ED 352 323 SP 034 111

AUTHOR Marsh, Herbert W.
TITLE The Multidimensional Structure of Physical Fitness: Invariance over Gender and Age.
PUB DATE Jul 92
NOTE 26p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Aerobics; Age; Child Health; *Construct Validity; Elementary Secondary Education; *Factor Analysis; Foreign Countries; *Goodness of Fit; Health Education; *Physical Fitness; Sex
IDENTIFIERS Australia; *Health Related Fitness; *Physical Fitness Tests

ABSTRACT

The present investigation extends the factor analytic approach pioneered by Fleishman (1964), incorporating subsequent developments in the application of confirmatory factor analysis and the physical fitness literature (e.g., an emphasis on maximum oxygen intake). More specifically, the ability of an a priori factor structure of physical fitness to fit (i.e., account for) data from the 1985 Australian Health and Fitness Survey based on 25 indicators of fitness (field exercises, technical measures, and laboratory measures) is tested for 2,817 boys and girls aged 9, 12, and 15. An 8-factor model derived from previous research fit the data well for each of the 6 age/gender groups considered separately. Based on tests of factorial invariance, factor loadings and factor correlations were reasonably invariant across the six groups. This substantively important finding indicates that all 25 indicators were equally valid for boys and girls aged 9, 12, and 15. The results provided clear support for the multidimensionality of physical fitness and call into question attempts to summarize fitness with a single indicator (e.g., aerobic power) or a total score representing different components of physical fitness. (Contains 43 references.) (Author/IAH)

# The Multidimensional Structure of Physical Fitness: Invariance Over Gender and Age

Herbert W. Marsh

University of Western Sydney (Macarthur)

9 July, 1992

Running Head: **Physical Fitness**

Acknowledgements: Data used in this study is from the ACHPER Australian Health and Fitness Survey (1985), jointly funded by the Commonwealth Department of Sport Recreation and Tourism, the Commonwealth Schools Commission, the Commonwealth Department of Health, and the National Heart Foundation of Australia. The author would like to thank Janet Pyke, Barry Ridge, Patsy Tremayne, and Gordon Treble for their helpful comments at various stages of the research. Correspondence in relation to this article should be sent to Professor Herbert W. Marsh, University of Western Sydney, Macarthur, PO Box 555, Campbelltown, NSW 2560, Australia.

## The Multidimensional Structure of Physical Fitness: Invariance Over Gender and Age

### Abstract

The present investigation extends the factor analytic approach pioneered by Fleishman (1964), incorporating subsequent developments in the application of confirmatory factor analysis and the physical fitness literature (e.g., an emphasis on $VO_2$max). More specifically, the ability of an a priori factor structure of physical fitness to fit (i.e., account for) data based on 25 indicators of fitness (field exercises, technical measures, and laboratory measures) is tested for 2,817 boys and girls aged 9, 12, and 15. An 8-factor model derived from previous research fit the data well for each of the 6 age/gender groups considered separately. Based on tests of factorial invariance, factor loadings and factor correlations were reasonably invariant across the six groups. This substantively important finding ir 'icates that all 25 indicators were equally valid for boys and girls aged 9, 12 and 15. The results provided clear support for the multidimensionality of physical fitness and call into question attempts to summarize fitness with a single indicator (e.g., $VO_2$max) or a total score representing different components of physical fitness.

# The Multidimensional Structure of Physical Fitness: Invariance Over Gender and Age

Physical fitness is a widely valued goal for men and women of all ages. Of particular relevance to the present investigation, there is a growing concern about youth fitness; young people's poor physical fitness, sedentary life style, and levels of obesity. Related concerns are also evident in the shift in emphasis in sport/exercise research and physical education from a narrow focus on sport and elite athletes towards a broader focus on health-related fitness and epidemiological studies of youth fitness. Physical fitness is also posited as a mediating variable that contributes to health-related outcomes, social skills, and a variety of measures of psychological well being.

## A Construct Validity Approach

Despite the importance of the physical fitness construct, theoretical and empirically tested models of the structure of physical fitness have not been given adequate attention. On the one hand, there is a growing use of reasonably distinct, narrowly defined indicators of physical fitness without clarifying how they fit into the overall structure of physical fitness. On the other hand, there is an increasing number of physical fitness batteries based on implicit, typically untested assumptions about the structure of physical fitness and its generality across age, gender, and other individual characteristics. Because physical fitness is a hypothetical construct, its construct validity must be established. In a construct validity approach, investigations can be classified as within-construct studies that evaluate the internal structure of physical fitness using techniques such as factor analysis or between-construct studies that attempt to establish a theoretically consistent, logical pattern of relations between measures of physical fitness and other constructs. The resolution of at least some within-construct issues should be a logical prerequisite to between-construct research. This emphasis on construct validity, factor analysis, and within-construct studies of the structure of physical fitness is the focus of the present investigation.

Physical fitness tests are typically compared with age and gender norms or, perhaps, more sophisticated norms that also take into account body composition (height, weight, body fat) or biological maturity (e.g., Malina, 1989). Such comparisons, however, may confound the influences of skill, motivation, compensatory growth that reflects demands placed on the body by physical exertion, and genetically determined developmental growth (see discussion by Krahenbuhl, 1980). Of greater relevance to the present investigation, there has not been adequate attention given to the question of whether a given test measures the same component of physical fitness with equal validity for boys and for girls, and across different ages. This is a particularly relevant concern in tests of youth fitness during the early-adolescent period that is so potentially turbulent -- biologically and psychologically. It is important to emphasize that this concern is not one of differences in levels of performance that can be evaluated in relation to appropriately constructed norms. Rather, the concern is more fundamental, asking whether the same physical fitness indicator has the same meaning across subjects who differ in age and gender. If the same indicators reflect different components of fitness depending on age and gender, then the task of interpreting each indicator and presenting a profile of different components of fitness would be considerably more complicated. Such a finding would also call into question many current practices in assessing fitness which implicitly assume that the underlying meaning of a particular indicator is relatively invariant. This critical concern can only be addressed within the context of a construct validity approach and an evaluation of whether -- and how -- the structure of physical fitness varies according to individual characteristics such as gender and age. This issue, because of its importance to theory and practice, and because it has been given little attention in the physical fitness literature, will be a major focus of the present investigation.

In the physical fitness literature a distinction is typically made between the large sample epidemiology-like studies of youth fitness that rely primarily on easily administered field exercises that do not require expensive equipment and the small-sample laboratory studies of adult (or elite athlete) fitness that emphasize technically sophisticated measures which require expensive equipment. This distinction, however, invites the potential for confusing the indicators of physical fitness with the physical fitness construct and for confusing the technological sophistication required to obtain a measure with the construct validity of a measure. The inexpensively collected field exercises should not be viewed as "poor cousins" of the more expensive laboratory measures and the technologically sophisticated measures are not necessarily more valid indicators of the physical fitness construct. The purpose of the field exercises is not to provide a necessarily imperfect prediction of the laboratory measures that could be achieved if only adequate resources were available to test all subjects in a laboratory setting. Rather, both the field exercises and the laboratory measures are merely indicators of the physical fitness construct whose validity should be systematically evaluated within a construct validity approach. Hence there is a need to evaluate the structure of physical fitness in studies that include a wide array of field exercises and sophisticated laboratory measures, and that are based on sufficiently large samples to appropriately apply statistical techniques such as factor analysis.

## Physical Fitness: A Multidimensional Construct

Critical concerns in the study of Physical fitness are the definition of the construct and the selection of appropriate indicators. The position taken here is that physical fitness is a multidimensional construct and that physical fitness cannot be adequately understood if this multidimensionality is ignored.

Although there are many definitions of fitness, Clarke's general definition is widely accepted: "the ability to carry out daily tasks with vigor and alertness, without undue fatigue, and with ample energy to enjoy leisure-time pursuits and to meet unforeseen emergencies" (1976, p. 12) and "Physical fitness is the ability to last, to bear up, to withstand stress, and persevere under difficult circumstances where an unfit person would quit. It is the opposite to becoming fatigued fromm ordinary efforts, to lacking energy to enter zestfully into life's activities, and to becoming exhausted from unexpected, demanding physical exertion" (Clarke, 1979, p. 28). Safrit (1981, p. 213) stressed that physical fitness is a multidimensional construct that cannot be adequately reflected by a single measure, and that physical fitness tests should measure the full range of functional capacities and accurately reflect changes in appropriate physical activity and altered capacity. Similarly, based on his extensive literature review of physical fitness tests, Fleishman (1964, p. 37) concluded that: "There is no such thing as general physical proficiency. The problem is a multidimensional one." Baumgartner and Jackson (1987, p. 277) noted that "as the concept of physical fitness has moved away from athletic participation toward health-related fitness, there has been greater emphasis on cardiovascular function, body composition (leaness/fatness), strength, endurance, and lower-back flexibility, traits shown by medical and exercise scientists to promote health and reduce the risk of disease."

The original American Association for Health, Physical Education and Recreation (AAHPER) test battery had considerable impact on theory, measurement, research and practice of youth fitness assessment. Items were selected according to the requirements of: (a) minimal equipment, (b) student familiarity, (c) ease of administration by classroom teachers, (d) appropriateness across gender and (adolescent) ages, and (e) broad selection of different fitness components. The original battery consisted of 7 items: pullups, situps, shuttle run, standing long jump, 50 yard dash, 600 yard run, and a softball throw. Its major advantages were ease of administration, objectivity, face validity, and the availability of nationally representative normative data. The AAHPER battery in various forms has been repeatedly administered to large representative samples of school children in many Western countries including the US, UK, Australia, and New Zealand since the late 1950s. Evaluations of the construct validity (e.g., Baumgarten and Jackson, 1987; Cureton, 1980; Safrit, 1981) of the items, however, has called into question some aspects of the test and led to its subsequent revision. Ponthieux and Barker (1963) factor analyzed the AAHPER items and reported three factors defined primarily by the 600 yard run, pull-ups, and sit-ups, by the long jump, shuttle run, and 50 yard dash, and a single-item factor defined by the softball throw. In his classic factor analysis study of a wide variety of physical activities including some of the AAHPER items, Fleishman (1964) reported that 4 of the 7 items (shuttle run, softball throw, 50-yard dash, and standing long jump) loaded on a factor that he called explosive strength. Reflecting these and other concerns, the battery was subsequently altered, for example, by replacing the pull-up for girls with the flexed-arm hang, eliminating the softball throw that apparently has a substantial skill component, replacing the 600 yard event with a longer one mile or 9 minute run (Disch, Jackson, and Frankiewicz, 1975), modifying the sit-up test, and adding the skin fold measurements to assess body composition, and the sit-and-reach test to assess flexibility (AAHPERD, 1980).

### Aerobic Power and $VO_2$max

Increasingly -- in apparent contrast to the multidimensional perspective emphasized here -- researchers have adopted an implicitly unidimensional approach in which physical fitness is defined in terms of aerobic power, reflecting the integration of the cardiovascular, pulmonary, vascular, and muscular systems. In field exercises it is measured indirectly by items such as running a moderately long distance (e.g., 1.6 km run or 12 minute run) or step tests. With more sophisticated equipment it is measured by the working capacity or power produced at a given heart rate of 170 beats/minute (PWC 170) using a bicycle ergometer or treadmill. In laboratories with sophisticated equipment the "gold standard" measure of cardiovascular fitness is maximum oxygen intake ($VO_2$max). For adults, the usual criteria that $VO_2$max has been achieved (e.g., Boutcher, 1990; Cunningham, 1980; Schell & Leelarthaepin, 1990) are: the engagement in strenuous exercise involving continuous, rhythmeic movement lasting at least 15 minutes (e.g., running, cycling, swimming) that use large muscles and depend on oxidative energy systems; anaerobic metabolism as indicated by high levels of lactic acid; and a plateau in oxygen intake ($VO_2$max) with increasing peak work loads. Zwiren, Freedson, Ward, Wilke and Rippe (1991) recently compared direct measures of $VO_2$max with estimates based on five exercises for young adult females. They reported that $VO_2$max was more highly correlated with performance on a 1.5 mile run (r=.79) than with a step test (r=.55) or from heart rates on a submaximal cycle ergometer (r=.66). These results are consistent with a number of studies reporting high correlations between $VO_2$max and running speed over distances of 1.6 km or more in which correlations as high as .9 have been reported (Cooper, 1968). There is, however, considerable variation in these results and Cunningham (1980) indicated stronger relations are typically reported when students are more successfully motivated.

The trend towards defining physical fitness exclusively in terms of $VO_2$max is unfortunate, and appears to reflect a confusion between physical fitness which is a hypothetical construct and $VO_2$max which is only one indicator of this construct. Furthermore, this reliance on $VO_2$max as the "gold standard" measure of fitness leads to implicit assumptions about the structure of physical fitness that may be unwarranted. In particular, this situation seems to imply a relatively unidimensional construct of physical fitness that is inferred by $VO_2$max and that other indicators of fitness are important primarily in terms of how they relate to $VO_2$max and cardiovascular endurance. In contrast to this implicitly unidimensional perspective, Safrit (1981) argues that physical fitness is a multidimensional construct that cannot be adequately represented by a single indicator. Bar-Or (1987) argued that this emphasis of $VO_2$max ignores other components of youth fitness -- particularly childhood obesity that initiates a vicious circle that includes decreasing physical activity, poor self-esteem, and the inability to socialize. Baumartner and Jackson (1987) noted that adult obesity is a serious health problem and that 85% of adult obesity is linked to childhood obesity. Cureton (1987, p. 319) argued that "less attention should be given to cardiovascular fitness and more to the relation of physical activity and fitness to health/disease risk." Seefeldt and Vogel (1987) noted the dangers associated with over-reliance on a single-indicator approach, arguing instead for more broadly based, multidimensional definitions of fitness. Sallis (1987) argued that from the perspective of public health, physical activity is more important than fitness per se, and that the major benefit of physical activity in childhood is to establish patterns that are carried into adulthood. The American College of Sports Medicine (1990) recognized that moderate levels of physical activity that are insufficient to influence $VO_2$max may have important benefits on physical health. Similarly, Sharkey (1991) noted that health benefits apparently plateau at relatively low levels of $VO_2$max and proposed alternative tests that place more emphasis on what he referred to as endurance fitness. Also, Boutcher (1990) noted that as much as 90% of the variance of $VO_2$max may be genetically determined, calling into question the usefulness of $VO_2$max measures taken at a single point in time. Hence, particularly in relation to health-related fitness and multidimensional perspectives of physical fitness, the trend toward $VO_2$max as the "gold standard" measure of fitness is premature and unwarranted. More generally, it is important to evaluate $VO_2$max within a broader context of within- and between-network studies that incorporate a wide range of different fitness indicators and to establish its location within the multidimensional structure of physical fitness. Due in part to the typically small sample sizes in $VO_2$max studies, there apparently has been no large-scale factor analyses that included $VO_2$max and indicators of a wide variety of other components of physical fitness.

### The Structure of Physical Fitness: A Factor Analysis Approach

Edwin Fleishman's classic research (1964) on the structure and measurement of physical fitness provides an important basis for the present investigation. He applied factor analytic techniques to identify the components of physical fitness and to select appropriate indicators to include in a comprehensive fitness battery. Based on his review of factor analytic studies of physical fitness, Fleishman proposed specific components of fitness that he broadly classified into factors of strength that are the primary focus of the present investigation and factors of speed, flexibility, balance and coordination.

In the strength area, his results supported a priori predictions of separate components of strength and, perhaps, strength factors that are specific to different parts of the body. In particular, he demonstrated the distinctions between dynamic strength, static strength, and explosive strength. Dynamic Strength was defined by items such as pull-ups, push-ups, bent arm hang, rope climb, dips and squat thrusts "in which the arms are required repeatedly or continuously, to move to support the weight of the body" (p. 64), although short running tests, vertical and broad jumps, and sit-ups also loaded on this factor. Static Strength was defined primarily by the use of dynamometers and items reflecting the capacity to apply force to lift or push weights; a maximum force is exerted for a brief period of time where the force is exerted continuously up to this maximum (p. 65). Explosive Strength was defined by items such as short dashes, long and vertical jumps, and a softball throw that emphasize "the ability to extend maximum energy in one explosive act ... rather than continuous strain, stress, or repeated exertion" (p. 66). Fleishman also noted some evidence for separate components that are specific to particular parts of the body (arm, leg, trunk) or particular activities (e.g., running). Depending on the length of the test battery, Fleishman recommended the inclusion of tests of dynamic strength (push-ups, pull-ups), static strength (hand grip, arm pull dynamometer), explosive strength (50-yard dash or shuttle run, long jump, and softball throw), and trunk strength (leg lifts and hold half sit-up).

Fleishman's (1964) corresponding analysis of speed, flexibility, balance and coordination tests is less relevant to the present investigation because only one such test (sit-and-reach) was included in the test battery used here. He reported 6 factors that only partially supported his hypotheses: extent flexibility, dynamic flexibility, gross body equilibrium, balance with visual cues, speed of limb movements, and an explosive strength factor (like that in the analysis of strength tests).

The major emphasis of Fleishman's research was on physical fitness as a multidimensional construct. This emphasis was also reflected in his Performance Record for Basic Fitness Tests (Fleishman, 1964) in which a multidimensional profile of physical fitness components is presented. For each component, raw, percentile, and stanine scores are presented. Fleishman (1964, p. 136) specifically noted that "the most useful information is provided by separate tests, since this allows the pinpointing of strong and weak areas." Apparently bowing to popular demand, however, he also included a total fitness score -- the sum of the stanine scores -- because "many instructors and students feel the need for a single index to summarize a student's over-all performance" (p. 141). He justified this total score in part because the battery of tests reflected separate factors so that no one factor was unduly weighted. Fleishman also noted, however, that it may be better to differentially weight each indicator in relation to a particular criterion using statistical techniques such as multiple regression, but argued that in the fitness area such external criteria are seldom available. Further de-emphasizing the total score, Fleishman noted that it was primarily useful for quick comparisons but that "one should not lose sight of the fact that an FI [total fitness index] of 50 could be achieved by average performance on all tests, or by a combination of exceptional and inferior performances on a number of different tests" (p. 142). In this sense, overall fitness according to Fleishman (1964) is best represented as a multidimensional profile of different components of physical fitness.

There was little emphasis on cardiovascular endurance in Fleishman's (1964) study. Although he specifically noted the possibility of a cardiovascular endurance factor measured by long distance runs and prolonged body exertion, he did not consider relevant indicators in his factor analysis. Noting that "in retrospect, it would have been useful to include some variant of the 600 yard run-walk test" (p. 70), he concluded that the relation between cardiovascular endurance and the strength factors that he did consider "remains to be seen" (p. 71). However, in the large normative study of his Basic Fitness Tests, he indicated that "while it was not feasible to include the 600-Yard Run Walk in our experimental studies, this test was added in later stages of our normative study to provide a measure of Stamina or Cardio-Vascular Endurance" (p. 104). Thus, although he does not evaluate the relation between cardiovascular fitness and other fitness indicators, Fleishman does provide normative comparisons for boys and girls of differing ages on one indicator of this factor.

Fleishman (1964) did not specifically address the issue of how well his a priori structure of physical fitness generalizes across age and gender. This is not, perhaps, surprising since both his major factor analysis studies involved United States Navy recruits who were relatively homogeneous in terms of age and gender. In his much larger "national study" boys and girls between the ages of 12 and 18 were tested. In this study, considerable attention was given to age and gender differences in levels of performance. However, apparently because only one indicator per factor was used to define physical fitness in the national study, Fleishman (1964) did not present factor analyses of these data or report the consistency of relations among the indicators over gender and age. Hence the critically important issue of the generalizability of the structure of physical fitness over age and gender was not evaluated by Fleishman (1964).

Fleishman's (1964) factor analytic research of physical fitness reflected the "state of the art" in the early 1960s. In the ensuing three decades, however, there have been important new developments in the application of factor analysis and indices used to infer physical fitness. Thus, it would seem likely that the physical fitness literature should contain a large number of factor analytic studies following the important tradition established by Fleishman. Remarkably, however, a review of the current physical fitness literature reveals a surprising dearth of factor analytic studies. A computer search of several major indices (ERIC, PschInfo, AUSSPORT, MEDLINE) revealed thousands of studies with "factor analysis" or "physical fitness" as a descriptor, but almost no studies with both descriptors. Searches using these descriptors revealed no factor analyses of a comprehensive selection of physical fitness indicators published between 1980 and 1992 (except, perhaps, Hagan, Parrish, and Licciardone, 1991), and only a few relevant studies published between 1966 and 1980. Whereas this search does not mean that there have been no comprehensive factor analytic studies of physical fitness measures in the last decade, it seems reasonable to conclude that there has been limited recent work in this area. It is not that the relevance of this factor analytic approach has waned in that: (a) textbooks in physical education and related areas almost universally emphasize components of fitness based in part on factor analytic studies such as Fleishman's (1964) research; (b) there currently exists a heated debate about the relative importance of $VO_2$max as the "gold standard" measure of physical fitness and a multidimensional perspective to physical fitness; and (c) there is a growing number of physical fitness batteries that are based on implicit, apparently untested, underlying factor structures that are assumed to generalize across age, gender, and other individual characteristics. In a related concern about fitness batteries, Sharkey (1991; pp. 5-6; also see Sharkey, 1988) noted that "these fitness batteries are based on a concept of generality, wherein factor analysis indicates common clusters or groups of items associated with a component of fitness" and that "it is clear that no single test or component of fitness adequately represents the entire component." It seems that researchers and practicioners have relied on classic factor analytic studies such as Fleishman's research and intuition to classify an ever increasing number of specific physical fitness indicators into largely untested categories of physical fitness, implicitly assuming that each indicator reflects the same component of physical fitness with equal validity for subjects differing in age, gender, and other individual characteristics. Whereas such blind faith may be justified by intuition and expert opinion, it is also important to pursue empirical tests of these implicit assumptions.

## The Present Investigation

The purpose of the present investigation is to test the ability of an a priori factor structure of physical fitness to account for data based on the **Australian Health and Fitness Survey** (**AHAFS**) and the extent to which the same factor structure fits data for boys and girls aged 9, 12 and 15. More generally, the present investigation extends the factor analytic approach pioneered by Fleishman (1964), incorporating subsequent developments in the methodology of factor analysis. In particular, in "exploratory" factor analysis that was the "state of the art" at the time Fleishman did his research, the researcher has little control over the resulting factors. Whereas researchers may predict an a priori factor structure, they have no way of testing the ability of their a priori factor structure to fit their data. Instead, support for a priori predictions are based on the extent to which the factors that "come out" match those that were predicted or match those obtained in other factor analyses. This exploratory mode of factor analysis is being replaced by a "confirmatory" approach to factor analysis in which the researcher specifically tests the ability of an a priori factor structure to fit the data, thus providing a much stronger basis for testing theory. Furthermore, tests of factorial invariance allow researchers to constrain any one, any set, or all parameter estimates to be constant across two or more groups. The present investigation is a within-network study of construct validity in which I test an a priori structure of physical fitness. This is not to deny the critical importance of between-network studies that relate components of physical fitness to other constructs such as health, athletic accomplishments, and psychological well-being. Rather, consistent with Fleishman's underlying premise, it is argued that the relation between physical fitness and external constructs can not be adequately understood unless there is reasonable resolution of within-construct issues about the structure of physical fitness.

Data for the present investigation are based on the Australian Health and Fitness Survey (**AHAFS**) that was conducted in 1985 with a nationally representative sample of Australian school children aged 7 to 15 (Pyke, 1987). The survey consisted of a comprehensive array of health and fitness measures including: (a) an extensive survey of sport/physical activities and health-related behaviours (e.g., smoking), (b) field exercises including measures of cardiovascular endurance (1.6K run), dynamic strength (situps, pushups), explosive strength/power (standing long jump, 50M dash), flexibility (sit-and-reach), and body composition (height, weight, and body girths); (c) technical measures (PWC170, dynamometer strength tests, skinfolds, blood pressure, and lung function), and sophisticated laboratory measures ($VO_2$max). In contrast to most youth fitness research, the **AHAFS** included technical and laboratory measures of fitness. In contrast to most laboratory research the **AHAFS** was based on a large, representative sample and included a wide array of non-laboratory measures. This data provides a strong basis for the present investigation because: considerable expertise was called upon in selecting tests for inclusion in the battery; the tests and protocols for their administration were extensively pilot-tested (including the production of an instructional video); and the nationally representative data provide a unique basis for evaluating the underlying structure of physical fitness and its generality across gender and age. (Surprisingly, given the comprehensiveness of this data, almost no analyses based on it have been published in major international journals and none have used this data to evaluate the structure of physical fitness.)

## Methods

### Sample and Procedures.

A detailed description of the sampling design, test selection, testing protocols, and collection of the data is presented by Pyke (1987) and is summarized here only briefly. The target population consisted of all students aged 7 to 15 enrolled in Australian schools in September, 1985. A two-stage probability sample was used in which 52 primary and 52 secondary schools were randomly chosen, and samples of 10 boys and 10 girls from each age/sex category were selected from each school. For present purposes only 6 groups are considered (boys and girls aged 9, 12, and 15) that were administered technical and laboratory measures as well as the field exercises that were completed by all participants. The ages were selected "to approximate the pre-pubertal, pubertal, and postpubertal growth stages" (Pyke, 1987, p. 10). Because of the added expense in obtaining $VO_2$max measures, subsamples of students within each of the 6 groups were selected. Students were selected who had previously completed other items from schools that were within a reasonable proximity of laboratory facilities where this testing was conducted.

Insert Table 1 About Here

All measures considered here (see Table 1) except for $VO_2$max were collected by data collection teams that went to participating schools. A total of 10 data collection teams, each consisting of 10 data collectors and a supervisor, were used. In the initial testing session, students completed height, weight, girth, sit and reach, long jump, push-up, sit-ups, skin fold, lung function, and dynamometer strength tests. These tests were conducted indoors and the order of presentation was varied. Following the completion of the indoors testing, the 50 M dash and then the 1.6 K run were conducted out of doors. Blood pressure and PWC170 were measured the following day. The $VO_2$max testing was independently conducted by staff in exercise physiological laboratories in tertiary institutions. In all tests, students were encouraged to do their best without any pressure being applied. Parental consent was obtained prior to collection of data and students could withdraw from the study at any time.

### Statistical Analyses.

An overview of the CFA approach. A detailed presentation of the conduct of confirmatory factor analysis (CFA) is beyond the scope of the present investigation, and is available elsewhere (e.g., Bollen, 1989; Byrne, 1989; Hayduk, 1987; Long, 1983; Joreskog & Sorbom, 1989; Marsh, 1987; Marsh & Hocevar, 1985; McDonald, 1985; Pedhazur & Schmelkin, 1991). Briefly, the researcher posits an a priori structure and tests the ability of a solution based on this structure to fit the data. In a CFA study, the parameters typically consist of factor loadings (the relations between measured variables and the latent factors), factor variances and covariances (relations among the factors) and factor uniquenesses (a combination of specific and error variances like 1 minus the communality estimates in traditional, exploratory factor analyses). As in exploratory factor analyses, the factor loadings are of central importance in the "definition" of each factor. In the present investigation, for example, a total of 25 physical fitness indicators (see Table 1) are hypothesized to represent 9 physical fitness factors. A particularly parsimonious model would be one in which each indicator had a non-zero factor loading on only the factor that it was hypothesized to measure. Thus, for example, factor loadings relating $VO_2$max, the 1.6K run, and PWC170 to the first latent factor -- Cardiovascular Endurance -- would be freely estimated, but the factor loadings relating these indicators to other factors would be fixed to be zero.

In addition to factor loadings, there are also uniquenesses associated with each of the indicators that reflect a combination of error and specific variance. These uniqueness terms are often hypothesized to be uncorrelated, but it is also possible to fit correlated uniquenesses to reflect relations between individual indicators that cannot be explained in terms of the a priori factors. In the present application, for example, correlated uniquenesses were posited for the following pairs of indicators: the two running measures (1.6 K run and 50 M dash); the two static strength indicators involving the shoulder (shoulder push and shoulder pull); the two static strength measures involving hand grip strength (right grip strength and left grip strength); and the two skinfold measures involving the arm (biceps and triceps). These correlated uniquenesses are apparently consistent with the general "running" factor and fitness associated with specific parts of the body proposed by Fleishman (1964).

In the CFA approach, diagnostics such as the modification index in LISREL (Joreskog & Sorbom, 1988) indicate whether freely estimating a parameter that has been constrained (e.g., a factor loading that has been fixed to be zero) will substantially improve the fit of the model. Using a step-wise process, this information can be used to improve the model. Because this process capitalizes on chance (in much the same way as step-wise approaches to multiple regression), it should be used cautiously. Thus, Joreskog and Sorbom recommend that researchers should only free parameters that can be justified from a substantive point of view. If many parameters are freed, then it is important to compare critical parameter estimates in solutions based on the original a priori model and the final a posteriori model. Ultimately, a posteriori models that differ substantially from a priori models should be cross-validated with new data.

Tests of factorial invariance. When parallel data exists for more than one group, CFA provides a particularly powerful test of the equivalence of solutions across the multiple groups. The researcher is able to fit the data subject to the constraint that any one, any set, or all parameters are equal in the multiple groups. The minimal condition for "factorial invariance" is the equivalence of the factor loadings in multiple groups, although Byrne, Muthen and Shavelson (1989) argued for the usefulness of "partial invariance" in which the factor loading for at least one indicator per factor is invariant across groups. It is also of interest to test for the invariance of factor correlations (see Marsh & Hocevar, 1985) that reflect relations among the different factors. Of less relevance is the invariance of factor variances and the uniquenesses associated with individual indicators. Particularly when the focus of the CFA is to test the invariance of solutions across multiple groups, it is critical that analyses are conducted with covariance matrices in which variables are scaled along a common metric across the multiple groups (e.g., the original score values or the same transformation across all groups) and not correlation matrices in which each group is scaled in relation to its own mean and standard deviation (i.e., a different transformation for each group; for further discussion see Joreskog & Sorbom, 1988).

Goodness of fit. A critical issue in the application of CFA is how to determine whether a solution based on an a priori model adequately fits the data or how to compare the relative fit of competing models. The evaluation of goodness of fit is not fully resolved, but a general approach is to: (a) determine that the iterative procedure converges to a proper solution that is well-defined (e.g., the solution has no parameter estimates that have impossible values such as negative variances or correlations greater than 1.0); (b) establish that parameter estimates are substantively reasonable in relation to the a priori model and common sense (e.g., if 4 of 5 indicators of static strength have positive factor loadings but the factor loading for the remaining indicator is negative, then the solution does not make sense and should be interpreted with extreme caution); and (c) evaluate the $X^2$ test statistic and various fit indices in relation to rules of thumb and values from competing models. Whereas, there is an emphasis on goodness of fit indices in CFA studies, it should be noted that the first two criteria are logical prerequisites to evaluating goodness of fit indices. If the empirical solution is improper, than the parameter estimates and fit indices should only be evaluated with extreme caution -- if at all. If the parameter estimates are not consistent with the a priori model and make no sense, then goodness of fit indices may be irrelevant.

In an evaluation of goodness-of-fit indices typically used in CFA, Marsh, Balla and McDonald (1988; also see McDonald & Marsh, 1990) noted that the Tucker Lewis Index (TLI) was the only widely used index that was relatively independent of sample size and relatively unaffected by the inclusion of additional parameter estimates that were known to have zero values in simulated data, and so it is emphasized here. McDonald and Marsh noted that the widely used Bentler-Bonett index is biased, and presented an alternative to it based on noncentrality -- the relative noncentrality index (RNI) -- that is not biased (also see Bentler, 1990). McDonald and Marsh recommended that parsimony should be considere. in evaluating goodness of fit. Following Mulaik, et al. (1989), they used the parsimony ratio defined as the ratio of the degrees of freedom in the model to be tested and a suitably defined null model (here taken to be a model which produces a diagonal reproduced covariance matrix in which all measured variables are assumed to be uncorrelated and the degrees of freedom is equal to the number of measured variables). Thus the parsimony ratio reflects the complexity/parsimony of the model and not the ability of the model to fit the data. All other things being equal, more parsimonious models are preferable to more complex models. However, McDonald and Marsh (also see Marsh & Balla, 1992) questioned the apparently arbitrary operationalization of parsimony in parsimony indices that are defined as the product of the parsimony ratio and some other index of fit, but argued that if parsimony indices are to be used then they should be based on an unbiased index such as the RNI. For present purposes I emphasize the TLI, but also present the $X^2$, RNI, parsimony ratio, and the parsimony index based on the RNI (PRNI) and note that most other indices of fit can be derived from the information that is presented.

The present application. The data for the present investigation are covariance matrices (or, equivalently, correlation matrices supplemented by standard deviations; see Appendix) for each of the 6 groups (boys and girls aged 9, 12, and 15). All statistical analyses were conducted with the commercially available "mainframe" version of LISREL 7 (Joreskog & Sorbom, 1988). In order to facilitate interpretations each indicator was standardized in relation to the "total group" mean and standard deviation of scores across all six groups and the two running indicators (1.6K run and 50 M dash) were reverse scored so that higher values reflect better levels of fitness. (Note that all subgroups were standardized in relation to the same mean and SD rather than standardizing each group in relation to its own mean and standard deviation.) Because there was relatively little missing data (except for $VO_2max$ for which only a subsample of students were tested), students with missing values for more than 2 indicators (other than $VO_2max$) were excluded, but all students with $VO_2max$ scores were automatically retained. The total N across all 6 groups was 2,817 (an average of 469.5 per group), but only 277 students (an average of 46 per group) had $VO_2max$ scores. Furthermore, because of the two-stage clustered sampling design, standard errors based on the assumption of simple random sampling substantially overestimate sampling variability in summary statistics and distort tests of statistical significance (see NCES, 1986, for related discussion of the High School and Be, ond Data that also used a two-stage sampling scheme). To compensate for this bias, the effective sample size was estimated to be 1800 (an average of 300 students per group). It is important to note that this correction has no effect on any of the parameter estimates; it only effects the degrees of freedom used in tests of statistical significance.

## Results

### The Initial Model

In the first stage of the analyses, I tested the ability of the a priori model (see Table 1) to fit the data separately for each of 6 groups (boys and girls aged 9, 12 and 15). Several features of the present investigation, however, require special attention. Ideally, for purposes of CFA, there should be three or more good indicators of each factor. In the present application, however, 4 hypothesized factors have only 2 indicators and Flexibility has only 1 indicator. Single-indicator factors can be considered, but they provide a weak basis for testing construct validity of a factor and for the appropriate correction for error that are possible when there are multiple indicators. Two-indicator factors, although globally identified in most applications, may result in unstable or improper solutions. One expedient approach to such problems is to require the two indicators to load equally on each factor, thus reducing the number of estimated parameters and typically producing a more stable solution that is less likely to be improper.

In the present investigation, nine-factor solutions did not result in proper solutions when tested separately for each group. In an attempt to resolve this problem, factor loadings for all two-indicator factors were required to be equal. (Also, because PWC170 had consistently small loadings on the Cardiovascular Endurance factor, the remaining two indicators were also required to load equally.) Even these equality constraints, however, did not result in proper solutions for all groups. One problem was that the correlation between Explosive Strength and Dynamic Strength was consistently close to 1.0 and sometimes exceeded 1.0 (which is, of course, an improper solution). When these two factors were combined to form a single factor -- subsequently called Explosive/Dynamic Strength -- the 8-factor model resulted in proper solutions for all 6 groups. The implications of this initial decision -- in terms of the application of CFA and the structure of physical fitness -- is evaluated subsequently in greater detail.)

Insert Tables 2 and 3 About Here

In further refinements of this model, several other improvements were made (based on LISREL's modification indices). First, as indicated earlier, correlated uniquenesses between several pairs of indicators were added. Also, although the PWC170 factor had only a small loading on the Cardiovascular Endurance factor, it had a substantial loading on the Static Strength factor. Finally, the number of push-ups loaded negatively on the Body Girth factor (indicating that individuals with larger bodies are less proficient at push-ups). Parameter estimates based on this model are presented in Table 2. This model resulted in fully proper solutions for all 6 groups in that the iterative process converged, no parameter estimates for any of the groups fell outside their permissible values, and matrices of parameter estimates were positive definite. The factors are well-defined in that -- with the exception of the PWC170 -- all indicators load substantially on the factor that they were hypothesized to represent. Furthermore, the goodness of fit indices (Table 3) are reasonable and consistent for each of the 6 groups considered separately and for the total across the six groups (e.g., TLI and RNI indices all approximate the .9 value that is typically interpreted to reflect an adequate goodness of fit). An inspection of the results for the six groups suggested that at least the factor loadings are reasonably consistent across groups, but the CFA approach offers much stronger tests of the equivalence of solutions across groups.

The Invariance of Solutions Across Gender and Age.

A substantively important issue in the present investigation is to evaluate the extent to which the physical fitness factor structure is the same for boys and girls of different ages. With the CFA approach it is possible to constrain any one parameter, any set of parameters, or all parameters to be the same across any two groups, any set of groups, or all groups. To the extent that a more parsimonious solution with such invariance constraints is able to fit the data, then there is support for the invariance constraints. If, however, the imposition of such invariance constraints results in a substantially poorer fit, then there is evidence against the invariance constraints.

In the present investigation I evaluate factorial invariance in relation to the six groups (boys and girls of three ages) and four sets of parameters (factor loadings, factor variances, factor correlations, and uniquenesses). I begin with tests of the equality of factor loadings across all groups, followed by tests of factor correlations, and then consider tests of uniquenesses and factor variances that are substantively less important. As expected, the results vary logically depending on the goodness of fit index. The RNI index is monotonic with model complexity, in that requiring any parameters to be equal in two groups cannot result in an improved index and will result in a poorer index unless the two parameters happen to be exactly equal when no constraints are imposed. For the RNI, there is support for an equality constraint if the decrement in fit resulting from its introduction is small. The TLI typically behaves similarly to the RNI, but contains a penalty for model complexity such that it is technically possible for the introduction of invariance constraints to result in an improved TLI. The PRNI (the parsimony index based on the RNI) severely penalizes model complexity and leads to the selection of more parsimonious models than the other indices.

For all indices there is strong support for the invariance of factor loadings and factor correlations. In fact, the TLI that is emphasized here, is slightly better for the model imposing the complete invariance of factor loadings and factor correlations across all six groups (.890; Table 3) than the corresponding model with no invariance constraints (.888; Table 3). Because the model with factor loading and factor correlation invariance is so much more parsimonious than the model with no invariance constraints (see parsimony ratio in Table 3), the PRNI parsimony index is substantially better for the model with invariance constraints (.842 vs. .742). Whereas the RNI always favors the model with no invariance constraints, the difference is small (.897 vs .908) in relation to the substantial difference in parsimony. Thus, the results of these invariance tests provide good support for the invariance of factor loadings and factor correlations across gender and age.

In contrast, there is not such good support for the invariance of factor variances or the uniquenesses (Table 3). These results are consistent with the observation that the variances associated with each measured variable differs systematically with gender and particularly age (see Appendix), and these differences must be reflected in larger factor variances, larger uniquenesses, or both. As noted earlier, the invariance of factor variances and uniquenesses is substantively less important than the invariance of factor correlations and particularly the factor loadings. Furthermore, even though there is not support for the invariance for factor variances and uniquenesses across all 6 groups, it is possible that invariance constraints are supported within more specific sets of groups. In particular, it is relevant to test the invariance of solutions over age separately for boys and girls, and to test the invariance of solutions over gender separately for students aged 9, 12, and 15.

Invariance Across Age For Each Gender and Across Gender For Each Age.

Consistent with results based on tests across all 6 groups, there is good support for the invariance of factor loadings and factor correlations within each of the more specific tests summarized in Table 4. For each age considered separately, there is support for the invariance of these parameter estimates across scores for boys and girls. Similarly, for boys and girls considered separately, there is good support for the invariance of these parameters over age. These results further substantiate interpretations based on all 6 groups.

Insert Tables 4 and 5 About Here

Also consistent with earlier analyses of the 6 groups, there is poorer support for the invariance of factor variances and uniquenesses. This lack of invariance, however, varies consistently depending on the comparison. The lack of invariance is evident in tests of invariance over age considered separately for girls and for boys (Table 4). In contrast, there is better support for the invariance of factor variances and, perhaps, uniquenesses over gender for the separate analyses of each age group. In particular, there is good support for the complete invariance of all parameter estimates across gender for 12 year olds (TLIs of .896 vs. .889; Table 4) and for the invariance of at least the factor variances for 9 year olds (TLIs of .878 vs. .878; Table 4). Support for the invariance of these parameter estimates across gender is weaker for 15 year olds. These results suggest that the lack of invariance in factor variances and factor uniquenesses is due primarily to comparisons across different ages, although there are also gender differences in these parameter estimates for 15 year olds.

Finally, combining these results, I examined various combinations of invariance constraints to find the most parsimonious model (see Table 5) that is best able to fit the data. Beginning with factor variances, the most parsimonious model able to fit the data constrains factor variances to be: equal across 9 year old boys and girls; equal across 12 year old boys and girls and 15 year old girls; freely estimated for 15 year-old boys. For uniquenesses, only the invariance across 12 year old boys and girls is supported. This "final" model is remarkably parsimonious as indicated by the large parsimony index (.968). It requires only 208 parameters to be estimated across the six groups which is about half of the 408 parameter estimates required by the original model with no invariance constraints. Despite this substantial reduction in the number of estimated parameters, the goodness of fit evaluated by the TLI is marginally better for the more constrained model (TLIs of .890 vs .888). The RNI that is monotonic with model complexity is necessarily better for the original unconstrained model (.908 vs. .892) but the difference is small in relation to the change in parsimony. The parsimony index more severely penalizes model complexity, and so the advantage of the final constrained model over the original unconstrained model is even more extreme according to this index (.864 vs. .742). In fact, the parsimony index leads to the selection of even more highly constrained models than the "final" model, although the differences are small (see Table 4).

Substantive Evaluation of Parameter Estimates.

The final constrained model (Table 2) differs from the original a priori model (Table 1) in two major respects. First, the correlation between the Dynamic and Explosive Strength factors was so large that the separation of the two factors could not be substantiated. Second, the PWC170 was posited to reflect the Cardiovascular factor, but it loads more highly on Static Strength. Also, whereas pushups also loads on Girth this is not unexpected (see related finding for weight reported by Fleishman, 1964) and pushups loads more highly on the Dynamic/Explosive Strength factor that it was intended to measure. The correlated uniquenesses are also theoreticaliy relevant. The correlated uniqueness relating the two running measures -- 1.6K run and 50M dash -- are consistent with the general running measure proposed by Fleishman. The correlated uniqueness relating the two shoulder strength tests, the two hand grip tests, and the two arm skinfold measures may also be consistent with Flieshman's suggestion that fitness is specific to particular parts of the body.

Although no specific pattern of correlations among the different facets was posited a priori, these results are a potentially important contribution of the present investigation -- particularly since there was such good support for the invariance of the correlations over boys and girls and across the three ages. The Cardiovascular factor is substantially correlated with Static Strength (.578), but not with Dynamic/Explosive Strength (.067) -- even though Dynamic/Explosive a: d Static Strength are substantially correlated with each other (.549).

Not surprisingly, the correlation between Girth and Skinfold (.871) is the largest of all the factor correlations. These two factors could not, however, be combined into a single factor without substantially hurting goodness of fit. Consistent with this finding is the observation that Dynamic/Explosive strength is more negatively correlated with Skinfold (-.443) than with Girth (-.175), whereas Static strength and Lung Function are more positively correlated with Girth (.572 and .467) than with Skinfold (.210 and .168). This contrasting pattern of relations involving the Girth and Skinfold factors is theoretically reasonable and demonstrates why it may be inappropriate to combine the two factors. Both the Girth and Skinfold factors, however, are similarly related to the Cardiovascular (-.422 and -.492) and Blood Pressure (.346 and .240) factors.

Flexibility is not substantially correlated with any of the other fitness factors, although it has small positive correlations with the two strength factors (rs of .272 and .229). Similarly, Blood Pressure is not substantially correlated with the other fitness factors, although it is positively related to Girth, Skinfold, Lung Function, and Static Strength factors, but negatively correlated with the Cardiovascular factor.

Lung Function factor is substantially correlated with Girth (.467) and Static Strength (.698), less substantially correlated with other fitness factors, and nearly uncorrelated with the Cardiovascular factor (-.029). Whereas the relation between Lung Function and Static Strength reflects in part the relation of both these factors to Girth, the size of this relation seems surprisingly large.

Further Tests of the "Final" Model.

It is also useful to provide further tests of the "final" model and to demonstrate additional features of the CFA approach. Whereas constraints on the a priori predictions were required to achieve a model that resulted in proper solutions when tests were conducted separately for each group, considerable robustness to the solution is added by the introduction of invariance constraints across the different groups. Thus, for example, it is not necessary to impose equality constraints on the two-indicator factors in order to achieve a proper solution, although relaxing this constraint did not substantively influence the overall pattern of parameter estimates and resulted in the same TLI=.890 as the "final" model.

Insert Table 6 About Here

Of greater substantive interest was an attempt to fit the original 9-factor model with the added stability of the invariance constraints in the final model. The resulting nine-factor model fit the data marginally better than the final model (TLIs of .891 vs. .890). Whereas the very large correlation between the Dynamic Strength and Explosive Strength factors (.925; see Table 6) was less than 1.0, the solution was technically improper in that the factor correlation matrix was not positive definite (see Joreskog & Sorbom, 1988). Also, the pattern of relations between these two Strength factors and the other fitness factors is very similar. Furthermore, when this model was fit separately to each of the 6 groups (i.e., there were no between-group invariance constraints), every solution was improper and the estimated correlation between Dynamic Strength and Explosive Strength was greater than 1.0 in some of the solutions. These observations -- particularly the finding that the 9-factor solution was still technically improper -- apparently provides support for the initial decision to combine the two factors for purposes of the present investigation. Substantively, the similarity in the pattern of correlations between these two factors and the remaining factors also supports this decision. More generally, other parameter estimates based on the 8-factor (Table 2) and 9-factor (Table 6) solutions are very similar, suggesting that combining or not combining these two factors is not a critical concern in the present investigation. It may be, however, that the two factors could be better differentiated in other studies that have more and, perhaps, better indicators of these two strength factors.

## Summary and Discussion

The focus of the present investigation is both substantive and methodological. Methodologically, the study demonstrates the CFA approach to testing the structure of physical fitness. This acknowledges and extends the important factor analytic tradition in physical fitness testing established by Fleishman (1964) that has apparently been neglected in the last decade. Substantively, the study is important in that it supports -- with some exceptions -- the a priori structure of physical fitness and the invariance of this structure across gender and age. These findings are important because the present investigation examined a more diverse sample of physical fitness indicators than is typically considered, and because the size and representativeness of the sample are better than most studies that include technical and laboratory measures of fitness.

Fundamental premises underlying the present investigation are that physical fitness is a hypothetical construct that must be validated within a construct validity approach, and that physical fitness is a multidimensional construct that cannot be adequately understood if this multidimensionality is ignored. Factor analysis, particularly the CFA approach demonstrated here, is an important statistical tool for evaluating both these premises. The final model (Table 2), although it differs somewhat from the originally posited model (Table 1), provides clear support for the multidimensionality of physical fitness. Support for this multidimensional structure is particularly strong in that at least the factor loadings and correlations among the factors are reasonably invariant for boys and girls aged 9, 12 and 15. This provides much stronger support for the empirical factor structure than would a test based on a single group.

This invariance of the factor structure also has important practical implications for physical fitness testing. As noted earlier, inadequate attention has been given to the question of whether a given indicator measures the same component of physical fitness with equal validity for boys and for girls, and across different ages. If the underlying meaning of a particular indicator differs depending on gender or age, then the task of interpreting each indicator and assessing physical fitness would be much more difficult. The invariance of the factor loadings and the factor correlations bear on two different aspects of this issue. The factor loadings reflect the relation between a particular indicator and the underlying latent constructs that it is posited to represent -- the validity of the indicator. For example, the fact that $VO_2max$ and the 1.6K run (but not PWC170) load substantially on the Cardiovascular factor support their construct validity as indicators of this factor. Support for the invariance of these factors implies that these indicators are equally valid for boys and girls of different ages. The invariance of the factor correlations indicates that relations among the different factors are the same for boys and girls of different ages. Thus, for example, Skinfold is positively related to some components of physical fitness and negatively related to others, but the size and direction of these correlations are similar for boys and girls of different ages.

Despite some important strengths of the present investigation and the CFA approach, there are also important limitations. Critical limitations inherent in the CFA approach and, to some extent, the present investigation, are the number of indicators needed to infer each factor and the sample sizes. Obviously, a factor cannot be identified if there are no indicators of the factor. Thus, for example, Fleishman (1964) reported six dimensions related to speed, flexibility, balance, and coordination. In the present investigation, only one relevant measure of Flexibility (the sit-and-reach test) was included, and so this aspect of Fleishman's research could not be evaluated with the data considered here. More generally, in the CFA approach there should be at least three good indicators of each hypothesized factor. Particularly in relation to the original a priori model that posited nine factors (Table 1), this recommendation was only satisfied for only 3 of 9 hypothesized factors. This limitation of the existing data apparently contributed to the improper solutions based on the original a priori model and, perhaps, to the failure to distinguish between the Explosive Strength and Dynamic Strength factors. As demonstrated here, expedient solutions to this problem include imposing invariance constraints within or between groups, and fitting more parsimonious models with fewer latent factors and estimated parameters. A second limitation that was not such a serious problem in the present investigation is the large sample size required of CFA studies. There are no absolute guidelines about the minimum sample size that is required and the requirements may be somewhat idiosyncratic to particular applications. Nevertheless, sample sizes of at least 200 subjects (per group) are typically recommended and considerably more subjects may be required for models involving many measured variables, latent factors, and estimated parameters (e.g., Tanaka, 1987). Problems associated with small sample sizes can, perhaps, also be offset by fitting more parsimonious models or by imposing invariance constraints. Ultimately, however, these sample size requirements mean that the CFA approach may not be appropriate to many small-scale laboratory studies of physical fitness.

It is also relevant to consider the implications of the present investigation to notions of "overall" physical fitness. The clear support for the multiple dimensions of physical fitness and the small correlations among many of physical fitness factors imply that it is inappropriate to simply average the different factors -- or the indicators used to infer the factors -- to obtain an overall index of fitness. It is obvious that considerable information in the specific factors will be lost in the formation of single total score. A much more useful summary of physical fitness is a profile of scores in which each score is compared to standards established in relation to appropriate norm groups, criterion references, or multiple sets of scores for the same individual collected over an extended period of time (e.g., achieving a "personal best"). In relation to a particular criterion, it may be appropriate to provide a single summary score that represents an optimally weighted combination of the multiple dimensions in which the weights are established on the basis of theory, empirical researrh, and, perhaps, expert opinion. Even here, however, the weight assigned to each dimension is likely to vary considerably depending on the particular criterion (e.g., performance in different athl·tic tasks, physical health, or psychological well-being) and, perhaps, the manner in which the weights are established. Implicit in this weighted average approach is the recognition that not all dimensions of physical fitness may be relevant to all situations (e.g., the appropriate weight for a particular dimension in a given situation may be zero). Thus, consistent with the multidimensional perspective of physical fitness emphasized here (also see Fleishman, 1964), the most generally useful summary of physical fitness dimensions is a multidimensional profile of scores rather than a single indicator (e.g., $VO_2$max) or a total that is based on the implicit assumption that the importance of all dimensions is the same for all intended purposes of the physical fitness test.

## References

American Alliance for Health, Physical Education, Recreation, and Dance (1980). Health related physical fitness manual. Washington, D. C.: AAHPERD.

American College of Sports Medicine (1990). Guidelines for exercise testing and prescriptions (4th ed.). Philadelphia: Lea and Febiger.

Bar-Or, O. (1987). A commentary to children and fitness: A public health perspective. Research Quarterly for Exercise and Sport, 58, 304-307.

Baumartner, T. A. & Jackson, A. S. (1987). Measurement for evaluation in physical education and exercise science (3rd ed). Dubuque, Iowa: William C. Brown.

Bentler, P. M. (1990). Comparative fit indices in structural models. Psychological Bulletin, .

Bentler, P. M. & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88, 588-606.

Bollen, K. A. (1989. Structural equations with latent variables. New York: Wiley.

Boutcher, S. H. (1990). Aerobic fitness: Measurement and issues. Journal of Sport & Exercise Psychology, 12, 235-247.

Byrne, B. M. (1989b). A primer of LISREL: Basic applications and programming for confirmatory factor analytic models. New York: Springer Verlag.

Byrne, B. M., & Shavelson, R. J., & Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial invariance. Psychological Bulletin, 105, 456-466.

Clarke, H. H. (1976). Application of measurement to health and physical education. Englewood Cliffs, NJ: Prentice-Hall.

Clarke, H. H. (1979). Definition of physical fitness. Journal of applied physical education and recreation, 50, 28.

Cooper, K. H. (1968). A means of assessing maximum oxygen intake. Journal of the American Medical Association, 203, 201-204.

Cunningham, D. A. (1980). Physical working capacity of children and adolescents. In G. A. Stull & T. K. Cureton (Eds,), Encyclopaedia of physical education, fitness, and sports (pp. 481-494 ). Salt Lake City, Utah: Brighton.

Cureton, K. J. (1980). The AAHPER Youth Fitness Test. In G. A. Stull & T. K. Cureton (Eds,), Encyclopaedia of physical education, fitness, and sports (pp. 425-443 ). Salt Lake City, Utah: Brighton.

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

Fleishman, F. A. (1964) The structure and measurement physical fitness. Englewood Cliffs, NJ: Prentice-Hall.

Hagan, R. D., Parrish, G., & Licciardone, J. C. (1991). Physical fitness is inversely related to heart disease risk: A factor analytic study. American Journal of Preventive Medicine, 7, 237-243.

Hayduk, L. A. (1987). Structural equation models with LISREL: Essentials and advances. Baltimore: John Hopkins University Press.

Joreskog, K. G., & Sorbom, D. (1988). LISREL 7: A guide to the program and applications. Chicago: SPSS, Inc.

Krahenbuhl, G. S. (1980). Individual differences and the assessment of youth fitness. In G. A. Stull & T. K. Cureton (Eds,), Encyclopaedia of physical education, fitness, and sports (pp. 470-480 ). Salt Lake City, Utah: Brighton.

Long, J. S. (1983). Confirmatory factor analysis. Newbury Park, CA: Sage.

Malina, R. M. (1989). 1988 C. H. McCloy Research Lecture: Children in the exercise sciences. Research Quarterly for Exercise and Sport, 60, 305-317.

Marsh, H. W. (1987). The factorial invariance of responses by males and females to a multidimensional self-concept instrument: Substantive and methodological issues. Multivariate Behavioral Research, 22, 457-480.

Marsh, H. W. & Balla, J. (April, 1992). Goodness of fit in confirmatory factor analysis: The effects of sample size and model parsimony. Paper presented at the 1992 Annual Meeting of the American Educational Research Association, San Francisco.

Marsh, H. W., Balla, J. R. & McDonald, R. P. (1988). Goodness-of-fit indices in confirmatory factor analysis: The effect of sample size. Psychological Bulletin, 102, 391-410..

Marsh, H. W., & Hocevar, D. (1985). The application of confirmatory factor analysis to the study of self-concept: First and higher order factor structures and their invariance across age groups. Psychological Bulletin, 97, 562-582.

McDonald, R. P. (1985). Factor analysis and related methods. Hillsdale, NJ: Erlbaum.

McDonald, R. P, & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness-of-fit. Psychological Bulletin, 107, 247-255.

Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. Psychological Bulletin, 105, 430-445.

National Center for Educational Statistics (1986). High School and Beyond, 1980: Sophomore cohort second follow-up (1984). Data file user's manual. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

Pedhazur, E. J. & Schmelkin, L. P. (1991). Measurement, design, and analysis: An integrated approach. Hillsdale, NJ: Erlbaum.

Ponthieux, N. A., & Barker, D. G. (1963). An analysis of the AAHPER Youth Fitness Test. Research Quarterly for Exercise and Sport, 34, 525-526.

Pyke, J. E. (1987). Australian Health and Fitness Survey. Parkside, South Australia: The Australian Council for Health, Physical Education and Recreation.

Safrit, M. J. (1981). Evaluation in physical education. Englewood Cliffs, NJ: Prentice-Hall.

Sallis, J. F. (1987). A commentary to children and fitness: A public health perspective. Research Quarterly for Exercise and Sport, 58, 326-330.

Schell, J. & Leelarthaepin, B. (1990). Physical fitness assessment in exercise and sport science. Matraville, NSW, Australia: Leelar Biomediscience Services.

Seefeldt, V., & Vogel, P. (1987). Children and fitness: A public health perspective. Research Quarterly for Exercise and Sport, 58, 331-333.

Sharkey, B. J. (1988). Specificity for testing. In W. Granna, J. Lombardo, B. Sharkey, & J. Stone (Eds). Advances in sports medicine and fitness (pp. 25-40). Chicago: Year Book Medical.

Sharkey, B. J. (1991). New dimensions in aerobic Fitness. Champaign, IL: Human Kinetics Books.

Tanaka, J. S. (1987). "How big is big enough?": Sample size and goodness of fit in structural equation models with latent variables. Child Development, 58, 134-146.

Taylor, W., & Baranowski, T. (1991). Physical activity, cardiovascular fitness, and adiposity in children. Research Quarterly for Exercise and Sport, 62, 157-163.

Zwiren, L. D., Freedson, P. S., Ward, A., Wilke, S. & Rippe, J. M. (1991). Estimation of $VO_2$max: A comparative analysis of five exercise tests. Research Quarterly for Exercise and Sport, 62, 73-78.

Table 1

A Priori Fitness Factors and a Description of Indicators

Cardiovascular Endurance

Maximal Oxygen Uptake ($VO_2max$). A continuous direct measure of maximal oxygen uptake (in mL./kg minute) was taken using a treadmill. The initial treadmill speed varied according to age and sex, and was increased 2% every two minutes until criteria of steady state were achieved. All results were compiled at the central survey office.

1.6 K Run. Measured on an oval track of 200 m or 400 m.

Physical Work Capacity (PWC170). Measured in Kmg/kg x min using a Monark bicycle as a continuous test with 3 workloads of 3 minutes each, each being at higher workload that the previous workload. Direct measures of heart rate were taken with a stethoscope and stopwatch, and PWC170 scores were generated by computer.

Explosive Strength

50 M dash. Measured (in sec.), after a warm-up, in a single sprint over a flat, cross-wind track.

Standing Long Jump. The longest of two jumps (in cms) done from a standing take-off.

Dynamic Strength

Sit-ups. These were done with knees were bent to 140° in a cadence of 20/minute up to a maximum of 100.

Push-ups. The number done in 30 seconds, using a 46 cm chair with student's feet behind a line set at their elbow height from the front of the chair.

Static Strength (in Kg.)

Right Grip Strength. The best of two trials with the dynamometer resting on the opposite shoulder.

Left Grip Strength. The best of two trials with the dynamometer resting on the opposite shoulder.

Shoulder Push Strength. The best of two trials with the dynamometer at the level of nipples.

Shoulder Pull Strength. The best of two trials with the dynamometer at the level of nipples.

Leg Strength. The best of two trials with back against the wall and knees bent at 115°.

Flexibility/Joint Mobility

Sit and reach. The student was seated, stretched as far as possible to hold for 3 seconds. Score is the cms. beyond their toes (or negative scores if they do not reach their toes)..

Blood Pressure (mm Hg measured after 5 min. rest with a mercury sphygmomanometer)

Systolic Blood Pressure. The Korotkoff sound I.

Diastolic Blood Pressure. The mean of Korotkoff sounds IV and V.

Lung Function (in L using a Vitalograph adjusted to student's height).

FEV1. The 1 second Forced Expiratory Volume

FVC. Forced Vital Capacity

Body Girth (cm. assessed using a constant tension tape)

Mid-arm Girth

Waist Girth

Hip Girth

Skinfolds (in mm measured with a Holtain calliper).

Biceps Skinfold

Triceps Skinfold

Subscapular Skinfold

Suprailiac Skinfold

Midabdominal Skinfold

Note. The a priori categories are based on the design of the battery (Pyke, 1987) and previous research -- particularly Fleishman (1964).

Table 2

Eight Factor Solution For 12 Year Old Boys Standardized To a Common Metric

| Variables | Factor Loadings | | | | | | | | | Corr[a] |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Uniq | Uniq |
| VO2M | .711 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .582 | |
| 1.6K Run | .723 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .448 | |
| PWC170 | .078 | 0 | .522 | 0 | 0 | 0 | 0 | 0 | .683 | |
| 50M Dash | 0 | .659 | 0 | 0 | 0 | 0 | 0 | 0 | .541 | .159 |
| Long Jump | 0 | .699 | 0 | 0 | 0 | 0 | 0 | 0 | .556 | |
| Situp | 0 | .436 | 0 | 0 | 0 | 0 | 0 | 0 | .953 | |
| Pushup | 0 | .427 | 0 | 0 | 0 | 0 | -.241 | 0 | .723 | |
| Rt Grip | 0 | 0 | .780 | 0 | 0 | 0 | 0 | 0 | .357 | |
| Lft Grip | 0 | 0 | .771 | 0 | 0 | 0 | 0 | 0 | .351 | .196 |
| Shld Pull | 0 | 0 | .710 | 0 | 0 | 0 | 0 | 0 | .339 | |
| Shld Push | 0 | 0 | .644 | 0 | 0 | 0 | 0 | 0 | .516 | .082 |
| Leg | 0 | 0 | .644 | 0 | 0 | 0 | 0 | 0 | .529 | |
| Sit/reach | 0 | 0 | 0 | 1.000 | 0 | 0 | 0 | 0 | 0 | |
| Systolic | 0 | 0 | 0 | 0 | .769 | 0 | 0 | 0 | .398 | |
| Diastolic | 0 | 0 | 0 | 0 | .680 | 0 | 0 | 0 | .530 | |
| FVC | 0 | 0 | 0 | 0 | 0 | .959 | 0 | 0 | .043 | |
| FEV1 | 0 | 0 | 0 | 0 | 0 | .912 | 0 | 0 | .118 | |
| Arm Girth | 0 | 0 | 0 | 0 | 0 | 0 | .890 | 0 | .098 | |
| Waist Grth | 0 | 0 | 0 | 0 | 0 | 0 | .877 | 0 | .186 | |
| Hip Girth | 0 | 0 | 0 | 0 | 0 | 0 | .885 | 0 | .191 | |
| Skinfold1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .874 | .221 | |
| Skinfold2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .820 | .294 | .039 |
| Skinfold3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .901 | .188 | |
| Skinfold4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .905 | .169 | |
| Skinfold5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .913 | .152 | |

Factor Correlations

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Caridovas | 1 | | | | | | | |
| Explo/Dyn | .578 | 1 | | | | | | |
| Static | .067 | .549 | 1 | | | | | |
| Flexibil | .111 | .272 | .229 | 1 | | | | |
| Blood Pr | -.172 | .106 | .306 | .100 | 1 | | | |
| Lung | -.029 | .269 | .698 | .123 | .257 | 1 | | |
| Girth | -.422 | -.175 | .572 | .005 | .346 | .467 | 1 | |
| Skinfold | -.492 | -.443 | .210 | -.069 | .240 | .168 | .871 | 1 |

Factor Variances

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | .882 | 1.060 | .950 | 1.019 | .983 | .988 | 1.119 | 1.159 |

Note: See Table 1 for a description of the measured variables. This solution based on data for 12 year old boys was standardized in relation to a common metric across all 6 groups to facilitate interpretations (see Joreskog & Sorbom, 1989). The factor loadings and factor correlations that are the major focus of the present investigation were invariant across all 6 groups, but factor variances and uniquenesses varied depending on the group (see discussion of results). Thus factor loadings and factor correlations presented here are the same for all 6 groups. The average factor variance across the 6 groups is 1.0 but differs from group to group.

[a] Correlated uniquenesses suggest that the relationship between a pair of measured variables is greater than can be explained by the posited structure. The four correlated uniqueness posited here represent relations between the two running measures (1.6 k run and 50 m dash); the two static strength indicators involving the shoulder (shoulder push and shoulder pull); the two static strength measures involving hand grip strength (right grip strength and left grip strength); and the two skinfold measures involving the arm (bicep and tricep).

Table 3

Goodness of Fit For Separate Solutions For Each Group With No Invariance Constraints and For Selected Invariance Constraints Imposed Across All Groups

| Model | P-Ratio | $X^2$ | DF | TLI | RNI | PRNI |
|---|---|---|---|---|---|---|
| No Invariance Constraints | | | | | | |
| Boys Age=9 | .817 | 600 | 245 | .894 | .914 | .746 |
| Boys Age=12 | .817 | 789 | 245 | .899 | .918 | .749 |
| Boys Age=15 | .817 | 664 | 245 | .906 | .924 | .754 |
| Girls Age=9 | .817 | 834 | 245 | .865 | .890 | .727 |
| Girls Age=12 | .817 | 791 | 245 | .876 | .899 | .734 |
| Girls Age=15 | .817 | 629 | 245 | .881 | .902 | .737 |
| 6-Group Total | .817 | 4309 | 1470 | .888 | .908 | .742 |
| Invariance Constraints | | | | | | |
| (the same constraint over all 6 groups) | | | | | | |
| No Invar | .817 | 4309 | 1470 | .888 | .908 | .742 |
| FL Invar | .861 | 4594 | 1550 | .886 | .902 | .776 |
| FL,FCr Invar | .939 | 4884 | 1690 | .890 | .897 | .842 |
| FL,FCr, FV Invar | .961 | 5506 | 1730 | .873 | .878 | .844 |
| FL,Fcr, U Invar | 1.017 | 7349 | 1830 | .824 | .821 | .835 |
| Total (FL,FCr,FV,U) Invar | 1.039 | 8188 | 1870 | .803 | .796 | .827 |

Note. P-Ratio = Parsimony Ratio. TLI =Tucker-Lewis Index. RNI = Relative Noncentrality Index. PRNI = Parsimony RNI. FL = Factor loadings. FCr = Factor Correlations. FV = Factor Variances. U = uniquenesses. The total $X^2$ and degrees of freedom summed across the 6 groups considered separately is necessarily the same as for the corresponding analysis across the 6 groups in which no invariance constraints are imposed.

Table 4
Goodness of Fit For Solutions With Selected Invariance Constraints Imposed Across Age Within Gender and Across Gender Within Age

| Model | P-Ratio | $X^2$ | DF | TLI | RNI | PRNI |
|---|---|---|---|---|---|---|
| No Invariance Constraints | | | | | | |
| Across Age Within Gender | | | | | | |
| Across Age For Boys | .817 | 2054 | 735 | .900 | .919 | .750 |
| Across Age For Girls | .817 | 2256 | 735 | .857 | .883 | .721 |
| Total | .817 | 4309 | 1470 | .888 | .908 | .742 |
| Across Gender Within Age | | | | | | |
| Across Gender For Age=9 | .817 | 1435 | 490 | .878 | .900 | .735 |
| Across Gender For Age=12 | .817 | 1580 | 490 | .889 | .909 | .743 |
| Across Gender For Age=15 | .817 | 1294 | 490 | .896 | .915 | .747 |
| Total | .817 | 4309 | 1470 | .888 | .908 | .742 |
| Factor Loading Invariance | | | | | | |
| Across Age Within Gender | | | | | | |
| Across Age For Boys | .852 | 2155 | 767 | .899 | .914 | .779 |
| Across Age For Girls | .852 | 2365 | 767 | .856 | .877 | .747 |
| Total | .852 | 4520 | 1534 | .887 | .903 | .770 |
| Across Gender Within Age | | | | | | |
| Across Gender For Age=9 | .843 | 1495 | 506 | .876 | .896 | .755 |
| Across Gender For Age=12 | .843 | 1601 | 506 | .892 | .909 | .766 |
| Across Gender For Age=15 | .843 | 1390 | 506 | .889 | .906 | .764 |
| Total | .843 | 4487 | 1518 | .886 | .904 | .762 |
| Factor Loading, Factor Corr Invariance | | | | | | |
| Across Age Within Gender | | | | | | |
| Across Age For Boys | .914 | 2251 | 823 | .904 | .912 | .834 |
| Across Age For Girls | .914 | 2555 | 823 | .854 | .867 | .792 |
| Total | .914 | 4763 | 1646 | .890 | .899 | .822 |
| Across Gender Within Age | | | | | | |
| Across Gender For Age=9 | .890 | 1541 | 534 | .881 | .894 | .796 |
| Across Gender For Age=12 | .890 | 1636 | 534 | .897 | .908 | .808 |
| Across Gender For Age=15 | .890 | 1469 | 534 | .889 | .901 | .802 |
| Total | .890 | 4637 | 1602 | .890 | .902 | .803 |
| Factor Loading, Factor Corr, Factor Variance Invariance | | | | | | |
| Across Age Within Gender | | | | | | |
| Across Age For Boys | .932 | 2651 | 839 | .880 | .888 | .828 |
| Across Age For Girls | .932 | 2646 | 839 | .851 | .861 | .802 |
| Total | .932 | 5229 | 1678 | .877 | .885 | .825 |
| Across Gender Within Age | | | | | | |
| Across Gender For Age=9 | .903 | 1589 | 542 | .878 | .890 | .804 |
| Across Gender For Age=12 | .903 | 1643 | 542 | .899 | .908 | .821 |
| Across Gender For Age=15 | .903 | 1545 | 542 | .882 | .894 | .807 |
| Total | .903 | 4777 | 1626 | .887 | .898 | .811 |
| Total (Factor Loading, Corr, Var, and Unique) Invariance | | | | | | |
| Across Age Within Gender | | | | | | |
| Across Age For Boys | .994 | 3827 | 895 | .818 | .819 | .814 |
| Across Age For Girls | .994 | 3596 | 895 | .791 | .792 | .788 |
| Total | .994 | 7423 | 1790 | .817 | .818 | .813 |
| Across Gender Within Age | | | | | | |
| Across Gender For Age=9 | .950 | 1810 | 570 | .862 | .869 | .826 |
| Across Gender For Age=12 | .950 | 1757 | 570 | .896 | .901 | .856 |
| Across Gender For Age=15 | .950 | 2115 | 570 | .827 | .836 | .794 |
| Total | .950 | 5683 | 1710 | .865 | .871 | .828 |

Note. P-Ratio = Parsimony Ratio. TLI =Tucker-Lewis Index. RNI = Relative Noncentrality Index. Ck = Cross-validation Index. PRNI = Parsimony RNI. FL = Factor loadings. FCr = Factor Correlations. FV = Factor Variances. U = uniquenesses.

Table 5

Goodness of Fit For Solutions With Selected Invariance Constraints Imposed On Specific Groups of Boys (B) and Girls(G) Aged 9, 12 and 15.

| Model | P-Ratio | $X^2$ | DF | TLI | RNI | PRNI |
|---|---|---|---|---|---|---|
| Factor loadings, Factor Correlations Invariant Across All Groups;<br>Uniquenesses Non-Invariant (free) For all groups;<br>Patterns of Factor Variance Invariances as Follows: | | | | | | |
| FV: B9=B12=G9=G12=G15; B15=NI | .957 | 5182 | 1722 | .883 | .888 | .850 |
| FV: B9=B12=G9=G12; G15=NI; B15=NI | .952 | 5140 | 1714 | .884 | .889 | .847 |
| FV: B9=G9; B12=G12; G15=NI; B15=NI | .948 | 4944 | 1706 | .890 | .895 | .849 |
| FV: B9=G9; B12=G12=G15; B15=NI | .952 | 4958 | 1714 | .890 | .895 | .852 |
| Factor loadings, Factor Correlations Invariant Across All Groups;<br>Patterns of Uniqueness (U) and Factor Variance (FV) Invariances as Follows: | | | | | | |
| U: B9=G9=B12=G12; G15=NI; B15=NI;<br>FV: B9=G9=B12=G12; G15=NI; B15=NI; | .999 | 5810 | 1798 | .870 | .870 | .869 |
| U: B9=G9; B12=G12; G15=NI; B15=NI;<br>FV: B9=G9; B12=G12; G15=NI; B15=NI; | .979 | 5263 | 1762 | .884 | .887 | .868 |
| U: B12=G12; B9=NI; G9=NI; G15=NI; B15=NI;<br>FV: B9=G9; B12=G12; G15=NI; B15=NI. | .963 | 5052 | 1734 | .889 | .893 | .860 |
| U: B9=G9; B12=G12=G15; B15=NI;<br>FV: B9=G9; B12=G12=G15; B15=NI; | .983 | 5492 | 1770 | .878 | .880 | .865 |
| U: B12=G12; B9=NI; G9=Ni; G15=NI; B15=NI;<br>FV: B9=G9; B12=G12=G15; B15=NI; | .968 | 5068 | 1742 | .890 | .892 | .864 |

Note. P-Ratio = Parsimony Ratio. TLI =Tucker-Lewis Index. RNI = Relative Noncentrality Index. PRNI = Parsimony RNI. FL = Factor loadings. FCr = Factor Correlations. FV = Factor Variances. U = uniquenesses. For all models summarized in this table, factor loadings and factor correlations were invariant across all six gender/age groups. In the first set of models, factor variances -- but not uniquenesses -- were constrained to be equal across various combinations of groups. Thus, for example, in the final factor variances were held to be invariant in solutions for boys and girls aged 9, and for boys and girls aged 12 and girls aged 15; factor variances for boys aged 15 were not constrained to be invariant with any other groups. In the second set of models, factor variances and uniquenesses were constrained to be equal across different combinations of groups.

Table 6

Nine Factor Solution For 12 Year Old Boys Standardized To A Common Metric

| | Factor Loadings | | | | | | | | | | Corr[a] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Variables | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Uniq | Uniq |
| VO2M | .676 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .622 | |
| 1.6K Run | .755 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .415 | |
| PWC170 | .083 | 0 | 0 | .523 | 0 | 0 | 0 | 0 | 0 | .682 | |
| 50M Dash | 0 | .695 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .492 | .154 |
| Long Jump | 0 | .725 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .523 | |
| Situp | 0 | 0 | .432 | 0 | 0 | 0 | 0 | 0 | 0 | .940 | |
| Pushup | 0 | 0 | .423 | 0 | 0 | 0 | 0 | -.209 | 0 | .732 | |
| Rt Grip | 0 | 0 | 0 | .780 | 0 | 0 | 0 | 0 | 0 | .360 | |
| Lft Grip | 0 | 0 | 0 | .771 | 0 | 0 | 0 | 0 | 0 | .353 | .199 |
| Shld Pull | 0 | 0 | 0 | .712 | 0 | 0 | 0 | 0 | 0 | .337 | |
| Shld Push | 0 | 0 | 0 | .644 | 0 | 0 | 0 | 0 | 0 | .515 | .081 |
| Leg | 0 | 0 | 0 | .645 | 0 | 0 | 0 | 0 | 0 | .529 | |
| Sit/reach | 0 | 0 | 0 | 0 | 1.000 | 0 | 0 | 0 | 0 | 0 | |
| Systolic | 0 | 0 | 0 | 0 | 0 | .920 | 0 | 0 | 0 | .144 | |
| Diastolic | 0 | 0 | 0 | 0 | 0 | .560 | 0 | 0 | 0 | .676 | |
| FVC | 0 | 0 | 0 | 0 | 0 | 0 | .982 | 0 | 0 | .001 | |
| FEV1 | 0 | 0 | 0 | 0 | 0 | 0 | .888 | 0 | 0 | .158 | |
| Arm Girth | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .890 | 0 | .165 | |
| Waist Grth | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .877 | 0 | .185 | |
| Hip Girth | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .885 | 0 | .191 | |
| Skinfold1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .873 | .222 | |
| Skinfold2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .821 | .294 | .039 |
| Skinfold3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .901 | .189 | |
| Skinfold4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .904 | .170 | |
| Skinfold5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .914 | .151 | |
| Factor Correlations | | | | | | | | | | | |
| Caridovas | 1 | | | | | | | | | | |
| Explosive | .547 | 1 | | | | | | | | | |
| Dynamic | .702 | .925 | 1 | | | | | | | | |
| Static | .070 | .517 | .558 | 1 | | | | | | | |
| Flexibil | .109 | .242 | .341 | .229 | 1 | | | | | | |
| Blood Pr | -.168 | .095 | .145 | .278 | .092 | 1 | | | | | |
| Lung | -.019 | .256 | .183 | .690 | .125 | .231 | 1 | | | | |
| Girth | -.420 | -.165 | -.256 | .573 | .005 | .321 | .465 | 1 | | | |
| Skinfold | -.499 | -.418 | -.529 | .210 | -.069 | .216 | .170 | .871 | 1 | | |
| Factor Variances | | | | | | | | | | | |
| | .875 | 1.032 | 1.158 | .945 | 1.020 | .974 | .989 | 1.122 | 1.162 | | |

Note: See Table 1 for a description of the measured variables. This solution based on data for 12 year old boys was standardized in relation to a common metric across all 6 groups to facilitate interpretations (see Joreskog & Sorbom, 1989). The factor loadings and factor correlations that are the major focus of the present investigation were invariant across all 6 groups, but factor variances and uniquenesses varied depending on the group (see discussion of results). This 9-factor solution is technically improper in that the factor correlation matrix is not positive definite; even though none of correlations exceeds 1, the correlation between the dynamic and explosive strength factors (r=.925) approaches 1.

[a] Correlated uniquenesses suggest that the relationship between a pair of measured variables is greater than can be explained by the posited structure. The four correlated uniqueness posited here represent relations between the two running measures (1.6 K run and 50 M dash); the two static strength indicators involving the shoulder (shoulder push and shoulder pull); the two static strength measures involving hand grip strength (right grip strength and left grip strength); and the two skinfold measures involving the arm (bicep and tricep).

Appendix 1A

Descriptive Statistics for the 25 Physical Fitness Variables: Means and Standard Deviations For Six Groups

| | Boys | | | | | | Girls | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variables | Age= 9 | | Age= 12 | | Age= 15 | | Age= 9 | | Age= 12 | | Age= 15 | |
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| VO2M | .39 | .94 | .26 | .84 | .79 | .87 | -.52 | .73 | -.50 | .92 | -.62 | .78 |
| 1.6K Run | .07 | .87 | .45 | .75 | .87 | .69 | -.67 | .96 | -.33 | .86 | -.41 | .96 |
| PWC170 | -.54 | .47 | .15 | .66 | 1.57 | .89 | -.91 | .40 | -.28 | .57 | .13 | .63 |
| 50M Dash | -.50 | .75 | .22 | .76 | 1.20 | .64 | -.89 | .87 | -.04 | .75 | .18 | .78 |
| Long Jump | -.56 | .62 | .16 | .75 | 1.37 | .83 | -.84 | .65 | -.16 | .72 | .16 | .71 |
| Situp | -.43 | .75 | .39 | 1.08 | .81 | .99 | -.52 | .70 | -.09 | .89 | -.12 | .87 |
| Pushup | .19 | .76 | .32 | .87 | 1.00 | .91 | -.29 | .76 | -.51 | .84 | -.78 | .74 |
| Rt Grip | -.75 | .38 | -.01 | .57 | 1.57 | .87 | -.97 | .34 | -.12 | .52 | .51 | .53 |
| Lft Grip | -.72 | .37 | .03 | .58 | 1.60 | .89 | -.95 | .34 | -.17 | .51 | .44 | .53 |
| Shld Pull | -.70 | .39 | -.03 | .59 | 1.49 | 1.06 | -.87 | .34 | -.17 | .50 | .54 | .64 |
| Shld Push | -.84 | .30 | -.21 | .58 | 1.46 | 1.02 | -.79 | .34 | -.02 | .53 | .68 | .58 |
| Leg | -.66 | .42 | .15 | .63 | 1.58 | .95 | -.87 | .38 | -.23 | .53 | .22 | .62 |
| Sit/reach | -.41 | .76 | -.53 | .86 | -.13 | 1.07 | .14 | .83 | .23 | .92 | .81 | .96 |
| Systolic | -.49 | .86 | -.09 | .89 | .81 | .94 | -.46 | .89 | -.02 | .90 | .37 | .87 |
| Diastolic | -.24 | 1.05 | -.14 | .99 | .29 | .95 | -.16 | .97 | .03 | .99 | .29 | .90 |
| FVC | -.75 | .38 | .01 | .55 | 1.54 | .84 | -1.00 | .37 | -.09 | .53 | .57 | .55 |
| FEV1 | -.77 | .38 | -.05 | .54 | 1.48 | .85 | -.99 | .38 | -.05 | .55 | .66 | .58 |
| Arm Girth | -.79 | .68 | -.01 | .87 | .96 | .83 | -.67 | .67 | -.03 | .76 | .71 | .78 |
| Waist Grth | -.63 | .69 | .16 | .97 | .91 | .78 | -.73 | .72 | -.08 | .84 | .51 | .88 |
| Hip Girth | -.91 | .57 | -.09 | .75 | .88 | .70 | -.85 | .60 | .17 | .71 | 1.01 | .71 |
| Skinfold1 | -.41 | .73 | -.11 | 1.06 | -.58 | .73 | .15 | .87 | .21 | .92 | .82 | 1.03 |
| Skinfold2 | -.35 | .75 | -.11 | 1.01 | -.56 | .73 | .21 | .92 | .28 | .99 | .55 | 1.12 |
| Skinfold3 | -.46 | .66 | -.10 | 1.12 | -.14 | .78 | -.02 | 1.01 | .17 | 1.03 | .63 | 1.00 |
| Skinfold4 | -.42 | .70 | .02 | 1.19 | -.13 | .92 | -.02 | 1.00 | .15 | 1.00 | .47 | .90 |
| Skinfold5 | -.46 | .78 | -.01 | 1.14 | -.14 | .87 | -.04 | .98 | .18 | 1.00 | .56 | .90 |

Note: See Table 1 for a description of the variables. All variables were standardized (Mean=0 SD=1) across the total group. Thus, for example, VO$_2$max for 9-year-old boys is .39 (total group) standard deviations above the mean for the total group, and the scores are slightly less variable (SD=.94 vs. 1.0 for the total group). Times for the running measures (1.6K run and 50M dash) were multiplied by -1 so that positive scores reflected more positive fitness.

Appendi. 1B

Descriptive Statistics for the 25 Physical Fitness Variables: Correlations for Three Groups of Boys

SEX=BOYS AGE=9

1

| | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| VO2M | 1 | | | | | | | | | | | | | | | | | | | | | | | | |
| 16K Run | 65 | 1 | | | | | | | | | | | | | | | | | | | | | | | |
| PWC170 | 19 | 13 | 1 | | | | | | | | | | | | | | | | | | | | | | |
| 50M Dash | 19 | 44 | 06 | 1 | | | | | | | | | | | | | | | | | | | | | |
| Long Jump | 36 | 35 | 08 | 41 | 1 | | | | | | | | | | | | | | | | | | | | |
| Situp | 32 | 19 | 07 | 32 | 25 | 1 | | | | | | | | | | | | | | | | | | | |
| Pushup | 38 | 31 | 06 | 26 | 32 | 21 | 1 | | | | | | | | | | | | | | | | | | |
| Rt Grip | 08 | -04 | 26 | 24 | 14 | 21 | 09 | 1 | | | | | | | | | | | | | | | | | |
| Lft Grip | 09 | -04 | 31 | 24 | 16 | 20 | 08 | 80 | 1 | | | | | | | | | | | | | | | | |
| Shld Pull | 24 | 02 | 25 | 29 | 20 | 20 | 10 | 49 | 51 | 1 | | | | | | | | | | | | | | | |
| Shld Push | 34 | 00 | 18 | 16 | 07 | 10 | 04 | 37 | 38 | 47 | 1 | | | | | | | | | | | | | | |
| Leg | -18 | 03 | 16 | 16 | 16 | 13 | 09 | 47 | 43 | 39 | 19 | 1 | | | | | | | | | | | | | |
| Sit/reach | 15 | 15 | -05 | 14 | 24 | 12 | 17 | 12 | 10 | 17 | 05 | 13 | 1 | | | | | | | | | | | | |
| Systolic | -22 | -18 | 03 | 13 | -04 | 03 | -04 | 27 | 28 | 15 | 20 | 10 | 02 | 1 | | | | | | | | | | | |
| Diastolic | 02 | -13 | 08 | 01 | -02 | 05 | 05 | 17 | 20 | 07 | 11 | 00 | 00 | 57 | 1 | | | | | | | | | | |
| FVC | -09 | 01 | 35 | 12 | 07 | 13 | -10 | 41 | 41 | 41 | 41 | 25 | -03 | 23 | 16 | 1 | | | | | | | | | |
| FEV1 | -20 | 04 | 32 | 12 | 05 | 11 | -08 | 36 | 35 | 37 | 35 | 20 | -06 | 18 | 11 | 88 | 1 | | | | | | | | |
| Arm Girth | -25 | -33 | 23 | -06 | -15 | -02 | -25 | 40 | 39 | 38 | 36 | 23 | -06 | 28 | 12 | 39 | 31 | 1 | | | | | | | |
| Waist Grth | -26 | -34 | 23 | -11 | -18 | -03 | -29 | 34 | 35 | 33 | 32 | 20 | -10 | 25 | 15 | 41 | 32 | 81 | 1 | | | | | | |
| Hip Girth | -21 | -35 | 27 | -09 | -22 | 02 | -28 | 40 | 39 | 38 | 36 | 20 | -08 | 31 | 18 | 48 | 40 | 84 | 82 | 1 | | | | | |
| Skinfold1 | -34 | -45 | 13 | -28 | -32 | -11 | -33 | 24 | 23 | 17 | 20 | 10 | -12 | 23 | 14 | 24 | 20 | 75 | 71 | 76 | 1 | | | | |
| Skinfold2 | -25 | -37 | 09 | -23 | -25 | -13 | -34 | 22 | 21 | 19 | 20 | 11 | -09 | 20 | 11 | 23 | 18 | 75 | 70 | 68 | 85 | 1 | | | |
| Skinfold3 | -26 | -47 | 07 | -24 | -26 | -13 | -32 | 16 | 17 | 14 | 16 | 06 | -08 | [illegible] | 14 | 18 | 12 | 74 | 74 | 68 | 81 | 81 | 1 | | |
| Skinfold4 | -26 | -43 | 10 | -18 | 22 | -05 | -32 | 22 | 21 | 23 | 23 | 11 | -05 | [illegible] | 19 | 21 | 17 | 74 | 75 | 73 | 82 | 78 | 84 | 1 | |
| Skinfold5 | -39 | -41 | 09 | -21 | -24 | -10 | -31 | 21 | 22 | 20 | 21 | 13 | -09 | 22 | 13 | 22 | 18 | 79 | 80 | 74 | 82 | 84 | 86 | 89 | 1 |

SEX=BOYS AGE=12

| | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| VO2M | 1 | | | | | | | | | | | | | | | | | | | | | | | | |
| 16K Run | 54 | 1 | | | | | | | | | | | | | | | | | | | | | | | |
| PWC170 | 00 | 13 | 1 | | | | | | | | | | | | | | | | | | | | | | |
| 50M Dash | 34 | 53 | 21 | 1 | | | | | | | | | | | | | | | | | | | | | |
| Long Jump | 25 | 33 | 14 | 53 | 1 | | | | | | | | | | | | | | | | | | | | |
| Situp | 38 | 26 | 12 | 33 | 35 | 1 | | | | | | | | | | | | | | | | | | | |
| Pushup | 66 | 41 | -01 | 37 | 40 | 26 | 1 | | | | | | | | | | | | | | | | | | |
| Rt Grip | 03 | 02 | 46 | 28 | 25 | 21 | 03 | 1 | | | | | | | | | | | | | | | | | |
| Lft Grip | -05 | -01 | 46 | 27 | 28 | 24 | 05 | 84 | 1 | | | | | | | | | | | | | | | | |
| Shld Pull | -15 | 04 | 38 | 27 | 28 | 23 | 06 | 57 | 59 | 1 | | | | | | | | | | | | | | | |
| Shld Push | -11 | 04 | 37 | 23 | 25 | 18 | -02 | 46 | 49 | 59 | 1 | | | | | | | | | | | | | | |
| Leg | 04 | 07 | 35 | 24 | 29 | 26 | 02 | 59 | 59 | 54 | 37 | 1 | | | | | | | | | | | | | |
| Sit/reach | 16 | 03 | -06 | 09 | 18 | 13 | 19 | 08 | 13 | 15 | 14 | 17 | 1 | | | | | | | | | | | | |
| Systolic | -10 | -04 | 09 | 14 | 08 | 18 | -01 | 23 | 25 | 21 | 11 | 15 | 06 | 1 | | | | | | | | | | | |
| Diastolic | -02 | -10 | 07 | 00 | 02 | 11 | -04 | 13 | 10 | 07 | 06 | 07 | 01 | 51 | 1 | | | | | | | | | | |
| FVC | -14 | -01 | 43 | 22 | 25 | 19 | -05 | 57 | 57 | 54 | 48 | 50 | 03 | 22 | 10 | 1 | | | | | | | | | |
| FEV1 | -12 | -01 | 37 | 20 | 24 | 18 | -04 | 55 | 55 | 50 | 42 | 47 | 02 | 23 | 09 | 92 | 1 | | | | | | | | |
| Arm Girth | -49 | -37 | 29 | -13 | -16 | -08 | -32 | 47 | 46 | 45 | 38 | 32 | 01 | 25 | 10 | 42 | 39 | 1 | | | | | | | |
| Waist Grth | -43 | -41 | 31 | -20 | -20 | 13 | -42 | 40 | 40 | 36 | 34 | 28 | -02 | 18 | 09 | 42 | 39 | 88 | 1 | | | | | | |
| Hip Girth | -42 | -36 | 36 | -10 | -09 | -03 | -38 | [illegible] | 52 | 47 | 44 | 38 | 02 | 28 | 16 | 52 | 49 | 87 | 88 | 1 | | | | | |
| Skinfold1 | -46 | -49 | 07 | -37 | -32 | -26 | -44 | 22 | 20 | 20 | 18 | 08 | 08 | 14 | 10 | 17 | 15 | 80 | 82 | 76 | 1 | | | | |
| Skinfold2 | -43 | -46 | 12 | -36 | 31 | -26 | 42 | 16 | 15 | 16 | 13 | 09 | -09 | 08 | 06 | 15 | 13 | 74 | 79 | 68 | 84 | 1 | | | |
| Skinfold3 | -38 | -47 | 11 | -33 | -31 | -24 | 41 | 20 | 19 | 18 | 16 | 10 | 07 | 14 | 06 | 18 | 16 | 78 | 86 | 74 | 87 | 84 | 1 | | |
| Skinfold4 | -49 | -43 | 14 | -30 | -29 | -17 | 43 | 22 | 18 | 20 | 16 | 09 | 08 | 17 | 11 | 18 | 17 | 78 | 84 | 74 | 87 | 81 | 90 | 1 | |
| Skinfold5 | -47 | -51 | 13 | -33 | -32 | -24 | -46 | 21 | 19 | 18 | 17 | 14 | -06 | 11 | 05 | 19 | 18 | 81 | 87 | 76 | 87 | 86 | 88 | 90 | 1 |

SEX=BOYS AGE=15

| | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| VO2M | 1 | | | | | | | | | | | | | | | | | | | | | | | | |
| 16K Run | 46 | 1 | | | | | | | | | | | | | | | | | | | | | | | |
| PWC170 | 09 | 16 | 1 | | | | | | | | | | | | | | | | | | | | | | |
| 50M Dash | 18 | 49 | 27 | 1 | | | | | | | | | | | | | | | | | | | | | |
| Long Jump | 13 | 24 | 27 | 59 | 1 | | | | | | | | | | | | | | | | | | | | |
| Situp | 31 | 30 | 11 | 20 | 22 | 1 | | | | | | | | | | | | | | | | | | | |
| Pushup | 05 | 24 | 19 | 40 | 39 | 13 | 1 | | | | | | | | | | | | | | | | | | |
| Rt Grip | -15 | 06 | 51 | 44 | 40 | 04 | 30 | 1 | | | | | | | | | | | | | | | | | |
| Lft Grip | -13 | 04 | 50 | 41 | 38 | 05 | 26 | 87 | 1 | | | | | | | | | | | | | | | | |
| Shld Pull | -19 | 12 | 44 | 41 | 40 | 06 | 23 | 63 | 61 | 1 | | | | | | | | | | | | | | | |
| Shld Push | -05 | 07 | 49 | 34 | 37 | 08 | 16 | 57 | 58 | 66 | 1 | | | | | | | | | | | | | | |
| Leg | -03 | 10 | 45 | 36 | 38 | 13 | 26 | 64 | 67 | 55 | 52 | 1 | | | | | | | | | | | | | |
| Sit/reach | 07 | 08 | 17 | 21 | 33 | 10 | 22 | 31 | 31 | 28 | 28 | 23 | 1 | | | | | | | | | | | | |
| Systolic | -20 | -06 | 11 | 12 | 10 | 01 | 06 | 27 | 24 | 20 | 20 | 20 | 20 | 1 | | | | | | | | | | | |
| Diastolic | 05 | -05 | 09 | -01 | -03 | -02 | 00 | 18 | 15 | 11 | 11 | 14 | 09 | 43 | 1 | | | | | | | | | | |
| FVC | -12 | 11 | 57 | 35 | 38 | 04 | 12 | 62 | 61 | 58 | 59 | 56 | 28 | 27 | 15 | 1 | | | | | | | | | |
| FEV1 | -09 | 10 | 53 | 37 | 36 | 04 | 14 | 58 | 56 | 55 | 58 | 55 | 26 | 25 | 18 | 90 | 1 | | | | | | | | |
| Arm Girth | -25 | -17 | 32 | 14 | 15 | -13 | 03 | 57 | 56 | 52 | 45 | 47 | 18 | 30 | 10 | 46 | 42 | 1 | | | | | | | |
| Waist Grth | -10 | -21 | 31 | 01 | 00 | -17 | -23 | 40 | 43 | 37 | 36 | 36 | 04 | 27 | 11 | 46 | 39 | 78 | 1 | | | | | | |
| Hip Girth | -27 | -19 | 35 | 08 | 06 | -18 | -15 | 48 | 49 | 45 | 41 | 38 | 11 | 31 | 14 | 47 | 44 | 76 | 81 | 1 | | | | | |
| Skinfold1 | -22 | -41 | -08 | -34 | -32 | -23 | -34 | 01 | 02 | 05 | 02 | 01 | -09 | 19 | 06 | -01 | -03 | 55 | 63 | 56 | 1 | | | | |
| Skinfold2 | -07 | -43 | -14 | -40 | -31 | -22 | -36 | -07 | -06 | -02 | -04 | -07 | -10 | 13 | 07 | -11 | -13 | 43 | 49 | 47 | 81 | 1 | | | |
| Skinfold3 | -15 | -37 | 02 | -24 | -24 | -24 | -33 | 14 | 15 | 12 | 07 | 08 | -01 | 22 | 11 | 08 | 05 | 60 | 73 | 66 | 76 | 75 | 1 | | |
| Skinfold4 | -09 | -30 | -02 | -21 | -23 | -19 | -32 | 09 | 10 | 10 | 06 | 05 | -04 | 25 | 14 | 10 | 06 | 53 | 68 | 57 | 78 | 65 | 79 | 1 | |
| Skinfold5 | -16 | -36 | -01 | -24 | -25 | -29 | -37 | 11 | 10 | 10 | 07 | 07 | -06 | 24 | 07 | 12 | 08 | 59 | 74 | 63 | 84 | 70 | 83 | 83 | 1 |

Note. All correlations are presented without deciعal points. See Table 1 for a description of the variables. The actual analyses were conducted on covariance matrices that can be constructed (by LiSREL) from the correlations presented here and the standard deviations presented at the start of this appendix.

Appendix 1C

Descriptive Statistics for the 25 Physical Fitness Variables: Correlations For Three Groups of Girls

SEX=GIRLS AGE=9

| | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| VO2M | 1 | | | | | | | | | | | | | | | | | | | | | | | | |
| 16K Run | 66 | 1 | | | | | | | | | | | | | | | | | | | | | | | |
| PWC170 | -01 | 15 | 1 | | | | | | | | | | | | | | | | | | | | | | |
| 50M Dash | 47 | 46 | 18 | 1 | | | | | | | | | | | | | | | | | | | | | |
| Long Jump | 54 | 35 | 16 | 55 | | | | | | | | | | | | | | | | | | | | | |
| Situp | -02 | 22 | -01 | 37 | . | 1 | | | | | | | | | | | | | | | | | | | |
| Pushup | 20 | 25 | -08 | 25 | 25 | 17 | 1 | | | | | | | | | | | | | | | | | | |
| Rt Grip | 09 | 08 | 21 | 26 | 22 | 17 | 03 | 1 | | | | | | | | | | | | | | | | | |
| Lft Grip | 08 | 09 | 23 | 25 | 20 | 18 | 03 | 76 | 1 | | | | | | | | | | | | | | | | |
| Shld Pull | -01 | 16 | 29 | 31 | 30 | 19 | 14 | 48 | 47 | 1 | | | | | | | | | | | | | | | |
| Shld Push | 07 | 13 | 21 | 28 | 30 | 24 | 11 | 49 | 44 | 58 | 1 | | | | | | | | | | | | | | |
| Leg | -09 | 08 | 16 | 22 | 27 | 19 | -01 | 43 | 46 | 32 | 31 | 1 | | | | | | | | | | | | | |
| Sit/reach | -05 | 03 | -02 | 09 | 17 | 16 | 17 | 11 | 13 | 15 | 15 | 11 | 1 | | | | | | | | | | | | |
| Systolic | -09 | 01 | -06 | 17 | 06 | 05 | 00 | 25 | 24 | 15 | 18 | 11 | 09 | 1 | | | | | | | | | | | |
| Diastolic | 08 | -09 | 02 | 05 | -02 | -01 | -02 | 21 | 23 | 07 | 11 | 14 | 03 | 57 | 1 | | | | | | | | | | |
| FVC | -02 | 06 | 23 | 17 | 15 | 23 | -05 | 50 | 44 | 41 | 40 | 37 | 07 | 20 | 16 | 1 | | | | | | | | | |
| FEV1 | -02 | 09 | 18 | 19 | 18 | 21 | -01 | 47 | 41 | 36 | 43 | 37 | 04 | 21 | 16 | 90 | 1 | | | | | | | | |
| Arm Girth | -26 | -21 | 26 | -10 | -13 | -10 | -31 | 38 | 37 | 35 | 28 | 23 | 00 | 27 | 22 | 38 | 34 | 1 | | | | | | | |
| Waist Grth | -35 | -22 | 18 | -16 | -14 | -13 | -43 | 27 | 23 | 20 | 19 | 18 | -06 | 22 | 19 | 28 | 25 | 77 | 1 | | | | | | |
| Hip Girth | -38 | -28 | 26 | -13 | -13 | -07 | -36 | 39 | 37 | 28 | 28 | 24 | -03 | 29 | 28 | 42 | 39 | 82 | 79 | 1 | | | | | |
| Skinfold1 | -35 | -36 | 10 | -34 | -34 | -20 | -40 | 18 | 18 | 10 | 12 | 09 | -08 | 17 | 21 | 21 | 18 | 78 | 66 | 72 | 1 | | | | |
| Skinfold2 | -28 | -33 | 11 | -32 | -30 | -23 | -39 | 13 | 15 | 08 | 08 | 09 | -07 | 15 | 18 | 18 | 17 | 74 | 63 | 64 | 83 | 1 | | | |
| Skinfold3 | -33 | -35 | 09 | -32 | -29 | -27 | -41 | 13 | 11 | 05 | 09 | 05 | -03 | 22 | 21 | 19 | 18 | 74 | 75 | 72 | 79 | 75 | 1 | | |
| Skinfold4 | -26 | -32 | 09 | -30 | -30 | -22 | -39 | 15 | 13 | 12 | 15 | 08 | -03 | 24 | 28 | 20 | 16 | 74 | 71 | 70 | 81 | 74 | 85 | 1 | |
| Skinfold5 | -31 | -33 | 10 | -32 | -27 | -26 | -44 | 15 | 16 | 10 | 10 | 09 | -06 | 17 | 19 | 19 | 18 | 75 | 76 | 72 | 79 | 78 | 87 | 84 | 1 |

SEX=GIRLS AGE=12

| | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| VO2M | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | | | | | | | | | | | | | | | | | | | | | | | | | |
| 16K Run | 32 | 1 | | | | | | | | | | | | | | | | | | | | | | | |
| PWC170 | 04 | 10 | 1 | | | | | | | | | | | | | | | | | | | | | | |
| 50M Dash | 12 | 39 | 13 | 1 | | | | | | | | | | | | | | | | | | | | | |
| Long Jump | 14 | 29 | 21 | 48 | 1 | | | | | | | | | | | | | | | | | | | | |
| Situp | 51 | 25 | 05 | 24 | 22 | 1 | | | | | | | | | | | | | | | | | | | |
| Pushup | 37 | 27 | -03 | 27 | 31 | 25 | 1 | | | | | | | | | | | | | | | | | | |
| Rt Grip | 03 | -05 | 41 | 34 | 27 | 14 | 05 | 1 | | | | | | | | | | | | | | | | | |
| Lft Grip | 07 | -03 | 41 | 34 | 25 | 14 | 00 | 83 | 1 | | | | | | | | | | | | | | | | |
| Shld Pull | 00 | 10 | 43 | 29 | 22 | 12 | 03 | 60 | 61 | 1 | | | | | | | | | | | | | | | |
| Shld Push | 10 | 03 | 40 | 23 | 22 | 15 | 01 | 51 | 50 | 59 | 1 | | | | | | | | | | | | | | |
| Leg | 04 | 02 | 40 | 24 | 25 | 15 | 03 | 46 | 45 | 47 | 38 | 1 | | | | | | | | | | | | | |
| Sit/reach | 13 | 02 | 05 | 10 | 24 | 15 | 18 | 17 | 14 | 19 | 17 | 22 | 1 | | | | | | | | | | | | |
| Systolic | -15 | -03 | 02 | 14 | 01 | 11 | -06 | 30 | 29 | 20 | 21 | 13 | 11 | 1 | | | | | | | | | | | |
| Diastolic | -10 | -07 | 08 | 11 | 10 | 14 | -05 | 25 | 23 | 11 | 15 | 10 | 14 | 57 | 1 | | | | | | | | | | |
| FVC | -14 | -07 | 43 | 15 | 15 | 02 | -14 | 58 | 53 | 53 | 53 | 44 | 16 | 20 | 20 | 1 | | | | | | | | | |
| FEV1 | -06 | -02 | 37 | 17 | 18 | 03 | -05 | 54 | 48 | 50 | 51 | 40 | 13 | 17 | 16 | 90 | 1 | | | | | | | | |
| Arm Girth | -37 | -39 | 28 | -12 | -18 | -17 | -28 | 40 | 42 | 40 | 30 | 24 | 00 | 23 | 14 | 43 | 38 | 1 | | | | | | | |
| Waist Grth | -28 | -31 | 25 | -18 | -23 | -21 | -40 | 33 | 35 | 32 | 26 | 11 | -08 | 22 | 12 | 39 | 32 | 79 | 1 | | | | | | |
| Hip Girth | -20 | -36 | 39 | -07 | -11 | -16 | -32 | 51 | 52 | 44 | 44 | 30 | 05 | 30 | 21 | 59 | 53 | 81 | 80 | 1 | | | | | |
| Skinfold1 | -36 | -44 | 11 | -29 | -27 | -24 | -33 | 20 | 20 | 18 | 14 | 10 | -07 | 15 | 11 | 21 | 16 | 78 | 68 | 65 | 1 | | | | |
| Skinfold2 | -38 | -35 | 11 | 33 | -27 | -31 | -35 | 07 | 07 | 10 | 03 | 04 | -18 | 03 | -01 | 13 | 11 | 70 | 64 | 54 | 78 | 1 | | | |
| Skinfold3 | -15 | -44 | 09 | -31 | -28 | -25 | -31 | 17 | 18 | 13 | 09 | 06 | -13 | 15 | 11 | 16 | 10 | 72 | 74 | 64 | 80 | 78 | 1 | | |
| Skinfold4 | -21 | -35 | 12 | -26 | -29 | -25 | -37 | 20 | 20 | 22 | 15 | 06 | -17 | 18 | 11 | 19 | 14 | 75 | 75 | 64 | 82 | 77 | 84 | 1 | |
| Skinfold5 | -24 | -41 | 10 | -30 | -31 | -36 | -39 | 12 | 15 | 14 | 06 | 01 | -17 | 14 | 06 | 16 | 13 | 72 | 75 | 63 | 78 | 78 | 83 | 86 | 1 |

SEX-GIRLS AGE=15

| | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| VO2M | 1 | | | | | | | | | | | | | | | | | | | | | | | | |
| 16K Run | 48 | 1 | | | | | | | | | | | | | | | | | | | | | | | |
| PWC170 | 19 | 26 | 1 | | | | | | | | | | | | | | | | | | | | | | |
| 50M Dash | 06 | 51 | 23 | 1 | | | | | | | | | | | | | | | | | | | | | |
| Long Jump | 18 | 38 | 25 | 57 | 1 | | | | | | | | | | | | | | | | | | | | |
| Situp | 28 | 25 | 24 | 32 | 33 | 1 | | | | | | | | | | | | | | | | | | | |
| Pushup | -13 | 17 | 09 | 29 | 27 | 19 | 1 | | | | | | | | | | | | | | | | | | |
| Rt Grip | -04 | 07 | 31 | 26 | 31 | 23 | 06 | 1 | | | | | | | | | | | | | | | | | |
| Lft Grip | 04 | 08 | 34 | 23 | 27 | 16 | 07 | 75 | 1 | | | | | | | | | | | | | | | | |
| Shld Pull | -11 | 07 | 32 | 26 | 29 | 19 | 17 | 51 | 49 | 1 | | | | | | | | | | | | | | | |
| Shld Push | 00 | 07 | 31 | 22 | 29 | 26 | 18 | 48 | 45 | 53 | 1 | | | | | | | | | | | | | | |
| Leg | 11 | 15 | 35 | 23 | 34 | 26 | 12 | 41 | 41 | 43 | 43 | 1 | | | | | | | | | | | | | |
| Sit/reach | 15 | 09 | 17 | 09 | 23 | 17 | 01 | 15 | 14 | 23 | 30 | 20 | 1 | | | | | | | | | | | | |
| Systolic | -24 | -09 | -07 | 04 | -02 | -04 | -01 | 16 | 14 | 19 | 09 | -02 | 02 | 1 | | | | | | | | | | | |
| Diastolic | -21 | -07 | 03 | -02 | -05 | 07 | -04 | 13 | 10 | 14 | 10 | 01 | 05 | 53 | 1 | | | | | | | | | | |
| FVC | -01 | -02 | 25 | 05 | 17 | 10 | -13 | 39 | 31 | 34 | 39 | 32 | 22 | 16 | 12 | 1 | | | | | | | | | |
| FEV1 | 04 | -04 | 19 | 11 | 14 | 10 | -08 | 34 | 25 | 25 | 34 | 28 | 15 | 11 | 03 | 79 | 1 | | | | | | | | |
| Arm Girth | -27 | -20 | 14 | -13 | -15 | -07 | -16 | 35 | 33 | 36 | 20 | 19 | 08 | 33 | 25 | 28 | 22 | 1 | | | | | | | |
| Waist Grth | -07 | -19 | 07 | -12 | -14 | -12 | -24 | 26 | 26 | 25 | 10 | 11 | -04 | 27 | 21 | 19 | 16 | 70 | 1 | | | | | | |
| Hip Girth | -01 | -17 | 16 | -14 | -12 | -06 | -30 | 33 | 34 | 25 | 15 | 18 | 05 | 32 | 29 | 23 | 21 | 70 | 71 | 1 | | | | | |
| Skinfold1 | -35 | -31 | 05 | -30 | -28 | -14 | -23 | 22 | 23 | 18 | 13 | 16 | 07 | 27 | 20 | 19 | 16 | 75 | 55 | 65 | 1 | | | | |
| Skinfold2 | -19 | -23 | 00 | -27 | -29 | -21 | -28 | 09 | 11 | 07 | -02 | 02 | 01 | 28 | 13 | 10 | 08 | 65 | 44 | 55 | 72 | 1 | | | |
| Skinfold3 | -20 | -25 | 03 | -28 | -26 | -18 | -25 | 17 | 17 | 13 | 04 | 09 | 04 | 26 | 20 | 11 | 09 | 63 | 53 | 62 | 77 | 71 | 1 | | |
| Skinfold4 | -19 | -19 | 04 | -21 | -23 | -17 | -32 | 23 | 20 | 13 | 03 | 03 | 06 | 16 | 17 | 08 | 04 | 66 | 55 | 64 | 73 | 68 | 74 | 1 | |
| Skinfold5 | -28 | -19 | -02 | -24 | -19 | -24 | -30 | 11 | 11 | 05 | -03 | 06 | -07 | 16 | 08 | 06 | 05 | 56 | 53 | 56 | 64 | 66 | 68 | 74 | 1 |