afternoon time period, which is important since this is the period that best reflects children's free living physical activity. Overall, this study provided convergent and criterion evidence for the validity of using ACTIVITYGRAM® to assess children's physical activity patterns.

The ACTIVITYGRAM® provides some significant advantages for school-based assessments. The computerized version self-report instrument offers convenience for data collection since the data are entered into the software and do not require manual data entry from a paper form to the computer. The ACTIVITYGRAM® also provides built in feedback (on screen and with printed reports) to enable teachers to help teach children (and parents) about appropriate levels of physical activity. A limitation of the ACTIVITYGRAM® format is that it requires 20-30 minutes to complete and some preparation time to teach children how to complete

⇦TOC ⇦Chapter

it. Another limitation is that some younger students might have trouble accurately recalling the characteristics of their physical activity in certain time frames.

Bibliography

Ainsworth, B. E,, Caspersen, C. J.,, Matthews, C. E., Masse, L. C., Baranowski, T. & Zhu W. (2012). Recommendations to improve the accuracy of estimates of physical activity derived from self-report. *Journal of Physical Activity and Health, 9(Suppl. 1),* S76-S84.

Bassett, D. R., Jr., Ainsworth, B. E., Leggett, S. R., Mathien, C. A., Main, J. A., Hunter, D. C. & Duncan, G. E. (1996). Accuracy of five electronic pedometers for measuring distance walked. *Medicine and Science in Sports and Exercise*, 28, 1071-1077.

Bauman, A., Phongsavan, P., Schoeppe, S., Owen, N. (2006). Physical activity measurement - a primer for health promotion. *Promoting Education*, 13, 92-103.

Blair, S. N., Cheng, Y., & Holder, J. S. (2001). Is physical activity or physical fitness more important in defining health benefits? *Medicine and Science in Sports and Exercise*, 33, S379-S399.

Bowles, H.R. (2012). Measurement of active and sedentary behaviors: closing the gaps in self-report methods. *Journal of Physical Activity and Health, 9(Suppl. 1),* S1-S4.

Butte, N. F., Ekelund, U., & Westerterp, K. R. (2012). Assessing physical activity using wearable monitors: measures of physical activity. *Medicine and Science in Sports and Exercise, 44(Suppl. 1),* S5-S12.

Caspersen, C. J., Pereira, M. A., & Curran, K. M. (2000). Changes in physical activity patterns in the United States, by sex and cross-sectional age. *Medicine and Science in Sports and Exercise*, 32, 1601-1609.

Canada, Public Health Agency of. (2011). *Canadian Physical Activity Guidelines*. Toronto.

Centers for Disease Control and Prevention. (2001). Increasing physical activity. A report on recommendations of the Task Force on Community Preventive Services. *Morbidity and Mortality Weekly Report*, 50, 1-14.

Centers for Disease Control and Prevention. (2011). School Health Guidelines to Promote Healthy Eating and Physical Activity. *Morbidity and Mortality Weekly Report*, 60(No. 5).

Centers for Disease Control and Prevention. (2012). Youth Risk Behavior Surveillance — United States. *Morbidity and Mortality Weekly Report*, 61(No. 61-4).

Centers for Disease Control and Prevention. (2013). Methodology of the Youth Risk Behavior Surveillance System - Recommendations and Reports. *Morbidity and Mortality Weekly Report*, 62(RR01), 1-23

Chinapaw, M. J., Mokkink, L. B., van Poppel, M. N., van Mechelen, & W., Terwee, C. B. (2010). Physical activity questionnaires for youth: a systematic review of measurement properties. *Sports Medicine*, 40, 539-563.

Corder, K., Ekelund, U., Steele, R. M., Wareham, N. J., & Brage, S. (2008). Assessment of physical activity in youth. *Journal of Applied Physiology*, 105, 977-987.

Crocker, P. R., Bailey, D. A., Faulkner, R. A., Kowalski, K. C., & McGrath, R. (1997). Measuring general levels of physical activity: preliminary evidence for the Physical Activity Questionnaire for Older Children. *Medicine and Science in Sports and Exercise*, 29, 1344-1349.

Crocker, P. R., Eklund, R. C., & Kowalski, K. C. (2000). Children's physical activity and physical self-perceptions. *Journal of Sports Science*, 18, 383-394.

Crouter, S. E. (2003). Validity of 10 electronic pedometers for measuring steps, distance, and energy cost. *Medicine and Science in Sports and Exercise*, 35, 1455-1460.

Cumming, S. P. (2008). Sex differences in exercise behavior during adolescence: is biological maturation a confounding factor? *Journal of Adolescent Health*, 42, 480-485.

De Vries, S. I., Van Hirtum, H. W., Bakker, I., Hopman-Rock, M., Hirasing, R. A., & Van Mechelen, W. (2009). Validity and reproducibility of motion sensors in youth: a systematic update. *Medicine and Science in Sports and Exercise*, 41, 818-827.

Freedson, P., Bowles, H. R., Troiano, R., & Haskell, W. (2012). Assessment of physical activity using wearable monitors: recommendations for monitor calibration and use in the field. *Medicine and Science in Sports and Exercise*, 44(Suppl. 1), S1-S4.

Freedson, P., Pober, D., & Janz, K. F. (2005). Calibration of accelerometer output for children. *Medicine and Science in Sports and Exercise*, 37(Suppl. 11), S523-S530.

Jago, R. & T. Baranowski. (2004). Non-curricular approaches for increasing physical activity in youth: a review. *Preventive Medicine*, 39(1), 157-163.

Janz, K. F., Lutuchy, E. M., Wenthe, P., & Levy, S. M. (2008). Measuring activity in children and adolescents using self-report: PAQ-C and PAQ-A. *Medicine and Science in Sports and Exercise*, 40, 767-772.

Kowalski C., Crocker, P. R. E., & Kowalski, N. P. (1997). Convergent validity of the Physical Activity Questionnaire for Adolescents. *Pediatric Exercise Science*, 9, 342-352.

Kowalski K., Crocker, P. R. E., & Faulkner, R. A. (1997). Validation of the Physical Activity Questionnaire for Older Children. *Pediatric Exercise Science*, 9, 174-186.

Lee, S. M., Burgeson, C. R., Fulton, J. E., & Spain, C. G. (2007). Physical education and physical activity: Results from the school health policies and programs study 2006. *Journal of School Health*, 77, 435-463.

Mahar, M. T., & Rowe, D. A. (2008). Practical guidelines for valid and reliable youth fitness testing. *Measurement in Physical Education and Exercise Science*, 12, 126-145.

Malina, R. M. (1996). Tracking of physical activity and physical fitness across the lifespan. *Research Quarterly for Exercise and Sport*, 67, 48-57.

Mathews, C. E. (2002). Use of self-report instruments to assess physical activity. In G. J. Welk (Ed.), *Physical activity assessments for health-related research* (pp. 107-123). Champaign, IL: Human Kinetics.

McKenzie T. L. (2006). *SOPLAY: System for Observing Play and Leisure Activity in Youth. Description and Procedures Manual*. San Diego, CA.

McKenzie, T. L., Marshall, S. J., & Sallis, J. F. (2000). Leisure-time physical activity in school environments: an observational study using SOPLAY. *Preventive Medicine*, 30), 70-77.

McKenzie, T. L., Crespo, N. C., & Baquero, B. (2010). Leisure-time physical activity in elementary schools: analysis of contextual conditions. *Journal of School Health*, 80), 470-477.

McKenzie. T. L. (2002). Use of direct observation to assess physical activity. In G. J. Welk (Ed.). *Physical activity assessments for health-related research* (pp. 179-195). Champaign, IL: Human Kinetics; pg. 179-195.

Meredith, M. D., & Welk, G. J. (2010). *FITNESSGRAM & ACTIVITYGRAM Test Administration Manual*. Champaign, IL: Human Kinetics.

Moore, J. B., Hanes, J. C., Jr., Barbeau, P., Gutin, B., Trevino, R. P., & Yin, Z. (2007). Validation of the Physical Activity Questionnaire for Older Children in Children of Different Races. *Pediatric Exercise Science*, 19, 6-19.

Morrow, J. R., Jr. (2005). 2004 C. H. McCloy Research Lecture: Are American children and youth fit? it's time we learned. *Research Quarterly for Exercise and Sport*, 76, 377-388.

Morrow, J. R., Jr. & Ede, A. (2009). Statewide Physical Fitness Testing: A BIG waist or a big waste? *Research Quarterly for Exercise and Sport*, 80, 5-14.

Morrow, J. R., Jr., Tucker, J. S., Jackson, A. W., Martin, S. B., Greenleaf, C. A., & Petrie, T. A. (2013). Meeting physical activity guidelines and health-related fitness in youth. American *Journal of Preventive Medicine*, 44, 439-444.

National Association for Sport and Physical Education. (2004). *Moving into the future. National standards for physical education*. St. Louis: Mosby.

Rowlands, A. V., & Eston, R. G. (2005). Comparison of accelerometer and pedometer measures of physical activity in boys and girls, ages 8-10 years. *Research Quarterly for Exercise and Sport*, 76, 251-257.

Rowlands, A. V. (2007). Accelerometer assessment of physical activity in children: an update. *Pediatric Exercise Science*, 19, 252-266.

Schneider, P. L., Crouter, S. E., & Bassett, D. R. (2004). Pedometer measures of free-living physical activity: Comparison of 13 models. *Medicine and Science in Sports and Exercise*, 36, 331-335.

Scruggs, P. W., Beveridge, S. K., Eisenman, P. A., Watson, D. L., Schultz, B. B., & Ransdell, L. B. (2003). Quantifying physical activity via pedometry in elementary physical education. *Medicine and Science in Sports and Exercise*, 35, 1065-1071.

Sherar, L. B. (2007). Age and gender differences in youth physical activity: does physical maturity matter? *Medicine and Science in Sports and Exercise*, 39, 830-835.

Silverman, S., Keating, X. D., & Phillips, S. R. (2008). A lasting impression: A pedagogical perspective on youth fitness testing. *Measurement in Physical Education and Exercise Science*, 12, 146-166.

Sirard, J. R., & Pate, R. R. (2001). Physical activity assessment in children and adolescents. *Sports Medicine*, 31, 439-454.

Saint-Maurice, P. F., Welk, G. J., Silva, P, Siahpush, M., & Huberty, J. (2011). Assessing children's physical activity behaviors at recess: a multi-method approach. *Pediatric Exercise Science*, 23, 585-599.

Saint-Maurice, P. F., Welk, G., Ihmels, M. A., & Krapfl, J. R. (2011). Validation of the SOPLAY direct observation tool with an accelerometry-based physical activity monitor. *Journal of Physical Activity and Health*, 8, 1108-1116.

Saint-Maurice, P. F., Welk, G., & Heelan, K. (2013). Calibration of self-report methods against objective measures of physical activity: a new approach using the PAQ. *Journal of Physical Activity and Health*. In Press.

Strong, W. B., Malina, R. M., Blimkie, C. J., Daniels, S. R., Dishman, R. K., Gutin, B. et al. (2005). Evidence based physical activity for school-age youth. *Journal of Pediatrics*, 46, 732-7.

Thompson, D. R., Baxter-Jones, A. D., Mirwald, R. L., & Bailey, D. A. (2003). Comparison of physical activity in male and female children: Does maturation matter? *Medicine and Science in Sports and Exercise*, 35, 1684-1690.

Thompson, D. R. (2007). Childhood overweight and cardiovascular disease risk factors: the National Heart, Lung, and Blood Institute Growth and Health Study. *Journal of Pediatrics*, 150(1), 18-25.

Troiano R. P. (2005). A timely meeting: objective measurement of physical activity. *Medicine and Science in Sports and Exercise*, 37, S487-S489.

Trost, S. G. (1999). Validity of the Previous Day Physical Activity Recall (PDPAR) in fifth-grade children. *Pediatric Exercise Science*, 11, 341-348.

Trost, S. G. (2001). Objective measurement of physical activity in youth: current issues, future directions. *Exercise and Sport Science Reviews*, 29, 32-36.

Trost, S. G., McIver, K. L., & Pate, R. R. (2005). Conducting accelerometer-based activity assessments in field-based research. *Medicine and Science in Sports and Exercise*, 37(Suppl. 11), S531-S43.

Tucker, J. M., Welk, G., Nusser, S. M., Beyler, N. K., & Dzewaltowski, D. (2011). Estimating minutes of physical activity from the previous day physical activity recall: validation of a prediction equation. *Journal of Physical Activity and Health*, 8, 71-78.

Tudor-Locke, C., & Bassett, D.R., Jr. (2004). How many steps/day are enough? Preliminary pedometer indices for public health. *Sports Medicine*, 34), 1-8.

Tudor-Locke C., Craig, C. L., Beets, M. W., Belton, S., Cardon, G. M., & Duncan, S. et al. (2011). How many steps/day are enough? *Children and Adolescents. International Journal of Behavioral Nutrition and Physical Activity*, 8, 78-92.

U.S. Department of Health and Human Services (2008). *Physical Activity Guidelines Advisory Committee Report*. Washington DC: U.S. Department of Health and Human Services.

Wall, M. I., Carlson, S. A., Stein, A. D., Lee, S. M., & Fulton, J. E. (2011). Trends by age in youth physical activity: Youth Media Campaign Longitudinal Survey. *Medicine and Science in Sports and Exercise*, 43, 2140-2147.

Ward, D. S., Evenson, K. S. & Vaughn, A. (2005). Accelerometer use in physical activity: best practices and research recommendations. *Medicine and Science in Sports and Exercise*, 37(Suppl. 11), S582-S588.

Welk, G. J., Corbin, C. B., & Dale, D. (2000). Measurement issues for the assessment of physical activity in children. *Research Quarterly for Exercise and Sport*, 71, 59-73.

Welk, G. J., & Wood, K. (2000). Physical activity assessments in physical education: A practical review of instruments and their use in the curriculum. *Journal of Physical Education, Recreation and Dance*, 71(1), 30-40.

Welk, G. J., Dzewaltowski, D. A., Ryan, G. J., Sepulveda-Jowers, E. M., & Hill, J. L. (2001) Convergent validity of the Previous Day Physical Activity Recall and the ACTIVITYGRAM assessment. *Medicine and Science in Sports and Exercise*, 33, S144.

Welk, G. J. (2002). *Physical activity assessments for health-related research* Champaign, IL: Human Kinetics.

Welk, G. J. (2008). The role of physical activity assessments for school-based physical activity promotion. *Measurement in Physical Education and Exercise Science*, 12, 184-206.

Wiersma, L. D., & Sherman, C. P. (2008). The responsible use of youth fitness testing to enhance student motivation, enjoyment, and performance. *Measurement in Physical Education and Exercise Science*, 12, 167-183.

Weston, A.T., Petosa, R., & Pate, R. R. (1997). Validation of an instrument for measurement of physical activity in youth. *Medicine and Science in Sports and Exercise*, 29, 138-143.

Chapter 6
Aerobic Capacity Assessments

Kirk J. Cureton, Sharon A. Plowman, Matthew T. Mahar

The Fitnessgram Reference Guide is intended to provide answers to some common questions associated with use and interpretation of FITNESSGRAM® assessments. This chapter, devoted to Aerobic Capacity Assessments, describes the issues associated with the assessment of aerobic capacity in children, including the validity and reliability of the field assessments used in FITNESSGRAM®, conversion of field test scores into aerobic capacity, the aerobic capacity standards, and interpretation of scores. The section specifically addresses the following questions:

⇦TOC ⇦Chapter

What Is Aerobic Capacity?

Aerobic capacity (VO_2max) reflects the maximum rate that oxygen can be taken up and utilized by the body during exercise. The magnitude of VO_2max depends on the capacity of the lungs to exchange oxygen between the air and blood in lung capillaries, the capacity of the cardiovascular system to transport oxygen to the muscles, and the muscles' capacity to use oxygen. The highest rate of oxygen uptake and use reflects the upper limit in the ability of the body to supply energy via aerobic metabolism to the active muscles during strenuous exercise. Aerobic capacity is most commonly expressed relative to body weight to account for differences in body size and to reflect a person's ability to carry out weight-bearing tasks.

Why Is Aerobic Capacity Important?

Aerobic capacity is an important component of physical fitness because it reflects the overall capacity of the cardiovascular and respiratory systems (Mitchell, Sproule, & Chapman, 1958; Taylor, Buskirk, & Henschel, 1955) and the ability to carry out prolonged strenuous exercise (Astrand et al., 2003; Taylor et al, 1955). From a health perspective, good aerobic capacity has been shown to reduce all-cause mortality and the risk of hypertension, coronary heart disease, obesity, diabetes, some forms of cancer, and other health problems (Blair et al., 1989; LaMonte & Blair, 2006) in adults, and clinical risk factors for cardiovascular disease and metabolic syndrome in children and adolescents (Barge et al., 2004; Ortega, Ruiz, Castillo, & Sjostrom, 2008).

How Does "Aerobic Capacity" Differ from Terms Such as "Cardiovascular Fitness" or "Cardiorespiratory Endurance"?

Many terms have been used to describe this dimension of physical fitness, including cardiovascular fitness, cardiorespiratory fitness, cardiorespiratory endurance, aerobic fitness, maximal aerobic power, aerobic work capacity, and physical work capacity. For all practical purposes, these terms are used interchangeably. A subtle distinction is that cardiorespiratory endurance, aerobic work capacity, and physical work capacity are typically used to refer to performance ability (the capacity to perform large-muscle activity for a prolonged period of time), whereas aerobic capacity refers to a functional (physiological) capacity. Because the underlying functional capacity is the construct of most interest in relation to health, and because field tests are actually validated against VO_2max measured in the laboratory, the term aerobic capacity has been used in the FITNESSGRAM® materials.

How Is Aerobic Capacity Measured in the Laboratory?

Aerobic capacity is measured in the laboratory using a graded exercise test during which the rate of oxygen uptake is measured continually using sophisticated equipment. A graded exercise test is a test typically administered on a treadmill or cycle ergometer in which the intensity of exercise is progressively increased. The rate of aerobic metabolism and oxygen uptake increases as intensity of exercise increases up to the point at which the aerobic capacity is reached. At this point, even though the exercise intensity can be increased, the oxygen uptake no longer increases proportionally and there is a plateau in the relation of the rate of oxygen uptake to work rate (exercise intensity). The rate of oxygen uptake at the plateau is aerobic capacity.

Measurement of aerobic capacity in the laboratory is technically demanding, requiring expensive equipment and highly-trained technicians. It also is time consuming; a test requires about 30 minutes and only one person can be measured at a time. Therefore, the direct measurement of aerobic capacity is not possible or practical for most field settings, such as schools where large numbers of people must be tested.

What Types of Field Tests Are Used in FITNESSGRAM® to Assess Aerobic Capacity?

Three field tests are used in FITNESSGRAM® to assess aerobic capacity: the PACER (Progressive Aerobic Cardiovascular Endurance Run), the One-Mile Run, and a walk test (for adolescents 13 years of age or older). Each assessment is briefly described below:

- The PACER is a multistage test adapted from the 20-meter shuttle run test published by Leger and Lambert (1982) and revised in 1988 (Leger, Mercier, Gadoury, & Lambert). It involves running back and forth across a 20-meter course in time to music played from an audio recording. Beeps on the sound track indicate when a person should reach the ends of the course. The test begins at a slow pace, and each minute the pace increases. A participant continues running until the pace can no longer be maintained. This test is like a graded exercise test on the treadmill in which the treadmill speed is increased at regular intervals. The longer a person continues, the higher the rate of estimated oxygen uptake. In the FITNESSGRAM® software, VO_2max is predicted from the number of laps completed during the test and a test equating procedure (Zhu, Plowman & Park, 2010), which converts PACER laps into comparable one-mile run times, which are then used to predict VO_2max. The PACER is a fun alternative to distance run tests, and is recommended for children, adolescents, and young adults. A 15-meter modified test may be substituted for the 20-meter test in elementary- and middle-school-age children in situations in which insufficient indoor space is available for the 20-meter test (McClain, Welk, Ihmels, & Schaben, 2006).
- In the One-Mile Run test, the objective is to run a mile as fast as possible. Because the rate of oxygen uptake is related in part to the pace sustained, it is possible to estimate the highest rate of oxygen uptake possible from the average pace sustained. Age, gender, and body fatness also affect the prediction of aerobic capacity. Therefore, in the FITNESSGRAM® software, aerobic capacity is predicted from mile time, age, gender, and body mass index using an equation of Cureton et al. (1995) developed on a large sample of children and adolescents.
- In the One-Mile Walk test, the objective is to walk one mile as fast as possible. The heart rate is determined immediately after the walk. By knowing body weight and the walk speed, the primary determinants of the oxygen uptake during walking, and the heart rate at the end of the walk, indicative of the percentage of the aerobic capacity being used, it is possible to estimate the aerobic capacity. In the FITNESSGRAM® software, aerobic capacity is estimated from age, gender, weight, mile walk time, and heart rate at the end of the walk using the equation of Kline et al. (1987), which has been shown to be accurate for high school students (McSwegin, Plowman, Wolff, & Guttenburg, 1998). The walk test has the advantage of not requiring a maximal effort as is required in the two running tests.

How Reliable Is the Measurement of Maximal Oxygen Uptake in Youth?

Aerobic capacity (VO_2max) expressed relative to body weight ($mL \cdot kg^{-1} \cdot min^{-1}$) measured on the treadmill is the criterion against which FITNESSGRAM® field tests of aerobic capacity have been validated. Its reliability is important because it affects the magnitude of validity coefficients assessing the accuracy of the field tests for predicting VO_2max. Although a range of reliability coefficients has been reported, the consensus is that the reliability of measuring

VO_2max in youth is high and acceptable for a criterion measure of physical fitness. Table 1 found in an Appendix to this chapter summarizes the results of studies reporting the test-retest reliability coefficients for VO_2max ($mL \cdot kg^{-1} \cdot min^{-1}$) determined on the treadmill in youth. The values have generally varied from moderate to high; the two low coefficients in the table may not be comparable to the other results. One (.56) represents two measurements separated by 4-5 months, which is too long to represent the true reliability of the test. The other low coefficient (.47) was attributed to a long walking protocol in which leg fatigue or boredom may have affected the test outcome. In studies in which shorter walking protocols were used, high reliability was obtained. Results of the studies reporting reliability of VO_2max measurement in children and adolescents are reported in Table 1 in the Appendix to this chapter.

How Reliable Are the Field Tests of Aerobic Capacity?

The reliability of the three field tests of aerobic capacity is, for the most part, high. Consistently high reliability coefficients have been reported for the PACER and One-Mile Walk test. High coefficients also have been reported for children over nine years of age for distance runs such as the One-Mile Run. However, reliability of distance runs in younger children is lower, probably because of variation in motivation and pacing.

- ***One-Mile Run***. The reliability of distance run tests in youth was summarized by Safrit (1990). Reliability coefficients for 600-yd, 1600-m, 9-min, and 12-min runs ranged from approximately .60 to .90. Safrit concluded that the reliability of distance runs in children is for the most part high, but not uniformly so. Results of the relatively few studies that have reported reliability coefficients for the mile run test in youth are summarized in Table 2 in the Appendix to this chapter. In general, for children 9 years of age (third grade) and older, the reliability is moderate or high, with reliability coefficients above .66. For younger children, reliability coefficients are mixed, with those of Krahenbuhl, Pangrazi, Petersen, Burkett, and Schneider (1978) being high and those of Rikli, Petray, and Baumgartner (1992) being relatively low. Lower reliability on distance runs in young children may be due to variation in motivation and pacing strategy. Practice of steady pacing can improve run performance in children (Saltarelli & Andres, 1993). Low test-reliability due to the influence of behavioral variables may limit the validity of the One-Mile Run as a field test of VO_2max in young children.
- ***PACER***. Five studies have reported that the reliability of the PACER test in youth is moderate or high (see Table 3 in the Appendix to this chapter). Reliability coefficients were above .64 with no significant mean differences between two tests. Additional reliability studies with samples differing in age, gender, and fitness level would be useful to confirm the results of the studies cited here.
- ***One-Mile Walk***. McSwegin et al. (1998) reported that the reliability of VO_2max estimated from the One-Mile Walk test using the Kline et al. (1987) equation was high. They reported an intraclass correlation of .91 for repeat measures on 21 boys and girls 14-18 years of age.

How Valid Are the Field Tests of Aerobic Capacity in Children for Estimating Aerobic Capacity?

The three field tests used in the FITNESSGRAM® battery for estimating VO_2max have moderately good and approximately equal validity in children 10 years of age and above. VO_2max ($mL \cdot kg^{-1} \cdot min^{-1}$) is estimated with an error of 10-15% of the mean for most children.

Review of Validity Evidence for the One-Mile Run

The rationale (content and construct validity) for using the One-Mile Run to estimate VO_2max is based on the fact that for exhaustive exercise lasting longer than two minutes, energy is provided primarily through aerobic metabolism (Astrand et al., 2003). Therefore, performance on an event such as the One-Mile Run is determined, in large part, by the highest rate of aerobic metabolism (VO_2max) that can be maintained for the duration of the event. The highest rate of VO_2 that can be maintained during a distance run, in turn, is determined in large part by the VO_2max. Thus, distance run performance and VO_2max are correlated and a distance run performance can be used to estimate VO_2max. Moderately strong correlations between VO_2max and performances on distance run tests in adults and youth support this rationale (Safrit, Hooper, Ehlert, Costa, & Patterson, 1988). The construct validity evidence for use of distance run tests to estimate VO_2max depends on the extent to which variance in run performance is determined by VO_2max compared to other physiological and behavioral factors. The underlying factors that determine running performance are in part dependent on the distance or duration of the run. Balke (1963) found that in young adult trained runners, the estimated energy demand of the highest pace that could be maintained for 12 minutes equaled the VO_2max. The duration would probably be less for untrained youth because few can sustain 100% of the VO_2max for 12 minutes (Krahenbuhl, Morgan, & Pangrazi, 1989; McCormack, Cureton, Bullock, & Weyand, 1991; Sloniger, Cureton, & O'Bannon, 1994). A study with college students found that distance runs of 1 mile and longer measure the same underlying factors, whereas the factors underlying shorter runs were different (Disch, Frankiewicz, & Jackson, 1975). A study with elementary school children obtained similar results (Jackson & Coleman, 1976). These studies suggest that if VO_2max is the primary determinant of distance running, that runs of one mile and longer should be used to assess VO_2max. Correlations between distance runs of different distances and VO_2max support this deduction (Baumgartner & Jackson, 1991; Disch et al., 1975; Jackson & Coleman, 1976; Krahenbuhl, Pangrazi, Petersen, Burkett, & Schneider, 1977; Krahenbuhl et al., 1978; Safrit et al., 1988).

Variables other than aerobic capacity, including body fatness, running skill and economy, physiological variables that affect the $\%VO_2max$ that can be maintained, effort given on the test, appropriate pacing, and environmental conditions also affect distance running performance in youth (Cureton, 1982; Cureton, Boileau, Lohman, & Misner, 1977; Krahenbuhl et al., 1989; McCormack et al., 1991). With the exception of body fatness, the influence of these variables reduces the correlation (concurrent validity evidence) of distance run tests with VO_2max ($mL \cdot kg^{-1} \cdot min^{-1}$). The confounding effect of behavioral variables such as motivation and proper pacing may be more important in younger than in older children (McCormack et al., 1991). Excess body fat reduces VO_2max expressed relative to body weight (Buskirk & Taylor, 1957; Welch, Reindeau, Crisp, & Isenstein, 1957) and performances on field tests that involve prolonged running (Cureton et al., 1977, 1978; Cureton, Baumgartner, & McManis, 1991; Sparling & Cureton, 1983). Therefore, part of the association of VO_2max ($mL \cdot kg^{-1} \cdot min^{-1}$) with the field tests reflects the influence of body fatness on both variables. This is reflected by the fact that correlations of distance run tests with VO_2max expressed relative to fat-free weight are lower than those with VO_2max expressed relative to body weight (Cureton, 1982). Therefore, validity coefficients of running field tests with VO_2max ($mL \cdot kg^{-1} \cdot min^{-1}$) should not be interpreted only in terms of cardiovascular-respiratory capacity; they also reflect the influence of differences in %fat.

The concurrent validity of distance run tests has been evaluated by correlating distance run performance with VO_2max (mL·kg^{-1}·min^{-1}). Studies performed on adults and children are summarized in several sources (Baumgartner & Jackson, 1991; Safrit et al., 1988). For studies on youth involving runs of 1-1.5 miles or 9 to 12 minutes and in which VO_2max was measured on the treadmill, validity coefficients have ranged from approximately .60 to .80 (with one exception). Studies on the concurrent validity of the One-Mile Run are summarized in Table 4 in the Appendix to this chapter.

Review of Validity Evidence for the PACER Test

An attractive feature of the PACER is its high content (logical) validity. The PACER is a progressive, multistage maximal exercise test that closely simulates a graded, speed-incremented treadmill test used in the laboratory to directly measure VO_2max. The VO_2max required is submaximal at earlier stages and increases progressively each minute up to maximal (Leger & Gadoury, 1989; Leger & Lambert, 1982). Because the speed of running is controlled, variation in pacing has little influence on test outcome. Because a maximal effort is required only at the end of the test, motivation is probably less of a problem than with the One-Mile Run, in which a sustained, near-maximal intensity is required throughout.

The concurrent validity evidence for the PACER test has been established in numerous studies by correlating the VO_2max at the end of the test or the highest test stage (running speed) attained with VO_2max directly measured on the treadmill. In two studies on adults (Leger & Gadoury, 1989; Leger & Lambert, 1982), VO_2 measured by backward extrapolation immediately after the test was highly correlated with and did not differ significantly from the VO_2max measured during the final minute of a walking graded exercise test on the treadmill. In studies on adults, validity coefficients correlating test performance with VO_2max (mL·kg^{-1}·min^{-1}) have ranged from .83 to .93 with standard errors of estimate ranging from 3.6 to 5.4 mL·kg^{-1}·min^{-1} (Leger & Gadoury, 1989; Leger & Lambert, 1982; Leger et al., 1988; Paliczka, Nichols, & Boreham, 1987; Ramsbottom, Brewer, & Williams, 1988). Plowman and Liu (1999) found large differences in the accuracy of published regression equations predicting VO_2max in college students, with the Leger et al. (1988) adult equation being more accurate than the equation of Ramsbottom et al. (1988) or an equation of Leger et al. (1988) based on youth and young adults 8 to 19 years.

Studies that have investigated the concurrent validity evidence of the PACER in youth are summarized in Table 5 in the Appendix to this chapter. The range of validity coefficients and standard errors of estimate are similar to those for the One-Mile Run, indicating that the PACER has moderate evidence of concurrent validity as a field test of VO_2max. Some of these studies (Barnett, Chan, & Bruce, 1993; Leger et al., 1988; Mahar et al., 2006, 2011; Mercier, Gadoury, & Lambert, 1988) used age, sex, and anthropometric variables (skinfold thickness or body weight) in addition to PACER performance to improve the prediction of VO_2max and others did not. It is clear that age is an important predictor because it helps take into account the improvement in running economy that occurs during growth and development (Barnett et al., 1993; Leger et al., 1988). The change in running economy alters the relation between running performance (highest stage or speed on the test) and VO_2max. Sex is not always a significant predictor, but Mahar et al. (2011) found the age/sex interaction was significant, as would be expected, based on the established differences between boys and girls in age-related changes in VO_2max (mL·kg^{-1}·min^{-1}) (Krahenbuhl, Skinner, & Kort, 1985). Mahar et al. (2011) also found a significant quadratic relationship between PACER laps and VO_2max (mL·kg^{-1}·min^{-1}).

In general, the concurrent validity evidence for the PACER test appears to be approximately the same as distance run tests for estimating VO_2max. In one study in which the PACER and a 6- min run were correlated with VO_2max in the same sample, VO_2max was more highly correlated with the PACER test than with the distance run (r = .76 vs. .63) (van Mechelen, Hlobil, & Kemper, 1986). Dinschel (1994) reported that in 4th- and 5th-grade boys and girls, the laps completed on the PACER test and mile run time were moderately correlated (r = -.63 and -.57). Mahar et al. (1997) reported similar results for a large sample of 10- and 11-year-old boys and girls, with correlations between PACER laps and one-mile run time with VO_2max ranging from -59 to -.67. Plowman and Liu (1999) found that in a sample of college students, the validity coefficients and standard errors of estimate for VO_2max predicted from the One-Mile Run using the Cureton et al. (1995) equation and from the PACER using three different equations were similar, although the absolute accuracy of one of the Leger et al. (1988) equations was considerably better than predictions from two other equations.

Review of the Validity Evidence for the One-Mile Walk Test

McSwegin et al. (1998) reported that the validity of VO_2max estimated from the walk test was high. They reported a correlation of .84, a standard error of estimate of 4.5 $mL \cdot kg^{-1} \cdot min^{-1}$, and a total error of 5.2 $mL \cdot kg^{-1} \cdot min^{-1}$ between VO_2max estimated using the Kline et al. equation and directly-measured VO_2max in 44 boys and girls 14-18 years of age.

How Were the Standards for Aerobic Capacity in FITNESSGRAM® Established?

FITNESSGRAM® standards for aerobic capacity were first published in the 1987 FITNESSGRAM® User's Manual (CIAR, 1987). These standards were designed to represent the lowest levels of aerobic capacity consistent with minimizing disease risk and ensuring adequate functional capacity for daily living (Cureton & Warren, 1990). The levels of aerobic capacity were established by expert opinion, taking into account developmental changes. The standards were first presented as upper and lower boundaries of a Healthy Fitness Zone in 1992 (CIAR, 1992). The lower boundary and its interpretation were essentially the same as the original standards. More specific rationale linking the lower-boundary aerobic capacity values to reduced disease risk was developed and first published in the FITNESSGRAM® Technical Reference Manual (Morrow et al., 1994). The upper-boundary standards were designed to represent a "good" level of aerobic capacity, one that is associated with lower risk of disease and higher work capacity than the lower-boundary standards. The rationale for the upper and lower boundaries of the Healthy Fitness Zone was based on data linking VO_2max with disease risk in adults. At the time the FITNESSGRAM® standards were developed, no comparable data linking aerobic capacity to disease risk existed for children. In recent years, studies linking aerobic capacity to disease risk in children have confirmed that the FITNESSGRAM® aerobic capacity standards had utility for detecting health risk (Ruiz, Ortega, Rizzo et al., 2007; Lobello, Pate, Dowda, Liese, & Ruiz, 2009; Adegboye, Anderssen, Froberg et al., 2009).

The current FITNESSGRAM® criterion-referenced standards for aerobic capacity were developed in 2010, introduced with the version 9 software, and retroactively included in version 8 of the software. The procedures used in developing the standards for VO_2max have been described in detail (Welk, Laurson, Eisenmann, & Cureton, 2011). The standards were designed to indicate the level of aerobic capacity associated with increased risk of the metabolic syndrome in youth. The metabolic syndrome is a cluster of symptoms, including abdominal obesity, insulin

⇦TOC ⇦Chapter

resistance, disordered blood lipids, hypertension and glucose intolerance that increase the risk of cardiovascular disease and diabetes. The clinical diagnosis of metabolic syndrome is based on measures of waist circumference, resting blood lipids, blood pressures, and blood glucose (Grundy et al., 2005). To develop the standards, available data on aerobic capacity estimated from heart rate during a treadmill graded exercise test and the clinical measures used to diagnose metabolic syndrome were obtained from a nationally-representative sample of U.S. children and adolescents gathered during the National Health and Nutrition Examination Survey (NHANES) between 1999 and 2002. Sophisticated statistical analyses (including Receiver Operating Characteristic curves) were used to identify two thresholds, below which risk was increased by different degrees. These two thresholds allowed for the identification of three separate zones, a healthy fitness zone and two where improvement is needed. The advantage of three zones over two is that it provides a more prescriptive message about a youngster's fitness level.

The ***"Healthy Fitness Zone (HFZ)"*** was established by emphasizing cut-point sensitivity (Se) (percentage of children with metabolic syndrome who are correctly identified as having the condition) over specificity (Sp) (percentage of healthy children who are correctly identified as not having the condition). The high sensitivity of this cut-point should ensure that most children with metabolic syndrome have fitness levels below this threshold. A child with a fitness level (i.e., VO_2max value) above this cut-point should have a very low risk of metabolic syndrome and has a good level of fitness. The sensitivity threshold was set at a higher value for boys (Se ~ .85) than girls (Se ~ .75) because there is a stronger link between fitness and metabolic syndrome in boys. Achieving the same level of diagnostic classification accuracy in girls would have necessitated setting standards at an exceptionally high level (values higher than boys for most age groups). The final values were set to provide equivalent VO_2max values for boys and girls less than 12 years of age.

The ***"Needs Improvement–Health Risk (NI-HR)"*** zone was established by emphasizing specificity over sensitivity. The high specificity of this cut-point (>95%) should ensure that youth with low levels of fitness (VO_2max values below this threshold) would get appropriate feedback about potential risk. The diagnostics suggest that 95% of children without metabolic syndrome will have fitness levels above this threshold. It is possible that some children with metabolic syndrome could fall above this threshold (due to lowered Se) but the goal of this threshold is to identify youth who may have increased risk due to being below this threshold. The "Needs Improvement–Health Risk" zone provides youth/parents with an appropriate warning of health risk. The final values were set to provide equivalent VO_2max values for boys and girls less than 12 years of age.

The ***"Needs Improvement (NI)"*** zone is an intermediate zone between the calculated thresholds of the bottom (lowest acceptable VO_2max) of the HFZ and the top (highest VO_2max) of the NI–HR zones. This intermediate zone represents levels of aerobic capacity associated with moderate risk of the metabolic syndrome. Students whose scores place them in this zone receive a message encouraging them to strive to achieve the HFZ.

The standards are age and sex specific. The standards are empirically derived using contemporary statistical methods from clinical disease risk data in youth and take into account developmental changes in aerobic capacity and disease risk factors. They are truly health-related.

Why Are the Standards for the One-Mile Run, PACER, and One-Mile Walk Tests All Expressed as VO_2max?

The primary reason for expressing the standards for the One-Mile Run, PACER, and the One-Mile Walk tests as VO_2max is that VO_2max is the measure of interest related to health. The statistical process through which the revised standards were developed identified the level of VO_2max that corresponded to a higher or lower risk for metabolic syndrome. A single set of VO_2max standards serves all three field tests. While One-Mile Run performance, PACER laps, and heart rate response to walking at a given speed are related to VO_2max, they are not measures of VO_2max. Scores on the field tests must be converted into the common currency of VO_2max in order to be related to health risk.

In the previous FITNESSGRAM® standards for aerobic capacity, One-Mile Run times and PACER laps equivalent to the aerobic capacity standards were provided. This was possible by using different, but less accurate, methods of linking performance scores to VO_2max. Using the old approach, the chances of misclassifying fitness and disease risk is increased. The most accurate estimates of VO_2max are obtained when demographic measures such as age, sex, and BMI or weight are combined with performance on the mile run or PACER, or in the case of mile walk test, walk time and heart rate, in a complex formula to predict VO_2max (Cureton et al, 1994; Kline et al., 1987; Mahar et al., 2010). In revision of the standards, the most accurate estimates of VO_2max from the field tests were used to optimize classification of health risk.

How Is the Aerobic Capacity Reported in FITNESSGRAM® Calculated?

In the software used to produce reports of physical fitness test results in FITNESSGRAM®, aerobic capacity is predicted VO_2 max from a statistical (regression) equation that relates performance on or responses to the test. Conversion of field test performances to VO_2 max allows comparison of scores on the three field tests and permits changes in the relation of the test performance to VO_2 max that occur with age to be taken into account. Details on the prediction equations used for the various aerobic capacity assessments are provided below:

Prediction of VO_2max from the One-Mile Run test

The equation used to predict VO_2 max ($mL \cdot kg^{-1} \cdot min^{-1}$) from the One-Mile Run was based on work by Cureton et al. (1995). The equation was based on a sample of 753 males and females, 8-25 years of age and uses age (years), sex (coded 0=F and 1=M), body mass index (BMI in units of $kg \cdot m^{-2}$) and mile run time (minutes) for the prediction ($R = .72$, $SEE = 4.8$ $mL \cdot kg^{-1} \cdot min^{-1}$).

$$VO_2max = .21\ (Age \times Sex) - .84\ (BMI) - 8.41\ (Time) + .34\ (Time2) + 108.94$$

The relation between VO_2 max and One-Mile Run times is curvilinear. There is an inverse, relatively-linear relation between VO_2 max and One-Mile Run time for times below about 11 minutes, but virtually no relation for times above about 11 minutes. For One-Mile Run times above about 13 minutes, predicted VO_2 max values are actually higher than for lower One-Mile Run times. Therefore, any times above 13 minutes should be set to 13 before predicting VO_2 max with this equation.

Prediction of VO_2max from the PACER Test

A number of equations are available for predicting VO_2 max from the PACER test, and the accuracy of these equations has been studied in detail (Mahar, Guerieri, Hanna, & Kemble, 2011). The inclusion of many variables in the equations can help to increase predictive accuracy, but it can complicate assessments and decrease utility for field-based assessments. Therefore, consideration was given to finding an equation that could produce reasonable predictions without requiring the collection of too many supplemental variables. Data for the analyses were combined from a number of studies and this made it possible to develop prediction equations that work across multiple age ranges.

The final selected equation used to estimate aerobic capacity includes age and the number of laps performed (Mahar et al., 2013). Consideration was given to including a gender term in the equation but the standards are already gender specific so this was not necessary. Consideration was also given to including BMI in the equation. While inclusion of a BMI term has been shown to improve the prediction of estimates from the mile run, this was not the case with the PACER. The regression equation with age and laps yielded reasonable predictive utility while also facilitating use in school based programs. Boys and girls have to perform more laps as they get older. Boys have to perform more laps than girls at a given age (after the age of 12 years).

Prediction of VO_2max from the One-Mile Walk Test

The equation of Kline et al. (1987) is used to predict VO_2 max ($mL \cdot kg^{-1} \cdot min^{-1}$) for the walk test. The equation was based on 343 men and women, 30-69 years of age and uses the person's age (y), gender (F=0, M=1), weight (lb), walk time (min) and heart rate at the end of the mile walk (bpm) for the prediction (R = .88, SEE = 5.0 mL.kg-1.min-1). McSwegin et al. (1998) have shown this equation to be valid in high school age individuals.
$VO_2max = -.3877\ (Age) + 6.315\ (Gender) - .0769\ (Weight) - 3.2649\ (Time) - .1565\ (bpm)$

Do the PACER, One-Mile Run, and One-Mile Walk Tests Give the Same Classification of Fitness?

The PACER, One-Mile Run, and One-Mile Walk tests are all designed to estimate VO_2max, but due to differences in the nature of the assessments and means through which they are converted into an estimate of VO_2max, they may not always yield the same classification of fitness. This is because there is error in predicting directly measured (actual) VO_2max with each of the field tests. Thus, it is possible a child could be classified as being within the Healthy Fitness Zone by one test, but in the Some-Risk or High-Risk Needs Improvement Zones by another test. Summary data from schools may also vary depending on the choice of assessment that is used. It is not possible to determine the exact pattern of agreement since it would vary by age and gender and would be influenced by other variables such as the degree of motivation as well as environmental conditions. Teachers and school officials should be aware that the results from the three assessments cannot be directly compared. Regardless of what test is used the focus should be on the relative differences in fitness achievement from one year to the next (either on an individual level or a group level).

How Can We Best Motivate Students to Perform on the Aerobic Capacity Measure?

To obtain accurate information about aerobic capacity it is important that students provide their best effort. This must be reinforced to the students prior to the test. Some teachers may prefer to provide a target or goal for students to strive for but this cannot be directly determined for all tests. Therefore, the best recommendation is to encourage students to do their best so that they get the most accurate score. Aerobic fitness tests are not unlike intellectual aptitude tests in which an individual's absolute score only assumes meaning when evaluated relative to standards.

The estimates of aerobic capacity from the PACER test are dependent on the child's age and the number of laps that are completed. The number of laps required to achieve the Healthy Fitness Zone for boys and girls ranging in age from 10-18 are available to teachers and students using FG software.

With the mile run, the prediction of aerobic capacity depends on other factors (including the child's BMI). Therefore, it is not possible to provide a direct goal time for which students should aim. Students should be encouraged to cover the distance as quickly as possible. Similarly, with the One-Mile Walk test it is not possible to produce estimated or goal times. This is because performance on the assessment depends on the child's heart rate relative to the time it took to complete the walk. One student may prefer a faster walking pace while others may use a slower pace. An advantage of the Walk Test is that it is possible to estimate aerobic capacity regardless of the pace that is chosen to complete the walk. A brisk walking pace is recommended to obtain the most valid data since it produces a more pronounced change in heart rate.

How Does Body Size and Composition (Weight, Percent Body Fat, BMI) Impact Aerobic Capacity?

Aerobic capacity reflects the highest rate oxygen can be taken up and used by the body. When it is measured in the laboratory, the rate of oxygen uptake is expressed in liters per minute ($L \cdot min^{-1}$). Other things being equal, children with higher fat-free body mass, who have bigger hearts, blood volumes, lungs, and muscles involved in the uptake, transport and use of oxygen, tend to have higher values for oxygen uptake than smaller children (Astrand, 1952; Norman, Drinkard, McDuffie, Ghorbani, Yanoff, & Yanovski, 2005). To adjust for the size effect, VO_2max values in units of $L \cdot min^{-1}$ have traditionally been divided by body weight in kg (1 kg = 2.2 lb). When expressed relative to body weight in $mL \cdot kg^{-1} \cdot min^{-1}$, the effect of body size is reduced but the influence of body fatness is introduced (Cureton, 1982). Body fat does not contribute to the body's ability to use oxygen, but it increases body weight and BMI, and thus decreases the VO_2max when it is expressed relative to body weight (Buskirk & Taylor, 1957; Welch, Reindeau, Crisp, & Isenstein, 1957). Other things being equal, leaner children with lower body weights will have higher VO_2max ($mL \cdot kg^{-1} \cdot min^{-1}$) values than children with more body fat or higher BMI. Overweight children are at a disadvantage on tests of aerobic capacity. Excess fat is associated with poorer performances on the One-Mile Run and PACER tests (Cureton et al., 1977, 1982, 1991, 1995; Ihasz et al., 2006; Lloyd et al., 2003; Rowland et al., 1999), and lower values for VO_2max ($mL \cdot kg^{-1} \cdot min^{-1}$) estimated from all three field tests and measured in the laboratory (Buskirk & Taylor, 1957; Cureton et al., 1977; Rowland et al., 1999). The lower scores on tests of aerobic capacity do not necessarily mean that cardiovascular-respiratory capacity in an absolute sense is low (although it may be), but relative to body weight, it is. The lower VO_2max values are, however, associated with reduced capacity for weight-bearing

⇦TOC

⇦Chapter

physical activity and exercise, and increased health risk (Brage et al., 2004; Lobello et al., 2009; Ortega et al., 2008). Adjustment of the PACER protocol to start at a lower speed to better accommodate overweight children does not improve test scores (Ihasz, Finn, Meszaros, & Zsidegh, 2006). Procedures for adjusting the field test scores for body fatness have been proposed (Cureton et al., 1991; Lloyd et al., 2003), but these would be difficult for teachers to implement and interpret. Children with quite high levels of body fat and BMI will tend to have quite low levels of aerobic capacity and will have difficulty achieving the Healthy Fitness Zone without reducing body weight. The influence of body weight and composition on aerobic capacity and on risk of the metabolic syndrome is a common underlying factor and accounts for part of the relationship between aerobic capacity and disease risk.

Why Are Standards for Boys Generally Higher than for Girls?

It is not known for certain why the aerobic capacity standards for boys and girls are different at most ages (although the new standards are the same at ages 10 and 11). Hormonal and other biological sex differences and environmental factors may result in different risks of the metabolic syndrome due to factors other than those associated with aerobic capacity. Also, inherent, gender-related differences in body composition and in hemoglobin concentration cause VO_2max ($mL \cdot kg^{-1} \cdot min^{-1}$) values for boys and girls who have the same level of physical activity to be different. The differences prior to puberty are very small or nonexistent (for hemoglobin concentration), but they increase during puberty and adolescence. These differences are linked in part to differences in the reproductive hormones. Regardless of the reason, the standards for boys and girls reflect the different levels of VO_2max that are associated with increased risk for metabolic syndrome.

Why Aren't Criterion-Referenced Standards Available for the One-Mile Run and PACER for Children Under 10 Years of Age?

Standards were not developed for children under age 10 because of concerns over the reliability and validity of the test results. Even with practice, it is difficult to assure that young children will pace themselves appropriately on a One-Mile Run, and give a maximal effort on the One-Mile Run and PACER tests. This is reflected in the fact that the reliability and validity of the one-mile run, and the validity of the PACER for estimating VO_2max in young children are not consistently good. Therefore, there is the danger that aerobic capacity will be inappropriately evaluated (underestimated) in a considerable number of children. By practicing these tests several years before actually being compared to standards, there is a greater probability fewer misclassifications will occur. The One-Mile Walk test reduces these problems, although it still requires maintaining a focus on walking as fast as possible, but it has not been validated for young children.

To What Extent Is Aerobic Capacity Determined by Genetics Versus Physical Activity?

There is a genetic component to aerobic capacity. Some people inherit characteristics that give them a naturally higher level of aerobic capacity than other people. However, the genetic component is thought to be relatively small, accounting for less than 30% of the differences between people (Bouchard et al., 1992). Thus, aerobic capacity mostly reflects the level of habitual physical activity. In particular, aerobic capacity reflects the intensity and amount of dynamic, moderate-to-vigorous, sustained (aerobic) physical activity in which youth participate.

However, even the improvement in VO_2max has a genetic component, with some people capable of much more improvement than others (Bouchard et al., 1999; Prud'homme, Bouchard, LeBlanc, Landrey, & Fontaine, 1984).

How Can Aerobic Capacity Be Improved?

Aerobic capacity of youth can be improved with sustained periods of higher-intensity exercise (Pate & Ward, 1990). Although the exact dose of exercise needed in youth has not been identified, three or more sessions per week in which moderately-high-intensity exercise is sustained for 30 minutes or more are probably required. Any dynamic exercise involving large muscle groups is suitable, such as vigorous walking, jogging/running, cycling, swimming, and vigorous games. Improvements are proportional to the amount of moderately-high-intensity exercise completed per week.

Appendix

Table 1. Reliability of VO_2max ($mL \cdot kg^{-1} \cdot min^{-1}$) in Children and Adolescents

Source	Sample	Test Type	Reliability Coefficient[a]
Boileau et al. (1977)	21 M, 11-14 y	Walk	r = .87
Cunningham et al. (1977)	66 M, 10 y	Walk/Run	r = .56
Cureton (1976)	27 M & F, 7-12 y	Walk	r = .88
Paterson et al. (1981) 8 M, 10-12 y		Walk	R = .47
		Jog	R = .87
		Run	R = .95
Pivarnik et al. (1996)	32 F, 10-16	Walk Run	R =. 93

Note. [a]*r = interclass reliability; R = intraclass reliability*

Table 2. Reliability of the One-Mile Run Test in Children and Adolescents

Source	Sample	Reliability Coefficient
Beets and Pitetti (2006)	114 M & 66 F, 13-18 y	R = .66, .77
Bono et al. (1991)	15 M & 15 F, 5th grade	r = .91
	15 M & 15 F, 8th grade	r = .93
	15 M & 15 F, 11th grade	r = .98
Krahenbuhl et al. (1978)	34 F, 1st grade	r = .82[a]
	49 M, 3rd grade	r = .92[a]
Rikli et al. (1992)[b]	20 M & 16 F, Kindergarten	R = .53, .39
	15 M & 17 F, 1st grade	R =.56, .54
	45 M & 52 F, 2nd grade	R =.70, .71
	53 M & 63 F, 3rd grade	R =.84, .90
	44 M & 37 F, 4th grade	R =.87, .85

Notes. r = interclass reliability; R = intraclass reliability for a single trial

[a]*1600-m run*

[b] *First coefficient is for males, second is for females*

Table 3. Reliability of the PACER Test in Children and Adolescents

Source	Sample	Reliability Coefficient
Beets and Pitetti (2006)	123 M, & 62 F 13-18 y	R = .68, .64
Dinschel (1994)	57 M & 44 F, 4-5th grade	R = .84
Leger et al. (1988)	139 M & F, 6-16 y	r = .89
Liu et al. (1992)	20 M & F, 12-15 y	R = .93
Mahar et al. (1997)	137 M & 104 F, 10-11 y	R = .90

R = interclass reliability; R = intraclass reliability for a single trial

Table 4. Concurrent Validity of the One-Mile Run in Children, Adolescents, and College Students

Source	Sample	Validity Coefficient	SEE ($mL \cdot kg^{-1} \cdot min^{-1}$)
Bono et al. (1991)	15 M & 15 F, 5th grade	-.76	4.6
	15 M & 15 F, 8th grade	-.80	4.9
	15 M & 15 F, 11th grade	-.85	4.3
	45 M & 45 F, 5-11th grade	-.73	5.3
	45 M & 45 F, 5-11th grade	-.84[a]	4.3
Cureton et al. (1977)	140 M & 56 F, 7-11th grade	-.66	4.9
Cureton (1995)	490 M & 263 F, 8-25 yrs	.72[b]	4.8
Krahenbuhl et al. (1978)	49 M, grades 1-3	-.60[c]	5.1
	34 F, grades 1-3	-.74[c]	4.4
Krahenbuhl et al. (1977)	38 M & F, 3rd grade	-.62[d]	5.3
	18 F, 3rd grade	-.26[d]	5.5
	20 M, 3rd grade	-.71[d]	5.0
Plowman and Liu (1999)	94 M & F, 18-30 yrs	.82[e]	4.6
Rowland et al. (1999)	36 M, 6th grade	.77[f]	3.7

a Prediction from age, gender, weight, sum of two skinfolds, and One-Mile Run/Walk

b Prediction from age × gender, BMI, MRW (Mile Run/Walk), and MRW^2

c 1600-m run

d 1609-m run

e Correlation between VO_2max predicted from using Cureton et al. (1995) equation and measured VO_2max

Table 5. Concurrent Validity of the PACER Test in Children and Adolescents

Source	**Sample**	**Validity Coefficient**	**SEE ($mL \cdot kg^{-1} \cdot min^{-1}$)**
Armstrong et al. (1988)	77 M, 11-14 y	.54	5.3
Barnett et al. (1993)	27 M & 28 F, 12-17y	.74	4.6
		.82[b]	4.0
		.85[c]	3.7
		.72[a]	5.4
Boreham et al. (1990)	23 M, 14-16 y	.64	4.5
	18 F, 14-16 y	.90	2.5
	23 M & 18 F, 14-16 y	.87	3.9
Leger et al. (1988)	188 M & F, 8-19 y	.71	5.9
Liu et al. (1992)	22 M, 12-15 y	.65	5.3
	26 F, 12-15 y	.51	5.2
	48 M & F, 12-15 y	.69	5.5
	48 M & F, 12-15 y	.72[a]	5.3

Mahar et al. (2006)	135 M & F, 12-14 y	.65[d]	6.4
Mahar et al. (2011)	174 M & F 10-16 y	.75[e]	6.2
Matsuzaka et al. (2004)	132 M & F 8-17 y	.74[f]	5.5
Ruiz et al. (2008)	193 M & F 13-19 y	. 76[g]	5.3
van Mechelen et al. (1986)	41 M, 12-14 y	.68	4.0
	41 F, 12-14 y	.69	3.5
	82 M & F, 12-14 y	.76	4.4

[a] *Cross-validation of the Leger et al. (1988) equation*

[b] *Prediction from age, sex, and maximal shuttle speed*

[c] *Prediction from triceps skinfold, sex, and maximal shuttle speed*

[d] *Prediction from gender, body mass, and PACER laps*

[e] *Prediction from age, gender, age × gender, BMI, PACER laps, and PACER laps squared*

[f] *Prediction from age, gender, BM,I and PACER speed*

[g] *Prediction from age, gender, weight, height, and PACER stage*

Bibliography

Adegboye, R.A., Anderssen, S.A., Froberg, K., Sardina, L. B., Heitmann, B.L., Steene-Johannessen, J., Kolle, E., & Andersen, L.B. (2011). Recommended aerobic fitness level for metabolic health in children and adolescents: a study of diagnostic accuracy. *British Journal of Sports Medicine*, *45*, 722-728.

Armstrong, N., Williams, J., & Ringham, D. (1988). Peak oxygen uptake and progressive shuttle run performance in boys aged 11-14 years. *British Journal of Physical Education, 19 (Suppl. 4)*, 10-11.

Astrand, P.O. (1952). Experimental studies of physical working capacity in relation to sex and age. Copenhagen: Ejnar Munksgaard.

Astrand, P.O., Rodahl, K., Dahl, H., & Stromme, S. (2003). *Textbook of work physiology*. New York: McGraw-Hill.

Balke, B. (1963). *A simple field test for the assessment of physical fitness* (Publication No. 63-6). Oklahoma City: Federal Aviation Agency, Aeromedical Research Institute.

Barnett, A., Chan, L.Y.S., & Bruce, I.C. (1993). A preliminary study of the 20-m multistage shuttle run as a predictor of peak VO_2 in Hong Kong Chinese students. *Pediatric Exercise Science*, *5*, 42-50.

Baumgartner, T.A., & Jackson, A.S. (1991). *Measurement for evaluation in physical education and exercise science*. Dubuque: Wm. C. Brown.

Beets, M.W., & Pitetti, K.H. (2006). Criterion-referenced reliability and equivalency between the PACER and 1-mile run/walk for high school students. *Journal of Physical Activity and Health*, *3* (Suppl. 2), S17-S29.

Blair, S.N., Kohl, H.W., III, Paffenbarger, R.S., Jr., Clark, D.G., Cooper, K.H., & Gibbons, L.W. (1989). Physical fitness and all-cause mortality: A prospective study of healthy men and women. *Journal of the American Medical Association*, *262*, 2395-2401.

Boileau, R.A., Bonen, A., Heyward, V.H., & Massey, B.H. (1977). Maximum aerobic capacity on the treadmill and bicycle ergometer of boys 11-14 years of age. *Journal of Sports Medicine*, *17*, 153-162.

Boiarskaia, E.A., Boscolo, M.S., Zhu, W., & Mahar, M.T. (2011). Cross-validation of an equating method linking aerobic FITNESSGRAM® field tests. *American Journal of Preventive Medicine*, *41*, S124-S130.

Bono, M.J., Roby, J.J., Micale, F.G., Sallis, J.F., & Shepard, W.E. (1991). Validity and reliability of predicting maximum oxygen uptake via field tests in children and adolescents. *Pediatric Exercise Science, 3*, 250-255.

Boreham, C.A.G., Paliczka, V.J., & Nichols, A.K. (1990). A comparison of the PWC_{170} and 20-MST tests of aerobic fitness in adolescent schoolchildren. *Journal of Sports Medicine and Physical Fitness, 30*, 19-23.

Bouchard, C., An, P., Rice, T., Skinner, J.S., Wilmore, J.H., … Rao, D.C. (1999). *Journal of Applied Physiology*, *87*, 1003-1008.

Bouchard, C., Dionne, F.T., Simoneau, J.-A., & Boulay, M.R. (1992). Genetics of aerobic and anaerobic performances. In J. O. Holloszy (Ed.), *Exercise and sport sciences reviews*, Vol. 20 (pp. 27-58). Baltimore: Williams and Wilkins.

Brage, S., Wareham, N.J., Wedderhopp, N., Andersen, L.B., Ekelund, U., Froberg, K., & Franks, P.W. (2004). Features of the metabolic syndrome are associated with objectively measured physical activity and fitness in Danish children. *Diabetes Care*, *27*, 2004.

Buskirk, E.R., & Taylor, H.L. (1957). Maximal oxygen intake and its relation to body composition, with special reference to chronic physical activity and obesity. *Journal of Applied Physiology*, *11*, 72-78.

Chun, D.M., Corbin, C.B., & Pangrazi, R.P. (2000). Validation of criterion-referenced standards for the mile run and progressive aerobic cardiovascular endurance tests. *Research Quarterly for Exercise and Sport, 71*, 125-134.

CIAR. (1987). *FITNESSGRAM® test administration manual.* Dallas: The Cooper Institute for Aerobics Research.

CIAR. (1992). *The Prudential FITNESSGRAM® test administration manual*. Dallas: The Cooper Institute for Aerobics Research.

Cunningham, D.A., van Waterschoot, B.M., Paterson, D.H., Lefcoe, M., & Sangal, S.P. (1977). Reliability and reproducibility of maximal oxygen uptake measurement in children. *Medicine and Science in Sports and Exercise*, *9,* 104-108.

Cureton, K. (1982). Distance running performance tests in children: What do they mean? *Journal of Physical Education, Recreation and Dance*, *53*, 64-66.

Cureton, K.J. (1976). *Determinants of running and walking endurance performance in children: Analysis of a path model*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign, Champaign, IL.

Cureton, K.J., Baumgartner, T.A., & McManis, B. (1991). Adjustment of 1-mile run/walk test scores for skinfold thickness in youth. *Pediatric Exercise Science*, *3,* 152-167.

Cureton, K.J., Boileau, R.A., Lohman, T.G., & Misner, J.E. (1977). Determinants of distance running performance in children: Analysis of a path model. *Research Quarterly*, *48,* 270-279.

Cureton, K.J., Sloniger, M.A., O'Bannon, J.P., Black, D.N. & McCormack, W.P. (1995). A generalized equation for prediction of VO_2 peak from one-mile run/walk performance in youth. *Medicine and Science in Sports and Exercise*, *27*, 445-451.

Cureton, K.J., Sparling, P.B., Evans, B.W., Johnson, S.M., Kong, U.D., & Purvis, J.W. (1978). Effect of experimental alterations in excess weight on aerobic capacity and distance running performance. *Medicine and Science in Sports*, *10*, 194-199.

Cureton, K.J., & Warren, G.L. (1990). Criterion-referenced standards for youth health-related fitness tests: A tutorial. *Research Quarterly for Exercise and Sport*, *61*, 7-19.

Dinschel, K.M. (1994). *Influence of agility on the mile run and PACER tests of aerobic endurance in fourth and fifth grade school children.* Unpublished master's thesis, Northern Illinois University, DeKalb, IL.

Disch, J., Frankiewicz, R., & Jackson, A. (1975). Construct validation of distance run tests. *Research Quarterly*, *46,* 169-176.

Grundy, S.M., Cleeman, J.I., Daniels, S.R., Donato,K.A., Eckel,R.H., Franklin, B.A. Costa, F. (2005). Diagnosis and management of the metabolic syndrome. *Circulation, 112*, 2735-2752.

Ihaz, F., Finn, K.J., Meszaros, J., & Zsidegh, M. (2006). Does a modified progressive aerobic cardiovascular endurance run test protocol benefit overweight children? *Research Quarterly for Exercise and Sport*, *77* (Suppl.), A19.

Jackson, A.S., & Coleman, A.E. (1976). Validation of distance run tests for elementary school children. *Research Quarterly*, *47*, 86-94.

Kline, G.M., Porcari, J.P., Hintermeister, R., Freedson, P.S., Ward, A., McCarron, R.F., Ross, J., & Rippe, J.M. (1987). Estimation of VO_2max from a one-mile walk, gender, age, and body weight. *Medicine and Science in Sports and Exercise*, *19*, 253-259.

Krahenbuhl, G.S., Morgan, D.W., & Pangrazi, R.P. (1989). Longitudinal changes in distance running performance of young males. *International Journal of Sports Medicine*, *10*, 92-96.

Krahenbuhl, G.S., Pangrazi, R.P., Burkett, L.N., Schneider, M.J., & Petersen, G.W. (1977). Field estimation of VO_2max in children eight years of age. *Medicine and Science in Sports*, *9*, 37-40.

Krahenbuhl, G.S., Pangrazi, R.P., Petersen, G.W., Burkett, L.N., & Schneider, M.J. (1978). Field testing of cardiorespiratory fitness in primary school children. *Medicine and Science in Sports and Exercise*, *10*, 208-213.

Krahenbuhl, G.S., Skinner, J.S., & Kohrt, W.M. (1985). Developmental aspects of maximal aerobic power in children. In R. L. Terjung (Ed.), *Exercise and sport science reviews, Vol. 13* (pp. 503-538). New York: MacMillan.

LaMonte, M.J., & Blair, S.N. (2006). Physical activity, cardiorespiratory fitness, and adiposity. *Current Opinion in Clinical Nutrition and Metabolic Care*, *9*, 540-546.

Leger, L., & Gadoury, C. (1989). Validity of the 20 m shuttle run test with 1 min stages to predict VO_2max in adults. *Canadian Journal of Sport Sciences*, *14*, 21-26.

Leger, L.A., & Lambert, J. (1982). A maximal multistage 20-m shuttle run test to predict VO_2max. *European Journal of Applied Physiology and Occupational Physiology*, *49*, 1-12.

Leger, L.A., Mercier, D., Gadoury, C., & Lambert, J. (1988). The multistage 20 metre shuttle run test for aerobic fitness. *Journal of Sports Sciences*, *6*, 93-101.

Liu, N.Y-S., Plowman, S.A., & Looney, M.A. (1992). The reliability and validity of the 20-meter shuttle test in American students 12 to 15 years old. *Research Quarterly for Exercise and Sport*, *63*, 360-365.

Lloyd, L.K., Bishop, P.A., Walker, J.L., Sharp, K.R., & Richardson, M.T. (2003). The influence of body size and composition on FITNESSGRAM® test performance and adjustment of the FITNESSGRAM® test scores for skinfold thickness in youth. *Measurement in Physical Education and Exercise Science*, *7*, 205-226.

Lobello, F., Pate, R.R., Dowda, M., Liese, A.D., & Ruiz, J.R. (2009). Validity of cardiorespiratory fitness criterion-referenced standards for adolescents. *Medicine and Science in Sports and Exercise*, *41*, 1222-1229.

Looney, M.A., & Plowman, S.A. (1990). Passing rates of American children and youth on the FITNESSGRAM® criterion-referenced physical fitness standards. *Research Quarterly for Exercise and Sport*, *61*, 215-223.

Mahar, M.T., Guerieri, A.M., Hanna, M.S., & Kemble, C.D. (2011). Estimation of aerobic fitness from 20-m multistage shuttle run test performance. *American Journal of Preventive Medicine*, *41*, S117-S123.

Mahar, M.T., Rowe, D.A., Parker, C.R., Mahar, F.J., Dawson, D.M., & Holt, J.E. (1997). Criterion-referenced and norm-referenced agreement between the mile run/walk and PACER. *Measurement in Physical Education and Exercise Science*, 1, 245-258.

Mahar, M.T, Welk, G.J., Rowe, D.A., Crotts, D.J., & McIver, K.L. (2006). Development and validation of a regression model to estimate VO2peak from PACER 20-m shuttle run performance. *Journal of Physical Activity and Health*, *3*(Suppl. 2), S34- S46.

Mahar, M. T. et al. (2013). Estimation of aerobic fitness from PACER performance. Manuscript in preparation.

Matsuzaka, A., Takahashi, Y., Yamazoe, M., Kumakura, N. Ikeda, A., Wilk, B., … (2004). Validity of the multistage 20-m shuttle run test for Japanese children, adolescents and adults. *Pediatric Exercise Science*, *16*, 113-125.

McClain, J.J., Welk, G.J., Ihmels, M., & Schaben, J. (2006). Comparison of two versions of the PACER aerobic fitness test. *Journal of Physical Activity and Health*, *3*(Suppl. 2), S476-S57.

McCormack, W.P., Cureton, K.J., Bullock, T.A., & Weyand, P.G. (1991). Metabolic determinants of 1-mile run/walk performance in children. *Medicine and Science in Sports and Exercise*, *23*, 611-617.

McSwegin, P.J., Plowman, S.A., Wolff, G.M., & Guttenberg, G.L. (1998).
Measurement in Physical Education and Exercise Science, *2*, 47-63.

Mitchell, J.H., Sproule, B.J., & Chapman, C.B. (1958). The physiological meaning of the maximal oxygen intake test. *Journal of Clinical Investigation*, *37*, 538-547.

Morrow, J. R., Jr., Falls, H. B., Kohl, H. W., III. (Eds.). (1994). *The Prudential FITNESSGRAM® technical reference manual*. Dallas: The Cooper Institute for Aerobics Research.

Norman, A.-C., Drinkard, B., McDuffie, J.R., Ghorbani, S., Yanoff, L.B., & Yanovski. (2005). Influence of excess adiposity on exercise fitness and performance in overweight children and adolescents. *Pediatrics*, *115*, e690-e696.

Ortega, F.B., Ruiz, J.R., Castillo, M.J., & Sjostrom. (2008). Physical fitness in childhood and adolescence: A powerful marker of health. *International Journal of Obesity*, *32*, 1-11.

Paliczka, V.J., Nichols, A.K., & Boreham, A.G. (1987). A multi-stage shuttle run as a predictor of running performance and maximal oxygen uptake in adults. *British Journal of Sports Medicine*, *21*, 163-165.

Pate, R.R., & Ward, D.S. (1990). Endurance exercise trainability in children and youth.
In Grana, W. A. et al. (Eds.). *Advances in sports medicine and fitness*, Vol. 3. Chicago: Yearbook Medical Publishers.

Paterson, D.H., Cunningham, D.A., & Donner, A. (1981). The effect of different treadmill speeds on the variability of VO_2max in children. *European Journal of Applied Physiology and Occupational Physiology*, *47*, 113-122.

Pivarnik, R.M., Dwyer, M.C., & Lauderdale, M.A. (1996). The reliability of aerobic capacity (VO2max) testing in adolescent girls. *Research Quarterly for Exercise and Sport*, *67*, 345-348.

Plowman, S.A. & Liu, N.Y. (1999). Norm-referenced and criterion-referenced validity of the one-mile run and PACER in college age individuals. *Measurement in Physical Education and Exercise Science*, *3*, 63-84.

Prud'homme, D. Bouchard, C., LeBlanc, C., Landrey, F., & Fontaine, E. (1984). Sensitivity of maximal aerobic power to training is genotype-dependent. *Medicine and Science in Sports and Exercise*, *16*, 489-493.

Ramsbottom, R., Brewer, J., & Williams, C. (1988). A progressive shuttle run test to
estimate maximal oxygen uptake. *British Journal of Sports Medicine*, *22*, 141-144.

Rikli, R.E., Petray, C., & Baumgartner, T.A. (1992). The reliability of distance run tests for children in grades K-4. *Research Quarterly for Exercise and Sport*, *63*, 270-276.

Rowland, T., Kline, G., Goff, D., Martel, L., & Ferrone, L. (1999). One-mile run performance and cardiovascular fitness in children. *Archives of Pediatric and Adolescent Medicine*, *153*, 845-849.

Ruiz, J.R., Ortega, F.B., Rizzo, N. S. … (2007). High cardiovascular fitness is associated with low metabolic risk score in in children: the European Youth Heart Study. *Pediatric Research*, *61*, 350-355.

Ruiz, J.R., Ramirez-Lechuga, J., Ortega, F.B., Castro-Pinero, J., Binitez, J.M., Arauzo-Azofra, A. et al. (2008). Artificial neural network-based equation for estimating VO2max from the 20-m shuttle run test in adolescents. *Artificial Intelligence in Medicine*, *44*, 233-245.

Safrit, M.J. (1990). The validity and reliability of fitness tests for children: A review. *Pediatric Exercise Science*, *2*, 9-28.

Safrit, M.J., Hooper, L.M., Ehlert, S.A., Costa, M.G., & Patterson, P. (1988). The validity generalization of distance run tests. *Canadian Journal of Sport Sciences*, *13*, 188-196.

Saint-Maurice, P.F., Welk, G.J., Laurson, K., & Brown, D. (In press). Measurement agreement between estimates of aerobic fitness in youth: Emphasis on the impact of body mass index. *Research Quarterly for Exercise and Sport*.

Saltarelli, W.A., & Andres, F.F. (1993). Teaching steady pacing to students - a practical method. *Journal of Physical Education, Recreation and Dance*, *68*, 67-70.

Sloniger, M.A., Cureton, K.J., & O'Bannon, P.J. (1994). One-mile run/walk performance in young men and women: Role of anaerobic metabolism. *Canadian Journal of Applied Physiology*, *22*, 337-350, 1997.

Sparling, P.B., & Cureton, K.J. (1983). Biological determinants of the sex difference in12-min run performance. *Medicine and Science in Sports and Exercise*, *15*, 218-223.

Taylor, H.L., Buskirk, E., & Henschel, A. (1955). Maximal oxygen uptake as an objective measure of cardiorespiratory performance. *Journal of Applied Physiology*, *8*, 73-80.

van Mechelen, W., Hlobil, H., & Kemper, H.C.G. (1986). Validation of two running tests as estimates of maximal aerobic power in children. *European Journal of Applied Physiology and Occupational Physiology*, *55*, 503-506.

Welch, B.E., Reindeau, R.P., Crisp, C.E., & Isenstein, R.S. (1957). Relationship of maximal oxygen consumption to various components of body composition. *Journal of Applied Physiology*, *12*, 395-398.

Welk, G.J., Going, S. B., Morrow, J.R., & Meredith, M.D. (2011). Development of new criterion-referenced fitness standards in the FITNESSGRAM® program: Rationale and conceptual overview. *American Journal of Preventive Medicine*, *41*, S63-S67,

Welk, G.J., Laurson, K. R., Eisenmann, J.C., & Cureton, K.J. (2011). Development of youth aerobic capacity standards using receiver operating characteristic curves. *American Journal of Preventive Medicine*, *41*, S106-S110.

Welk, G.J., De Saint-Maurice Maduro, P.F, Laurson, K.R., Brown, D. (2011). Field evaluation of the new FITNESSGRAM® criterion-referenced standards. *American Journal of Preventive Medicine*, *41*, S131-S142.

Zhu, W. (1998). Test equating: What, why, how. *Research Quarterly for Exercise and Sport*, *69*, 11-23.

Zhu, W., Plowman, S. A., & Park, Y. (2010). A primer-centered equating method for setting cut-off scores. *Research Quarterly for Exercise and Sport*, *81*, 400-409.

Chapter 7
Body Composition Assessments

Scott B. Going, Timothy G. Lohman, Joey C. Eisenmann

The FITNESSGRAM® Reference Guide is intended to provide answers to some common questions associated with use and interpretation of this chapter. Devoted to Body Composition Assessment, this chapter provides the rationale for including body composition assessments in the FITNESSGRAM® program and reviews the basis for the tests and standards that are used. The following questions are specifically addressed:

General Information about Body Composition

What Makes Up a Person's Body Composition?

Body composition refers to the components that make up one's total body weight. Approximately 50 elements combine to make up 100,000 chemical components, approximately 200 cell types, and 4 main tissues of the body. The major contributors to body weight are the fluid, muscle, bone, and fat content. Also included are organs, skin, and nerve tissue. Typically, in simplified body composition assessment models, all lean components (e.g., fluid, muscle and bone) are combined into what is called the fat-free mass (FFM). FFM accounts for about 80-85% of body weight on average in boys and 70-85% in girls, depending on age. Average fat content is 15-20% in boys and 15-30% in girls (Laurson, Eisenmann, & Welk, 2011a).

Why Is Body Composition Important?

Body composition is a critical component of one's ability to perform functional activities and also one's health. Skeletal muscle, the major component of FFM along with fluid, provides the propulsive force for movement and accounts for much of total daily energy expenditure. Bones are the supporting framework and provide protection for vital organs. Fluids are the medium for transport of oxygen, nutrients, and other vital chemicals and metabolites. Adipose tissue serves a vital energy storage role and secretes a variety of products that are essential for regulation of energy balance and other tissues.

Why Is Body Composition an Essential Part of Health Related Fitness Assessment?

Research has shown that excessive fatness (i.e., obesity) is associated with higher levels of cardiovascular disease risk factors (e.g., blood pressure and blood lipids) (Going et al., 2011; Laurson, Eisenmann, & Welk, 2011c; Williams et al., 1992) and risk of Type 2 diabetes in children and adolescents, as well as adults (Aristimuno, Foster, Voors, Srinivasan, & Berenson, 1984; Berenson, McMahon, & Voors, 1980; Berenson et al., 1982). Furthermore "tracking" studies that follow youth over time show a relationship between childhood and adult obesity with the relationship being stronger as children become adolescents. Together these studies indicate that excess body fatness in children and youth increase the likelihood of obesity and obesity-related adult diseases including coronary heart disease, hypertension, hyperlipidemia, and Type 2 diabetes.

What Changes Occur in Body Composition During Childhood and Adolescence?

Both total muscle and fat mass increase during childhood. During adolescence, boys continue to increase muscle mass, whereas in girls the increase in muscle mass slows significantly and plateaus at about age 15. The increase in total body fat is greater in girls than boys. In terms of percent body fat, the patterns are quite different between boys and girls. In girls, the percent of body fat remains relatively steady into mid-childhood and then begins to increase during late-childhood and throughout adolescence. In boys, there is a pre-pubertal "blip" increase and then percent body fat actually declines during puberty due to the rapid increase in muscle mass.

What Is the Latest Estimate of Obesity in Children?

Obesity has increased dramatically in both children and adults in the past twenty years. It has reached alarming levels and has not spared any region of the United States, age group, or ethnic group. However, the prevalence varies depending on age, ethnicity, and geographic region. According to recent national surveys 78 million U.S. adults (more than one-third) are obese. Among children and adolescents aged 2-19 years, more than 5 million girls and ~7 million boys are obese, which is almost 1 out of 5 boys and girls (Ogden, Carroll, Kit, & Flegal, 2012).

Over the past three decades, the childhood obesity rate has more than doubled for preschool children aged 2-5 years and adolescents aged 12-19 years, and it has more than tripled for children aged 6-11 years. Given the relationship between childhood and adult obesity, these statistics predict a disturbing trend for future greater levels of adult obesity unless effective treatment and prevention programs are developed.

What Is the Gold Standard for Body Composition?

Most body composition methods, whether lab- or field-based, have errors of 2.5% to 4.0% for estimation of body fatness. The laboratory approach that combines body density, total body water, and total bone mineral (called a multicomponent approach) is the most accurate with an error of ~2%. Densitometry (e.g., underwater weighing and air displacement plethysmography) and dual energy x-ray absorptiometry (DXA) have errors of 2.5% to 3.0% for estimating fatness. Skinfolds and bioelectric impedance analysis (BIA) have errors of 3 to 4% fat, and BMI estimates fatness with an error of >5%.

What Field Methods Are Available in FITNESSGRAM®?

FITNESSGRAM® uses percent fat from skinfolds and BIA as the preferred field methods to estimate body fatness. Measurement of two skinfolds (triceps plus calf) can be successfully used to estimate percent fat in children of all ages. Skinfolds have proved to be one of the most effective field methods for estimating body fatness, with standard errors of estimate of 3 to 4% body fat (in the hands of a well-trained technician). The errors associated with BIA are similar to skinfolds assuming the participant is normally hydrated. A second method, based on height and weight, called body mass index (BMI), is also available for being a proxy for body fatness; however, the prediction error is considerably larger (5-6%) and therefore this approach is not as effective for estimating body fat (Going & Lohman, 1998).

Why Does FITNESSGRAM® Recommend the Use of Percent Body Fat Rather than BMI?

Excess fat, rather than weight, increases the risk of chronic disease. Weight for height indices like the body mass index (BMI) provide information about the amounts of various tissues that together make up body weight. Nevertheless, because BMI is correlated with percent fat, it is used as a surrogate measure of body composition.

Percent Body Fat Measured by Skinfold Assessments

How Valid and Reliable Are Skinfold Assessments?

Skinfolds are reliable (give similar results with repeated measures) measures of body composition, providing the teacher or nurse has sufficient training and experience in the skinfold measurement approach and has followed the standardized protocols for triceps and medial calf skinfold measurements.

What Factors Improve the Reliability and Validity of Skinfolds?

The best way to obtain reliable and valid skinfolds is to train with an expert or with a videotape demonstration. The average skinfolds for 6 to 10 subjects should agree within 15% of the expert for each skinfold site if the training is effective, and no individual difference should be larger than 20%.

Are There Differences in the Quality and Accuracy of Skinfold Calipers?

There are several skinfold calipers available. Studies have shown general agreement between commonly used Harpenden, Lange, and Lafayette calipers, designed for research, and FITNESSGRAM® and Ross calipers designed for field testing. A one to two millimeters difference between calipers for a single site is typical; however, if you are not familiar with a particular caliper the difference can be greater. Therefore, whatever caliper you use, you should practice measures on 30 or more children before conducting evaluations that will be reported to children and their parents. FITNESSGRAM® recommends using any of the following calipers. Contact information for companies that sell skinfold calipers is contained in Table 1.

Table 1. Contact Information for Companies that Sell Skinfold Calipers

Accu-Measure Fitness 3000 Plastic Skinfold Caliper ($19.99)

www.accumeasurefitness.com
Accu-Measure, LLC
P.O. Box 4411 • Greenwood Village, CO 80155-4411
Phone: 303-799-4721 • Fax: 303-799-4778
Toll-free: 800-866-2727
Information E-mail: info@accufitness.com

Also found online at:

General Nutrition Center (GNC):
http://www.gnc.com/product/index.jsp?productId=2134356
Amazon.com: http://www.amazon.com/Accu-Measure-Fitness-3000-Personal-Tester/dp/B000G7YW74

Harpenden Skinfold Caliper ($359.00-$379.00)

http://www.harpenden-skinfold.com/
USA Distributor
Mediflex Surgical Products
250 Gibbs Road
Islandia, NY 11749
Tel: 631-582-8440
Fax 631-582-8487
Email: sales@mediflex.com
Website: www.mediflex.com

Also found online at:

Healthcheck Systems:
http://www.healthchecksystems.com/harpenden_skinfold_calipers.htm
Amazon.com: http://www.amazon.com/Harpenden-Skinfold-Caliper-With-Software/dp/B000BK30W4

Lafayette Skinfold Calipers ($100.00-$300.00)

http://www.lafayetteevaluation.com
LAFAYETTE INSTRUMENT, WORLDWIDE OFFICE
PO Box 5729
Lafayette, IN 47903 USA
Phone: (765) 423-1505
US Toll Free: (800) 428-7545
Fax: (765) 423-4111
sales@lafayetteinstrument.com
info@lafayetteinstrument.com

Also found online at:

Amazon.com: http://www.amazon.com/Lafayette-Instrument-Skinfold-Caliper-II/dp/B007G4S6L8
Medco Sports Medicine: https://www.medco-athletics.com/Supply/Product.asp?Leaf_Id=260961

Lange Skinfold Calipers ($200.00-$300.00)

http://www.beta-technology.com
Beta Technology
2841 Mission St., Santa Cruz CA 95060 USA
Customer Service: (831) 426-0882. Toll Free in USA: (800) 858-2382 Fax: (831) 423-4573
Technical Support: (262) 631-4460 or (262) 631-4461.Toll Free in USA: (800) 468-4893
Fax: (410) 943-1545

Also found online at:

Amazon.com: http://www.amazon.com/Lange-Skinfold-Caliper-Includes-Deluxe/dp/B000PC667E
Quick Medical: http://www.quickmedical.com/calipers/lange_skinfold.html
Power Systems: http://www.power-systems.com/p-2711-lange-skinfold-caliper-with-case.aspx

Slim Guide Skinfold Caliper ($15.00-$40.00)

http://www.linear-software.com
Linear Software
info@linear-software.com
support@linear-software.com
sales@linear-software.com

Also found online at:

Amazon.com: http://www.amazon.com/Creative-Health-6575XXXX-Skinfold-Caliper/dp/B000NN9SDO
Healthcheck Systems:
http://www.healthchecksystems.com/product/index.cfm?product_id=3462

How Much Training Is Recommended to Perform Skinfolds?

Research has found that training by an expert or through an audiovisual tape (Human Kinetics, Champaign, IL) is essential to measure skinfolds accurately. Training for the triceps and calf skinfolds can be done in a one hour workshop. Practice on 20 to 30 subjects is recommended, with feedback from an expert on your measurement techniques. To certify for accurate skinfold measurements you should measure the same skinfolds as the expert on 6 to 10 children or adults. Your agreement should be within the limits specified above.

Percent Body Fat Measured with Bioelectrical Impedance Analysis

How Does Bioelectric Impedance Analysis Work?

Bioelectric impedance analysis (BIA) is based upon the physical principles of Ohm's Law. BIA is a function of the resistance (pure opposition to current flow) and reactance (opposition caused by capacitance produced by the cell membrane) to the flow of a low-level electric current passed through the body. Total body water can be estimated from impedance because the electrolytes in the body's fluids are excellent conductors of electrical current. When the volume of water is large, the current flows more easily through the body with less resistance. The resistance is greater in individuals with large amounts of body fat since adipose tissue is a poor conductor because of its low water content. Because of the high water fraction of lean tissue (~73%-74% in adults), impedance is useful for estimating fat-free mass and, by difference, fat mass (body weight – fat-free mass).

How Reliable and Valid Are Measurements Done with Bioelectric Impedance Devices?

Resistance, reactance and impedance can be measured with excellent precision and the results are reproducible. Bioelectrical Impedance is used to estimate body composition (e.g., fat-free mass and percent fat) using equations that relate resistance and reactance to the body composition component of interest. The reliability of impedance measures may be better than skinfolds. The validity and accuracy for estimating percent fat and FFM depend on the validity of the assumptions on which the equations are based. Under appropriate conditions (namely hydration status of the participant), the validity and accuracy of BIA is similar to the skinfold method.

What Are the Issues Associated with Using Bioelectric Impedance Devices?

The volume of the body's FFM or TBW is estimated indirectly from BIA, which requires certain basic assumptions to be made, for example, that the body is a cylinder with a uniform length and cross-sectional area, and that the impedance to the current flow is directly related to cross-sectional area. Because the body segments (trunk, arms, legs) are not uniform in length or cross-sectional area, resistance to the flow of current through these segments will differ. Thus, application of the equation relating impedance to area and length introduces error because of the complex geometric shape of the body. Nevertheless, other sources of error are of more concern, e.g., differences between instruments, subject factors, technician skill, environmental factors, and the equation used to estimate FFM and percent fat. Hand held devices and BIA scales do not necessarily give the same results since the path taken by the current will vary ("path of least resistance") and they will likely give different results compared to the tetrapolar (electrodes on both hands and feet) method. Factors such as eating, drinking, dehydrating, and exercising alter an individual's hydration status, an important source of error, since most equations assume normal hydration. Also, cool temperatures that decrease skin temperature cause increased resistance and a decreased estimate of FFM and higher percent fat. The prediction equation used to estimate FFM and percent fat can be a major source of prediction error if it is inappropriate for the individual being measured. It is important to follow the manufacturer's recommendation for testing procedures. BIA prediction equations should be selected based on the age, gender, and race/ethnicity of the individual being measured.

Are There Differences Between Bioelectric Impedance Analyzers?

Although there is a high correlation between resistance values measured by different analyzers, differences exist in the measured resistances as well as the estimates of percent fat and fat-free mass (Heyward & Wagner, 2004). The equation used to estimate percent fat and fat-free mass is an important source of difference across manufacturers' instruments. To control these potential differences, it is important to use the same manufacturers' instrument, and even the same instrument if possible, with all students who are being measured, especially if a goal is to monitor changes in body composition over time. Choose an instrument with equations that have been developed for the population of interest (e.g., boys, girls, adolescents, etc.).

Contact information for companies that sell bioelectric impedance analyzers:

Tanita Corporation of America, Inc.
2625 South Clearbrook Drive
Arlington Heights, Illinois 60005, USA
Phone: (847) 640-9241
Fax: (847) 640-9261
eMail: 4health@tanita.com
Web: http://www.tanita.com
Models: BF-689 Children's Body Fat Monitor ($90.00)
can be used with children ages 5-17 years
Found online at:

Amazon.com: http://www.amazon.com/Tanita-BF-689-Body-Monitor-Children/dp/B0057IO0O2

BF-2000 IronKids Radio Wireless Body Fat Monitor ($180.00)
FDA cleared body fat measurements for ages 5-17 years
Found online at:

Amazon.com: http://www.amazon.com/BF-2000-IronKids-Wireless-Body-Monitor/dp/B0057IKZP0

OMRON Healthcare Co., Ltd.
United States
Omron Healthcare, Inc.
1925 W. Field Court
Lake Forest, IL 60045
Consumer Support: 877-216-1333
Phone: 847-680-6200
Media Inquiries: 847-247-5637

Fax: 847-680-6269
http://www.omronhealthcare.com
Models: HBF 510 Body Composition Monitor with scale ($80.00)
can be used with children aged 10 and up
Found online at:

Amazon.com: http://www.amazon.com/Omron-HBF-510W-Composition-Monitor-Scale/dp/B001IV61J4

BF 516 Body Composition Monitor with scale ($130.00)
can be used with children ages 6 and up
Found online at:

Amazon.com: http://www.amazon.com/Omron-Body-Composition-Monitor-Scale/dp/B001803OS6

Stayhealthy, Inc.
724 East Huntington Drive
Suite A-D
Monrovia, California 91016
Tel. (626) 256-6152
Email: inquiries@stayhealthy.com
Model: BC3 Body Composition Analyzer ($140.00)
can be used with children aged 10 and up

Body Composition Standards

How Were FITNESSGRAM® Standards Developed for Body Composition?

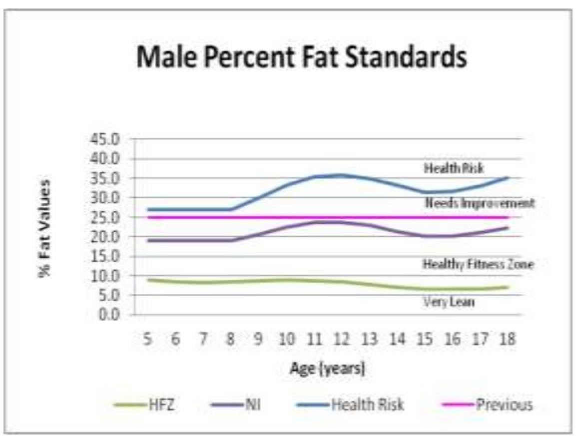
As with the aerobic capacity, FITNESSGRAM® body composition standards are based on a criterion-referenced, health-related approach. The procedures used in developing the FITNESSGRAM® body composition standards have been described in detail (Laurson, Eisenman, Welk, 2011a; Laurson, Eisenmann, & Welk, 2011b). The standards were designed to indicate the level of percent body fat and then body mass index associated with increased risk of the metabolic syndrome in youth. The metabolic syndrome is a cluster of adverse cardio-metabolic risk factors including: elevated abdominal obesity indicated by waist girth, poor control of blood glucose and insulin, disordered blood lipids, and high blood pressure that increase the risk of cardiovascular disease and diabetes. The approach taken by Laurson and colleagues used nationally representative data from the National Health and Nutrition Examination Survey (NHANES) collected between 1999 and 2004. First age and gender-specific percent body fat growth curves were established. Then age-and gender-specific thresholds for diagnosis of metabolic syndrome were determined using a statistical procedure called Receiver Operating Characteristic Analysis (Laurson, et al., 2011a). Two thresholds were identified; the first was the level of body fat that best identified youth with metabolic syndrome, and the second was the level of body fat that best identified those youth without metabolic syndrome combined with the fewest misclassifications in each case. These two thresholds allowed for the identification of three separate zones, a healthy fitness zone and two where improvement is needed. The advantage of three zones over two is that it provides a more prescriptive message about the youngster's body composition level. The following standards were established for percent body fat.

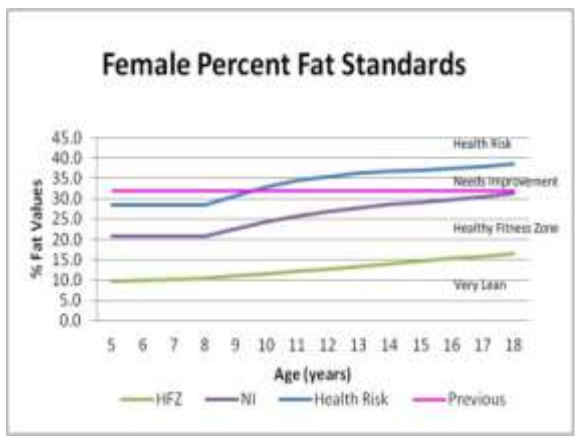
The “***Healthy Fitness Zone (HFZ)***” was established by emphasizing sensitivity (Se; percentage of children with metabolic syndrome who are correctly identified by high percent fat values as having the condition) over specificity. A sensitivity threshold of ≥90 was selected as the low risk (HFZ) value, indicating that ≥90% of the youth with metabolic syndrome have a percent body fat above this level. These percent body fat values (age and gender specific) represent the top of the HFZ. Individuals equal to or below these percent body fat values should have a very low risk of metabolic syndrome.

The “***Needs Improvement–Health Risk (NI-HR)***”zone was established by emphasizing specificity (Sp; percentage of youth with acceptable percent body fat values who are correctly identified as not having the metabolic syndrome) over sensitivity. The percent body fat values with a Sp of ≥ 90 (indicating that ≥90% of youth without metabolic syndrome had a percent body fat below these values) that still maintained the highest possible Se were selected. Individuals above these age and gender specific threshold percent fat values are likely to exhibit unfavorable metabolic profiles.

The “***Needs Improvement (NI)***” zone is an intermediate zone that marks the transition between the HFZ and the NI-HR zones. Since at least 90% of youth with metabolic syndrome have a percent body fat higher than the HFZ and 90% of the youth without metabolic syndrome have a percent fat lower than the NI-HR, this NI-SR zone comprises a mix of youth with and without the syndrome and carries a moderate risk of the condition.

FITNESSGRAM® body composition standards also includes a category called “Very Lean”. The Very Lean zone has not been evaluated against Metabolic Syndrome since it is excess fatness that increases risk of metabolic syndrome. The Very Lean zone was set to be equivalent to the age-and gender-specific 5th percentile of BMI, which is the accepted definition of underweight and an indication of possible under-nutrition and the potential for impaired growth.





Standards were also established for Body Mass Index (BMI). Initially, Receiver Operating Characteristic analysis was used to determine levels of BMI that best corresponded with the percent fat thresholds and that would classify boys and girls into the same fitness zone as would be achieved based on their percent body fat (Laurson et al., 2011b). The resulting BMI thresholds were similar to existing CDC BMI standards endorsed by the American Academy of Pediatrics. Further analysis showed very little difference in classification into fitness zones between the FITNESSGRAM® BMI standards and the CDC BMI standards. Consequently, FITNESSGRAM® adopted the CDC BMI thresholds to align with existing standards and avoid the potential confusion caused by competing standards.

Why Is BMI Used as an Alternative Within FITNESSGRAM®?

BMI is offered as an alternative because teachers may not be trained to measure skinfolds and in some school districts there may be regulations limiting skinfold measurements. Also, schools may not have an appropriate BIA device. Body mass index is fairly well correlated with percent body fat and use of BMI, although a surrogate for body composition, does yield useful information for body composition estimation in children if the standards are used as presented.

Does BMI Provide a Better Index than Weight Charts for Children?

Growth charts have been published by the Center for Disease Control and Prevention (CDC) for body mass index (BMI) in boys and girls, 2 to 20 years of age. These charts are percentiles showing the distribution of BMI at a given age and can be used to identify children who are overweight (BMI >85th percentile) or obese (BMI >95th percentile). BMI, the ratio of weight over height expressed as kg per meter squared, is a better indicator of fatness than weight tables alone, which gives no indication of body composition. The BMI growth charts offer an improvement over the weight tables. The FITNESSGRAM® BMI standards were derived from the percent fat standards and have been shown to discriminate boys and girls with metabolic syndrome from boys and girls who do not have metabolic syndrome (Laurson, Eisenmann, & Welk, 2011b).

Other Issues with Body Composition Assessments

How Should Body Composition Results Be Interpreted?

Children and especially adolescents who remain above the recommended ranges for body fat are at greater risk to remain overfat as an adult and consequently at greater risk of chronic diseases such as higher blood pressure, a poorer lipid profile, cardiovascular disease and Type 2 diabetes. It is important to recognize that BMI is a less accurate indicator of body fatness than skinfolds and BIA. Children and adolescents who fall 1-2 units above their respective standard may not be overfat. Instead, they could have more muscle and bone weight and thus be heavier for their height because of higher than average lean mass, not excess fat. For these children and youth, it would be appropriate to estimate fatness using skinfolds or BIA.

Will Body Composition Testing Increase Risks for Eating Disorders?

There has been concern by some teachers and parents that skinfold testing will make a child overly focused on their body weight and lead to eating disorders. There is no empirical evidence to suggest that this is likely to happen. In fact, lack of awareness and the lack of appropriate perceptions of body image are probably far greater contributors to the development of eating disorders. Body composition testing offers an opportunity for teachers to discuss with students perceptions and cultural obsessions with thinness, unrealistic expectations, and misleading body images portrayed in media that prevail in our society. The teacher can set a tone of acceptance of different body types and the importance of genetic contribution to body composition, including body shape and body weight. With greater tolerance for variation in fitness levels, children can better determine the relation of their body composition to health without fear of ridicule. Avoiding the assessment of body composition does nothing to address the cultural norms to be thin or the tendency for many children and adolescents to gain excess weight and fatness as they become adults.

In recent years an increase in eating disorders, including binging and bulimia, has occurred in adolescents and young adults, especially females, while the prevalence of anorexia nervosa has been relatively stable (Fairburn & Wilson, 1993). This increase is seen in England, New Zealand, and the U.S. populations where studies have been reported. In the national school-based Youth Risk Behavior Survey (1992), a high prevalence of body weight dissatisfaction was found, especially in the high school female population, although concern about eating shows up in elementary school surveys as well. In the study by Mellin, Irwin, and Scully (Mellin, Irwin, & Scully, 1992), for example, dieting, fear of fatness, and binge eating were reported by 31 to 46% of 9-year-old and 46 to 81% of 10-year old girls. In addition, 58% of girls perceived themselves to be overweight. Using FITNESSGRAM® standards can help young children set realistic standards for their body fatness and avoid the overemphasis on leanness that is prevalent in our culture.

Body composition testing (like any fitness assessment) is a personal matter. Because body composition is a particularly sensitive issue, additional care should be taken to ensure that it is conducted in a setting in which the child feels safe, accepted, and his/her privacy is respected. Assessments should be done by a trained professional (e.g., PE teacher, school nurse, health educator) in a private setting and only the measurer, child, and parent should be privy to the result. In a P.E. setting, measurements should be made behind a screen to maintain privacy.

What Are Some Tools and Resources to Use in Developing Education Programs About Body Composition?

Heyward VH, Wagner DR. *Applied Body Composition Assessment*, Second Edition. Champaign, IL: Human Kinetics; 2004.

American Alliance of Health PE, Recreation, and Dance. *Physical best activity guide. Preparing for a lifetime of fitness through physical education*. Champagne, IL: Human Kinetics; 1999.

Meredith, M.D., and Welk, G.J., editors. *FITNESSGRAM® & ACTIVITYGRAM® Test Administration Manual*. Updated Fourth Edition. Dallas, TX: The Cooper Institute; 2010.

Houtkooper, L.B. and Going, S.B. Body composition: How should it be measured? Does it affect performance? *Sports Science Exchange*, 7(#52), 1994.

Bibliography

Aristimuno, G. G., Foster, T. A., Voors, A. W., Srinivasan, S. R., & Berenson, G. S. (1984). Influence of persistent obesity in children on cardiovascular risk factors: the Bogalusa Heart Study. *Circulation*, 69(5), 895-904.

Berenson, G. S., McMahon, C. A., & Voors, A. W. (1980). *Cardiovascular risk factors in children: The early natural history of athersclerosis and essential hypertension*. New York: Oxford University Press.

Berenson, G. S., Webber, L. S., Srinivasan, S. R., Voors, A. W., Harska, D. W., & Dalferes, E. R. (1982). Biochemical and anthropometric determinants of serum B- and pre-B-lipoproteins in children: The Bogalusa Heart Study. *Arteriosclerosis*, 2, 325-334.

Fairburn, C. G., & Wilson, G. T. (1993). *Binge Eating, Nature, Assessment and Treatment*. New York: The Guilford Press.

Going, S. B., & Lohman, T. G. (1998). Assessment of body composition and energy balance. In D. R. Lamb & R. Murray (Eds.), *Perspectives in Exercise Science and Sports Medicine* (Vol. 11). Carmel, IN: Cooper Publishing Group.

Going, S. B., Lohman, T. G., Cussler, E. C., Williams, D. P., Morrison, J. A., & Horn, P. S. (2011). Percent body fat and chronic disease risk factors in U.S. children and youth. *American Journal of Preventive Medicine*, 41(4 Suppl 2), S77-86. doi: 10.1016/j.amepre.2011.07.006

Heyward, V. H., & Wagner, D. R. (2004). *Applied Body Composition Assessment*. 2nd Edition. Champaign, IL: Human Kinetics.

Laurson, K. R., Eisenmann, J. C., & Welk, G. J. (2011a). Body fat percentile curves for U.S. children and adolescents. *American Journal of Preventive Medicine*, 41(4 Suppl 2), S87-92.

Laurson, K. R., Eisenmann, J. C., & Welk, G. J. (2011b). Body Mass Index standards based on agreement with health-related body fat. *American Journal of Preventive Medicine*, 41(4 Suppl 2), S100-105. doi: 10.1016/j.amepre.2011.07.004

Laurson, K. R., Eisenmann, J. C., & Welk, G. J. (2011c). Development of youth percent body fat standards using receiver operating characteristic curves. *American Journal of Preventive Medicine*, 41(4 Suppl 2), S93-99.

Mellin, L. M., Irwin, C. E., Jr., & Scully, S. (1992). Prevalence of disordered eating in girls: a survey of middle-class children. J Am Diet Assoc, 92(7), 851-853.

Ogden, C. L., Carroll, M. D., Kit, B. K., & Flegal, K. M. (2012). Prevalence of obesity in the United States, 2009-2010. *NCHS Data Brief*(82), 1-8.

Williams, D. P., Going, S. B., Lohman, T. G., Harsha, D. W., Srinivasan, S. R., Webber, L. S., & Berenson, G. S. (1992). Body fatness and risk for elevated blood pressure, total cholesterol, and serum lipoprotein ratios in children and adolescents. *American Journal of Public Health*, 82(3), 358-363.

Chapter 8
Muscular Strength, Endurance, and Flexibility Assessments

Sharon A. Plowman

The FITNESSGRAM® Reference Guide is intended to provide answers to some common questions associated with the use and interpretation of FITNESSGRAM® assessments. This chapter, devoted to Muscular Strength, Endurance, and Flexibility Assessments, presents the rationale, reliability, and validity for each item, as well as how the criterion referenced standards were set. The following questions are specifically addressed:

Why Is Muscular Fitness Important?

Proper functioning of the musculoskeletal system requires that muscles be able to exert force or torque (measured as strength), resist fatigue (measured as muscular endurance), and move freely through a full range of motion (measured as flexibility). The benefits of musculoskeletal fitness (sometimes called neuromuscular fitness) have long been acknowledged in terms of sport performance by individuals of all ages and for activities of daily living, maintenance of independent functionality, and prevention of falls in the elderly (Brill, Macera, Davis, Blair & Gordon, 2000; Kell, Bell & Quinney, 2001; Pizzigalli, Filippini, Ahmaidi, Jullien, & Rainoldi, 2011; Warburton, Gledhill & Quinney, 2001a, 2001b; Warburton Nicol & Bredin, 2006; Wolfe 2006). There is now increasing evidence for children/adolescents and adults that enhanced musculoskeletal fitness is associated with an improvement in overall health status and, conversely, a reduction of risk for chronic disease, disability (Payne, Gledhill, Katzmarzyk, Jamnik & Ferguson, 2000b; Warburton, et al., 2006; Westcott, 2012) and, in adults, mortality. Mortality rates have been found to be lower in adult males and females with moderate to high muscular fitness (primarily measured by grip strength, sit-ups, leg and bench press) compared to individuals with low muscular fitness, even after adjusting for cardiorespiratory fitness, body composition, and other potentially confounding variables (FitzGerald, et al., 2004; Gale, Martyn, Cooper & Sayer, 2007; Katzmarzyk & Craig, 2002; Rantanen, et al., 2003; Ruiz, et al., 2008; Sasaki, Kasagi, Yamada & Fujita, 2007).

High levels of muscular strength and muscular endurance and/or resistance training improvements positively impact or predict long term changes in body composition (Hasselstrøm, Hansen, Froberg, & Andersen, 2002; Mason, Brien, Craig, Gauvin & Karzmarzyk, 2009; Ruiz, et al., 2009; Twisk, Kemper, & van Mechelen, 2000; Warburton, et al, 2001b; Warburton, et al., 2006), some cardiovascular risk factors (Artero, et al.,2012; Barnekow-Bergkvist, Hedberg, Janlert, & Jansson, 2001; Garcia-Artero, et al., 2007; Janz, Dawson, & Mahoney, 2002; Magnussen, Schmidt, Dwyer & Venn, 2012; Martinez-Gomez, et al., 2012; Olson, Dengel, Leon, & Schmitz, 2007; Ortega, Ruiz, Castillo, & Sjöström, 2008 b; Ruiz, et al., 2008; Warburton, et al., 2001a) and bone health (Boreham, & McKay, 2011; Warburton, et al., 2001a, 2001b).

Indeed, the optimal prevention strategy for osteoporosis as an adult is the attainment of a strong, dense skeleton during the growing years. Despite a large (~70-85%) genetic contribution to bone mass, resistance and high impact exercise can contribute an additional 5-15% to bone formation (Boreham & Riddoch, 2001; Faigenbaum, et al., 2009). The positive effects of high impact activity loading are most evident during the prepubertal and early pubertal years (Gunter, Almstedt, & Jantz, 2012) and gains achieved then can be maintained into adulthood (Baxter-Jones, Kontulainen, Faulkner, & Bailey, 2008). Performance on musculoskeletal physical fitness tests in childhood and adolescence have been shown to be related to bone mass (Gracia-Marco, et al., 2011; Heinonen et al., 2000; Kontulainen et al., 2002; Morris et al., 1997 and van der Heijden, et al., 2010), and predictive of bone health in adolescence and adulthood, respectively (Barnekow-Bergkvisk, Hedberg, Pettersson, & Lorentzon, 2006; Kemper, et al.,2000; Vicente-Rodriguez, et al., 2004). This relationship may be mediated by the independent association between lean body mass and bone mass (Baptista, et al., 2012; Fonseca, de Franca, & van Praagh, 2008; Vicente-Rodriguez, et al., 2008).

Lean body mass is, of course, part of body composition. Among other things resistance training specifically and physical activity generically improves muscular strength through increases in fat-free (lean) body mass (Baxter-Jones, Eisenmann, Mirwald, Faulkner & Bailey, 2008; Moliner-Urdiales, Ortega, Vicente-Rodriguez, Rey-Lopez, Gracia-Marco, Widhalm, et al., 2010). The recent Institute of Medicine Report (2012) cited six high quality studies that provide direct evidence of a link between changes in muscle strength (particularly bench press, leg press, and squat) and power (vertical jump) and favorable changes in percent body fat, lean or fat-free body mass, waist circumference, and body mass index from late childhood to adulthood in both male and female normal and overweight individuals. Lower body muscular strength (as measured by vertical and standing long jumps) has been shown to be associated negatively with total and central body fat in male and female adolescents whereas higher levels of upper body strength were associated with higher levels of central body fat (Moliner-Urdiales, Ruiz, Vicente-Rodriguez, Ortega, Rey-Lopez, Espana-Romero, et al., 2009). Musculoskeletal fitness measured as push-up, sit-ups, grip strength, and trunk flexibility has been shown to be a significant predictor of weight gain (lower fitness at baseline leads to more weight gain) during a 20-year follow-up (Mason, et al., 2007).

Evidence is also emerging that indicates a positive impact of musculoskeletal fitness on metabolic syndrome/metabolic health risk factors (Metabolic Syndrome) in both adults (Churilla, Magyari, Ford, Fitzhugh & Johnson, 2012; Jurca, Lamonte, Barlow, Kampert, Church, & Blair., 2005; Jurca, Lamonte, Church, Earnest, FitzGerald, Barlow, et al., 2004; Magyari & Churilla, 2012; Strasser, Siebert, & Schobersberger, 2010; Wijndaele, et al., 2007) and youths (Artero, et al., 2011; Benson, Torode, & Singh, 2006; Benson, Torode, & Singh, 2008; Moreira, Santos, Vale, Soares-Miranda, Marques, Santos, et al., 2010; Mota, Vale, Martins, Gaya, Moreira, Santos, et al., 2010; Steene-Johannessen, Anderssen, Kolle & Andersen, 2009). This relationship appears to operate independently of, or in addition to, cardiorespiratory fitness and/or body mass/body composition.

Although a direct link between flexibility and health as defined by the traditional cardiovascular disease risk factors or Metabolic Syndrome has not been established, high levels of flexibility are associated with improved ability to complete activities of daily living (ADL), increased functional independence, and unrestricted mobility (Kell, et al., 2001) in adults. Two recent studies (Cortez-Cooper, Anton, DeVan, Neidre, Cook, & Tanaka, 2008; Yamamoto, Kawano, Gando, Lemitsu, Murakami, Sanada, et al., 2009) have reported a connection between flexibility and arterial stiffening (Arterial stiffness). That is, both a stretching training program and high sit-and-reach values have been linked with less arterial stiffening. Arterial stiffening is associated with impaired cardiovascular health. The linkage between muscular resistance exercise/training and arterial stiffness is still under debate.

A definitive connection between musculoskeletal flexibility, strength, endurance or power, and low back pain (LBP) remains elusive. In a healthy back, lumbar flexibility allows the lumbar curve to be almost reversed in forward flexion; hamstring flexibility allows anterior rotation (tilt) of the pelvis in forward flexion and posterior rotation in the sitting position; and hip flexor flexibility allows achievement of the neutral pelvic position. Inflexibility restricts these motions and causes increased compression of the disks. In a healthy back strong, fatigue-resistant abdominal muscles maintain proper pelvic position and reinforce the back extensor fascia providing support during forward flexion. Similarly, strong, fatigue-resistant back extensor muscles provide stability for the spine, maintain erect posture, and control forward flexion. Weak, easily fatigued muscles allow abnormal alignments, increase strain on the opposing muscle group, increase loading on the spine, and potentially cause disk compression (Plowman, 1992b). Although the anatomical rationale is strong for healthy back function, the research base for prevention of LBP is weak.

Prospective studies in adults are split between those that do predict first time or recurrent

LBP from flexibility (Beiring-Sorensen, 1984b; Nordgren, Schéle, & Linroth, 1980; Troup, Martin, & Lloyd, 1981) and those that do not predict either (Battie, Bigos, Fisher, Spengler, Hansson, Nachemson, et al., 1990; Jackson, Morrow, Brill, Kohl, Gordon, & Blair, 1998; Troup, Foreman, Baxter, & Brown, 1987). Studies involving muscle strength or muscle endurance measures are similarly split between those showing significant prediction of first time or recurrent LBP (Beiring-Sorensen, 1984a; Luoto, Heliövaara, Hurri, & Alaranta, 1995; Nordgren, et al., 1980; Suni, Oja, Miilunpalo, Pasanen, Vuori, & Bos, 1998; Taanila, Suni, Pihlajamaki, Mattila, Ohrankammen, Vuorinen, et al., 2012; Troup, et al., 1981) and those that do not predict (Jackson, et al., 1998; Leino, Aro, & Hasan, 1987).

The results for children and adolescents are much the same. There are studies that show significant prediction of LBP from impaired flexibility (Feldman, Shrier, Rossignol, & Abenhaim, 2001; Kujala, Taimela, Salminen, & Oksanen, 1994; Kujala, Taimela, Oksanen & Salminen, 1997) and studies that show no significant prediction (Burton, Clarke, McClune, & Tillotson, 1996; Mikkelsson, Nupponen, Kaprio, Kautiainen, Mikkelsson & Kujala, 2006; Salminen, Erkintalo, Laine, & Pentti, 1995; Sjölie, & Ljunggren, 2001). The results are similar for the predictive ability of muscular strength and endurance and LBP with several studies showing a significant prediction (Barnekow-Bergkvist, Gudrun, Janlert, & Jansson, 1998; Newcomer & Sinaki, 1996; Sjölie, & Ljunggren, 2001) and others showing no significant predictive ability (Mikkelsson, et al., 2006; Salminen, et al., 1995).

The tracking (maintenance of a characteristic over time) of musculoskeletal fitness has been shown to be moderately high (and higher than cardiovascular respiratory fitness) from adolescence to young adulthood (Twisk, et al., 2000). Taken together, muscular strength, muscular endurance, and flexibility are viewed as important dimensions of health related fitness and a means of improving the overall quality of life (Kell, et al., 2001).

What Field Tests Are Used to Assess Musculoskeletal Fitness in FITNESSGRAM®?

A number of different field tests have been used to assess muscular strength, muscular endurance, and flexibility. There is also considerable variability in the measurement protocols used for these assessments and these variations can greatly influence the safety and purpose of the assessment as well as the reliability and validity of the assessments. Considerable effort was spent to select items (and protocols) that were safe, reliable, and valid for the FITNESSGRAM® battery. The selected FITNESSGRAM® assessments and corresponding muscular functions are shown in Table 1. Instructions for the administration of each of these items, guidelines for interpreting the results, and the criterion referenced standards are described in the FITNESSGRAM® Test Administration Manual (Meredith & Welk, 2010).

Table 1. Musculoskeletal Assessments Used in the FITNESSGRAM® Battery

Function	Recommended Test	Optional Test Item(s)
Abdominal strength and endurance	Curl-up	
Trunk extensor strength and flexibility	Trunk lift	
Upper body strength and endurance	90^0 Push-up	Modified pull-up Flexed arm hang
Hamstring flexibility	Back-save sit and reach	
Shoulder flexibility	Shoulder stretch	

⇦TOC ⇦Chapter

Abdominal Strength and Endurance—The Curl-Up Assessment

What Is the Rationale for the Curl-Up Assessment?

A cadence-based curl-up test is recommended for abdominal strength and endurance testing in the FITNESSGRAM® battery. The selection of this test over a full sit-up assessment was based on extensive research and biomechanical analyses of arm placement, leg position, feet support, and range of motion of the movement (Plowman, 1992b).

The use of a cadence (20 reps per minute) with the curl-up was found to eliminate many of the concerns about the ballistic nature of one-minute all-out speed tests (Jette, Sidney, & Cicutti, 1984; Liemohn, Snodgrass, & Sharpe, 1988). Such timed tests with legs straight or bent often result in bouncing, jarring movements and reflect more power than strength or endurance properties and/or allow the use of accessory muscles (Sparling, Milard-Stafford, & Snow, 1997). The use of a pace helps to avoid early fatigue based on starting too fast, standardizes the movement from person to person, and makes it easier to judge whether a full proper repetition has been completed. In addition, the use of a cadence allows students to focus on their own performance. There can be no competitive speeding up. In practice the 3 second is slow enough to accomplish the intended goals described above and fast enough to allow for efficient testing of large groups in school settings. Liehmon, et al. (1996) found that high school girls performed fewer (34.07 ±21.93) curl-ups as a metronome paced test than without rhythmical pacing (37.72 ±12.06). Hui (2002) compared the effect of 5 different cadences (20, 25, 30, 35 reps per minute and free) on the Georgia Tech curl-up test in high school boys. Unlike Liehmon et al.'s results, more repetitions were achieved for the slower rhythms. However, the mean differences were small.

There has been considerable research on the various protocols for curl-up assessments and abdominal exercises. Readers interested in a more detailed review of the anatomical, electromyographical, and biomechanical considerations in the curl-up selection are referred to a section in the Appendix to this chapter titled "Supplemental Information on the Curl-Up Assessment Protocols."

What Is the Reliability of the Curl-up Test?

A number of studies have investigated the reliability of the curl-up assessment (Anderson, Zhang, Rudisill, & Gaa, 1997; Hyytiäinen, Salminen, Suvitie, Wickström, & Pentti, 1991; Jetté, et al., 1984; Knudson & Johnston, 1995; Patterson, Benninton, & De La Rosa 2001; Robertson & Magnusdottir, 1987; Vincent & Britten, 1980). Due to considerable differences in measurement protocol, only three studies are directly comparable. The Robertson and Magnusdottir results indicate a high degree of consistency (R = .97) among a college student population but the number of subjects is small. Values from the Anderson, et al. study with younger children (ages 6-10) were lower (R = .70), but this is not unexpected for this young age group. No matter which abdominal assessment is used, better values are consistently found for older students (high school and college), but even those for the younger students are generally deemed acceptable. Patterson et al. reported test-retest reliability of R = .89 and R = .86 for 10-12 year old boys and girls, respectively. Reliability for a single trial was reduced to R = .80 for boys and R = .75 for girls when the values were obtained from teacher-counted scores. Reliability of child-reported scores were R = .82 and R = .81 (test-retest) and R = .70 and R = .69 (single trial) for boys and girls, respectively. Child-reported scores were significantly higher than teacher-reported scores. Additional research is needed on elementary through high school age students of both sexes. A more detailed review of the reliability of abdominal assessments is available in the Appendix to this chapter in a section titled "Test-Retest Reliability of Field Tests of Abdominal Strength/Endurance."

What Is the Validity of the Curl-Up Test?

The curl-up test possesses logical (i.e., content and construct) validity (Axler & McGill, 1997; Flint, 1965; Godfrey, Kindig & Windell, 1977; Juker, McGill, Kropf, & Steffen, 1998; Mutoh, Mori, Nakamura, & Miyashita, 1981; Noble, 1981) as a test of abdominal strength and endurance. This observation is supported on the basis of anatomical and biomechanical analyses and through electromyography studies.

Despite their popularity and relative acceptance, few studies have compared sit-up performance with a criterion endurance test (Ball, 1993; DeWitt, 1944; Kjorstad, Hoeger, Harris, & Vaughn, 1998). The best results indicate that only 16% of the variance in abdominal muscle endurance is accounted for by sit-up performance. Other studies have reported lower validity evidence but the challenges in validation are due in large part to the lack of definitive criterion measures of abdominal strength. A detailed review of this literature is available in the Appendix to this chapter titled "Results of Concurrent Validity Studies for Various Forms of Sit-Ups and Curl-Ups."

Several studies have compared performances of full sit-ups and curl-ups (Diener, Golding, & Diener, 1995; Liemohn, et al., 1996; Lloyd, Walker, Bishop, & Richardson, 1996; Robertson & Magnusdottir, 1987; Sparling, et al., 1997; Vincent & Britten, 1980). Usually such a comparison of a new field test to a more established field test (one that has presumably been validated against a criterion measure) is done in an attempt to demonstrate convergent validity. The assumption then is that the field tests can be used interchangeably. The degree of association between sit-ups and curl-ups was found to account for only 7 to 42% of the variance. This means that the tests cannot be used interchangeably. This was interpreted as being positive, however. The curl-ups are intended to utilize different muscles over a more restricted range of motion than the sit-ups.

Trunk Extensor Strength—Trunk Extension Assessment

What Is the Rationale for the Trunk Extension Test?

Low back pain is a major source of disability and discomfort in our society. Risks are greater with advancing age but awareness and attention to trunk musculature at early ages is important to reduce future risks. Of the five anatomical and physiological areas which have been identified as critical for the development and maintenance of low back function (low back lumbar, hamstring, and hip flexor flexibility, plus abdominal and trunk extensor strength and endurance) only trunk extension strength and endurance has been shown to predict both first time and recurrent low back pain (LBP) (Plowman, 1992b). Retrospective studies of low back pain which have included a measure of trunk extension strength and endurance have shown significant relationships between them, including three in which electromyographic records were able to distinguish between those who did and did not have low back pain (DeVries, 1968; Hultman, Nordin, Saraste & Ohlsen, 1993; Roy, DeLuca, & Casavant, 1989; Roy et al., 1990).

The assessment of static extensor endurance known as the 240s over a table edge or Biering-Sorensen test is the only strength and endurance item that has been shown consistently in prospective studies to be predictive of LBP (Biering-Sorensen, 1984a, 1984b; Luoto, et al., 1995; Sjolie & Ljuggren, 2001; Suni, et al., 1998; Taanila, et al., 2012). Of the back extensor tests used in the research studies, the 240s over the table edge test is the only one not requiring sophisticated laboratory equipment. However, it does require a table and straps or personnel to hold the individual's lower body in place and it is time consuming. Several new tests have attempted to modify the Biering-Sorensen version (Albert, Bonneau, Stevenson, & Gledhill, 2001; Ito, et al., 1996; Moreau, Green, Johnson, & Moreau, 2001) but more research needs to be done on these before they are deemed acceptable for inclusion in FITNESSGRAM®. In the meantime the prone trunk extension lift is used to indicate both trunk extension flexibility and minimal strength and endurance.

Are There Risks Associated with Hyperextension on the Trunk Lift Assessment?

Hyperextension of the spine is often described as a contraindicated movement because of potential harm to the spinal cord. The greatest danger from excessive hyperextension is to athletes such as gymnasts, javelin throwers, weight lifters, and football linemen (Tittel, 1990), where speed and opposing forces (often in a rotational plane) can result in disc compression, nerve impingement, facet loading, and possibly fractures of the vertebrae. Nachemson and Elfström (1970) have shown that intradiscal pressure at the L3 level while performing active back hyperextension in which both the upper trunk and lower extremities were raised was equivalent to that of bent knee sit-ups, but less than lifting 20kg using correct biomechanics. Presumably, this pressure in back extension is lower when the legs are not also arching, but there are no data to support this assumption. Leimohn (1991) reported that "slow and controlled hyperextension movements are appropriate for inclusion in exercise programs. Spinal hyperextension is a natural and very functional movement. Moreover, maintenance of good spinal range of motion is in the best interest of the biomechanics of the spine" (p. 3). A restricted range is utilized to discourage excessive hyperextension. It is not intended as a test to identify hyperextension.

What Is the Reliability of the Trunk Lift Test?

Moreau, et al. (2001) presented a summary of 10 studies reporting the test-retest reliability for the Biering-Sorensen test in normal individuals. Four studies reported intraclass correlations of 0.54, 0.73, 0.98 and 0.99; five studies reported Pearson product-moment correlations of 0.20, 0.63, 0.74, 0.87, and 0.89; one study reported a Spearman rank-order correlation coefficient of 0.91 indicating a large spread of values. Other reliability studies (Hannibal, Plowman, Looney, & Brandenburg, 2006; Ito, et al., 1996; Jackson, Lowe, & Jensen, 1996; Johnson, Miller, & Liehmon,1997; O'Connell, et al.,2004; Patterson, Rethwisch, & Wiksten, 1997; Wear, 1963) utilized variations of a prone back extension task. In all cases test-retest reliability for a single trial was found to be high (.85-.998). However, sufficient reliability information is still not available for elementary aged individuals.

What Is the Validity of the Trunk Lift Test?

The trunk lift is intended to be a measure of both minimal trunk extensor strength and lumbar flexibility. As such it has logical (i.e., content) validity. However, there is limited research on both the 240s over the table edge test and the trunk lift. The low (.21, .25) correlations of the Biering-Sorenson (1984b) results contrasted with the high (.82) Jorgensen and Nicolaisen (1986) results seem to clearly indicate that the 240s test is an endurance as opposed to a strength test. Johnson, et al. (1997) and Liemohn, et al. (2000) have performed two studies investigating the contribution of selected variables to the performance of the trunk lift test.

Twelve college-aged males and females participated in the Johnson, et al. study. The best predictors of a passing performance were found to be isokinetic endurance (15 reps at 150 degrees per second), torso length, and body weight. Thirteen males and 23 females from 18-35 years of age participated in the Liemohn, et al. study. Multiple regression analysis revealed that the three most important ($R2 = .614$, $p < .001$) variables were passive trunk extension (floor to suprasternal notch measurement of flexibility achieved by pushing up with arms), 240s over the table edge test time (strength/endurance), and total work performed on a Cybex TEF unit at 120 degrees per second (strength). Patterson, et al. (1997) evaluated a modified version of the trunk lift (subjects were not stopped at 12 inches) in high school students and obtained concurrent validity correlations of .68 in females and .70 in males with goniometer measures of flexibility. Hannibal, et al. (2006) evaluated the validity of the FITNESSGRAM® (FG) trunk extension test (FG-TE) and the Box-90˚ dynamic trunk extension test (B-90˚ DTE) field tests in high school students 14-18 years. Parallel Roman Chair dynamic trunk extension (PRC-DTE), static trunk extension (PRC-STE), and dynamometer static back lift comprised the laboratory comparison tests. The amount of variance accounted for between the FG-TE and each of the laboratory tests ranged from 0-13%. Clearly, the FG-TE was not shown to be an acceptable test of either static or dynamic muscular endurance or static back extensor strength. The B-90˚ DTE was an attempt to find an alternative field test of trunk extension. This test did account for 38% of the variance with the PRC-DTE for the girls and 67% for the boys. As with reliability, validity data are still lacking for elementary aged individuals. More research is needed to develop an acceptable trunk extension test. A detailed review with tables summarizing this research is available in an Appendix to this chapter titled "Reliability and Validity of the Trunk Extension Assessment."

Upper Body Strength and Endurance

The 90◦ Push-Up Test

What Is the Rationale for Recommending the 90◦ Push-Up Test?

A number of assessments of upper arm and shoulder girdle strength and endurance have been used in various youth fitness batteries. Perhaps the most commonly used assessment is the pull-up test. The 90° push-up was selected as the recommended test item in FITNESSGRAM® because it has some very practical advantages over the pull-up. The most important advantages are that it requires no equipment and very few zero scores occur.

Data from the National Children and Youth Fitness Study I (NCYFS I) (Ross, Dotson, Gilbert, & Katz, 1985) revealed that 10-30% of the boys from 10 to 14 years of age and over 60% of the girls from 10 to 18 years of age could not do even one chin-up! The President's Council on Physical Fitness and Sports National School Population Fitness Survey (Reiff, et al., 1986) showed similar results: 40% of boys aged 6-12 years of age could not do more than one pull-up and 25% could not do even one; 70% of all girls 6-17 years of age could not do more than one pull-up and 55% could not do any. Furthermore, 45% of boys 6-14 years of age and 55% of the girls 6-17 years of age could not perform the flexed arm hang for more than 10 seconds. Obviously, such tests are not discriminating.

The majority of children can successfully perform the 90° push-up assessment and have a more favorable experience. In one study, only 5% of both boys and girls over 8 years of age, and only 10% of both boys and girls ages 6-8 years of age could not do even one 90° push-up (Massicotte, 1990). This number of zero scores is similar to those obtained with the modified pull-up in NCYFS II (Ross, Pate, Delpy, Gold, & Svilar, 1987). The primary difficulty with the modified pull-up is that it requires equipment that must be adjusted as each student is tested individually. Baumgartner and colleages (Baumgartner, Oh, Chung, & Hales, 2002; Wood & Baumgartner, 2004) have proposed a revised full push-up that requires body contact with the floor from the chest to the knees in the down position. However, even with students in university level fitness classes described as being accustomed to executing push-ups, zero scores in females (~27% of those used to calculate percentile norms) were a problem (Baumgartner, Hales, Chung, Oh, & Wood, 2004). The impact of body weight and body composition on upper body extremity test scores has long been recognized and recently reaffirmed (Lloyd, et al., 2000; Walker, Lloyd, Bishop, & Richardson, 2000). The reason the modified pull-up and 90° push-up provide a better range of scores is probably related to the fact that, in both, part of the body weight is supported (Pate, Ross, Baumgartner, & Sparks, 1987). Engelman and Morrow (1991), however, found that the modified pull-up does not negate the effect of body composition on upper body strength performance. Students need a realistic chance to be successful in testing and to improve with training in order to be motivated to try. For the majority of students, the 90° push-up provides this chance given appropriate instruction, training, and supervision. An additional advantage is that with adequate physical training push-up scores improve while this is not always the case for chin-ups, pull-ups, or flexed arm hang (Rutherford & Corbin, 1994). For these reasons the full length pull-up was dropped from the FITNESSGRAM® test battery in version 8.0 released in 2005.

What Is the Reliability of the Modified Pull-Up, the Flexed Arm Hang, and the 90° Push-Up?

Although high school students (grades 10-12) appear to have been overlooked, in all other school grades (including college) one or another of the field tests of upper body extremity strength and endurance have been found to be generally reliable. While many studies have evaluated full length push-ups without a cadence, several have investigated the reliability of the 90° push-up in elementary school children (Saint Romain & Mahar, 2001; Tomson, 1992; Zorn, 1992). These values (.64 to .99) are acceptable, although the total sample size is small. Jackson, Fronme, Plitt, and Mercer (1994) and Murr (1997) reported excellent reliability for the 90° push-up with college age subjects, although in the Jackson, et al. study, the females did the push-ups from their knees. McManis, Baumgartner, and West (2000) attempted to determine the reliability of the 90^0 push- up in three separate samples of elementary, high school, and college students. Intraclass stability reliability coefficients for the elementary and high school students were determined based on partner counts and ranged from .50 to .86. Similar values for the college students were obtained for each of 3 or 4 independent judges. With the exception of a probable outlier (.22), all of the other correlation coefficients were between .68 and .87. Lubans et al. (2011) found intraclass stability reliability coefficients of .90 for boys and .93 for girls on the 90◦ push-up in 9th graders. The typical error of the push-up test was lower in girls than boys, but in both groups there was evidence of systematic error, suggesting that a learning effect had occurred and that practice before testing is warranted. Objectivity of the scores from the elementary students ranged from .46 to .75, but student scores were consistently higher than adult counts as students tended simply to count each attempted 90° push-up and not evaluate whether it was completed with correct form. Objectivity between the four judges evaluating the college students ranged from .16 to .91 with 6 of the 16 coefficients being above .70. Tsigilis, Douda, and Tokmakidis (2002) reported an intraclass reliability of .89 for the flexed arm hang in college aged males and females, but with a large coefficient of variation (18.6%). Ortega, et al. (2008a) determined from heteroscedasticity analyses and Bland-Altman plots that the longer the time of performance in the bent arm hang test, the worse the degree of reliability agreement. A detailed review with tables summarizing this research is included in an Appendix to this chapter titled "Test-Retest Reliability of Upper Arm and Shoulder Assessments."

What Is the Validity of the Modified Pull-Up, the Flexed Arm Hang, and the 90° Push-Up?

The recommended test for upper body strength and endurance for The FITNESSGRAM® is the 90^0 push-up at a cadence of one repetition per every 3 seconds. The modified pull-up and flexed arm hang are optional items. Full pull-ups are an option only for 6.0 software users. Each test has a specific anatomical logical validity, but they are not necessarily anatomically interchangeable. For example, both the modified pull-up and the 90° push-up involve the pectoralis major, however the pull-up uses the latissimus dorsi and biceps as contributing muscles while the push-up involves the triceps and anterior deltoid. Hand position alters load and muscle activity in variations of the push-up (Freeman, Karpowicz, Gray & McGill, 2006; Gouvali & Boudolos, 2005). Correlations among the field tests have been found to vary from low (r = .31) to moderately high (r = .81) depending on the commonality of musculature. A detailed review of research on this topic is available in an Appendix to this chapter titled "Validity of Upper Arm and Shoulder Strength Field Assessments."

Flexibility Assessments

Back Saver Sit and Reach

What Is the Rationale for the Back Saver Sit and Reach Test?

The recommended item for lower body flexibility assessment is the Back Saver Sit and Reach Test. The assessment is conceptually similar to the more traditional Sit and Reach test but is intended to be safer on the back by restricting flexion somewhat. In the traditional sit and reach assessment, the forward flexion movement of the trunk with the legs extended causes the anterior portion of the vertebrae to come closer together such that the discs bulge posteriorly and the muscles, facia, and ligaments of the back are stretched. It also involves a forward rotation of the pelvis and sacrum which elongates the hamstrings. Cailliet (1988) has pointed out that stretching both hamstrings simultaneously results in "overstretching" the low back, especially in terms of excessive disc compression and posterior ligament and erector spinae muscle strain. He believes that stretching one hamstring at a time, by having the other leg flexed, "...'protects' the low back by avoiding excessive flexion of the lumbosacral spine" (Cailliet, 1988, p. 179). In addition, Cailliet points out that a lack of flexibility in one leg or the other causes asymmetrical restriction of the pelvis, pelvic rotation, and lateral flexion. This asymmetrical reaction is transmitted to the lumbosacral spine and "…has been considered a mechanical cause or aggravation of low back pain (Cailliet, 1988, p. 179). Liemohn, Sharpe, and Wasserman (1994b) experimentally investigated to determine whether there was less L1-S1 flexion in the back saver unilateral sit and reach than the traditional bilateral sit and reach. The amount of flexion occurring in the lumbar spine was quantified by resistance change signals using an Ady-Hall lumbar monitor. The amount of flexion did not differ between the two versions of the sit and reach. However, those subjects who indicated a preference said they were more comfortable holding the unilateral stretch than the bilateral version.

An additional advantage of the Back Saver Sit and Reach is that it allows the legs to be evaluated separately. This allows for the determination of symmetry (or asymmetry) in hamstring flexibility. In addition, testing one leg at a time eliminates the possibility of hyperextension of both knees. Patterson, Wiksten, Ray, Flanders, and Sanphy (1996) reported that 3 out of 40 boys and 4 out of 44 girls passed the back saver sit and reach on one side and failed it on the other side. Both Liemohn, Sharpe, and Wasserman (1994a) and Patterson, et al. (1996) emphasized that there is value in detecting such differences both when the asymmetry is a result of an injury or is an imbalance that might lead to a potential injury or postural disturbance. If identified, feedback can be given and remedial exercises prescribed.

What Is the Reliability of the Back Saver Sit and Reach Test?

Reliability data spanning a period of 50 years have shown that the stand and reach test, the sit and reach test, and the sit and reach test modified to accommodate anatomical differences are extremely consistent. Five studies (Hartman & Looney, 2003; Hui & Yuen, 2000; Hiu, Yuen, Morrow, & Jackson, 1999; Liemohn, et al., 1994a, 1994b; Patterson, et al., 1996) have established intraclass reliability for the Back Saver Sit and Reach with correlations of .93 to .99 and 95% confidence intervals of .89 to .99 at the widest. Subjects in these studies included both males and females from 6 to 41 years of age. A detailed review with summary tables is available in an Appendix to this chapter titles "Test-Retest Reliability of Field Tests of Hamstring Flexibility."

What Is the Validity of Back Saver Sit and Reach Test?

The various forms of stand or sit and reach test were originally intended to measure low back and hamstring flexibility. Early research (Broer & Galles, 1958; Mathews, Shaw, & Bohnen, 1957) validated these tests against Leighton flexometer measures of combined trunk and hip flexibility with reasonably acceptable results. Since then researchers have attempted to validate several version of the stand or sit and reach against criterion measures for both the low back and the hamstrings. The Modified Schober test (Macrae & Wright, 1969) is the most common criterion test of low back (so called lumbar or vertebral) flexibility. Both the passive straight leg raise and the active knee extension measured by flexometer, goniometer, or inclinometer have been used as criterion tests of hamstring (hip) flexibility. The overwhelming pattern has been that standing or sitting, classic or modified, one leg or two, parallel or V position, the sit and reach is moderately to highly related to hamstring flexibility and as such is a valid measure of hamstring flexibility (r=.39-.89). Conversely, correlations between the various versions of the sit and reach and low back (lumbar or vertebral) flexibility (r= -.003-.70) are with few exceptions so low that any sit and reach version cannot be considered a valid measure of low back flexibility. Recently, however, one study combined both hip and spine flexibility in an assessment of the Back Saver Sit and Reach using more modern technology. Chillón, et al. (2010) tested 138 adolescents (57 girls and 81 boys) on both versions of the sit and reach test while simultaneously measuring hip (sacral), lumbar (back), and thoracic (chest) angles with angular kinematic analysis. The difference between the two tests was 0.41 cm and was deemed meaningless from a practical point of view. There were significant differences between left and right values for both the hip and lumbar (but not the thoracic) angles for the back saver version. As has been the pattern, the strongest correlations were found between the flexibility scores and hip angle. When a stepwise linear regression was conducted using the average measures of the two legs, the hip angle independently explained 42% of the variance in Back Saver Sit and Reach performance. However, lumbar (back) angle explained an additional 30% and the thoracic angle a further 4%. Thus hip and lumbar angles together explained 77% of the variance in the test performance. Contributions were only slightly different when left and right leg data were analyzed separately. Although these results confirm that hamstring flexibility is the largest contributor to the Back Saver Sit and Reach they also suggest that the Back Saver Sit and Reach can be "considered an appropriate and accurate measure for hip and low-back flexibility" (Chillón, et al., 2010, p. 646). Obviously more studies using this technology over a wide range of ages would be beneficial. A detailed review of the validity of the test with summary tables is available in an Appendix to this chapter titles "Validity of Field Tests of Low Back and/or Hamstring Flexibility."

Should the Back Saver Sit and Reach Standards Be Adjusted for Body Dimensions?

The question of the influence of body dimensions, especially height and weight, has been a persistent one in the use of norms in physical fitness testing. Indeed, the original AAHPER Youth Fitness Test included two sets of norms from 1957-1965: one based on age alone and a second based on the Neilson-Cozens Index. This index included age to the nearest month, height to the nearest half-inch, and weight to the nearest pound. It was calculated based on "exponents" that were totaled into a "class" and percentile ranks were given for each class. The sit and reach was not part of the fitness battery at that point in time, but teachers found this system too time consuming and by the 1976 revision of the test the index had been dropped.

There is evidence that hamstring flexibility varies as children grow (Kendall, & Kendall, 1948). The fewest number (30%) of boys and girls can touch their toes (double leg) at age 12 years and 13 years, respectively. Comparable data are not available on the single leg sit and reach, but it would likely be similar. Cotton (1972b) summarized all of the available data from studies that had attempted to isolate the impact of body dimensions on flexibility. She concluded that "…in most cases, there is no relationship between anthropometric measures and trunk flexibility as measured by bobbing or sit-and-reach tests" (p.261). One study among those reviewed did find a significant difference when the extremes of the groups were considered, but another did not.

Hoeger and colleagues (Hoeger, Hopkins, Button, & Palmer, 1990) suggested a modified sit and reach which establishes a relative zero point designed to eliminate concern about disproportionate limb length bias. Thus, if a teacher believes a particular student has been unfairly evaluated after the initial testing using the standard box, the Hoeger method might be a reasonable option to try. As it is, the student is always only being asked to deal with his/her own body and the passing criterion is set at approximately the 25th percentile from the AAHPERD Health-Related Physical Fitness test and the National Children and Youth Fitness Survey normative data. Research has not shown the Hoeger system to be any more or less valid than the standard (or Back Saver) sit and reach in measuring hamstring flexibility (Castro-Piñero, et al., 2009a; Hui, et al., 1999).

Shoulder Stretch

What Is the Rationale for Including the Shoulder Stretch?

The shoulder stretch has been added as an option to try and illustrate that flexibility is important throughout the body—not just in the hamstrings—and that flexibility is very specific to each joint. It is intended to parallel the strength and endurance functional assessment of the upper arm and shoulder girdle. Too often, just assessing one flexibility item gives students the false impression that a single result indicates their total body flexibility, which, of course, may not be true. No validity or reliability data are available for the shoulder stretch.

What Is the Basis for the Criterion Referenced Standards for Muscular Strength, Endurance, and Flexibility?

Ideally, since the identified muscular strength, endurance, and flexibility items are part of a health related physical fitness test battery, the criterion referenced standards utilized would be linked to some specific status of a health factor. These standards should then represent an absolute desirable or protective level of that characteristic. This has now been done for both the aerobic capacity and body composition standards (Morrow, Going, & Welk, 2011). Unfortunately, this is not yet possible with the musculoskeletal test items and even the newest norms presented for several musculoskeletal items are based on percentiles, instead of criterion-reference values (Castro-Piñero, et al., 2009b).

Briefly, the problems are as follows (Plowman, 1992a, 1992b): The criterion health condition to which both general and specific measures of hamstring flexibility (the Back Saver Sit and Reach), low back flexibility (no separate field test available), abdominal strength and endurance (the curl-up), and trunk extension for flexibility and strength (the trunk lift) were originally linked with low back pain. The anatomical logic for this linkage remains strong (Plowman, 1993), but this theoretical link is, for unknown reasons, much stronger than the research evidence between low back function (in terms of measurable muscle strength, endurance, and flexibility) and low back pain onset or recurrence.

Individuals with low back pain typically show lower levels of truncal strength, muscular endurance, and flexibility than those who are pain free and an association between a history of low back pain and back extensor endurance has been shown in both adolescents and adults (Andersen, Wedderkopp, Leboeuf-Yde, 2006; Nourbakhsh & Arab, 2002; Payne, Gledhill, Katzmarzyk, & Jamnik, 2000a + see section on "Why Is Muscular Fitness Important?"). The standard of 240 seconds is inherent in the Biering-Sorensen trunk extensor strength test and other norms are available for modifications of this test (Payne, Gledhill, Katzmarzyk, Jamnik & Keir, 2000c; Johnson, Mbada, Akosile, & Agbeja, 2009; Mbada, Ayanniyi, & Adedoyin, 2009), but the current trunk lift is not comparable. No specific level of strength, muscular endurance, or flexibility has emerged as critical. Therefore, in the strictest sense of the word, true criterion referenced standards are not possible for these items at this time.

The criterion-referenced reliability and validity of the Back-Saver Sit and Reach cut-off scores for 6-12 year old children have recently been reported by Looney and Gilbert (2012). Pooled reliability data from 21 boys and 22 girls based on the current FITNESSGRAM® cut-off scores were P = .91 for the right leg and .95 for the left leg (indicating a high proportion of agreement for pass/fail decisions for trials one week apart) and Km = .82 for the right leg and .90 for the left leg (indicating the proportion of agreement in classification beyond what was expected by chance). Validity evidence (from 87 boys and 91 girls) indicating how well Back Saver Sit and Reach pass/fail decisions matched pass/fail decisions of the passive straight leg raise criterion measure showed that the best scores for 6-12 year olds are 8 and 9 inches for boys and girls, respectively. This supports the current standards for all ages of boys and girls 6-10 years old but differs from the current value of 10 inches for 10-11 year old girls.

The situation is even more difficult for the upper arm and shoulder girdle measures. Because muscle action is necessary for the proper mineralization of bone, it has been speculated that upper body strength is necessary as a protection against osteoporosis at advanced ages (Smith & Gilligan, 1987). Unfortunately, this has not been demonstrated experimentally. Therefore, there is neither a criterion health condition, a criterion test, nor criterion values against which to establish true criterion referenced standards for these tests. Furthermore, the variation in musculature utilized in the different test items means that all are not evaluating the upper arm and shoulder girdle precisely the same. This anatomical diversity further complicates the setting of equivalent standards among the tests. Data presented by Saint Romain and Mahar (2001) indicate that only approximately 70% of a sample of 5th and 6th grade boys and girls were classified the same way using results from the modified pull-up and 90 degree push-up. More work is needed on equating these tests as has been done with the aerobic fitness measures (Zhu, Plowman, & Park, 2010).

An alternative method for establishing criterion referenced scores is to compare the performance of individuals who have been instructed (trained) in a particular trait and hence should score high on any valid test of that trait against those who have not been instructed (untrained) in the same trait and hence should score lower on any valid test of the trait (Berk, 1976). An attempt to determine criterion referenced scores utilizing this technique and the NCYFS I and II survey results yielded phi coefficients that showed only weak relationships between instructional status (classified as physically active or inactive based on questionnaire data) and mastery status (classified as scoring above or below the criterion cut-off score). The validity or contingency coefficients were found to be little better than chance for achieving a correct classification of mastery or nonmastery categories. This was true for both males and females at all ages for the two legged sit and reach, timed 1-minute knee flex, feet held sit-ups, and free hanging pull-ups (Looney & Plowman, 1990). However, a shortcoming of this approach may be that the use of questionnaire data to establish physical activity (training) status was inadequate.

Rutherford and Corbin (1994) had more success when college women were actually put through a training program to determine instructional status. They established and cross-validated criterion referenced standards of 16 for the 90° push-ups, 5 seconds for the flexed arm hang, and .5 for pull-ups (which is obviously non-functional in practice). These scores had a probability of correct classification of .71, .68, and .71 respectively. It will be necessary to replicate Rutherford and Corbin's study for boys and girls at each age or grade level for each of the strength, endurance, and flexibility tests.

Despite the growing body of evidence (described in the "Why Is Muscular Fitness Important" section) linking higher levels of musculoskeletal fitness with positive health status throughout the age span, neither absolute values nor any true criterion-referenced standards have emerged. This area is fertile for research.

Currently the criterion referenced standards for all of these items are based on expert opinion from an analysis of normative data provided from NCYFS I and II, Canadian National Norms (Massicotte, 1990), and the Rutherford and Corbin (1994) data.

How Should Tests Be Done to Get Reliable and Valid Results?

To obtain accurate results from field tests it is important to adhere to specific guidelines. The following list is presented to assist with administering these assessments in physical education.

- The key to good test data is preparation. The instructor giving the test should carefully read and practice the test administration guidelines prior to any involvement with the students.
- Any equipment needed should be gathered and checked to be sure it is exactly what is called for and functioning properly.

- A testing plan should be devised and diagrammed to maximize efficiency and student involvement.
- Students should be instructed on proper techniques for each item. Emphasize slow controlled movements.
- The instructor should explain to students what each test is intended to measure and why that matters to them now.
- Students should practice each item and demonstrate proper form before the actual testing. For example, the curl-up without the feet being held may require a lot of practice for students to learn the technique.
- If several items are available try to guide students into selecting the most appropriate choice for success.
- If students are self-testing or testing each other, allow additional time for practice or do practice testing as part of the learning process. Guide students as to what to look for in order to count only those repetitions that are done properly.
- Provide an atmosphere that motivates each student to do his/her best.

The 2012 Institute of Medicine Fitness Measures and Health Outcomes Report and FITNESSGRAM®

The 2012 Fitness Measures and Health Outcomes in Youth Institute of Medicine Report's recommendations for musculoskeletal fitness and flexibility test items are different from the current test items in FITNESSGRAM®. Why is this, and what should my school do about it? The charge for the committee on Fitness Measures and Health Outcomes was to study the research literature to determine the relationship between specific musculoskeletal fitness test items and health in children and adolescents with the immediate goal being to identify test items for a future national survey of physical fitness in American youth.

Based on their evaluations and deliberations they recommended two musculoskeletal items for inclusion: hand grip (upper body isometric strength) and standing long jump (lower body strength/power). These items are included in the EuroFit (1988) and European Union ALPHA Health-Related Fitness Test Battery for Children and Adolescents (2011) and would allow direct international comparisons. These two items were also recommended for use in the schools but are not currently included in FITNESSGRAM®.

In addition, although direct health related impact has yet to be established, the modified pull-up and push-up (upper-body musculoskeletal strength/endurance) and curl-up (core strength/endurance) as well as the sit-and-reach or back-saver sit-and-reach (flexibility) were suggested for inclusion as fitness educational tools and items for continued research. At this point in time these items have simply not been studied well enough in relation to health to meet the committee's primary inclusion criteria. However, the committee found that these items were valid, reliable, feasible, and valuable educational tools. They all are, of course, current FITNESSGRAM® items.

In the future the hand grip and standing long jump may be added to FITNESSGRAM® as optional or primary items. For now, schools wishing to include these items for their own use certainly may do so. The AHPHA Test Manual (www.thealphaproject.net) is recommended for test administration instructions and normative values.

Are Muscular Fitness Tests Safe for Children?

Any exercise or physical activity if done improperly or excessively can lead to possible negative effects (e.g., injury). However, if correct movements are done in a controlled fashion and individual characteristics and limitation are taken into account, muscular fitness testing can be done safely by school children and adolescents.

As explained in the rationale section for each test, an attempt has been made to select the best (reliable and valid) and safest items based on current knowledge and practicality. The quality of the child's movement in performing the test is critical (Liemohn, Haydu, & Phillips, 1999). If an item cannot be done in a slow controlled fashion or if pain is experienced, then that item should not be performed by the individual.

Appendix

Metabolic Syndrome Information

The Metabolic Syndrome is a group of risk factors that collectively promote the development of cardiovascular disease and increase the risk of diabetes. Specifically these risk factors are: high fasting glucose, high waist circumference, high triglycerides, high blood pressure, and low high-density lipoprotein cholesterol. Some definitions include a proinflammatory state (Steinberger, Daniels, Eckel, Hayman, Lustig, McCrindle et al., 2009; Strasser, et al., 2010; Zimmet, George, Kaufman, Tajima, Silink, Arslanian, et al., 2007). Metabolic Syndrome is becoming more prevalent in children and adolescents, driven by the growing obesity epidemic in this young population. Without lifestyle changes, the risk factors for Metabolic Syndrome persist from childhood to adolescence to young adulthood (Saland, 2007).

Arterial Stiffness Information

Arteries become stiffer (lose their compliance, that is, the ability to expand and recoil with cardiac pulsation and relaxation) as individuals age. This occurs whether or not an individual has plaque build-up inside the arteries (atherosclerosis), high blood pressure (hypertension), or other diseases (Cortez-Cooper, et al., 2008). The resultant increased arterial stiffness is associated with impaired cardiovascular function, and it is an independent risk factor for hypertension, a variety of cardiovascular/coronary heart disorders, stroke, and mortality (Fernhall & Agiovlasitis, 2008; Seals, 2003; Yamamoto, et al., 2009). Thus, in terms of health, high arterial compliance (low arterial stiffness) is good.

Many factors including physical fitness, physical activity, and body composition affect arterial stiffness. (Fernhall & Agiovlasitis, 2008) Both cross sectional comparisons of endurance trained versus sedentary individuals and training studies of previously sedentary individuals have linked lower artery stiffness/higher arterial compliance with high aerobic fitness/physical activity in both sexes over a wide range of ages (Boreham, 2004; Havlik, et al., 2003; Jae, et al., 2010; Seals, 2003; Sugawara, et al., 2006; Tanaka, DeSouza, & Seals, 1998; Tanaka, et al., 2000). Increased body mass/%body fat and decreased aerobic capacity (as measured by the PACER 20 meter shuttle test) have been shown to be associated with arterial stiffening in otherwise healthy prepubescent children (Sakuragi, et al., 2009).

The linkage between muscular strength or resistance activity/training and arterial stiffness is still under debate. Although some studies (primarily those that involved high intensity strength training) suggest that either a single bout or chronic resistance training can increase arterial stiffening in adults(DeVan, et al., 2005) , there are currently more studies (primarily those that involved low or moderate resistance work or resistance work in conjunction with aerobic training in a circuit format) that show either no change or a reduction in arterial stiffness (Cortez-Cooper, et al., 2008; Fahs, Heffernan, Ranadive, Jae & Fernhall, 2010; Miura & Aoki, 2005; Seals, DeSouza, Donato, & Tanaka, 2008). Data are needed for youth.

Supplemental Information about Protocols for the Curl-Up Assessment

There are a number of different positions used for abdominal assessments. In particular, arm position, leg position, and the degree of trunk flexion have been varied. Each variation imposes different musculoskeletal demands on the body.

Arms placed across the chest or at the sides both offer approximately the same resistance to the abdominal flexion motion and avoid any excessive hyperflexion of the neck. However, arms placed at the sides offer the advantage of a convenient method of measurement (sliding forward 3 or 4.5 inches), which can be readily standardized between subjects.

Knees flexed instead of straight decreases movement of the fifth lumbar vertebra over the sacral vertebrae (Clarke, 1976). However, contrary to early evidence and common belief, the hip flexors are active whether the knees are flexed or not. This is especially true if the feet are held or the abdominals become fatigued (Anderson, Nillson, Ma, Thorstnensson, 1997; Flint, 1965; Godfrey, Kindig, & Windell, 1977; Mutoh, Mori, Nakamura, & Miyashita, 1981; Sparling, et al., 1997). A 1998 study by Juker, McGill, Kropf, and Steffen demonstrated that all forms of sit-ups tested (straight-leg with feet anchored, bent-knee with feet anchored, and bent-knee with feet anchored and heel press) activated the hip flexor (psoas) muscles more than a bent-knee, feet free curl-up. At the same time the curl-up was found to activate the external obliques, internal obliques, and transverse abdominals more than any of the sit-up variations. Needle biopsy results have shown that flexed knee sit-ups actually cause more intervertebral disc pressure than straight leg sit-ups (Nachemson & Elfström, 1970). Recently, Axler, and McGill (1997) confirmed this finding using electromyography (EMG) data. However, the values were both high and similar. More importantly, this study provided additional evidence that disc compression is much lower in both a feet anchored or feet free curl-up than for either the bent-knee or straight leg sit-up.

Among the 12 abdominal exercises studied by Axler and McGill, curl-ups resulted in the highest abdominal muscle activation to compression load in the upper and lower rectus abdominus. An electromographic (EMG) study (Parfrey, Docherty, Workman, & Behm, 2006) compared abdominal and hip flexor activation using 3 hand positions (5, 10, and 15 cm of movement), 2 knee positions [90˚ (FITNESSGRAM® uses 140˚) and straight], and 2 stabilization (feet held or not) conditions. The EMGs were monitored during isometric held positions to avoid the potential artifact as a result of movement. The 10 cm (~4 in and closest to the FITNESSGRAM®'s 4.5 inches for individuals >10 years), non-fixed feet, bent-knee position produced the highest activation in the upper rectus abdominis, lower rectus abdominis, and lower abdominal stabilizers with minimal activation of the hip flexors. Escamilla, Babb, DeWitt, et al. (2006) also performed an EMG analysis of 12 abdominal exercises. Upper and lower rectus abdominus muscle activity was shown to be greater in the curl-up exercise than in the bent-knee sit-up and rectus femoris and psoas muscle activity higher in the bent-knee sit-up than the curl-up. Contradicting earlier results, external and internal oblique activity was found to be higher in the bent-knee sit-up than the curl-up. The abdominals are responsible for only the first 30-45° of movement in the sit-up, with the hip flexors (psoas, iliacus, and rectus femoris) being responsible for the rest (Flint, 1965; Ricci, Marchetti, & Figura, 1981). If the flexion motion is continued beyond approximately 45°, the already shortened hip flexors are exercised through only a short arc which can lead to adaptive shortening. The psoas also attempts to hyperextend the spine as it flexes the hip and generates high compression and shear forces at the lumbar-sacral junction (Escamilla, et al., 2006). Thus, the curl-up should be a more specific and safer test than a full sit-up (Liemohn, Snodgrass, & Sharpe, 1988; Norris, 1993), especially for those who need to minimize lumbar spinal flexion or compressive forces because of low back instabilitiy or pathologies (Escamilla, et al., 2006). An item response theory model analysis of three variations (feet unanchored, feet anchored, feet unanchored on 30˚ inclined board) of six "sit-up" exercises scored as the number performed in 1 minute ranked the difficulty of each item. The participants were male and female college students. In general, the exercises with hands above the waist were found to be more difficult than those with hands at or below the waist; the tests with feet unanchored were harder than those with feet anchored, and no differences were found using the inclined board versus lying flat. The item difficulty values fell within a small range. Specifically, the partial curl-up with hands at the sides and feet unanchored was found to be more difficult than the curl-up with the arms across the chest and the feet anchored. However, the partial curl-up was recommended as the *"…most useful test for a national physical fitness battery…appropriate…for the average or perhaps low fit individual*," primarily because of the anatomical advantages described above (Safrit, Zhu, Costa, & Zhang, 1992, p.282). A 1996 review by Knudson concurred that "*the trunk-curl (TC) with unsupported feet appears to be the safest test and exercise for the abdominal muscles*" (p.27) and more recent research confirms this conclusion. A 2009 literature synthesis of electromyographic studies in abdominal exercises (Manfort-Pañego, Vera-García, Sánchez-Zuriaga, and Sarti-Martínez, 2009) concluded that in terms of safety, "the most important factors are (a) avoid active hip flexion and fixed feet, (b) do not pull with the hands behind the head, and (c) [utilize] a position of knees and hips flexion during upper body exercises [such as raising the shoulders off the floor] (p. 242)." The format of the curl-up used in FITNESSGRAM® meets all of these conditions.

Concerns that the spine has a finite number of flexion-extension cycles before disc damage occurs are based primarily on research by McGill and colleagues (Callaghan and McGill, 2001; Drake, Aultman, McGill, and Callaghan, 2005; Marshall and McGill, 2010; Tampier, Drake, Callaghan, & McGill, 2007) using in vitro specimens (dissected pig cervical vertebrae) subjected to moderate compressions loads and bending cycles ranging from 4400 to 86,400 in which half to all specimens had disc herniations. Contreras and Schoenfeld (2011) provide an excellent examination of this evidence pointing out the difficulty in extrapolating this evidence to an intact human doing crunch type exercises. As they point out these are certainly excessive numbers when compared with the way even the most enthusiastic individual does crunches; the removal of the muscles in preparing the cadaver pig spines for testing alters biomechanics; and, there is no fluid available to flow back into the discs as it does when tissue is alive. The pig spines were subjected to full range of motion and this is smaller than the range of motion of the human lumbar spine. When humans perform the curl-up/crunch properly (as noted above) it involves ~30 degrees of trunk flexion, much of it in the thoracic spine and not the lumbar spine. Given that the lumbar spine doesn't reach end range flexion, these studies are not really relevant. They also point out the benefits of spinal flexion exercises. They determined that "based on current research, it is premature to conclude that the human spine has a limited number of bending cycles" and that "the claim that dynamic flexion exercises are injurious to the spine in otherwise healthy individuals remains highly speculative…." (p. 14). McGill acknowledges that the level of spinal loading at which tissue damage occurs remains obscure and that there is probably a U-shaped relationship between spinal activity level and low back disorders. He is adamant that "sit-ups should not be performed at all by most people" (2007; p. 89), but a modified curl-up is one of his "Big Three" stabilization exercises (modified curl-up, side bridges, and quadruped bird-dog) for rehabilitation and training (McGill, 2001; McGill, 2007). Assuming there is no existing spinal pathology (disc herniation, prolapse, or flexion intolerance) spinal flexion movement in not contraindicated (Cantreras and Schoenfeld, 2011). The healthy fitness zone values for curl-ups are within sound recommended training limits.

Test-Retest Reliability of Field Tests of Abdominal Strength/Endurance

The table below summarizes results of studies on the reliability and validity of the abdominal strength/endurance assessments. Some of these articles were discussed in the chapter but readers interested in specific details should consult the original references.

Table 2. Test-Retest Reliability of Field Tests of Abdominal Strength/Endurance

	Subjects			**Reliability Coefficients [interclass (r) or intraclass (R)]**
Lead Author (Date)	**N**	**Sex**	**Age**	
Anderson (1997)	107 129	M F	6-10 y	R = .70 knees flexed, feet free 20 rpm curl-up
Buxton (1957)	53	M&F	6-15 y	r = .94 knees flexed, feet held, total N
Craven (1968)	63	M	college	r = .86 knees flexed, 1 min
Cureton (1975) Diener (1995)	49 11 21	M M F	8-11 y adults	r = .60 legs straight, feet held, N to max of 100 r = .98 knees flexed, feet free, curl-up, 1 min r = .97
DiNucci (1990)	43	M	college	r = .83 knees flexed R = .91 feet held, 1 min
	57	F		r = .85 R = .91
		M&F		r = .84 R = .91
Fleishman (1964)	201	M	adults	r = .72 knees flexed, timed
Glover (1962)	37	F	6-9 y	r = .78 knees flexed, 30s
	29	M	6-9 y	r = .91
Harvey (1967a)	60	F	college	r = .78 curl down test, 30s knees flexed, feet held
Hyytiäinen (1991)	30	M	35-44 y	r = .57 graded sit-up, 1RM r = .93 partial curl, 240s max hold
Jackson (1996)	31	M	college	R = .98 knees flexed, feet held, elbows to opposite knee
Jetté (1984)	43	M&F	school children	r = .88
Knudson (1995)	103	M	College	R=.88 bench trunk curl-up R = .94
Magnusson (1957)	~55	M&F	1st grade	r = .68 knees flexed, timed
	66		3rd & 4th grade	r = .82
			6th grade	r = .77
Noble (1975	48	M	College	r = .81 knees flexed, feet
	48	F		r = .91 free, oblique, 1 min
Patterson (2001)	36 48	M F	10-12y	R = .89 FG curl-up test-retest R = .86 R = .80 R = .75 single trial, teacher scored

Robertson (1987)	12 12	M F	college	R = .93 4 inch curl-up, min R = .97 knees flexed, feet free, 1 min
Safrit (1987)	27	M	11	r = .62 knees flexed, feet
	88	M	12	r = .83 held, 1 min
	104	M	13	r = .79
	58	M	14	r = .86
	44	F	11	r = .64
	92	F	12	r = .85
	85	F	13	r = .89
	43	F	14	r = .81
Scott (1959)	140	F	college	r = .94 knees flexed, timed
Sparling (1997)	167 38	M F	college	R = .92 two trials R = .86 single trial feet on bench, thighs perpendicular to floor, curl-up
Tomson (1992)	16	M	2nd grade	r = .75 knees flexed, feet
	7	F		r = .88 free, arms crossed, 1 min
		M		r = .68 knees flexed, feet
Tsigilis (2002)	98	F M&F	college	r = .00 free, arms straight, 1 min R = .83 knees flexed, feet held, 30 sec
Vincent (1980)	70	M	7-12 y	r = .62 knees flexed, feet free, curl 4s fwd, R&L
	40	F		
	138	M	junior	r = .53
	22	F	high school	
Waldhelm (2011)	19 15	M M	College College	r = .71 R = .92
Zorn (1992)	15	M	10-12 y	r = .83 knees flexed, feet
	13	F		r = .76 free, arms crossed, 1 min
		M		r = .79 knees flexed, feet
		F		r = .74 free, arms straight, 1 min

Results of Concurrent Validity Studies for Various Forms of Sit-Ups and Curl- Ups

The sit-up and the curl-up are the two most common assessments of abdominal strength and muscular endurance but it has proven difficult to fully evaluate the validity of the test. The amount of variance accounted for in the criterion strength scores ranges from less than 1% (DeWitt, 1944) to 32% (Ball, 1993) for various forms of the sit-up. The studies that specifically tested the curl-up (Diener, Golding, & Diener, 1995; Hall, Hetzler, Perrin, & Weltman, 1992; Kjorstad, et al., 1998; Knudson & Johnston, 1995; Knudson, 2001) did so against static or isokinetic measures of concentric and eccentric abdominal strength. Results from the Hall et al. study revealed weak relationships for the males tested and almost no relationships for the females. Many of these relationships were negative, indicating that better curl-up performance was associated with lower strength scores and vice versa. Hall, et al. speculated that both the use of an isokinetic criterion measure to validate a dynamic (isotonic) field test and the speed variation in the performance of the tests might have contributed to the poor results. Results were no better for the other studies where the criterion variable was a static abdominal crunch. Kjorstad, et al. (1998) did find that the curl-up was more highly related to isokinetic trunk flexion endurance than a flexed knee full sit-up, but even then only 5.3% of the variance was accounted for. These results and the wide variety of criterion tests that have been used by investigators point out the fact that no absolutely agreed upon criterion measure for abdominal strength and endurance exists, making statistical validation difficult. Thus, whatever it is that the curl-up test is measuring in terms of abdominal function, it is different from whatever the traditional sit-up is measuring. More validation work is needed for the curl-up. The table below summarizes some of the findings. Readers interested in specific details should consult the original references.

Table 3. Validity of Field Tests of Abdominal Strength and/or Muscle Endurance

	Subjects			**Field Test**[a]	**Criterion Test**		**Criterion Test**	
Lead Author (Date)	**N**	**Sex**	**Age**		**Strength**	**r**	**Muscle Endurance**	**r**
Ball (1993)	144	M	18-33 y	knees flexed, feet held, arms across chest, 1 min	1-RM trunk flexion	.57	60% 1-RM	.40
Berger (1966)	47	M	college	knees flexed, feet held, full sit-up, 2 min	1-RM sit-up, weighted	.51		
Biering-Sorensen (1984a)	449 479	M F	30-60 y	one sit-up, graded by arm position, legs straight	strain gauge, static MVC	-.34 (M) -.39 (F)		
Craven (1968)	61	M	college	1. straight leg sit-ups, 1 min;	Tensiometer, static MVC	.60		
				2. bent leg sit-ups 1 min;		.36		
				3. straight leg sit-ups, N		.53		
DeWitt (1944)	102	M	college	1. knee flexed, feet free,	1-RM abdominal lift,	.04	static sit-up with feet held	.25

⇦TOC ⇦Chapter

				oblique sit-up, N	dynamometer			
				2. knees flexed, feet held, oblique sit-up, N		.16		.37
				3. knees flexed, feet held, oblique sit-up, 2 min		.14		.26
Diener (1995)	15 21	M F	adults	1. knees flexed, feet free, curl-up, 1-min 2. straight leg sit-up, 1 min	static abdominal crunch	.14 (M) .43 (F) .04 (M) .07 (F)		
Hall (1992)	23 28	M F	M=23 y M=22 y	1. knees straight, feet held, hand behind head, 1 min	isokinetic dynamometer, peak torque single effort concentric (C) and eccentric (E)	M:-.18(C) M:-.21 (E) F: .42 (C) F: .40 (E)		
				2. knees flexed, feet free, 4 inch curl-up, 1 min		M:-.41 (C) M:-.38 (E) F:- .07 (C) F:-.08 (E)		
				3. knees flexed, feet held, arms across chest, 1 min		M:-.25 (C) M:-.28 (E) F: .27 (C) F: .32 (E)		

Harvey (1967a)	60	F	college	curl down, full ROM, feet held, 1 min	1-RM, dynamometer	.32		
Kjorstad (1998)	30 28	M F	College	1. knees flexed, full sit-up 2. curl-up	static abdominal crunch	-.19 .01	Isokinetic truck flexion	.08 .23
Knudson (1995)	10 10	M F	College	Bench trunk curl 2 min	Cybex Dynamometer Peak torque/BW isokinetic 30°·sec^{-1}	M+F=.08	Cybex isokinetic	M+F=.38
Knudson (2001)	22 22	M F	College	Curl-up 100 reps max	Cybex dynamometer static peak torque/BW	M=.07 F=-.19	Cybex static, 20sec torque/peak	M=.23 F=.10

[a]Convergent validity correlations between full range of motion sit-ups (knees flexed, feet held, arms crossed on chest, 1 min) and various forms of curl-ups (knees flexed, feet free, partial range of motion, Georgia Tech) have been reported between r = .27 and .67 (Diener, Golding & Diener, 1995; Lloyd, et al., 1996; Robertson & Magnusdottir, 1987; Sparling, Millard-Stafford, & Snow, 1997; Vincent & Britten, 1980).

Reliability and Validity of Field Tests of Trunk Extension

The table below summarizes results of studies on the reliability and validity of the trunk extension tests. Some of these articles were discussed in the chapter but readers interested in specific details should consult the original references.

Table 4. Reliability and Validity of Field Tests of Trunk Extension

	Subjects			**Reliability**		**Criterion Validity**	
Lead Author (Date)	**N**	**Sex**	**Age**	**Field Test**	**r**	**Test**	**r**
Biering-Sorensen (1984b	449 479	M F	30-60 y	strength/endurance 240 s prone extension hold		strain gauge static MVC[a]	-.26 -.31
Hannibal (2006)	40 32	M F	14-18y	90˚Dynamic Trunk Extension supported on box (B-90˚DTE)	.996 .99	Parallel Roman chair-dynamic trunk extension	.82 .62

						(PRC-DTE) Parallel Roman chair-static trunk extension (PRC-STE) Dynamometer static back lift (DSBL)	.55 .38 -.29 -.23
				Prone extension	.998 .998	PRC-DTE PRC-STE DSBL	.23 -.11 -.15 .33 -.04 -.36
Hyytiäinen (1991)	30	M	35-44 y	strength/endurance 240 s prone extension hold prone extension	.74		
Ito (1996)	37 53	M F	*M*=46.2y	prone sternum off floor; supported lower abdomin	.97 .94 R = .97		
Jackson (1996)	118 142	M F	College	best trial single trial	R = .96 R = .86		
Johnson (1997)	5 7	M F	20-30y		R = .89		
Jorgensen (1986)	53 23	M F	22-61 y	strength/endurance 240 s prone extension hold	.89	60% MVC	.82
Muller (2010)	18 11	M F	*M*= 25.1y	Ito test		Biering-Sorensen 240s prone extension	.52
O"Connell (2004)	31 38 22 20	M F M F	6-10y	Prone cadenced extensions on mat	.55	Static dynamometer	 .03
Patterson (1997)	43 45 43	M F M	high school	prone extension – best trial prone extension –	R = .95 (M) R = .93	Goniometer	.70 (M) .68

	45	F		single trial	(F) R = .90 (M) R = .85 (F)		(F)
Waldhelm (2011)	15	M	college	Prone extension	R= .79		
Wear (1963)	62	M	college	1. prone back extension 2. supine back extension	r = .96 r = .92		

[a]MVC=Maximal Voluntary Contraction

Test-Retest Reliability of Upper Arm and Shoulder Assessments

The table below summarizes results of studies on the upper arm and shoulder. Some of these articles were discussed in the chapter but readers interested in specific details should consult the original references.

Table 5. Test-Retest Reliability of Field Tests of Upper Arm and Shoulder Girdle Strength/Endurance

	Subjects					
Lead Author (Date)	**N**	**Sex**	**Age/Grade**	**Field Test**	**Reliability Coefficients interclass (r)/intraclass (R)**	
					Two Trials	*Single Trial*
Baumgartner (2002)	63 89	M F	College	Full length push-up, body from chest to knees contacts floor in down position		R = .95 R = .93
Cotten (1990)	8	M	K	modified pull-up	R = .72	R = .56
	11	F	K		R = .85	R = .74
	27	M	1		R = .76	R = .62
	29	F	1		R = .90	R = .85
	22	M	2		R = .88	R = .79
	22	F	2		R = .89	R = .81
	21	M	3		R = .75	R = .59
	27	F	3		R = .88	R = .78
	33	M	4		R = .90	R = .82
	37	F	4		R = .92	R = .86

	31	M	5		R = .79	R = .65
	33	F	5		R = .83	R = .71
	29	M	6		R = .90	R = .82
	33	F	6		R = .95	R = .90
DiNucci, (1990)	143 57	M F M&F	college	flexed arm hang	r = .93 r = .92 r = .94	R = .96 R = .96 R = .97
Engelman (1991)	70	M	3	pull-up	R = .95	R = .90
	87	F	3		R = .95	R = .91
	89	M	4		R = .96	R = .92
	74	F	4		R = .95	R = .91
	83	M	5		R = .91	R = .83
	67	F	5		R = .96	R = .92
	242	M	3, 4, & 5		R = .94	R = .88
	228	F	3, 4, & 5		R = .95	R = .91
	70	M	3	modified pull-up	R = .81	R = .68
	87	F	3		R = .90	R = .83
	89	M	4		R = .91	R = .83
	74	F	4		R = .87	R = .77
	83	M	5		R = .87	R = .77
	67	F	5		R = .90	R = .82
	242	M	3, 4, & 5		R = .87	R = .77
	228	F	3, 4, & 5		R = .89	R = .81
Jackson (1994)	40 23	M F	M=24.5 y F=24.7 y	90° push-up 90° push-up knees on floor	R = .96 R= .98	R = .92 R = .96
Lubans (2011)	42/26	M F	14.8 y	90° push-up		R = .90 R = .93
Kollath (1991)	61 44	M F	9 9	modified pull-up		R = .91 R = .72
McManis (2000)	83 73 36 34 40 44	M F M F M F	3, 4, & 5 3, 4, & 5 9 & 10 9 & 10 college college	90° push-up		R = .71* R = .64* R = .50* R = .86* R = .22,.68,.75** R=.75,.84,.84,.87**
McManis (1994)	25 20 45 32 23 55	M F M&F M F M&F	3, 4, & 5 3, 4, & 5 3, 4, & 5 9 & 10 9 & 10 9 & 10	90° push-up 90° push-up 90° push-up 90° push-up 90° push-up 90° push-up	R = .90 R = .91 R = .91 R = .59 R = .94 R = .75	R = .82 R = .84 R = .83 R = .42 R = .88 R = .60

⇦TOC ⇦Chapter

Murr (1997)	50	M	College	90° push-up (FG) 90° push-up (US Army)		R = .92 R = .98
Pate (1993)	38 56	M F M&F M F M&F M F M&F M F M&F M F M&F	9-10 y	pull-up flexed arm hang full push- up modified PU, overhand grip modified PU underhand grip		R = .80 R = .66 R = .79 R = .90 R = .85 R = .88 R = .83 R = .71 R = .85 R = .83 R = .81 R = .83 R = .85 R = .88 R = .87
Saint Romain (2001)	30 32	M F M+F		90° push-up	R = .99 R = .97R = .99	R = .99 R = .94 R = .98
	30	M		modified pull-up	R = .98	R = .96
	32	F			R = .98	R = .95
		M+F			R = .99	R = .97
Tomson (1992)	16 7	M F		90° push-up		r = .76 r = .78
Tsigilis (2002)	98	M+F		Flexed arm hang		R = .89
Wood (2004)	32	F	college	Bent-knee push-up		R = .83
Zorn (1992)	28	M F		90° push-up		r = .85 r = .64

** based on student counted scores*

** calculated separately for each of 3 (M) or 4 (F) “judges” from videotape analysis

Validity of Upper Arm and Shoulder Strength Field Assessments

A number of studies have examined the validity of the various field tests to upper body strength against criterion tests that should be close anatomical matches. The table summarizing the validity of upper arm and shoulder assessments is presented below. The best validity coefficients occurred between the revised full length floor contact push-up and a bench press based on a percentage of body weight (BW). Approximately 76% of the variance was accounted for in females and 64% in males. For other variations, the validity coefficients accounted for only 16% and 32% of the variance in the matched criterion test for muscular endurance. All of the rest of the comparisons account for much less. Thus, concurrent validity has not established these tests as absolute measures of strength or muscle endurance. The difficulty may be in the selection of the criterion measures or in the inability to isolate specific muscle groups in both sets of measures. Additionally, these test results may all be confounded by body weight and/or body composition to varying degrees. The Baumgartner, et al. study (2002) accounted for BW in the criterion measure. When Pate, et al.'s (1993) results were expressed relative to body weight (i.e., per kg) the observed validity coefficients for strength improved considerably to the .50 to .70 range. However, most of the muscle endurance correlations remained lower and statistically non-significant even when adjusted by weight.

Specific validation data are available for the 90˚ push-up as a strength measure in only one study conducted on college age females (Rutherford & Corbin, 1993) and one on college males (Jackson, Fromme, Plitt, & Mercer, 1994). These correlations (presented in the accompanying table) are of the same order of magnitude as the other studies where males were used as subjects, and, hence, much better than the lower values typically obtained with females. In addition, the 90˚ push-up test shows higher relationships with the criterion tests than the field tests that are supposedly anatomically matched (i.e., pull-up and latissimus pull-downs; flexed arm hang and biceps arm curl). When the 90˚ push-up test was correlated with the sum of the three criterion tests (bench press, latissimus pull-downs, and arm curl) divided by body weight in the Rutherford and Corbin study, the validity coefficient improved to .70, showing that body weight is a factor in this test. The validity coefficients between the 90˚ push-up and muscular endurance are better than most other items, but not good (Jackson, et al., 1994). More research is needed on the 90˚ push-up, especially with elementary and secondary school aged children and adolescents. The table below summarizes some of the findings. Readers interested in specific details should consult the original references.

Table 6. Validity of Upper Arm and Shoulder Strength Field Assessments

Lead author (Date)	Subjects N	Sex	Age	Field Test[a]	Criterion Tests Strength	R	Endurance	r
Ball (1993)	144	M	college	Push-up	Bench Press, 1-RM	.56	Bench Press, 60% 1-RM, 30 lifts x min-1, N	.17
				Pull-up	Latissimus pull downs, 1-RM	.40	Latissimus pull downs 60% 1-RM, 30 lifts x	.14
Baumgartner (2002)	58 48	M F	College	Full length push-up, chest to knees floor contact			min-1, N Bench press: M = 70%BW; F=40% BW	.80 .87
Jackson (1994)	40 23	M F	M=24.5 y M=24.7 y	90° Push-up knees on floor	Bench Press, 1-RM	.30 .23	Max Rep at 45.5 kg Max Rep at 22.7 kg	.41 .40
Pate (1993)	38	M	9-10 y	Pull-up	Latissimus	-.16	Latissimus	.25
	56	F			pull downs, 1-RM	.05	pull downs 50%-1-RM, N	.09
		M&F				.11		.08
		M		Push-up	Bench Press, 1-RM	.36	Bench Press, 50% 1-RM, N	.47
		M&F				.38		.17

  The Cooper Institute, Dallas, TX.

		M F		Flexed Arm Hang	Arm Curl 1-RM	-.23 -.12	Arm Curl 50% 1-RM,	-.15 -.15
		M&F				-.06	N	-.09
Rutherford (1993)	204	F	College	Pull-up 90°	Bench Press 1 -RM Bench Press	.27 .37		
				Push-up Pull-up	1 -RM Latissi mus	.19		
				90°	Pull-downs 1-RM Latissi mus	.47		
				Push-up Flexed	Pull-downs 1-RM Arm Curl	.26		
				Arm Hang 90°	1-RM Arm Curl	.46		
				Push-up	1-RM			
Baumgartner (2002)	58	M	College	Full			min-1, N Bench	.80
	48	F		length push-up, chest to			press: M = 70%BW; F=40% BW	.87

Wood (2004)	77	F	College	Bent-knee push-up	Bench press, 40%BW	.67

				push-up chest to knee floor contact		.68

Test-Retest Reliability of Field Tests of Hamstring Flexibility

The table below summarizes results of studies on the test retest reliability of field tests for hamstring flexibility. Some of these studies were discussed in the chapter but readers interested in specific details should consult the original references.

Table 7. Test-Retest Reliability of Field Tests of Hamstring Flexibility

	Subjects			**Reliability Coefficients**	
Lead Author (Date)	**N**	**Sex**	**Age**	**Assessment**	**intraclass (R)**
Allen (1988)	10			sit and reach	r = .97
Ayala (2012)	243	M&F	Mean age= 21.3 & 20.7y	sit and reach toe touch	R = .92 R = .89
Bozic (2010)	84	M	college	sit and reach	R = .94
Broer (1958)	50	F	18-31 y	stand and reach	r = .97
Buxton (1957)	50	M&F	6-15 y	stand and reach	r =.95
Cotten (1972)	37 38	M F	College	sit and reach	r = .88
Davis (2008)	5 5	M F	college	sit and reach	R = .94
DiNucci (1990)	143 57	M F M&F	college	sit and reach	r = .92 / R = .96 r = .95 / R = .97 r = .94 / R = .97
Gauvin (1990)	47 26	M F	18-73 y LBP patients	sit and reach	R = .98
Hartman (2003)	21 23	M F	6-12 y	back saver sit and reach	R=.97(L), .98(R) R= .96(L), .97(R) R=.99 R=.97
Harvey (1967b)	100	F	college	stand and reach	r = .92
Hoeger (1992)	31	F	adults	modified sit and reach	R = .89
Hui (2000)	62	M	17-41 y	modified back saver	R = .96 (L) / .97 (R)
	96	F			R = .97 (L) / .97 (R)

		M		back saver	R = .93 (L) / .98 (R)
		F			R = .97 (L) / .98 (R)
		M		sit and reach	R = .98
		F			R = .96
		M		V-sit and reach	R = .96
		F			R = .89
Hui (1999)	62 96	M F M F M F	17-41y	backsaver sit and reach sit and reach V-sit and reach	R = .93 (L), .98 (R) R = .97 (L), .97 (R) R = .98 R = .96 R = .96 R = .89
Hyytiäinen (1991)	30	M	35-44 y	stand and reach	r = .93
Jackson (1986)	100	F	13-15 y	sit and reach	R = .99
Jackson (1989)	52 52	M F	20-45 y	sit and reach sit and reach	r = .99 r = .99
Jackson (1996)	31	M	College	V-sit and reach	R = .98
Jones (2002)	43 46	M F	11-16 y	Sit and reach	r=.88
Kippers (1987)	16 17	M F	Average age = 21.6 yrs.	stand and reach	r = .97
Liemohn (1994 a & b)	40	M&F	college	sit and reach back saver modified sit and reach	R = .98 R = .98 R = .99
López-Miñarro (2009)	76 67	M F	college	sit and reach back saver modified sit and reach (R, L)	 R = .97 R = .98 R =.97, .96 R= .97, .97
Magnusson (1957)	53 66 53	M&F	1st grade 3rd/4th grade 6th grade	stand and reach	r = .70 r = .91 r = .84
Mier (2011)	30 30	M F	*M*= 25y	sit and reach	R =.97 R = .97
Minkler (1994)	99	M&F	18-35y	modified sit and	R = .99

				reach	
Patterson (1996)	42 46	M F	11-15y	back saver	R = .99(L), 99 (R) R = .99 (L), .99 (R)
Poley (1948)	63	F	college	stand and reach	R = .93
Safrit (1987)	27	M	11 y	sit and reach	r = .94 / R = .97
	88	M	12 y		r = .94 / R = .97
	104	M	13 y		r = .94 / R = .97
	58	M	14 y		r = .95 / R = .97
	44	F	11 y		r = .87 / R = .93
	92	F	12 y		r = .93 / R = .96
	85	F	13 y		r = .88 / R = .93
Tsigilis (2002)	43 98	F M & F	14 y college	sit and reach	r = .80 / R = .89 R = .94
Waldhelm	15	M	college	sit and reach	R = .98
Wear (1963)	53	M	college	sit and reach	r = .94
Wells (1952)	100	F	college	sit and reach stand and reach	r = .98 r = .96

Note: Correlation coefficients are specified as follows: interclass (r)/intraclass (R).

Validity of Field Tests of Low Back and/or Hamstring Flexibility

The table below summarizes results of studies on the validity of field tests for hamstring flexibility. Some of these studies were discussed in the chapter but readers interested in specific details should consult the original references.

Table 8. Validity of Field Tests of Low Back and/or Hamstring Flexibility

	Subjects			**Field Test[a]**	**Criterion Test**	
Lead Author (Date)	**N**	**Sex**	**Age**		**Hamstring r**	**Low Back r**
Ayala (2011)	55	M	Mean age = 26 y	Sit and reach Toe touch Back saver sit and reach Modified sit and reach	Passive straight leg raise *(B)* .61 .24 .47	
	48	F	= 23 y	Sit and reach Toe touch Back saver sit and reach Modified sit and reach	.75 .93 .92 .90 .73	

Ayala (2012)	156 87	M + F	Mean age = 21.3 y	Sit and reach Toe touch	Passive straight leg raise *(B)* .795 .704	
Baltaci	102	F	= 20.7y college	Back saver sit and reach	Goniometer .44[c], .37[d]	
Biering-Sorensen (1984b)	449 479	M F M F	30-60y	finger to floor distance[b]	straight leg -.65 raise (R) [c]-.70 active knee .55 extension (R) .56	modified Schober[e] -.35 modified Schober[e] -.20
Bozic (2010)	84	M	college	sit and reach	passive straight leg raise .63	
Broer (1958)	100	F	College	stand and reach	Leighton flexometer .81	
Castro-Piñero (2009)	45 42	M F	6-12y 13-17y	sit and reach	Passive straight leg raise .377 .337	
Cornbleet (1996)	199 211	M F	5-12y	Hoeger modified sit and reach sit and reach	.375 .259 Hip joint angle .76 (sacral)	
Davis (2008)	42 39	M F	college	sit and reach	Knee extension angle .57 Sacral angle .65 Straight leg raise .65	

Hartman (2003)	87 92	M F	6-12y	back saver Sit and reach	Passive straight leg raise .67[c], .69[d] .48[c], .42[d] Active knee extension .47[c], .50[d] .57[c], .54[d] Passive straight leg raise .66[c],.67[d] .49[c], .47[d] Active knee extension .40[c], .40[d] .54[c], .52[d]	Modified Schober .07[c],.003[d] -.06[c],-.06[d] Inclinometer .28[c],.26[d] .10[c],.10[d] Modified Schober .05 -.07 Inclinometer .29 .16
Hui (2000)	62	M	17-41	mod. back saver	goniometer .67[c], .61[d]	mod Schober .47 [c] [d]
				back saver	goniometer .44 [c], .46 [d]	mod Schober .27 [c] .24 [d]
				sit and reach	goniometer .67[c], .61[d]	mod Schober .27
				Vsit and reach	goniometer .67[c], .61[d]	mod Schober .42
	96	F		mod. back saver	goniometer .54[c], .50[d]	mod Schober .23 [c] .26 [d]
				back saver	goniometer .50 [c], .39 [d]	mod Schober .15 [c] .18 [d]
				sit and reach	goniometer .53[c], .46[d]	mod Schober .24
				Vsit and reach	goniometer .52[c], .44[c]	mod Schober .24
Hui (1999)	62 96	M F	17-41 y	back saver sit and reach	straight leg raise .46 (L) [d] goniometer .39 (L) .48 (L), .47 (R) .46 (L), .53 (R)	.24 (L), .27(R) .15 (L), .18 (R) .27 .24
				Vsit and reach	.58 (L), .63 (R) .44 (L), .52 (R)	.42 .24
Jackson (1986)	100	F	13-15 y	sit and reach	straight leg raise .64	modified Schober .28
Jackson (1989)	52 52	M F	20-45y	Sit and reach	straight leg raise .89 straight leg raise .70	modified Schober .59 modified Schober .12
Kippers (1987)	16 17	M F	M=21.6 y	stand and reach	photographic analysis -.79	photographic analysis .10

Liemohn (1994)	20 20 20 20	M F M F	College	back saver sit and reach	straight leg raise .76 straight leg raise .72 straight leg raise .70 straight leg raise .70	inclinometer .32 inclinometer .38 inclinometer .29 inclinometer .40
López-Miñarro (2009)	76 67 76 67	M F M F	college	sit and reach Back saver	straight leg raise .56 (L), .59 (R) .75 (L0, .76 (R) Sacral angle .52 (L), .59 (R) .69 (L), .64 (R) Straight leg raise .53 (L-L), .51 (R-R) .70 (L-L), .66 (R-R) Sacral angle .47 (L-L), .49 (R-R) .55 (L-L), .51(R-R)	
Mathews (1957)	66	F	College	stand and reach sit and reach	Leighton flexometer (trunk & hip) .80 Leighton flexometer (trunk & hip) .74	
Mier (2011)	30 30	M F	*M*=25 y	sit and reach	Passive straight leg raise .64 (day1), .66 (day2) .79(day 1), .81(day2)	
Minkler (1994)	48 51	M F	college	modified stand and reach	straight. leg raise (R)[c] .75 straight. leg raise (R)[c] .66	modified Schober .40 modified Schober .25
Nicolaisen (1985)	53 24	M F	21-61 y 27-60	finger to floor distance	active knee extension .60	modified Schober .12
Patterson, (1996)	42 46	M F	11-15 y	back saver	Straight leg raise .72 (L), .68 (R) goniometer .51 (L),	modified Schober.15 (L), .10 (R)
Rodriguez-Garcis (2008)	125 118	M F	college	sit and reach	.52 (R) Straight leg raise .56-.59 .72-.74	.17 (L), .25 (R)
Sinclair (1993)	52 48	M F	15-16 y	sit and reach	Pelvi-spinometer .79	Pelvi-spinometer .38

vanAdrichem (1973)	84 60 55 49	M M F F	6-12 y 6-12 y 13-18 y 13-18 y	stand and reach		modified Schober .23 modified Schober .14
Youdas (2008)	106 106	M F	20-79y	sit and reach hip joint angle (sacral)	Passive straight leg raise .59	modified Schober .33 modified Schober -.15

[a]*The correlation between the sit and reach and stand and reach scores has been reported to range between .73 and .95 (Mathews, Shaw, & Bohnen, 1957; Mathews, Shaw, & Woods, 1959; Wells & Dillon, 1952). The correlation between the two legged sit and reach and the one legged "back saver" sit and reach has been reported to be between .91 and .92 in 79 7-13 y old boys and girls (Gilbert & Plowman, 1993).*
[b]*The finger to floor distance differs from the stand and reach in that movement beyond the level of the feet is not possible.*
[c]*R=right leg.*
[d]*L=left leg*
[e]*The modified Schober was validated against radiologically determined back flexibility r=.97, N=342 (Macrae & Wright, 1969).*

Bibliography

Albert, W.J., Bonneau, J., Stevenson, J.M., & Gledhill, N. (2001). Back fitness and back health assessment considerations for the Canadian Physical Activity, Fitness and Lifestyle appraisal. *Canadian Journal of Applied Physiology, 26*, 291-317.

Allen, M.E. (1988). Clinical kinesiology: Measurement techniques for spinal disorders. *Orthopaedic Review, 17,* 1097-1104.

Andersen, B., Wedderkopp & Leboeuf-Yde, O. (2006). Association between back pain and physical fitness in adolescents. *Spine*, *31*, 1740-1744.

Anderson, E.A. Nilsson, J. Ma, Z. & Thorstensson, A. (1997). Abdominal and hip flexor muscle activation during various training exercises. *European Journal of Applied Physiology, 75,* 115-123.

Anderson, E.A., Zhang, J.J., Rudisill, M.E., & Gaa, J. (1997). Validity and reliability of a timed curl-up test: Development of a parallel form for the FITNESSGRAM abdominal strength test. *Research Quarterly for Exercise and Sport, 68* (Suppl.), A-51.

Artero, E.G., Lee, D-C., Lavie, C.J., Espana-Romero, V., Sui, X., Church, T.S., et al. (2012). Effects of muscular strength on cardiovascular risk factors and prognosis. *Journal of Cardiopulmonary Rehabilitation and Prevention, 32,* 351-358.

Artero, E.G., J.R.Ruiz, F.B. Ortega, V. España-Romero, G. Vicente-Rodríguez, D. Molnár, et al. (2011). Muscular fitness is independently associated with metabolic risk in adolescents: The HELENA study. *International Journal of Obesity,* 35, S151-S168

Axler, C.T., & McGill, S.M. (1997). Low back loads over a variety of abdominal exercises: Searching for the safest challenge. *Medicine and Science in Sports and Exercise*, *29*, 801-810.

Ayala, F., Sainz de Baranda, P., De Ste Croix, M., & Santonja, F. (2011). Criterion-related validity of four clinical tests us
ed to measure hamstring flexibility in professional futsal players. *Physical Therapy in Sport, 12,* 175-181.

Ayala, F., Sainz de Baranda, P., De Ste Croix, M., & Santonja, F. (2012). Reproducibility and criterion-related validity of the sit and reach test and toe touch test for estimating hamstring flexibility in recreationally active young adults. *Physical Therapy in Sport, 13,* 219-226.

Ball, T.E. (1993). The predictability of muscular strength and endurance from calisthenics. *ResearchQuarterly for Exercise and Sport*, *64* (Suppl.), A-39. [Abstract]

Baltaci, G., Un, N., Tunay, V., Besler, A., & Gerceker, S. (2003). Comparison of three different sit and reach tests for measurement of hamstring flexibility in female university students. *British Journal of Sports Medicine, 37,*59-61.

Baptista, F., Barrigas, C., Vieira, F., Santa-Clara, H., Homens, P.M., Fragoso, I., et al. (2012). The role of lean body mass and physical activity in bone health in children. *Journal of Bone and Mineral Metabolism, 30,* 100-108.

Barnekow-Bergkvist, M., G.E. Gudrun, U. Janlert, & E. Jansson. (1998). Determinants of self-reported neck-shoulder and low back symptoms in a general population. *Spine,* 23 (2), 235-243.

Barnekow-Bergkvist, M., G. Hedberg, U. Janlert, & E. Jansson. (2001). Adolescent determinants of cardiovascular risk factors in adult men and women. *Scandinavian Journal of Public Health,* 29, 208-217.

Barnekow-Bergkvist, M., G. Hedbery, U. Pettersson, R. Lorentzon. (2006). Relationships between physical activity and physical capacity in adolescent females and bone mass in adulthood. *Scandinavian Journal of Medicine and Science in Sports,* 16, 447-455.

Battie M.C., S.J. Bigos, L.D. Fisher, D.M. Spengler, T.H. Hansson, A.L. Nachemson ,et al. (1990). The role of spinal flexibility in back pain complaints within industry: A prospective study. *Spine,* 15,768-773.

Baumgartner, T.A., Hales, D., Chung, H., Oh, S., & Wood, H.M. (2004). Revised push-up test norms for college students. *Measurement in Physical Education and Exercise Science, 8*, 83-8.

Baumgartner. T.A., Oh, S., Chung, H., & Hales, D. (2002). Objectivity, reliability, and validity for a revised push-up test protocol. *Measurement in Physical Education and Exercise Science, 6*, 225-242.

Baxter-Jones, A.D. Kontulainen, S.A., Faulkner, R.A., Bailey, D.A. (2008). A longitudinal study of the relationship of physical activity to bone mineral accrual from adolescence to young adulthood. *Bone, 43*, 1101-1107.

Baxter-Jones, A.D., Eisenmann, J.C., Mirwald, R.L., Faulkner, R.A., & Bailey, D.A. (2008). The influence of physical activity on lean mass accrual during adolescence: a longitudinal analysis. *Journal of Applied Physiology, 105,* 734-741.

Benson, A.C., M.E. Torode, & M.A. F. Singh. (2008). Effects of resistance training on metabolic fitness in children and adolescents: A systematic review. *Obesity Reviews,* 9, 43-66.

Benson, A.C., M.E. Torode, & M.A. F. Singh. (2006). Muscular strength and cardiorespiratory fitness is associated with higher insulin sensitivity in children and adolescents. *International Journal of Pediatric Obesity,* 1, 222-231.

Berger, R. A. (1966). Evaluation of the 2-minute sit-up test as a measure of muscular endurance and strength. *Journal of the Association of Physical and Mental Rehabilitation, 20(4),* 140.

Berk, R.A. (1976). Determination of optimal cutting scores in criterion-referenced measurement. *Journal of Experimental Education, 45*, 4-9.

Biering-Sorensen, F. (1984a). A one-year prospective study of low back trouble in a general population.*Danish Medical Bulletin, 31*, 362-375.

Biering-Sorensen, F. (1984b). Physical measurements as risk indicator for low-back trouble over a one- year period. *Spine*, *9*, 106-119.

Boreham, C.A., Ferreira, I, Twisk, J. W., Gallagher, A.M., Savage, M. J., & Murray, L, J. (2004). Cardiorespiratory fitness, physical activity, and arterial stiffness: The Northern Ireland young hearts project. *Hypertension*, *44*,721-726.

Boreham, C.A.G., & McKay, H.A. (2011) Physical activity in childhood and bone health. *British Journal of Sports Medicine*, 45, 877-879.

Boreham, C., & C. Riddoch. (2001). The physical activity, fitness, and health of children. *Journal of Sport Sciences*, *19*, 915-929.

Bozic, P.R., Pazin, N.R., Berjan, B.B., Planic, N.M, & Cuk, I.D. (2010) Evaluation of the field tests of flexibility of the lower extremity: Reliability and the concurrent and factorial validity. *Journal of Strength and Conditioning Research,* 24 (9), 2523-2531.

Brill,P.A., Macera, C.A., Davis, D.R., Blair, S.N., & Gordon, N. (2000) Muscular strength and physical function. *Medicine and Science in Sports and Exercise,32*,412-416.

Broer, M.R., & Galles, N.R.G. (1958). Importance of relationship between various body measurements in performance of the toe-touch test. *Research Quarterly*, *29,* 253-263.

Burton, A.K., Clarke,R.D., McClune, T.D.& Tillotson, K.M. (1996). The natural history of low back pain in adolescents. *Spine*, 21(20), 2323-2328.

Buxton, D. (1957). Extension of the Kraus-Weber test. *Research Quarterly*, *28*, 210-217.

Cailliet, R. (1988). Low back pain syndrome (4th ed.). Philadelphia: F.A. Davis.

Callaghan, J.P., McGill, S.M. (2001). Intervertebral disc herniation: studies on a porcine model exposed to highly repetitive flexion/extension motion with compressive force. *Clinical Biomechanics, 16,(1),* 28-37.

Castro-Piñero, J., Chillón, P., Ortega, F.B., Montesinos, J.L., Sjöström, M., & Ruiz, J.R. (2009a) Criterion-related validity of sit-and-reach and modified sit-and-reach test for estimating hamstring flexibility in children and adolescents aged 6-17 years. *International Journal of Sports Medicine*, 30, 658-662.

Castro-Piñero, J, González-Montesinos, J.L., Mora, J., Keating, X.D., Girela-Rejón, M.J. Sjöström, M., et al. (2009b). Percentile values for muscular strength field tests in children aged 6 to 17 years: Influence of weight status. *Journal of Strength and Conditioning Research*, 23(8), 2295-2310.

Chillón, P., Castro-Piñero, J., Ruiz, J.R., Soto, V.M., Carbonell-Baeza, A., Dafos, J.,et al. (2010) Hip flexibility is the main determinant of the back-saver sit-and-reach test in adolescents. *Journal of Sports Sciences*, *28 (6)*, 641-648.

Churilla, J.R., Magyari, P.M., Ford, E.S., Fitzhugh, E.C., & Johnson, T.M. (2012). Muscular strengthening activity patterns and metabolic health risk among US adults. *Journal of Diabetes*, *4*, 77-84.

Clarke, H.H. (1976). Exercise and the abdominal muscles. *Physical Fitness Research Digest, 6(3).*Washington, DC: President's Council on Physical Fitness and Sport.

Contreras, M.A., & Schoenfeld, B. (2011a). To crunch or not to crunch: An evidence-based examination of spinal flexion exercises, their potential risks, and their applicability to program design. *Strength and Conditioning Journal, 33(4),* 8-18.

Contreras, B. & Schoenfeld, B. (2011b). To crunch or not to crunch. Available online at http://www.t-nation.com/free_online_ article/most recent/to_crunch_or not_to_crunch. Accesses 2/28/2013 .

Cornbleet, S.L., & Woolsey, N.B. (1996) Assessment of hamstring muscle length in school-aged children using the sit-and-reach test and the inclinometer measure of hip joint angle. *Physical Therapy*, 76 (8), 850-855.

Cortez-Cooper, M.Y., Anton, M.M., DeVan, A.E., Neidre, D.B., Cook, J.N. & Tanaka, H. (2008). The effects of strength training on central arterial compliance in middle-aged and older adults. *European Journal of Cardiovascular Prevention and Rehabilitation*, 15, 149-155.

Cotten, D.J. (1972a). A comparison of selected trunk flexibility tests. *American Corrective Therapy Journal*, *26,* 24-26.? ....

Cotton, D.J. (1972b). Selected anthropometric measures and trunk flexibility measurement-a study.*Journal of Physical Education* ,250-262.

Cotten, D.J. (1990). An analysis of the NCYFS II modified pull-up test, *Research Quarterly for Exercise and Sport*, *61*, 272-274.

Council of Europe Committee for the Development of Sport (1993). *EUROFIT: Handbook for the EUROFIT Tests of Physical Fitness*. Strasbourg: Council of Europe.

Craven, C.W. (1968). An evaluation of methods of administering the sit-up test. Unpublished master's thesis, University of Texas at Austin. In H.H. Clarke (Ed.), Exercise and the abdominal muscles. *Physical Fitness Research Digest, 6(3*). Washington, DC: President's Council on Physical Fitness and Sport.

Cureton, K.J., Boileau, R.A., & Lohman, T.G. (1975). Relationship between body composition measures and AAHPER test performances in young boys. *Research Quarterly, 46*, 218-229.

Davis, D.S., Quinn, R.O., Whiteman, C.T., Williams, J.D., & Young, C.R. (2008) Concurrent validity of four clinical tests used to measure hamstring flexibility. *Journal of Strength and Conditioning Research,* 22 (2), 583-588.

DeVan, A.E., Anton, M.M., Cook, J.N., Neidre, D.B., Cortez-Cooper, M.Y., & Tanaka, H. (2005). Acute effects of resistance exercise on arterial compliance. *Journal of Applied Physiology, 98(6)*, 2287-2291.

DeVries, H.A. (1968). EMG fatigue curves in postural muscles. A possible etiology for idiopathic low back pain. *American Journal of Physical Medicine, 47(4)*, 175-181.

DeWitt, R.J. (1944). A study of the sit-up type of test, as a means of measuring strength and endurance of the abdominal muscles. *Research Quarterly, 15,* 60-63.

Diener, M.H., Golding, L.A., & Diener, D. (1995). Validity and reliability of a one-minute half sit-up test of abdominal strength and endurance. *Sports Medicine, Training and Rehabilitation*, *6,*105-119.

DiNucci, J., McCune, D., & Shows, D. (1990). Reliability of a modification of the health-related physical fitness test for use with physical education majors. *Research Quarterly for Exercise and Sport,61*, 20-25.

Drake, J.D.M., Aultman, C.D., McGill, S.M., & Callaghan, J.P. (2005). The influence of static axial torque in combined loading on intervertebral joint failure mechanics using a porcine model. *Clinical Biomechanics, 20(10)*, 1038-1045.

Engelman, M.E., & Morrow, J.R., Jr. (1991). Reliability and skinfold correlates for traditional and modified pull-ups in children grades 3-5. *Research Quarterly for Exercise and Sport, 62*, 88-91.

Escamill, R.F., Babb, E., DeWitt, T., Jew, P., Kelleher, P., Burnham, T., et al. (2006). Electromyographic analysis of traditional and nontraditional abdominal exercises: Implications for rehabilitation and training. *Physical Therapy, 86*,656-671.

Fahs, C.A., Heffernan, K.S., Ranadive, S., Jae, S.Y., & Fernhall, B. (2010). Muscular strength is inversely associated with aortic stiffness in young men. *Medicine & Science in Sports & Exercise,* 42(9), 1619-1624.

Faigenbaum, A.D., Kraemer, W.J., Blimkie, C.J.R., Jeffreys, I., Micheli, L.J., Nitka, M., et al. (2009). Youth resistance training: Updated position statement paper from the National Strength and Conditioning Association. *Journal of Strength and Conditioning Research, 23(Supplement 5),* S60-S79.

Feldman, D.E., I. Shrier, M. Rossignol, & L. Abenhaim. (2001). Risk factors for the development of low back pain in adolescence. *American Journal of Epidemiology*, 154, 30-36.

Fernhall, B. & Agiovlasitis, S. (2008). Arterial function in youth: Window into cardiovascular risk. *Journal of Applied Physiology*, 105, 325-333.

FitzGerald,S.J., Barlow, C.E., Kampert, J.B., Morrow, J.R. Jr.,Jackson, A.W., & Blair, S.N. (2004).Muscular fitness and all-cause mortality: Prospective observations. *Journal of Physical Activity and Health, 1*,7-18.

Fleishman, E.A. (1964). The structure and measurement of physical fitness. Englewood Cliffs, NJ: Prentice-Hall.

Flint, M.M. (1965). Abdominal muscle involvement during the performance of various forms of sit-up exercise. *American Journal of Physical Medicine, 44,* 224-234.

Fonseca, R.M.C., deFranca, N.M., & vanPraagh, E. (2008). Relationship between indicators of fitness and bond density in adolescent Brazilian children. *Pediatric Exercise Science, 20* 40-49.

Freeman, S., Karpowicz, A., Gray, J., & McGill, S. (2006). Quantifying muscle patterns and spine load during various forms of the push-up. *Medicine & Science in Sports & Exercise, 38,*570-577.

Gale, C.R., C.N. Martyn, C. Cooper, & A.A. Sayer. (2007). Grip strength, body composition and mortality. *International Journal of Epidemiology,* 36, 228-235.

Garcia-Artero, E., Ortega, F.B., Ruiz, J.R., Mesa, J.L., Delgado, M. Gonzalez-Gross, M., et al. (2007). Lipid and metabolic profiles in adolescents are affected more by physical fitness than physical activity (AVENA study). *Revista Espanola de Cardiologia, 60 (6),* 581-588.

Gauvin, M.G., Riddle, D.L., & Rothstein, J.W. (1990). Reliability of clinical measurements of forward bending using the modified fingertip-to-floor method. *Physical Therapy, 70,* 443-447.

Gilbert, J.A., & Plowman, S.A. (1993). Beyond the sit and reach: Assessment of back function. The reliability and validity of the back saver sit and reach. Paper presented at AAHPERD National Convention, Washington, DC.

Glover, E.G. (1962). Physical fitness test items for boys and girls in the first, second, and third grades.Unpublished master's thesis, University of North Carolina at Greensboro, NC.

Godfrey, K.E., Kindig, L.E., & Windell, E.J. (1977). Electromyographic study of duration of muscle activity in sit-up variations. *Archives of Physical Medicine and Rehabilitation, 58*, 132-135.

Gouvali, M.K. & Boudolos, K. (2005). Dynamic and electromyographical analysis in variants of push- up exercise. *Journal of Strength and Conditioning Research, 19,* 146-151.

Gracia-Marco, L., Vicente-Rodriquez, G., Casajus, J.A., Molnar, D., Castillo, M.J., & Moreno, L.A. (2011). Effect of fitness and physical activity on bone mass in adolescents: the HELENA Study. *European Journal of Applied Physiology, 111,* 2671-2680.

Gunter, K. B., Almstedt, H.C. & Janz, K.F. (2012). Physical activity in childhood may be the key to optimizing lifespan skeletal health. *Exercise and Sport Sciences Reviews, 40 (1),* 13-21.

Hall, G., Hetzler, R.K. Perrin, D., & Weltman, A. (1992). Relationship of timed sit-up tests to isokinetic abdominal strength. *Research Quarterly for Exercise and Sport, 63,* 80-84.

Hannibal, N.S.III, Plowman, S.A., Looney, M.A., & Brandenburg, J. (2006). Reliability and validity of low back strength/muscular endurance field tests in adolescents. *Journal of Physical Activity & Health,3*(Suppl.2), S78-S89.

Hartman, J.G., & Looney, M.A. (2003). Norm-referenced and criterion-referenced reliability andvalidity of the back-saver sit-and-reach. *Measurement in Physical Education and Exercise Science,7,* 71-87.

Harvey, V.P., & Scott, G.D. (1967a). An investigation of the curl-down test as a measure of abdominal strength. *Research Quarterly, 38,* 22-27.

Harvey, V.P., & Scott, G.D. (1967b). Reliability of a measure of forward flexibility and its relation to physical dimensions of college women. *Research Quarterly, 38*, 28-33.

Hasselstrøm, H., S.E. Hansen, K. Froberg, L.B. Andersen. (2002). Physical fitness and physical activity during adolescence as predictors of cardiovascular disease risk in young adulthood. Danish Youth and Sports Study. An eight-year follow-up study. *International Journal of Sports Medicine*, 23, S27-S31.

Havlik, R.J., Simonsick, E.M., Sutton-Tyrell, K., Newman, A., Danielson, M.E., Brock, D.B., Pahor, M., Lakatta, E., Spurgeon, H., & Vaitkevicius, P.(2003). Association of physical activity and vascular stiffness in 70-79-year-olds: The health ABC study. *Journal of Aging and Physical Activity*, 11(2), 156-166.

Heinonen, A., Seivanen, H., Kannus, P., Oja, P., Pasanen, M., & Vuori, I. (2000). High-impact exercise and bones of growing girls: A 9-month controlled trial. *Osteoporosis International, 11*, 1010-1017.

Hoeger, W.W.K., & Hopkins, D.R. (1992). A comparison of the sit and reach and the modified sit and reach in the measurement of flexibility in women. *Research Quarterly for Exercise and Sport, 63*, 191-196.

Hoeger, W.W.K., Hopkins, D.R., Button, S., & Palmer, T.A. (1990). Comparing the sit and reach with the modified sit and reach in measuring flexibility in adolescents. *Pediatric Exercise Science, 2*,156-162.

Hui, S.C. (2002). Effect of cadence in curl-up for classifying the abdominal fitness levels of high school boys. *Research Quarterly for Exercise and Sport Supplement*, 73(1), A39-40.

Hui, S.C. & Yuen, P.Y. (2000). Validity of the modified back-saver sit-and-reach test: a comparison with other protocols. *Medicine & Science in Sports & Exercise*, *32*, 1655-1659.

Hui, S.C., Yuen, P.Y., Morrow, J.R., Jr., & Jackson, A.W. (1999). Comparison of the criterion-related validity of sit-and-reach tests with and without limb length adjustments in Asian adults. *Research Quarterly for Exercise and Sport, 70*,401-406.

Hultman, G., Nordin, M., Saraste, H., & Ohlsen, H. (1993). Body composition, endurance, strength, cross-sectional area, and density of MM erector spinae in men with and without low back pain. *Journal of Spinal Disorders*, 6(2), 114-123.

Hyytiäinen, K., Salminen, J.J., Suvitie, T., Wickström, & Pentti, J. (1991). Reproducibility of nine tests to measure spinal mobility and trunk muscle strength. *Scandinavian Journal of Rehabilitation Medicine*,*23*, 3-10.

IOM (Institute of Medicine) (2012). Fitness measures and health outcomes in youth. Washington, D.C.: The National Academics Press.

Ito, T., Shirado, O., Suzuki, H., Takahashi, M., Kaneda, K., & Strax, T.E. (1996). Lumbar trunk muscle endurance testing: An inexpensive alternative to a machine for evaluation. *Archives of Physical Medicine and Rehabilitation.*, *77*, 75-79.

Jackson, A.W., & Baker, A.A. (1986). The relationship of the sit and reach test to criterion measures of hamstring and back flexibility in young females. Research Quarterly for Exercise and Sport, 57, 183- 186.

Jackson, A.W., Fromme, C., Plitt, H., & Mercer, J. (1994). Reliability and validity of a 1-minute push- up test for young adults. *Research Quarterly for Exercise in Sport, 65*(Suppl.), A57-A58. [Abstract].

Jackson, A.W., Lowe, T.J., & Jensen, R.L. (1996). Reliability of the physical performance tests of theYMCA adult fitness test. *Research Quarterly for Exercise and Sport, 67*(Suppl.), A-43.

Jackson, A.W., & Langford, N.J. (1989). The criterion-related validity of the sit and reach test: Replication and extension of previous findings. *Research Quarterly for Exercise and Sport*, *60*, 384-387.

Jackson, A.W., J.R. Morrow, P.A. Brill, H.W. Kohl, N.F. Gordon, & S.N. Blair. (1998). Relations of sit-up and sit-and-reach tests to low back pain in adults. *Journal of Orthopaedic & Sports Physical Therapy*, 27 (1), 22-26.

Jackson, A.W., Morrow, J.R., Jr., Jensen, R.L., Jones, N.A., & Schultes, S.S. (1996). Reliability of thePrudential FITNESSGRAM trunk life in young adults. *Research Quarterly for Exercise and Sport,67,*115-117.

Jae, S.Y., Heffernan, K.S., Fernhall, B, Oh, Y.S., Park, W.H., Lee, M.K., et al. (2010). Association between cardiorespiratory fitness and arterial stiffness in men with the metabolic syndrome. *Diabetes Research and Clinical Practice*, 90(3), 326-332.

Janz, K.F., Dawson, J.D. & Mahoney. L.T.(2002). Increases in physical fitness during childhood improve cardiovascular health during adolescence: The Muscatine study. *International Journal of Sports Medicine*, 23, S15-S21.

Jetté, M., Sidney, K., & Cicutti, N. (1984). A critical analysis of sit-ups: A case for the partial curl-up as a test of abdominal muscular endurance. *Canadian Association of Health, Physical Education, and Recreation Journal*, *51 (1),* 4-9.

Johnson, K.R., Miller, M.A., & Liemohn, W.P. (1997). An examination of factors contributing to performance on the FITNESSGRAM trunk lift test. *Medicine and Science in Sports and Exercise*, *29* (Suppl.), S9.

Johnson, O.E., Mbada, C.E., Akosile, C.O., & Agbeja, O.A. (2009). Isometric endurance of back extensors in school-aged adolescents with and without low back pain. *Journal of back and musculoskeletal rehabilitation*, 22(4), 205-211.

Jones, M.A., Stratton, G., Reilly, T., & Unnithan, V.B. (2002). Measurement error associated with spinal mobility measures in children with and without low-back pain. *Acta Paediatrica, 91*, 1339-1343.

Jorgensen, K., & Nicolaisen, T. (1986). Two methods for determining trunk extensor endurance: Acomparative study. *European Journal of Applied Physiology*, *55*,639-644.

Juker, D., McGill, S., Kropf, P., & Steffen, T. (1998). Quantitative intramuscular myoelectric activity of lumbar portions of psoas and the abdominal wall during a wide variety of tasks. *Medicine & Science in Sports & Exercise*, *30*, 301-310.

Jurca, R., Lamonte, M.J., Barlow, C.E., Kampert, J.B., Church, T.S., & Blair, S.N. (2005). Association of muscular strength with incidence of metabolic syndrome in men. Medicine & Science in Sports & Exercise, 37, 1849-1855.

Jurca, R., Lamonte, M.J., Church, T.S., Earnest, C.P., FitzGerald, S.J., Barlow, C.E., et al. (2004). Associations of muscle strength and aerobic fitness with metabolic syndrome in men. *Medicine & Science in Sports & Exercise*, 36, 1301-1307.

Kanbur, N.Ö., Dűzgűn, İ., Derman, O., & Baltaci, G. (2005). Do sexual maturation stages affect flexibility in adolescent boys aged 14 years? *Journal of Sports Medicine and Physical Fitness*, *45*, 53-7.

Katzmarzyk, P.T., &Craig, C.L. (2002). Musculoskeletal fitness and risk of mortality. *Medicine & Science in Sports & Exercise*, *34*,740-744.

Kell, R.T., Bell, G., & Quinney, A. (2001). Musculoskeletal fitness, health outcomes and quality of life. *Sports Medicine*, *31*,863-873.

Kemper, H.C.G., Twisk, J.W.R., vanMechelen, W., Post, G.B., Roos, J.C., & Lips, P.(2000). A fifteen-year longitudinal study in young adults on the relation of physical activity and fitness with the development of the bone mass: The Amsterdam Growth and Health Longitudinal Study. *Bone,* 27, 847-853.

Kendall, H. & Kendall, F. (1948). Normal flexibility according to age groups. *Journal of Bone and JointSurgery, 39,* 690-694.

Kippers, V., & Parker, A.W. (1987). Toe-touch test: A measure of its validity. *Physical Therapy, 67,* 1680-1684.

Kjorstad, R.L., Hoeger, W.W.K., Harris, C., & Vaughn, R. (1998). Validity of two field tests of abdominal strength and endurance. *Medicine & Science in Sports & Exercise, 30* (Suppl.), S215.

Knudson, D. (1996). A review of exercise and fitness tests for abdominal muscles. SportsMedicine Update,11, 4-5. 25-30.

Knudson, D. (2001). The validity of recent curl-up tests in young adults. *Journal of Strength andConditioning Research,15,* 81-85.

Knudson, D., & Johnston, D. (1995). Validity and reliability of a bench trunk-curl test of abdominal endurance. *Journal of Strength and Conditioning Research, 9(3),* 165-169.

Kollath, J.A., Safrit, M.J., Zhu, W., & Gao, L.G. (1991). Measurement errors in modified pull-ups testing. *Research Quarterly for Exercise and Sport, 62,* 432-435.

Kontulainen, S.A., Kannus, P.A., Pasanen, M.E., Sievanen, H.T., Heinonen. A.O., Oja, P., et al. (2002). Does previous participation in high-impact training result in residual bone gain in growing girls? One year follow-up of a 9-month jumping intervention. *International Journal of Sports Medicine,* 23 (8), 575-581.

Kujala, U.M., Taimela, S., Oksanen, A.,& Salminen, J.J. (1997). Lumbar mobility and low back pain during adolescence: A longitudinal three-year follow-up study in athletes and controls. *The American Journal of Sports Medicine,* 25 (3), 363-368.

Kujala, U.M., Taimela, S., Salminen, J.J., Oksanen, A. (1994). Baseline anthropometry, flexibility and strength characteristics and future low-back pain in adolescent athletes and nonthletes: A prospective one-year follow-up study. *Scandinavian Journal of Medicine and Science in Sports,* 4, 200-205.

Leino, P., S. Aro, & J. Hasan. (1987). Trunk muscle function and low back disorders: A ten-year follow-up study. *Journal of Chronic Disease,* 40(4), 289-296.

Liemohn, W. (1991). Choosing the safe exercise. *ACSM certified news, 1(2),* 1-3.

Liemohn, W.P., Haydu, T., & Phillips, D. (1999). Questionable Exercises. *President's Council on Physical Fitness and Sports Research Digest, Series 3, No. 8*.

Liemohn, W.P., Miller, M., Haydu, T., Ostrowski, S., Miles, S., & Riggs, S. (2000). An examination of a passive and an active back extension range of motion (ROM) tests. *Medicine & Science in Sports & Exercise, 32* (Suppl.), S307.

Liemohn, W.P., Sharpe, G.L., & Wasserman, J.F. (1994a). Criterion related validity of the sit-and-reach test. *Journal of Strength and Conditioning Research, 8,* 91-94.

Liemohn, W.P., Sharpe, G.L., & Wasserman, J.F. (1994b). Lumbosacral movement in the sit-and-reach and in Cailliet's protective-hamstring stretch. *Spine,19,* 2127-2130.

Liemohn, W.P., Smith, J., Clapp, A., Farley, R., Muir, I., Ashley, C., & Howell, P. (1996). Evaluation of the utility of different abdominal endurance tests. *Research Quarterly for Exercise and Sport, 67* (Suppl.), A-61.

Liemohn, W., Snodgrass, L.B., & Sharpe, G.L. (1988). Unresolved controversies in back management: A review. *The Journal of Orthopaedic and Sports Physical Therapy, 9,* 239-244.

Lloyd, L.K., Walker, J.L., Bishop, P.A., & Richardson, M.T. (2000). The adjustment of FITESSGRAMtest scores for skinfold thickness in youth. *Research Quarterly for Exercise and Sport, 71* (Suppl.), A-52.

Looney, M.A., & Gilber, J. (2012). Validity of alternative cut-off scores for the back-saver sit and reach test. *Measurement in Physical Education and Exercise Science, 16,* 268-283.

Looney, M.A., & Plowman, S.A. (1990). Passing rates of American children and youth on the FITNESSGRAM criterion-referenced physical fitness standards. *Research Quarterly for Exercise and Sport, 61,* 215-223.

López-Miñarro, P.A., Andújar, P.S.B., & Rodríguez-Garcia, P.L. (2009). A comparison of the sit-and-reach test and the back saver sit-and-reach test in university students. *Journal of Sports Science and Medicine,* 8, 116-122.

Lubans, D.R., Morgan, P. Callister, R., Plotnikoff, R.C., Narelle, E, Riley, N., et al. (2011). Test-retest reliability of a battery of field-based health-related fitness measures for adolescents. *Journal of Sports Sciences, 29 (7),* 685-693.

Luoto, S. Heliövaara, M., Hurri, H., & Alaranta, H. (1995). Static back endurance and the risk of low- back pain. *Clinical Biomechanics,10,* 323-324.

Macrae, I.F., & Wright, V. (1969). Measurement of back movement. *Annals of Rheumatoid Disease,28,* 584-589.

Magnusson, L.I. (1957). The effect of a specific activity program on children with low muscular fitness.Unpublished doctoral dissertation, State University of Iowa, Ames.

Magnussen, C.G., Schmidt, M.D., Dwyer, T., & Venn. A. (2012). Muscular fitness and clustered cardiovascular disease risk in Australian youth. *European Journal of Applied Physiology, 112,* 3167-3171.

Magyari, P.M. & Churilla, J.R.(2012). Association between lifting weights and metabolic syndrome among U.S. Adults: 1999-2004 National Health and Nutrition Examination survey. *Journal of Strength and Conditioning Research, 26(11),* 3113-3117.

Marshall, L.W., & McGill, S.M. (2010). The tole of axial torque in disc herniation. *Clinical Biomechanics, 25(1),* 6-9.

Martinez-Gomez, D., Eisenmann, J.C., Gomez-Martinez, S., Veses, A., Romeo, J., Veiga, O.L., et al. (2012). Associations of physical activity and fitness with adipocytokines in adolescents. *Nutrition, Metabolism and Cardiovascular Diseases, 22 (3),* 252-259.

Mason,C., Brien, S.E., Craig, C.L., Gauvin, L., & Katzmarzyk, P.T. (2007). Musculoskeletal fitness and weight gain in Canada. *Medicine & Science in Sports & Exercise, 39,* 38-43.

Massicotte, D. (1990). Partial curl-ups, push-ups, and multistage 20 meter shuttle run, national norms for 6 to 17 year olds. Final report submitted to: Canadian Association for Health, Physical Education, and Recreation (CAHPER) and Fitness and Amateur Sport Canada. University of Quebec, Montreal.

Mathews, D.K., Shaw, V., & Bohnen, M. (1957). Hip flexibility of college women as related to length of body segments. *Research Quarterly, 28,* 352-356.

Mathews, D.K., Shaw, V., & Woods, J.B. (1959). Hip flexibility of elementary school boys as related to body segments. *Research Quarterly, 30,* 297-302.

Mbada, C.E., Ayanniyi, O., & Adedoyin, R.A. (2009). Reference values of static back extensor muscle endurance in healthy Nigerian adults. *Medical Principles and Practice,* 18(5), 345-350.

McGill, S.M. (2001). Low Back Stability: From formal description to issues for performance and rehabilitation. *Exercise and Sport Sciences Reviews, 29(1),* 26-31.

McGill, S.M. (2007). Low Back Disorders: Evidence-Based Prevention and Rehabilitation (2nd ed.). Champaign, IL: Human Kinetics.

McManis, B.G. Baumgartner, T.A. & West, D.A. (2000). Objectivity and reliability of the 90^{o} pushup test. *Measurement in Physical Education and Exercise Science, 4,* 57-67.

McManis, B.G., & Wuest, D. A. (1994). Stability reliability of the modified push-up in children.*Research Quarterly for Exercise in Sport, 65*(Suppl.), A58-A59. [Abstract].

Meredith, M.D., & Welk , G.J. (2010). *FITNESSGRAM® & ACTIVITYGRAM® Test Administration Manual*. (updated 4th ed.) Champaign, IL: Human Kinetics.

Mier, C.M. (2011). Accuracy and feasibility of video analysis for assessing hamstring flexibility and validity of the sit-and-reach test. *Research Quarterly for Exercise and Sport,* 82 (4), 617-623.

Mikkelsson, L.O., Nupponen, H., Kaprio, J., Kautiainen, H., Mikkelsson, M., & Kujala, U.M. (2006). Adolescent flexibility, endurance strength, and physical activity as predictors of adult tension neck, low back pain, and knee injury: A 25 year follow up study. *British Journal of Sports Medicine,* 40,107-113.

Minkler, S., & Patterson, P. (1994). The validity of the modified sit-and-reach test in college-age students. *Research Quarterly for Exercise and Sport, 65,* 189-192.

Miura, H., & Aoki, S. (2005). Influence of low-intensity circuit training on artery stiffness in females. *Japanese Journal of Physical Fitness and Sports Medicine,* 54(3), 205-210.

Moliner-Urdiales, D., Ortega, F.B., Vicente-Rodriquez, G., Rey-Lopez, J.P., Gracia-Marco, L., Widhalm, K., et al. (2010). Association of physical activity with muscular strength and fat-free mass in adolescents: the HELENA Study. *European Journal of Applied Physiology,published online. DOI 10.1007/s00421-010-1457-z.*

Moliner-Urdiales, D., Ruiz, J.R., Vicente-Rodriguez, G., Ortega, F.B., Rey-Lopez, J.P. Espana-Romero, V., et al. (2009). Associations of muscular and cardiorespiratory fitness with total and central body fat in adolescents: the HELENA Study. *British Journal of Sports Medicine 45*, 101-108.

Monfort-Pañego, M., Vera-García, F.J., Sánchez-Zunaga, D.& Sarti-Martínez, M.A. (2009). Electromyographic studies in abdominal exercises: A literature synthesis. *Journal of Manipulative Physiological Therapy*, 32, 232-244.

Moreau, C.E., Green, B.N., Johnson, C.D., & Moreau, S.R. (2001). Isometric back extension endurance tests: A review of the literature. *Journal of Manipulative and Physiological Therapeutics,* 24(2), 110-122.

Moreira, C., Santos, R., Vale, S., Soares-Miranda, L., Marques, A.I., Santos, P.C. et al. (2010). Metabolic syndrome and physical fitness in a sample of Azorean adolescents. *Metabolic Syndrome and Related Disorders,* 8(5), 443-449.

Morris, F.L., Naughton, G.A., Gibbs, J.L., Carlson, J.S., & Wark, J.D. (1997). Prospective ten-month exercise intervention in premenarcheal girls: Positive effects on bone and lean mass. *Journal of Bone and Mineral Research, 12,* 1453-1462.

Morrow, J.R., Going, S.B., & Welk, G.J. (Eds.) (2011). FITNESSGRAM® Development of criterion-referenced standards for aerobic capacity and body composition. *American Journal of Preventive Medicine*, 41(4) Supplement 2, S63-S143.

Mota, J., Vale, S., Martins, C., Gaya, A., Moreira, C., Santos, R., et al. (2010). Influence of muscle fitness test performance on metabolic risk factors among adolescent girls. *Diabetology & Metabolic Syndrome*, online at http://www.dmsjournal.com/content/2/1/42.

Müller, R., Strässle, K., & Wirth, B. (2010). *Journal of Electromyography and Kinesiology*, 20, 845-850.

Murr, M.S. (1997). Objectivity and reliability of two push-up test protocols for male college students.Unpublished doctoroal dissertation, University of Georgia, Athens.

Mutoh, Y., Mori, T., Nakamura, Y., & Miyashita, M. (1981). The relationship between sit-up exercises and the occurrence of low back pain. In H. Matsui & K. Kobayashi (Eds.). *Biomechanics VIII-A* (pp.180-185). Champaign, IL: Human Kinetics.

Nachemson, A., & Elfström, G. (1970). Intravital dynamic pressure measurements in lumbar discs: A study of common movements, maneuvers and exercises. Stockholm: Almqvist & Wiksell.

Newcomer K. & M. Sinaki. (1996). Low back pain and its relationship to back strength and physical activity in children. *Acta Paediatrica*, 85, 1433-1439.

Nicolaisen, T., & Jorgensen, K. (1985). Trunk strength, back muscle endurance and low-back trouble.*Scandinavian Journal of Rehabilitation Medicine*, *17,* 121-127.

Noble, L. (1975). A new curl-up test of abdominal endurance. Washington, DC: AAHPERD. (FromAAHPER Abstracts, Atlantic City, NJ)

Noble, L. (1981). Effects of various types of situps on iEMG of the abdominal musculature. Journal ofHuman Movement Studies, 7, 124-130.

Nordgren B., Schéle, R., & Linroth, K. (1980). Evaluation and prediction of back pain during military field service. *Scandinavian Journal of Rehabilitative Medicine*, 12, 1-8.

Norris, C.M. (1993). Abdominal muscle training in sport. *British Journal of Sports Medicine*, 27 (1), 19-27.

Nourbakhh, M., & Arab, A. (2002). Relationship between mechanical factors and incidence of low back pain. *Journal of Orthopaedics and Sports Physical Therapy*, 32, 447-460.

O'Connell, D.G., O'Connell, J.K., Garrett, M.L., Adams, N., Patterson, B., & Spencer, E. (2004). Isometric strength and dynamic back extensor endurance are unrelated in children ages 6-10 years: A pilot study. *Perceptual and Motor Skills, 99,* 1290-1294.

Oglesby, B., Pabst, P., Layes, J., & DiBrezzo, R. (1989). A comparison of children's upper body strength assessments: A preliminary study. *Medicine and Science in Sports and Exercise, 21(2)(*Suppl.), S111.

Olson, T.P., Dengel, D.R., Leon, A.S., & Schmitz, K.H. (2007). Changes in inflammatory biomarkers following one-year of moderate resistance training in overweight women. *International Journal of Obesity*, 31(6), 996-1003.

Ortega, F.B., Artero, E.G., Ruiz, J.R., Vicente-Rodriguez, G., Bergman, P., Hagströmer, M. et al. (2008a). Reliability of health-related physical fitness tests in European adolescents. The HELENA study. *International Journal of Obesity*, 32, 549-557.

Ortega, F.B., Ruiz, J.R., Castillo, M.J.,& Sjöström, M. (2008b). Physical fitness in childhood and adolescence: A powerful marker of health. *International Journal of Obesity*, 32, 1-11.

Parfrey, K.C., Docherty, D., Workman, R.C., & Behm, D.G. (2008). The effects of different sit- and curl-up positions on activation of abdominal and hip flexor musculature. *Applied Physiology, Nutrition and Metabolism,* 33, 888-895.

Pate, R.R., Burgess, M.L., Woods, J.A., Ross, J.G., & Baumgartner, T. (1993). Validity of field tests of upper body muscular strength. *Research Quarterly for Exercise and Sport, 64,* 17-24.

Pate, R.R., Ross, J.G., Baumgartner, T.A., & Sparks, E. (1987). The modified pull-up test. Journal of *Physical Education, Recreation & Dance, 58(9),* 71-73.

Patterson, P., Bennington, J., & de LaRosa, T. (2001). Psychometric properties of child-and teacher-reported curl-up scores in children ages 10-12 years. *Research Quarterly for Exercise and Sport*, 72(2), 117-124.

Patterson, P., Rethwisch, N., & Wiksten, D. (1997). Reliability of the trunk lift in high school boys and girls. *Measurement in Physical Education and Exercise Science, 1,* 145-151.

Patterson, P., Wiksten, D.L., Ray, L., Flanders, C., & Sanphy, D. (1996). The validity and reliability of the back saver sit-and-reach test in middle school girls and boys. *Research Quarterly for Exercise and Sport, 67,* 448-451.

Payne, N., Gledhill, N., Katzmarzyk P.T., & Jamnik,,V. (2000a). Health-related fitness, physical activity, and history of back pain. *Canadian Journal of Applied Physiology, 25,* 236-249.

Payne, N., Gledhill, N., Katzmarzyk P.T., Jamnik, V., & Ferguson, S. (2000b). Health implications of musculoskeletal fitness. *Canadian Journal of Applied Physiology, 25*,114-126.

Payne, N., Gledhill, N., Katzmarzyk, P.T., Jamnik, V.K., & Kier, P.J. (2000c). Canadian musculoskeletal fitness norms. *Canadian Journal of Applied Physiology,* 25(6), 430-442.

Pizzigalli, L., Filippini, A., Ahmaidi,S., Jullien, H., & Rainoldi, A. (2011). Prevention of falling risk in elderly people: The relevance on muscular strength and symmetry of lower limbs in postural stability. *Journal of Strength and Conditioning Research, 25 (2*), 567-574.

Plowman, S.A. (1992a). Criterion referenced standards for musculoskeletal physical fitness tests: An analysis. *Pediatric Exercise Science, 4,* 10-19.

Plowman, S.A. (1992b). Physical activity, physical fitness, and low back pain. In J.O. Holloszy (Ed.),*Exercise and sport sciences reviews, 20*, 221-242. Baltimore: Williams & Wilkins.

Plowman, S.A. (1993). Physical fitness and healthy low back function. In C. Corbin & R. Pangrazi (Eds.), *Physical activity and fitness research digest, 1(3)*. Washington, DC: President's Council on Physical Fitness and Sports.

Poley, M. (1948). Postural characteristics of college women as related to build. Unpublished doctoral dissertation, University of Iowa, Iowa City.

Rantanen, T., Volpato, S., Ferrucci, L., Heikkinen, E., Fried, L.P., & Guralnik, J.M. (2003). Handgrip strength and cause-specific and total mortality in older disabled women: Exploring the mechanism. *Journal of the American Geriatrics Society,* 51(5), 636-641.

Reiff, G.G., Dixon, W.R., Jacoby, D., Ye, G.X., Spain, C.G., & Hunsicker, P.A. (1986). The President's Council on Physical Fitness and Sports *1985 National School Population Fitness Survey* (HHS - Office of the Assistant Secretary for Health Research Project No. 282-84-0086). Ann Arbor, MI: The University of Michigan.

Ricci, B., Marchetti, M., & Figura, F. (1981). Biomechanics of sit-up exercises. Medicine and Science in Sports and Exercise, 13, 54-59.

Robertson, L.D., & Magnusdottir, H. (1987). Evaluation of criteria associated with abdominal fitness testing. *Research Quarterly for Exercise and Sport, 58,* 355-359.

Rodriguez-Garcia, P.L., López-Miñarro, P.A., Yuste, J.L., & de Baranda, P.S. (2008). Comparison of hamstring criterion-related validity sagittal spinal curvatures, pelvic tilt and score between sit-and-reach and toe-touch test in athletes. *Medicina Dello Sport*, 61(1), 11-20.

Ross, J.G., Dotson, C.O., Gilbert, G.G., & Katz, S.J. (1985). New standards for fitness measurement. *Journal of Physical Education, Recreation and Dance, 56(*1), 62-66.

Ross, J.G., Pate, R.R., Delpy, L.A., Gold, R.S., & Svilar, M. (1987). New health - related fitness norms.*Journal of Physical Education, Recreation and Dance, 58(9),* 66-70.

Roy, S.H., DeLuca, C.J., & Casavant, D.A. (1989). Lumbar muscle fatigue and chronic lower back pain.*Spine, 14*, 992-1001.

Roy, S.H., DeLuca, C.J., Snyder-Mackler, L., Emley, M.S., Crenshaw, R.L., & Lyons, J.P. (1990). Fatigue, recovery, and low back pain in varsity rowers. *Medicine and Science in Sports and Exercise, 22*, 463-469.

Ruiz, J.R., Castro-Piñero, J., Artero, E.G.,Ortega, F.B., Sjöström, M., Suni, J.,et al. (2009). Predictive validity of health-related fitness in youth: A systematic review. *British Journal of Sports Medicine,* 43, 909-923.

Ruiz, J.R., Castro- Piñero, J., España-Romero, V., Artero, E.G., Ortega, F.B., Cuenca, M.M. et al. (2011). Field-based fitness assessment in young people: The ALPHA health-related fitness test battery for children and adolescents. *British Journal of Sports Medicine, 45,* 518-524.

Ruiz, J.R., Ortega,F.B., Warnberg, J.,Moreno, L.A., Carrero, J.J., Gonzalez-Gross, M., et al. (2008). Inflammatory proteins and muscle strength in adolescents. *Archives of Pediatrics & Adolescent Medicine,* 162(5), 462-468.

Ruiz, J.R., Sui, X., Lobelo, F., Morrow, J.R., Jr., Jackson, A.W., Sjöström, M, et al. (2008). Association between muscular strength and mortality in men: Prospective cohort study. *British Medical Journal, 337*, a439.

Rutherford, W.J., & Corbin, C.B. (1993). Measuring upper body strength and endurance: Which test is best? *Kentucky AHPERD Journal*, Fall, 20-24.

Rutherford, W.J., & Corbin, C.B. (1994). Validation of criterion-referenced standards for tests of arm and shoulder girdle strength and endurance. *Research Quarterly for Exercise and Sport,65*, 110-119.

Saint Romain, B. & Mahar, M.T. (2001). Norm-referenced and criterion-referenced reliability of the push-up and modified pull-up. *Measurement in Physical Education and Exercise Science, 5*, 67-80.

Safrit, M.J., & Wood, T.M. (1987). The test battery reliability of the health-related physical fitness test.*Research Quarterly for Exercise and Sport, 58*, 160-167.

Safrit, M.J., Zhu, W., Costa, M.G., & Zhang, L. (1992). The difficulty of sit-up tests: An empirical investigation. *Research Quarterly for Exercise and Sport, 63*,277-283.

Sakuragi, S., Abhayaratna, K., Gravenmaker, K.J., O'Reilly, C., Srikusalanukul, W., Budge, M.M., et al. (2009). Influence of adiposity and physical activity on arterial stiffness in healthy children. *Hypertension* 53, 611-616.

Saland, J.M. (2007). Update on the metabolic syndrome in children. *Current Opinion in Pediatrics.* 19 (2), 183-191.

Salminen, J.J., Erkintalo, M. Laine, M., & Pentti, J.(1995). Low back pain in the young: A prospective three-year follow-up study of subjects with and without low back pain. *Spine,* 20 (19), 2101-2108.

Sasaki, H., Kasagi, F.,Yamada, M.,&Fujita, S.(2007). Grip strength predicts cause-specific mortality in middle-aged and elderly persons. *The American Journal of Medicine*, 120 (4), 337-342.

Scott, G., & French, E. (1959). *Measurement and evaluation in physical education*. Dubuque, IA: W.C. Brown.

Seals, D.R. (2003). Habitual exercise and the age-associated decline in large artery compliance. *Exercise and Sport Sciences Reviews*, 31(2), 68-72.

Seals, D.R., DeSouza, C.A., Donato, A.J., & Tanaka, H. (2008). Habitual exercise and arterial aging. *Journal of Applied Physiology*, 105(4), 1323-1332.

Sinclair, A., & Tester, G. (1993). The sit and reach test - what does it actually measure? *The Australian Council for Health, Physical Education, and Recreation National Journal, 115*, 8-13.

Sjölie, A.N. & Ljunggren, A.E. (2001).The significance of high lumbar mobility and low lumbar strength for current and future low back pain in adolescents. *Spine*, 26(23), 2629-2636.

Smith, E.L., & Gilligan, C. (1987). Effects of inactivity and exercise on bone. *The Physician and Sportsmedicine, 15(11),* 91-100.

Sparling, P.B., Millard-Stafford, M., & Snow, T.K. (1997). Development of a cadence curl-up test for college students. *Research Quarterly for Exercise and Sports, 68*, 309-316.

Steene-Johannessen, J., Anderssen, S.A., Kolle, E., & Andersen, L.B. (2009). Low muscle fitness is associated with metabolic risk in youth. *Medicine & Science in Sports & Exercise*, 41, 1361-1367.

Steinberger, J., Daniels, S.R., Eckel, R.H., L. Hayman, R.H. Lustig, B. McCrindle, et al. (2009). Progress and challenges in metabolic syndrome in children and adolescents: A scientific statement from the American Association Atherosclerosis, Hypertension, and Obesity in the young committee of the council on cardiovascular disease in the young; Council on cardiovascular nursing; and council on nutrition, physical activity and metabolism. *Circulation.* 119 (4), 628-647.

Strasser, B., Siebert, U. & Schobersberger, W. (2010). Resistance training in the treatment of the Metabolic Syndrome. *Sports Medicine*, 40 (5), 397-415.

Sugawara, J., Otsuku, T., Tanabe, T., Hayashi, K., Maeda, S., Kuno, S., et al. (2006). The effects of daily physical activity on the age-related carotid arterial stiffening in middle-aged and elderly people. *Japanese Journal of Physical Fitness and Sports Medicine*, 55, 11-14 S.

Suni, J.H., Oja, P., Miilunpalo, S.I., Pasanen, M.E., Vuori,I.M., & Bos. K.(1998). Health-related fitness test battery for adults: Association with perceived health, mobility, and back function and symptoms. *Archieves of Physical Medicine and Rehabilitation, 79 (5*),559-569.

Taanila, H.P., Suni, J.H., Pihlajamaki, H.K., Mattila, V.M., Ohrankammen, O., Vuorinen, P., et al. (2012). Predictors of low back pain in physically active conscripts with special emphasis on muscular fitness. *Spine Journal, 12 (9),* 737-748.

Tampier, C., Drake, J.D.M., Callaghan, J.P. & McGill, S.M. (2007). Progressive disc herniation: An investigation of the mechanisms using radiologic, histochemical, and microscopic dissection techniques on a porcine mode. *Spine, 32(25),* 2869-2874.

Tanaka, H., DeSouza, C.A., & Seals, D.R. (1998). Absence of age-related increase in central arterial stiffness in physically active women. *Arterisclerosis, Thrombosis, and Vascular Biology*, 18, 127-132.

Tanaka, H., Dinenno, F.A., Monahan, M.S., Clevenger, C.M., DeSouza, C.A., & Seals, D.R. (2000). Aging, habitual exercise, and dynamic arterial compliance. *Circulation*, 102, 1270-1275.

Tittel, K. (1990). The loadability and relieveability of the lumbo-sacral transition in sports. *The Journal of Sports Medicine and Physical Fitness, 30(2),* 113-121.

Tomson, L.M. (1992). A field trial of strength and endurance for primary grade children. Unpublished master's thesis, Arizona State University, Tempe, AZ.

Troup, J.D.G., Foreman, T.K., Baxter, C.E.,& Brown,D.,(1987). The perception of back pain and the role of psychophysical tests of lifting capacity. *Spine*, 12 (7), 645-657.

Troup, J.D.G., Martin, J.W.,& Lloyd, D.C. (1981). Back pain in industry: A prospective survey. *Spine,* 6(1), 61-69.

Tsigilis, N., Douda, H., & Tokmakidis, S.P. (2002): Test-retest reliability of the Eurofit test battery administered to university students. *Perceptual and Motor Skills, 95,* 1295-1300.

Twisk, J.W.R., Kemper, H.C.G., & van Mechelen, W. (2000). Tracking of activity and fitness and the relationship with cardiovascular disease risk factors. *Medicine & Science in Sports & Exercise,32,*1455-1461.

vanAdrichem, J.A.M., & vanderKorst, J.K. (1973). Assessment of the flexibility of the lumbar spine. Scandinavian Journal of Rheumatology, 2, 87-91.

van der Heijden, G.J., Wang, Z.Y.J., Chu, Z.L., Toffolo, G., Mancesso, E., Sauer, P. J.J., et al. (2010). Strength exercise improves muscle mass and hepatic insulin sensitivity in obese youth. *Medicine & Science in Sports & Exercise*, *42(11)*, 1973-1980.

Vicente-Rodriguez,G., Ara, I.,Perez-Gomez, Serrano-Sanches, J.A.,Dorado, C..& Calbet, J.A.L. (2004). High femoral bone mineral density accretion in prepubertal soccer players. *Medicine & Science in Sports & Exercise, 36 (10)*, 1789-1795.

Vicente-Rodriquez, G., Urzanqui, A., Mesana, M.I., Ortega, F.B., Ruiz, J.R., Ezquerra, J., et al. (2008). Physical fitness effect on bone mass is mediated by the independent association between lean mass and bone mass through adolescence: a cross-sectional study. *Journal of Bone and Mineral Metabolism, 26*,288-294.

Vincent, W.J., & Britten, S.D. (1980). Evaluation of the curl-up - a substitute for the bent knee situp.*Journal of Physical Education and Recreation, 51(2),* 74-75.

Waldhelm, A.. (2011) Assessment of core stability: Developing practical models. Doctoral dissertation. Louisiana State University, Department of Kinesiology.

Walker, J.L., Lloyd, L.K., Bishop, P.A., & Richardson, M.T. (2000). The influence of body size and composition on the successful completion of the FITNESSGRAM® pull-up test in fifth-and sixth-grade children. *Research Quarterly for Exercise and Sport*, *71* (Suppl.), A-54.

Warburton, D.E.R., Gledhill, N. & Quinney, A. (2001a). The effects of changes in musculoskeletal fitness on health. *Canadian Journal of Applied Physiology,* 26, 161-216.

Warburton, D.E.R., Gledhill, N. & Quinney, A. (2001b). Musculoskeletal fitness and health. *Canadian Journal of Applied Physiology,* 26, 217-237.

Warburton, D.E.R., Nicol, C.W. & Bredin, S.D. (2006). Health benefits of physical activity: The evidence. *Canadian Medical Association Journal,* 174, 801-809.

Wear, C.L. (1963). Relationship of flexibility measurements to length of body segments. *Research Quarterly, 34,* 234-238.

Wells, K.F., & Dillon, E.K. (1952). The sit and reach - a test of back and leg flexibility. *Research Quarterly,* 23, 115-118.

Westcott, W.L. (2012). Resistance Training is Medicine: Effects of strength training on health. *Current Sports Medicine Reports, 11(4),* 209-216.

Wijndaele, K., Duvigneaud, N., Matton, L., Duquet, W., Thomis, M., Beunen, G., et al.(2007). Muscular strength, aerobic fitness, and metabolic syndrome risk in Flemish adults. *Medicine & Science in Sports & Exercise,* 39, 233-240.

Wolfe, R.R. (2006). The underappreciated role of muscle in health and disease. *American Journal of Clinical Nutrition,* 84, 475-482.

Wood, H.M. & Baumgartner, T.A. (2004). Objectivity, reliability, and validity of the bent-knee push-up for college-aged women. *Measurement in Physical Education and Exercise Science, 8,*203-212.

Yamamoto, K., Kawano, H., Gando, Y., Lemitsu, M., Murakami, H., Sanada, K., et al. (2009). Poor trunk flexibility is associated with arterial stiffening. *American Journal of Physiology-Heart and Circulatory Physiology*, 297, H1314-H1318.

Youdas, J.W., Krause, D.A., & Hollman, J.H. (2008). Validity of hamstring muscle length assessment during the sit-and-reach test using an inclinometer to measure hip joint angle. *Journal of Strength and Conditioning Research,* 22 (1), 303-309.

Zhu, W., Plowman, S.A., Park, Y. (2010). A primer-test centered equating method for setting cut-off scores. *Research Quarterly for Exercise and Sport*, 81(4), 400-409.

Zimmet, P., A.K. George, F. Kaufman, N. Tajima, M. Silink, S. Arslanian, et al. (2007). The metabolic syndrome in children and adolescents—an IDF concensus report. *Pediatric Diabetes,* 8, 299-306.

Zorn, R.L. (1992). Selected tests of muscular strength and endurance for children. Unpublished master's thesis, Arizona State University, Tempe, AZ.

Chapter 9
Interpreting FITNESSGRAM® and ACTIVITYGRAM® Reports

**Marilu D. Meredith, Dolly Lambdin, Georgi Roberts,
Gregory J. Welk, Jim Morrow**

The FITNESSGRAM® Reference Guide is intended to provide answers to some common questions associated with use and interpretation of FITNESSGRAM® assessments. This chapter focuses on how to interpret information that is provided on the FITNESSGRAM® and ACTIVITYGRAM® reports. The following questions are specifically addressed:

⇦TOC ⇦Chapter

The FITNESSGRAM®/ACTIVITYGRAM® program was created 30+ years ago to provide students and parents with easily accessible information about a student's fitness levels. A key to the vision was the creation of the personalized FITNESSGRAM® Report. If fitness testing is to be effective in motivating efforts for change, students and their parents need to not only be informed of the results but also to be guided to behaviors that should result in improvement in the needed areas. The FITNESSGRAM® report fills this need by providing youth and parents with personalized information about the results of the FITNESSGRAM® assessment.

The report is specifically designed to communicate to students and their parents the student scores on various fitness components, how these scores relate to health, and what steps can be taken to improve in areas where a need for improvement is indicated. Having students participate in fitness testing without appropriate communication of the meaning of the results and plans for improvement is an inappropriate teaching practice (National Association for Sport and Physical Education, 2009). All students participating in fitness testing should be able to identify their strengths related to the various fitness components and activities that are best associated with improvement in areas where the need for improvement is indicated. The FITNESSGRAM® report is one way to communicate this information to students and parents. In addition teachers should help students with appropriate goal setting, regular opportunities to engage in appropriate physical activity, and formats for performance tracking to provide the best opportunity for appropriate fitness development.

How Does FITNESSGRAM® Evaluate Fitness Performance?

FITNESSGRAM® uses criterion-referenced standards to evaluate fitness performance. Many of these standards have been established to represent a level of fitness that is associated with some degree of protection against chronic disease. The FITNESSGRAM® Report communicates where a child's score on each fitness component falls in relation to the criterion standard. Performance on musculoskeletal components (strength, endurance, and flexibility) is classified in two general areas: the "Healthy Fitness Zone" (HFZ) and the "Needs Improvement" (NI) Zone. Performance on the body composition and aerobic capacity components is classified in three general areas: the "Healthy Fitness Zone" (HFZ), the "Needs Improvement" Zone, and a "Needs Improvement-Health Risk" Zone. Examples of each are provided here.

A score in the HFZ represents the level of fitness believed to provide some protection from the potential health risks imposed by a lack of fitness in this measure. It is not uncommon for children to achieve the HFZ for some dimensions of fitness but not for others. The FITNESSGRAM® report provides a clear depiction of whether a child's score is in the HFZ for each of the assessments. The printed portion of the report provides positive feedback on assessments in which the child achieved the HFZ.

The NI zone (the category below the HFZ) indicates a level of fitness that is below the minimal health standard. While the effect of low fitness often may not influence health until later in adulthood, it is important to identify potential risks early on so that adjustments can be made to improve in these areas. The text in the report provides this type of feedback in a constructive and prescriptive way so that children can set goals or targets to improve their fitness. It is important to clarify that the wording used for this category does not imply "bad fitness" or "poor fitness" but rather areas in which the child should seek improvement.

The third zone within the "Needs Improvement" zone is used for both Aerobic Capacity and Body Composition. There is still a HFZ that indicates a level that is associated with good health but the NI area is sub-divided into "Needs Improvement" and "Needs Improvement-Health Risk" (NI-HR). Students in both NI zones should aim to move into the HFZ, but youth in the NI-HR zone are at greater risk of chronic health conditions such as diabetes and cardiovascular disease in the future.

The standards for the HFZ represent minimal levels of fitness associated with good health. Students who desire to achieve a high level of athletic performance may need to consider setting goals well beyond the beginning levels of the HFZ. From a similar perspective, standards are not presented for students in grades K-3. This is both because of the challenges associated with determining standards and a philosophical decision by the FITNESSGRAM® Scientific Advisory Board. Performance levels are not the most important objective for young children in this age range. Instead, the emphasis for young children should be on enjoying activity, developing basic movement skills, and on learning to perform the test items successfully.

Research findings were used as the basis for establishing the FITNESSGRAM® health fitness standards.

---

1 Material adapted from the FITNESSGRAM®/ACTIVITYGRAM® Test Administration Manual.

How Do I Interpret the FITNESSGRAM Reports?

The goal in FITNESSGRAM® is to help youth develop lifelong habits of physical activity and to have sufficient fitness for good health. The feedback on the FITNESSGRAM® report is intended to help youth (and parents) learn about their personal level of health related fitness. The feedback is individualized in that the messages that are provided depend on the overall fitness profile for the given child. The feedback messages also vary depending on whether the physical activity questions are assessed [see section below]. In general, for each of the fitness components students that achieve the HFZ are provided with information on how to maintain their fitness over time while students in the NI zone are provided with information about how to improve their fitness. As described above, the revised standards for aerobic capacity and body composition have two different NI Zones. The additional NI-HR zone provides youth and parents with an appropriate warning that the level of fitness increases the child's risk of health problems.

It is hoped that parents will review the report with the child, celebrating the areas where scores indicate the HFZ and reviewing the advice provided on how to improve in areas of need. The next steps should be creating a plan to act on the advice provided, following the plan and, after a reasonable time, retesting to determine progress toward the HFZ.

It is important to recognize that the standards for the HFZ represent minimal levels of fitness associated with good health. Therefore, the standards should be attainable by most children that participate regularly in various types of physical activity. Because of this, we recommend that all students should strive to achieve a score that places them inside the HFZ. However, it is not uncommon for children to achieve the HFZ for some dimensions of fitness but not for others. Most children usually have areas that they excel in more than others.

⇦TOC ⇦Chapter

It is also important to point out that the reports do not provide information about a child's performance compared to other youth or information related to sports or athletic potential. Students who desire to achieve a high level of athletic performance may need to consider setting goals well beyond the beginning levels of the HFZ.

Performance levels are not the most important objective for young children in grades K-3. Instead the emphasis for young children should be on enjoying activity, developing motor skills, and on learning to perform the test items successfully.

Fitness is multidimensional and each dimension has different influences on health. Some background information is provided to assist in interpreting aerobic capacity assessments, body composition assessments, and musculoskeletal assessments

Feedback from the Aerobic Capacity Assessments

Aerobic capacity indicates the ability of the respiratory, cardiovascular, and muscular systems to take up, transport, and utilize oxygen during exercise and activity. A laboratory measure of VO_2max is generally the best measure of aerobic capacity. In addition to providing the actual score on the One Mile Run/Walk, the PACER, or the Walk Test, FITNESSGRAM® calculates an estimated VO_2max that may be used to compare performance from one test date to another on the two different test items. The VO_2max value is estimated using equations that take test performance, age, gender, and BMI into account. See Chapter 6, Aerobic Capacity Assessments, for details on the derivation of the aerobic capacity estimate and the processes used to establish and match the standards.

A low score on the field test estimates of aerobic capacity may be influenced by many factors. These include:

- actual aerobic performance level,
- body composition,
- running/walking efficiency and economy,
- motivation level during the actual testing experience,
- extreme environmental conditions,
- ability to pace on the one mile run/walk, and
- genetic makeup.

While genetic factors clearly cannot be changed, improvement in any of the other factors may lead to an improved test score. The amount of potential improvement is related to the beginning level of fitness and to the intensity, duration, and frequency of the training. Aerobic capacity can be improved substantially in an unfit person who begins to participate in sustained activities involving large muscle groups. However, the majority of the improvement will occur during the first six months. Thereafter, improvement will be much slower. It is also important to note that some individuals respond to training more quickly and easily than others.

Boys and girls who are over-fat may expect an improvement in the aerobic capacity measure with a reduction in body fat. For boys, aerobic capacity relative to body weight stays relatively constant during the growing years. For girls, aerobic capacity tends to remain constant between ages 5 and 10 years but decreases after age 10 years due to increasing gender-specific, essential fat.

One Mile Run and PACER test scores tend to improve progressively with age in boys even though VO_2max expressed relative to body weight tends to remain constant because running economy improves. In girls up to ages 10-12 years, these test scores also tend to improve due to improved running economy. Between ages 12 and 18, scores for girls tend to remain relatively constant because improved running economy is offset by declining VO_2max expressed relative to body weight. The differences in age-related changes in the relation of the One Mile Run or PACER test scores to running economy are taken into account in the formula that is used to estimate VO_2max in the FITNESSGRAM® program software.

Feedback from the Body Composition Assessments

Body composition standards have been established for both percent body fat and body mass index. It is important to note that the body fat and BMI exhibit different gender related changes during child development. Boys tend to gain muscle and lose fat as they move through puberty. Girls tend to see increases in body fat levels as they mature. These are normal developmental changes that are taken into account in establishing the standards. Values that are in the HFZ are considered to be appropriate for good health while those in the NI zones reflect values that would classify a child as being either overweight or obese according to the traditional growth charts. Students in the NI zones receive a message about potential risks associated with overweight status and tips to get into the HFZ. Students who are excessively lean receive a message about potential issues with insufficient body fatness. It is clearly possible for youth to be overweight or excessively lean and still be healthy. Therefore the feedback on the report is intended primarily to promote awareness about the child's current weight status and feedback on potential health issues if their values fall outside of the HFZ.

For boys, the HFZ for percent body fat begins at 18.8–22.2% depending on age (16.7–25.1 BMI). For girls, the HFZ for percent body fat begins at 20.8–31.3 % depending on age (16.7–25.1 BMI). Ideally students should strive to be within the HFZ for their age. A Body Mass Index in the "Needs Improvement" range indicates that the student is too heavy for his/her height. See Chapter 7, Body Composition Assessments, for details on the derivation of the body fat estimations and the processes used to establish and match the standards.

When interpreting body composition scores it is important to remember the following:

- Skinfold measurements and other body fat analysis methods (e.g. bioelectric impedance) provide estimates of body fatness, but there is considerable error in both methods (~2-5%).
- Body mass index provides an estimate of the appropriateness of the weight for the height.
- Body mass index may falsely identify a very muscular, lean person as being over fat (too heavy for height) or identify a light weight person with little muscular development but a large percent fat as being acceptable when they are actually over fat.

The limitations of BMI are well documented but the use of BMI provides a reasonable indicator of body composition and potential health risk for the majority of the population. The feedback should be interpreted carefully since individuals that are heavily muscled could be incorrectly classified as being overweight (false positive) and individuals with high fat and very low muscle tone could be classified in the HFZ when they should not be (false negative). Every attempt has been made to ensure that the number of misclassified scores is as minimal as possible and the information provided is valuable to individuals. These risks of false positives and false negatives are an inherent limitation of using BMI measures rather than percent body fat measures.

In general, students who have percent fat values indicating excessive body fat (i.e., not in the HFZ) should be encouraged to work toward the HFZ by slowly changing their body weight through increased physical activity and decreased consumption of high fat, high calorie, low nutritional value foods. Students with severe obesity or eating disorders generally need professional assistance in their attempts to modify these aspects of their lifestyle. It is important to note that health risks from obesity are greatly reduced if the child is physically active [See Chapter 3, Health Benefits of Physical Activity and Physical Fitness in Children].

It is important to remember in interpreting body composition results that most students who are over-fat may also have performances in other test areas that are outside the HFZ. An improvement in body composition will generally result in improved performance in aerobic capacity and also muscle strength and endurance, especially in the upper body, due to a reduction in excess weight and having to lift less weight.

FITNESSGRAM® also identifies students who are very lean. Feedback is provided on the FITNESSGRAM® report to indicate that being this lean may not be best for health. Parents and teachers should notice students who are categorized as being very lean and consider factors that may be responsible for their low level of body fat. Many students may naturally be very lean while others may have inappropriate nutritional patterns. The primary concern related to excessive leanness is that it could indicate malnutrition or signal a potential or current eating disorder. A factor to consider is whether the student's level of fat has suddenly changed from within the HFZ to a level identified as very lean. Severe changes may signal a potential problem. Creating awareness of a child's current status is the primary purpose in identifying lean students. Changes in status should be monitored.

FITNESSGRAM® results can be very helpful in allowing students to follow changes in their levels of body fat over time. Obesity is a health problem both for children and adults. Childhood is the most appropriate time to address problems or potential problems since the likelihood of being obese as an adult increases if one is obese as a child. Through proper referral to medical or weight loss specialists, obese children can be taught to make the necessary behavior changes to manage or control their level of body fatness.

Feedback from the Muscular Strength, Endurance, and Flexibility Assessments

Health-related standards have been established for the various assessments of muscular strength, endurance, and flexibility. Students who score below the HFZ on one or more areas of muscle strength, endurance, and flexibility should be encouraged to participate in exercises and other strengthening and stretching activities that will develop those areas. However, it is essential to remember that physical fitness training is very specific and the areas of the body being tested represent only a fraction of the total body.

To focus on activities that develop the extensors of the arms without equal attention to the flexors of the arms will not accomplish the important objective that is to develop an overall healthy musculoskeletal system. Remember, you must have strength and flexibility in the muscles on both sides of every joint. A useful activity for all students is to identify exercises to strengthen and stretch the muscles at every major joint of the trunk, upper body, and lower body.

Poor performance on the measures of abdominal strength, trunk extensor strength, and flexibility may merit special attention. Muscular strength, endurance, and flexibility are important attributes in a healthy, functioning back. See Chapter 8, Muscular Strength, Endurance, and Flexibility Assessments, for more details on the reliability and validity of these assessments and the rationale for inclusion in the FITNESSGRAM® battery.

How Are the Physical Activity Questions Used in the FITNESSGRAM® Reports?

Research has suggested that physical activity and physical fitness may exert independent effects on health. Because there are many factors influencing physical fitness, it is important to also focus attention on the more modifiable component of physical activity. To acknowledge the importance of physical activity in a child's overall health profile, the individualized feedback on the FITNESSGRAM® reports has been designed to integrate information about both physical activity and physical fitness. Three supplemental activity questions are included in the FITNESSGRAM® software to assess a child's level of involvement in aerobic, strength/endurance, and flexibility activity. The questions are:

1. On how many of the past 7 days did you participate in any physical activity for a total of 60 minutes or more over the course of the day? This would include moderate activities (walking, slow bicycling, or outdoor play), as well as vigorous activities (jogging, active games or active sports such as basketball, tennis, or soccer). (0, 1, 2, 3, 4, 5, 6, 7 days)
2. On how many of the past 7 days did you participate in exercises to strengthen and tone your muscles? This would include exercises such as push-ups, sit-ups, or weightlifting. (0, 1, 2, 3, 4, 5, 6, 7 days)
3. On how many of the past 7 days did you do stretching exercises to loosen up or relax your muscles? This would include exercises such as toe touches, knee bends, or leg stretches. (0, 1, 2, 3, 4, 5, 6, 7 days)

If the three physical activity questions are answered, the individualized feedback provided on the FITNESSGRAM® report factors in the child's specific answers. For example, if a child scores high on fitness but does not appear to be active, he/she receives encouraging information about the need to stay active to maintain his/her fitness. Alternately, if a child scores low on fitness but appears to be active, he/she receives messages encouraging him/her to keep up his/her efforts to be physically active. This information is intended to reinforce to children the importance of being physically active regardless of fitness level. A conceptual matrix that illustrates the basic decision-making algorithms is illustrated below: The actual feedback will be specific for each dimension of fitness (aerobic, musculoskeletal, and body composition) and will be more detailed. The chart is intended to illustrate the general concept used for integrating this information into the feedback algorithms.

Conceptual Framework Used to Integrate Information in the FITNESSGRAM® Report

	Is Child Physically Active?	
Fitness Results	Yes	No
Scores are in the Healthy Fitness Zone?	Congratulations, you are in the Healthy Fitness Zone. You are doing regular physical activity and this is keeping you fit.	Congratulations, you are in the Healthy Fitness Zone. To keep fit it is important that you do regular physical activity.
Scores are NOT in the Healthy Fitness Zone?..............	Even though your scores are not in the Healthy Fitness Zone, you are doing	Your scores were not in the Healthy Fitness Zone. Try to increase

	physical activity. Keep up the good work.	your activity levels to improve your fitness and health.

Note: It is important to note that these supplementary algorithms are activated only if the child completes the physical activity questions. If the questions aren't answered, the feedback is based only on the fitness levels. The actual feedback will be specific for each dimension of fitness (aerobic, musculoskeletal, and body composition) and is more detailed.

It should be pointed out that these questions are not required to produce a report. Children that do not complete the questions will receive feedback based solely on their fitness scores. For more information, see the content on physical activity questions within Chapter 5, Physical Activity Assessments, for descriptions and listings of these questions.

How Do I interpret the ACTIVITYGRAM® Report?[2]

The ACTIVITYGRAM® tool is a comprehensive 3-day record of physical activity behavior. Consistent with the personalized messages in the FITNESSGRAM® report, the ACTIVITYGRAM® report provides personalized information about the child's overall physical activity level (based on the 3 days that were assessed). The report provides information about the total amount of activity performed, a time profile of when they were most active or least active, and a diagram reflecting the types of activities they reported participating in as classified by the Activity Pyramid. The feedback can help children learn more about their activity habits and learn how they can become more physically active. For example, the task of reflecting on their activity habits provides children with experience in self-monitoring and self-evaluation, two important behavioral skills.

When interpreting the results of the ACTIVITYGRAM® it is important to acknowledge the limitations of this assessment. Assessing physical activity is very challenging. When interpreting the assessment, it is important to understand that the reports provide only estimates of activity behavior. In addition to problems with recall, there are additional difficulties that complicate this type of assessment. Children have inherently sporadic activity patterns that are difficult to capture with a self-report instrument. The instruments provide limited lists of possible activities and rely on categorization of activity into discrete time intervals. This may not reflect children's normal physical activity patterns. An additional limitation is that the results of this assessment may not generalize to the child's normal activity pattern. ACTIVITYGRAM® reflects only 2-3 days of activity and experts agree that about 7-14 days of monitoring are required to accurately represent normal activity habits. While these limitations may influence the accuracy of the test, they do not detract from the education value they contribute in the curriculum. While the ACTIVITYGRAM® instrument has been validated with data from objective physical activity monitors, it is not intended to provide precise estimates of the child's level of physical activity. Within the program the ACTIVITYGRAM assessment is viewed primarily as an educational tool to help a child learn about their personal activity patterns and understand the importance of daily physical activity. Click to see a sample ACTIVITYGRAM® report. Descriptions of the feedback on Minutes of Activity, the Time Profile, and the Activity Profile are also provided.

Feedback on Minutes of Activity

The Minutes of Activity section on the ACTIVITYGRAM® report shows the total minutes of activity the child reports on the three days of assessment. The Healthy Activity Zone is set at 2-4 bouts of activity or a total of 60 minutes a day for children and adolescents. No distinction is made between Moderate and Vigorous activity in this assessment (levels 3 and 4). This reinforces to children that physical activity is for everyone and that activity doesn't have to be vigorous to be beneficial.

Feedback on the Time Profile

The Time Profile indicates the times when students reported being physically active. Bouts of moderate and vigorous activity correspond to levels 3 and 4 on the graphical report. Because school time is often out of a student's control, the feedback for this section highlights activity patterns after school and on weekends. For a child to be considered "active" on this section of the report students must have at least one bout of activity after school and two on the weekends. Emphasis in the interpretation of the time profile data is placed on helping students identify times when they could be more active.

Feedback on the Types of Activities Performed

The Activity Profile reveals the different types of activities in which the child reported participation. Feedback is based on whether children participate in activities from different levels of the Activity pyramid. Ideally, children should have some lifestyle activity, aerobic activity, muscle strength and endurance activity, and flexibility activity. Lifestyle activity is recommended for all students (and adults). If students are not performing much activity, it is recommended to first try promoting lifestyle activity. From a health perspective, aerobic activity on the second level can make up for a lack of lifestyle activity on the first level but it is still desirable to promote lifestyle activity among all students. No distinctions are needed between the two types of aerobic activity on the second level. Some children may prefer aerobic activities whereas others may prefer aerobic sports. Participation in either of those categories would ensure that the student is receiving reasonable amounts of aerobic activity. At level 3, distinctions are made between musculoskeletal activity and flexibility activity and students are encouraged to perform some activity from each of these categories. Rest is coded at the top of the pyramid because levels of inactivity should be minimized. The feedback regarding this level does not mention non-discretionary activities like class, homework, eating, or sleeping. Rather, emphasis is placed on making children (and parents) aware of the child's use of discretionary time. For this reason, feedback is provided for the amount of time spent playing computer games or watching television. The cut point of two hours was selected as the standard to correspond with other national standards. Students reporting more than two hours would be provided with a message to recommend reducing the amount of inactive time in the day.

[2] Material adapted from the FITNESSGRAM®/ ACTIVITYGRAM® Test Administration Manual

What Other Reports Are Available in the FITNESSGRAM® Software That Provide Information on Individual Performance?

The FITNESSGRAM®/ACTIVITYGRAM® software provides many reports of individual information. In addition to the student and parent reports of assessment performance there are a number of other reports that are available in the software. Click on any link to see a sample of the report.

- ***FITNESSGRAM® Score Sheet***—these reports provide blank forms that can be used to record student scores during testing. The group score sheet can have the names of the students printed on the sheet if it is generated from the FITNESSGRAM® scores grid.

- ***FITNESSGRAM® Summary Report***—this report includes a summary listing of all student scores for every test event that the student has in the database.
- ***FITNESSGRAM® Achievement of Standards***—this report provides lists of students who achieved the HFZ for specific test items or for a specific number of test events.
- ***FITNESSGRAM® Longitudinal Tracking***—this report is of particular importance as it provides a small graph of each score that the student has in the database for every test event. This report would be excellent in a student portfolio.
- ***ACTIVITYGRAM® Data Sheet***-individual data recording form—this report provides a blank form for students to record their activity information for the three days.
- ***ActivityLog Student Report***—this report provides a print-out of the activity log calendar including number of steps and minutes of activity each day plus an indication of whether the student achieved his/her personal activity goals.
- ***Presidential Active Lifestyle Award (PALA)***—this report provides a list of students achieving the criteria for the PALA award for a specific time period.
- ***Student Certificate***—this report produces a certificate of achievement for each student. The teacher can specify the achievement.
- ***Student Information***—this report is a listing of student ID number, name, birthdate, gender, grade, and username and password.

What Group Reports Are Available in the FITNESSGRAM® Software?

The FITNESSGRAM®/ACTIVITYGRAM® software provides a number of reports that provide information on groups of students.

In the Details and Stats section of the software, the following reports of group data are available:

- ***FITNESSGRAM® Statistical Report***—this report is generated;
- ***ACTIVITYGRAM® Statistics Report***—minutes; and
- ***ACTIVITYGRAM® Statistics report***—% minutes by type of activity.

In the Data Overview section of the software, the following reports are available:

- ***Group Performance Reports***—this report indicates the total students tested and total in the HFZ (also by boys and by girls). It may be generated by administrative unit, grade level, and test component.
- ***Group Achievement of Standards Report***—this report indicates how many and what percentage of the student achieved the HFZ for one, two, three, four, five, or six assessment items.
- ***Percentage Tested Report***—this report assists the group coordinator to determine which schools have completed the input of their test data.

How Do I Use FITNESSGRAM® Data?

The data obtained from FITNESSGRAM® assessments can provide considerable value when used as part of a comprehensive physical education curriculum. The sections below highlight how FITNESSGRAM® assessments and reports can be used by teachers to enhance education of students, for the education of parents, and for programmatic curricular evaluation.

Enhancing Education of Students

1. Teaching students about the different dimensions of physical fitness and the importance of being physically fit (e.g. physical health and safety, mental and emotional health, and cognitive performance).
2. Teaching students how physical activity contributes to each dimension of physical fitness and why it is important to be physically active every day.
3. Teaching students how to measure the different fitness components using the FITNESSGRAM® assessments.
4. Teaching students how to interpret their own levels of physical fitness. For optimal effectiveness we recommend that teachers explain the details of the FITNESSGRAM® report in class as part of an educational activity rather than just distributing them to the students. The list below provides some tips for providing this educational component.
    a. Thoroughly read through the student and parent reports for a couple of students to make sure you understand the reporting format. Try explaining it to a friend until you feel you will be comfortable explaining and discussing the form with students and parents.
    b. Provide students with a copy of their FITNESSGRAM® report, their Longitudinal Tracking Chart, or a copy of their scores on a personalized card that indicate their scores and the healthy zone for each component (Aerobic Capacity HFZ values will only be available on the FITNESSGRAM® report since you need BMI calculation for that).
    c. If they have taken the tests before, have them identify the tests for which the score has improved and discuss what they have done that may have affected their improvement. If they did worse on any component, discuss reasons why they may not have performed as well as in the past (e.g. weather, time of day, not feeling well, etc.).
    d. Have students reflect on which tests they are in the HFZ on and which tests they are not. (Say "Look at each of your scores. If it is in the green section, you are in the Healthy Zone. Do you have one that is in the HFZ? Read the comment in that section. What does it tell you?")

Enhancing Knowledge of Parents

The following list summarizes benefits associated with distribution and education of FITNESSGRAM® to parents.

1. Promoting awareness among parents about the important role they have in helping their child be physically active and the importance of personal fitness for good health.
2. Building advocacy for the importance of physical education for their child. Many parents may not fully grasp how the physical development of their child influences every other aspect of their development as well as their self-esteem. The FITNESSGRAM® report provides a way to educate parents about the importance of physical fitness for good health and wellbeing.

3. Reminding the parents that they have a responsibility to help their child be physically active outside of school. Many parents neglect to remind their child to be physically active every day. The report will remind them that they play a critical role in promoting their child's participation in physical activity.
4. There are multiple ways to educate parents about FITNESSGRAM® and to share the reports. The list below provides some options:
   a. Have an educational session for Back to School Night or Family Fitness Night. Have an educational section (e.g. bulletin boards, PowerPoint, and student demonstration) of the importance of physical fitness, how the components are measured, goals set, and improvement worked toward through the FITNESSGRAM® program.
   b. Include printed parent reports in the parents' folders on teacher conference day.
   c. Provide a parent information session to help parents make sense of the FITNESSGRAM®/ACTIVITYGRAM® reports.
   d. Discuss the goal setting procedure used in class with the students to identify appropriate goals and procedures for working toward their goals.
   e. Help the parents brainstorm how they can help their child work toward their goals.

Enhancing Program Decisions for the Teacher

The following list summarizes ways to use the FITNESSGRAM® reports to help evaluate programming or to evaluate student learning for curricular decision.

1. Begin your analysis by reviewing the FITNESSGRAM® Unit Comparison Report and then looking at the Test Component Comparison Report and/or the Statistics report in the FITNESSGRAM® software. (This is most fun when done with a colleague or two because you begin to get excited about looking at data, but you can do it by yourself as well.) Look at the summary data about the percent of students in the HFZ for each component of fitness.
   a. Determine which components have the lowest percent of your students achieving the Healthy Fitness Zone overall and at each grade level/age.
   b. Reflect on what opportunities they have had to develop that component in your classes and during other times of the day (e.g. are there appropriate opportunities and encouragement at recess, before or after school, something they could do at home?).
   c. Identify activities or protocols that could be modified in your class routines to provide more practice in the areas of need.
   d. Identify how you can provide progression by gradually increasing the level of work in each area.
2. Review the FITNESSGRAM® Summary report.
   a. Identify students who are struggling with each item by highlighting scores that fall below the healthy fitness zone. You do this by looking up the Healthy Zone for each age and then looking at all the scores for that age student in the Summary Report.

b. Reflect on reasons why particular individuals are experiencing trouble in each component.
c. Identify any classes that have more students experiencing challenges than others.

3. Planning Action
   a. For classes that are struggling reflect reasons why they may be having more problems achieving healthy levels of fitness (e.g. time of day they come to PE [right after lunch?], number of children in the class, lack of physical activity time during the day).
   b. For students who are struggling in several components identify a priority for goal setting and strategies for motivating the students to persevere toward their goals. Arrange to talk with individuals and express confidence in their ability to improve and brainstorm ways to work toward their goals.
   c. Choose a few students to focus on and monitor their improvement closely to learn more about the factors that might be helpful in spurring their improvement.
4. Write down your overall goals and plans for improvement based on your reflections. Implement your plan.

What Do the FITNESSGRAM® Standards Actually Mean?

Aerobic Capacity Standards

Numerous studies have documented that physical fitness provides protection against health risks such as diabetes, cardiovascular disease, and some forms of cancer. The conditions tend to primarily affect adults but the conditions originate and can progress during adolescence. The current aerobic capacity standards were established based on a child's risk for developing metabolic syndrome, a precursor to cardiovascular disease and diabetes. Metabolic syndrome is characterized by a clustering of risk factors including abdominal obesity, high triglycerides, high blood pressure, glucose intolerance, and high levels of circulating insulin. Nationally-representative data on metabolic syndrome and aerobic fitness are available from the National Health and Nutrition Examination Survey (NHANES), so this data source was used to develop the standards. The process used made it possible to establish age and gender-specific standards that reflect different levels of risk while also taking into account normal changes during growth and maturation. Specifics for the development of the aerobic capacity, body composition, and musculoskeletal standards can be found in Chapters 6, 7, and 8 of this manual, respectively.

•The "***Healthy Fitness Zone (HFZ)***" represents a risk threshold identifying a level of aerobic capacity above which a child or adolescent should have a low risk of metabolic syndrome.

•The "***Needs Improvement (NI)***" zone is an intermediate zone between the HFZ and the Needs Improvement (NI-HR) zone. Students whose scores place them in the NI zone receive a message encouraging them to strive to achieve the HFZ. Aerobic capacity in this level is associated with a moderate risk of metabolic syndrome. The advantage of the two needs improvement zones is that it provides the opportunity to provide a more prescriptive message about the need to improve fitness.

•The "***Needs Improvement–HEALTH RISK (NI-HR)***" zone is a higher risk threshold identifying a level of aerobic capacity associated with a high risk of metabolic syndrome. The

"Needs Improvement-HEALTH RISK" fitness zone would provide youth/parents with an appropriate warning of health risk if the child has low fitness.

The aerobic fitness standards for boys tend to increase with age but they decrease for girls. These changes do not imply higher expectations for boys and lower expectations for girls. The changes are reflective of the natural developmental trends in aerobic capacity for boys and girls (boys gain muscle with age while girls tend to gain body fat). The lines actually reflect the same percentile score (same relative level of fitness) across age for both boys and girls.

It is important to note that the aerobic fitness standards are based on estimated aerobic capacity rather than on the actual fitness performance. Each of the primary assessments provides estimates of aerobic capacity, but differences in the tests and the associated prediction equations can lead to differences in fitness classification (depending on what test is used). To minimize misclassification, the PACER test score is equated to a corresponding mile run time to determine estimated aerobic capacity. This has been shown to improve the classification agreement between the two assessments. Detailed information on the derivation of these standards is available in Chapter 6, Aerobic Capacity Assessments.

Body Composition Standards

Excess body fat contributes to a number of health problems in adults as well as in youth. The most immediate risk for youth is developing diabetes so the healthy fitness zones were established based on risk for metabolic syndrome, a precursor to diabetes.

Metabolic syndrome is characterized by a clustering of risk factors including abdominal obesity, high triglycerides, high blood pressure, glucose intolerance, and high levels of circulating insulin. Nationally-representative data on metabolic syndrome and objective data on body composition are available from the National Health and Nutrition Examination Survey (NHANES), so this data source was used to develop the standards. The process used made it possible to establish age and gender-specific standards that reflect different levels of risk while also taking into account normal changes during growth and maturation.

•The "***Healthy Fitness Zone (HFZ)***" represents a risk threshold identifying a level of aerobic capacity above which a child or adolescent should have a low risk of metabolic syndrome.

•The "***Needs Improvement (NI)***" zone is an intermediate zone between the HFZ and the NI-HR zones. Students whose scores place them in the NI zone receive a message encouraging them to strive to achieve the HFZ. Body composition in this level is associated with a moderate risk of metabolic syndrome. The advantage of the two needs improvement zones (NI & NI-HR) is that it provides the opportunity to provide a more prescriptive message about the need to improve fitness.

•The "***Needs Improvement–HEALTH RISK (NI-HR)***" zone is a higher risk threshold identifying a level of aerobic capacity associated with a high of metabolic syndrome. The "Needs Improvement-HEALTH RISK" fitness zone would provide youth/parents with an appropriate warning of health risk if the child has low fitness.

The body composition standards are based on percent body fat. Although an assessment of percent body fat would be ideal, practical application in schools is very difficult. The majority of schools use body mass index (BMI) despite some well described limitations (e.g. it is unable to discern fat-mass from fat-free mass). To provide flexibility for use in schools, separate BMI standards were developed to correspond to the body fat values. The FITNESSGRAM® BMI standards were created so that they would agree with the %BF standards. The two assessments are very different and can't be expected to have perfect agreement. However, the resulting BMI standards can be interpreted in a similar way as the body fat standards described previously. A challenge in using and interpreting the BMI standards was that they did not correspond with the widely used CDC standards which are set at the 85th and 95th percentiles (for both boys and girls).

The FITNESSGRAM® values were based on the same CDC growth charts but were set at different percentiles based on the specificity and sensitivity cut-points. In boys, the values for the Healthy Fitness Zone and the Needs Improvement-Health Risk zone correspond with the 83rd percentile and 92nd percentiles in the CDC charts, respectively. In girls, the values for the Healthy Fitness Zone and the Needs Improvement-Health Risk zone correspond with the 83rd percentile and 90th percentiles in the CDC charts, respectively. The BMI values for ages 5-9 were set at the 85th percentile (essentially deferring to the CDC for standards for these ages due to lack of data to detect risk).

While the differences between the CDC values and the FITNESSGRAM® standards are small, it causes some children to be classified differently using the two methods. Therefore, the Cooper Institute commissioned an additional set of analyses to directly compare the predictive utility of the FITNESSGRAM® standards compared with the CDC values. The study used additional rounds of NHANES data and directly evaluated the Sensitivity and Specificity of the alternative classification schemes. The analyses revealed that the CDC standards were slightly better for boys but the FITNESSGRAM® standards were slightly better for girls. However, there were no statistically significant differences between the approaches. Because the two sets of standards were relatively similar it was determined advantageous to adopt the CDC values for the BMI health standards in FITNESSGRAM®. The advantage of this change is that youth receive consistent information from FITNESSGRAM® and the CDC/Growth Charts which are used by pediatricians. The disadvantage is that classification agreement may be slightly worse if comparisons are made with body fat estimates. The majority of schools now use BMI so the advantage of having consistent information about BMI far outweighed any lack of agreement between BMI and body fat estimates in those schools using both.

Detailed information on the derivation of the body fat and BMI standards is available in Chapter 7, Body Composition Assessment. Information about the supplemental analyses comparing the FITNESSGRAM® standards and the CDC values are available upon request.

Rationale for Musculoskeletal Fitness Standards

Little or no data exists to indicate levels of musculoskeletal fitness associated with good health. Therefore, it is difficult to determine objectively how much musculoskeletal fitness is necessary for children. Standards for these assessments were therefore based on a variety of criteria including expert opinion, previous data, and results from various research studies. See the Chapter 8, Musculoskeletal Fitness Assessments, specifically the topic, "What Is the Basis of Criterion Referenced Standards for Muscular Strength, Endurance and Flexibility?" A general discussion of criterion referenced standards is also available in Chapter 4, Physical Fitness Standards for Children.

What Is The Relationship Between BMI and the Aerobic Capacity Standards?

There have been questions about the Aerobic Capacity Healthy Fitness Zone Standards and why individuals with different BMI values have to perform differently to achieve the

Healthy Fitness Zone. This response is based on a paper written by J.R. Morrow and posted at The Cooper Institute website. Here is a link to a page that includes links to that document, FAQs, PowerPoint presentations, and recorded webinars on the topic of the aerobic capacity and body composition standards:

The explanation is in the formula used to estimate VO_2max (in the formula Mile Time includes either the time on the One Mile Run or the Equated Mile Time from the PACER).

$VO_2max = (.21\times(age\times sexcode)) - (.84\times BMI) - (8.41\times Mile\ Time) + (.34\times Mile\ Time \times Mile\ Time) + (108.94)$;

Importantly, the aerobic capacity standards all now relate back to VO_2max–the "criterion" measure of aerobic capacity.

Click http://www.cooperinstitute.org/hfz-standards to see a chart of the Healthy Fitness Zone standards.

So, regardless of which test one completes (Mile Run; PACER20 or PACER15), the results are translated into VO_2max for comparative purposes. Essentially, the various test performances are equated to arrive at a common VO_2max.

The absolute Aerobic Capacity (VO_2max) HFZ standard for a 10 year old girl is 40.2 ml/kg/min. Everyone with a VO_2max at or above 40.2 ACHIEVES the HFZ standard (40.2). Everyone below 40.2 does NOT ACHIEVE the HFZ standard (40.2).

A mile run time of 13:00 and PACER20 score of 10 and PACER15 of 13 are essentially "equivalent" with regard to VO_2max.

BUT WAIT–one's weight or body size (actually BMI here) also influences the estimated VO_2max. The equations for estimating VO_2max include the specific test performance AND BMI.

There is a negative relationship between body weight/BMI/Percent fat and VO_2max. That is accounted for in the equations for estimating VO_2max from the performance score (run time or PACER laps

$VO_2max = (.21\times (age\times sexcode)) - (.84\times BMI) - (8.41\times Mile\ Time) + (.34\times Mile\ Time\times Mile\ Time) + (108.94)$;

- There is a negative regression coefficient in the equation associated for weight (BMI).
- There is a negative regression coefficient in the equation associated with Mile Time (because a lower time is better).
- Thus, to achieve the Absolute Standard, one has to take into account the test performance (time or laps) AND BMI.
- The HIGHER the BMI, the more it negatively influences the results of the equation (reduces the estimated VO_2max) and the greater the test performance needs to be to overcome the influence of BMI.

When one takes into account BMI, a person with a lower BMI can achieve the criterion VO_2 more easily than one with a higher BMI.

The greater the BMI, the more that this adjustment comes into play.

To "counter" the influence of the larger body size influence on VO_2max, the person must achieve a better overall performance on the aerobic capacity test.

Can one who is overweight (or with a high BMI) achieve the HFZ standard? YES! However, it takes a much higher performance level for them to achieve the HFZ standard because of negative impact of their higher BMI on the calculation.

Another issue that comes up is that teachers have difficulty now because they cannot just say to everyone of a given age, “Here is what you have to do to achieve the HFZ.” This has positive and negative impacts. The new way is actually more valid but it is a bit more awkward to present to students and parents. Now, because achievement is a function of their test performance and BMI, this is more difficult to understand. However, students should be encouraged to do their best on the test and not think of “what is the minimum I must do so that I achieve the HFZ?”

Subsequent to completing the FITNESSGRAM®, teachers should be able to tell students who do not achieve the HFZ there are two ways in which their performance (and health) can improve—reduce your BMI and/or improve your test performance for your body size. Both of these will result in a higher estimated Aerobic Capacity (VO_2max) for one’s size (Morrow, 2012).

How Were Standards Established for the ACTIVITYGRAM® Assessment?

The goal in the ACTIVITYGRAM® assessment is for children to accumulate 60 minutes of physical activity a day. These goals are consistent with the most commonly used guidelines for physical activity in children and adolescents which recommend 60 minutes of moderate to vigorous intensity activity per day. (USHHHS, 2008)

See the Chapter 3, Health Benefits of Physical Activity and Physical Fitness in Children, for a more detailed description of appropriate physical activity for children.

Bibliography

Morrow, J.R., Jr. (2012). *The Relationship Between BMI and Aerobic Capacity Standards*. Document published at The Cooper Institute website, http://www.cooperinstitute.org/pub/file.cfm?item_type=xm_file&id=658

Morrow, J.R. Jr., Going, S.B., & Welk, G.J. (2011). FITNESSGRAM® Development of Criterion-Referenced Standards for Aerobic Capacity and Body Composition, *American Journal of Preventive Medicine*, 41.

National Association for Sport and Physical Education (2009). *Appropriate Instructional Practice Guidelines, K-12; A Side-by-Side Comparison*. Retrieved from http://www.aahperd.org/naspe

The Cooper Institute. Meredith, M.D. & Welk, G.J., eds. (2010). *FITNESSGRAM®/ACTIVITYGRAM® Test Administration Manual* (4th Ed.), Champaign, IL: Human Kinetics.

United States Department of Health and Human Services (2008). *HHS Announces Physical Activity Guidelines for Americans*. Retrieved from http://www.hhs.gov/news/press/2008pres/10/20081007a.html