Fitness Testing in two High School Physical Education Programs

A Thesis Submitted to the Faculty
of Graduate Studies and Research
in Partial Fulfillment of the Requirements
for the Degree of Master of Arts (Education)

by

Satomi Obara

Department of Physical Education
Division of Graduate Studies and Research
Faculty of Education
McGill University
Montreal, Quebec, Canada

July 1995

ISBN 0-612-12068-6

Abstract

The purpose of this study was to compare health-related physical fitness of high school students in two schools that have different physical education programs. Argyle Academy has physical education 6 classes out of 7 days, and Riverside Park Academy has 3 classes out of 7 days. These two high schools are similar (e.g., size, geographical location, common feeder schools, and social economic status) and affiliated with the same school board. Students in the seventh, eighth, and tenth grade physical education classes were selected for testing in each school. Data from 102 students from Argyle Academy and 96 students from Riverside Park Academy were collected during an eight week period. Health-related physical fitness was assessed using the following test battery: sum of five skinfolds, waist to hip ratio, 20 m shuttle run, curl-ups, 90 degree push-ups, back saver sit and reach, and trunk lift. School group, gender, and grade comparisons were performed using a MANOVA procedure. The students from Argyle Academy demonstrated significantly better scores for the 20 m shuttle run and trunk lift, while those from Riverside Park Academy demonstrated significantly better scores for the curl-ups and 90 degree push-ups. There was no significant difference in the sum of five skinfolds, waist to hip ratio, and back saver sit and reach. This preliminary study has described and analyzed health-related physical fitness of students at two high schools. Recommendations for determining fitness levels of high school students are provided.

Résumé

Le but de cette étude était de comparer le niveau de condition physique d'étudiants au niveau secondaire de deux différentes écoles ayant différents programmes d'éducation physique. La première école (Argyle Academy) a de l'éducation physique 6 jours d'un horaire de 7 jours et la deuxième école (Riverside Park Academy) pendant 3 jours d'un horaire de 7 jours. Les deux écoles sont similaires (grandeur, emplacement, status économique et social) et sont affiliées à la même commission scolaire. Les étudiants de la septième, huitième et dixième année de chaque école ont été sélectionné pour partciper à cette étude. Des données de 102 étudiants de Argyle et de 96 étudiants de Riverside Park ont été receuilli pendant une période de 8 semaines. Les variables mesurées comprenaient: la somme de 5 plis adipeux, le rapport de l'abdomen aux hanches, la puissance aérobie, la flexion du tronc, les redressements, les extensions des bras ainsi que l'élévation du tronc. Une analyze de variance multivariée a été complété pour comparer les deux écoles, les différents niveaux scolaires et les genres. Les étudiants de l'école Argyle ont demontré une supériorité au niveau de la puissance aérobie et de l'élévation du tronc; tandis que les étudiants de l'école Riverside Park ont demontré une supériorité pour les redressements et les extensions des bras. Aucune différence significative n'a été démontré pour la somme des plis adipeux, le rapport de l'abdomen aux hanches et la flexion du tronc. Cette étude préliminaire a décrit et analysé la condition physique d'étudiants de deux écoles secondaires. Des recommendations sont fournies afin de déterminer le niveau de condition physique d'étudiants du secondaire.

Acknowledgements

I would like to thank the students from Argyle Academy and Riverside Park Academy for their participation in this study. I wish they all continue to enjoy physical activity.

I would like to acknowledge the cooperation of Argyle Academy and Riverside Park Academy. Physical education teachers in both schools were particularly appreciated.

I am grateful to Adrienne, Melanie, Sammy, and Todd for their assistance in data collection. Their help enabled me to complete the data collection.

I thank my peers, M.C. and George, for their help in the pilot study.

I also thank Martin for providing equipments for fitness testing.

I wish to thank Dr. Rene Turcotte for his supervision in fitness test selection and statistical analyses and translation of the abstract in French.

Dr. Ben Dyson, my supervisor, offered me the opportunity to conduct this study. He has been generous, supportive, and understanding. I would like to express my appreciation to him for his guidance, support, encouragement, and his time and energy spent in supervision of my thesis as well as the data collection.

I also acknowledge the financial support of the Canadian Government, which provided me the opportunity to study in Canada.

I would never forget all of my friends who have been supportive and made me keep smiling.

Special thanks to my parents and family in Japan for their continuous support and encouragement.

Table of Contents

Page

List of Tables

List of Figures

Chapter 1

Introduction

Current research suggests that physical activity has positive effects on child and youth development. There has been much attention focused on the physical fitness levels of youth in North America. During adolescence, regular physical activity can produce marked physical changes (Corbin, 1987a), and knowledge about physical activity and its benefits can be meaningful (Goldfine & Nahas, 1993). Leading authorities such as Canadian Association for Health, Physical Education, Recreation, and Dance (CAHPERD) have recommended quality daily physical education (QDPE). In spite of the evidence for support of QDPE, many high schools still offer physical education only a few days a week or even reduce it to the elective category. Physical education is often viewed as a competitor with other curriculum subjects for time and resources (Dahlgram, 1987; Hansen & McKenzie, 1988; Verabioff, 1986).

There are a limited number of studies which investigated the effects of QDPE on high school students. These studies have not been successful in establishing physical education's credibility as an essential subject in school curriculum. In order to survive and grow, physical education needs to demonstrate its uniqueness and the essential benefits it brings to the school curriculum (Verabioff, 1986; O'Sullivan, Siedentop, & Tannehill, 1994). There is still a need for more quantitative and qualitative data to support the need for QDPE (Dahlgram, 1987).

1.1 Nature and Scope of the Problem

There have been numerous methods of defining physical fitness. Due to the ambiguity associated with definitions of physical fitness, many historical test batteries have assessed physical fitness partially using different tests. However, the current body of knowledge in exercise science and society's view of physical fitness suggest that a definition of physical fitness should focus on the health-related aspects of fitness (Blair, Falls, & Pate, 1983; Caspersen, Powell, & Christensen, 1985; Corbin, 1991; Gutin, Manos, & Strong, 1992; Pate, 1983, 1988), and most of the current youth fitness batteries are based on the testing of health-related physical fitness (Corbin, 1991). Health-related physical fitness includes cardiorespiratory endurance, muscular strength and endurance, flexibility, and body composition. Other components such as agility, power, speed, and balance are categorized as skill-related physical fitness and are not essential in day-to-day life (Blair et al., 1983; Caspersen et al., 1985; Corbin, 1991; Gutin et al., 1992; Pate, 1983, 1988) .

Current health-related physical fitness test batteries (e.g., AAHPERD Physical Best [American Alliance for Health, Physical Education, Recreation, & Dance (AAHPERD), 1988]; FITNESSGRAM [Cooper Institute for Aerobic Research, 1992]; Fit Youth Today [American Health and Fitness Foundation, 1986]) are composed of various test items. The individual test items differ in the reliability and validity of the test. Reliability and validity of individual tests are crucial issues in measurement and evaluation (Baumgartner & Jackson, 1991) and cannot be generalized across age groups and genders (Safrit, 1990). In general, very few studies have been conducted on the reliability and validity using children and youth in schools as subjects compared to those studies using adults.

Individual test items also differ in availability of standards. Criterion-referenced standards have been developed for some tests as an alternative to the traditional norm-referenced standards. However, these standards published for the recent version of fitness tests should be viewed as arbitrary, since criterion-referenced standards based on sound research are only available for selected fitness components (Blair, 1992; Corbin & Pangrazi, 1992). Consequently, it has been suggested that a properly developed test manual should include both criterion-referenced standards and tables of norms (Safrit, 1990). In selection of test items, these issues should be carefully examined in order to select the best test items available.

Assessment of physical fitness of youth is problematic. It can lead children and youth to discouragement, lack of peer approval, loss of self-esteem, and lack of motivation to be physically active (Corbin, 1986, 1987a). Lack of motivation would also prevent children and youth to exert their maximal efforts in fitness testing and influence the reliability of tests (Safrit, 1990). These aspects should be considered when evaluating physical fitness of the young population so that they can enjoy physical activity enough to optimize physical fitness levels and develop positive physical activity habits.

A limited number of studies have investigated the effects of QDPE on health-related physical fitness of high school students. It seems that there are several factors considered to be associated with the effects of QDPE on physical fitness of school children and youth: these factors include the teachers, motivation of students, stability of physiological body changes versus onset of puberty (Werner & Durham, 1988), optimal employment of duration, intensity, and frequency of exercise (Li & Dunham, 1993), and conceptual courses (Rider, Imwold, & Johnson, 1986). These factors deserve further inquiry.

1.2 Significance of the Study

A limited number of studies have investigated the effects of QDPE on health-related physical fitness of high school students. The results of these studies have been inconclusive. The earlier studies (Johnson 1969; Kemper, Verschuur, Ras, Snel, Splinter, & Tavecchio, 1976) examined the effects of increasing time in physical education instruction upon fitness and skill developments. These studies reported that significant improvements were seen only in boys in selected fitness test items such as push-ups, pull-ups, skinfold measurement, and other skill-related fitness component, while girls improved only in a skill-related component.

More recent studies were conducted to examine the effects of time as well as content differential of physical education instruction (i.e., QDPE). Dragicevick, Hill, Hopkins, and Walker (1987) found that initial significant improvements at the pilot school in comparison with the control school were seen for females and males in the sit and reach, for males in the curl-ups and shuttle run, and for females and males in the 500 m run. However, poor compliance with the program in the pilot school reduced the school difference in fitness during the second term. As a result, the students of control school were doing even better for females in the sit and reach and standing jump. Both schools showed a substantial loss in aerobic fitness over the summer holidays. There were no significant differences in the anthropometric measurements and the oxygen uptake at a heart rate of 170 beats per minute (VO_2170) on a bicycle ergometer.

Phillipp, Piland, Seidenwurm, and Smith (1991) found no significant effects of daily physical education courses during two consecutive summer sessions on physical fitness of high school students compared to that of control groups. The only statistically significant difference between the experimental and control groups in both years was for the 1.5 mile run, where the experimental group improved more. Participation in the two year experimental program did not verify statistically significant improvements.

Further investigation into the value of QDPE seems warranted. Therefore, this study investigated the effect of QDPE on health-related physical fitness of high school students.

1.3 Statement of The Problem

The purpose of this study was to compare the health-related physical fitness of high school students in grade seven, eight, and ten from two schools that have different physical education programs. Argyle Academy has a school curriculum which emphasizes fitness and physical education and has fifty minute daily physical education classes in the seventh and eighth grades (6 days per 7 day cycle) but not in the tenth grade (3 days per 7 day cycle). In comparison, Riverside Park Academy does not have daily physical education classes in any grade and has physical education classes only 3 days per 7 day cycle. The following hypotheses were examined:

(1) In comparison to students from Riverside Park Academy, students from Argyle Academy will demonstrate significantly better health-related physical fitness.

(2) There will be a significant interaction between school groups (Argyle and Riverside Park Academics) and genders for each health-related physical fitness performance variable.

(3) There will be a significant interaction between school groups and grades (grade 7, 8 and 10) for each health-related physical fitness performance variable.

(4) There will be a significant interaction among genders, grades, and school groups for each health-related physical fitness performance variable.

1.4 Limitations

In this study variation in subjects' motivation was considered to be a source of error. Precautions were taken to secure their maximal and similar motivation by applying the same procedures and instructions to each subject in all the tests. The errors produced from this source are not of magnitude to alter the general conclusions.

1.5 Delimitations

The external validity of the study was limited to high school students whose general characteristics and environmental influences are similar to those schools.

1.6 Definitions

Physical Activity: Any bodily movement produced by skeletal muscles that results in energy expenditure (Caspersen et al., 1985).

Exercise: Physical activity that is planned, structured, repetitive, and purposive in the sense that improvement or maintenance of one or more components of physical fitness is an objective (Caspersen et al., 1985).

Physical Fitness: A set of attributes that people have or achieve, which are either health- or skill-related. The degree to which people have these attributes can be measured with specific tests (Caspersen et al., 1985).

Health-Related Physical Fitness: (1) The ability to perform strenuous physical activity with vigor and without excessive fatigue, and (2) demonstration of physical activity traits and capacities that are consistent with minimal risk of developing hypokinetic diseases. The components of health-related physical fitness are 1) cardiorespiratory endurance, 2) muscular strength and endurance, 3) body composition, and 4) flexibility (Pate, 1983).

Chapter 2

Review of the Literature

The purpose of this study was to compare the health-related physical fitness of high school students from two schools which have different physical education programs in terms of the emphasis on fitness as a component of the curriculum. In order to assess health-related physical fitness of high school students, it was necessary to develop a test battery which was able to discriminate differences in students' physical fitness as well as for the future use of the test battery in schools. Currently available test items were examined so that the most appropriate tests in this study for each fitness component were selected, with safety, reliability, validity, applicability (i.e., procedure, instrumentation), and availability of standards of the test being the crucial selection factors. This review will be presented in two sections, (2.1) Health-Related Physical Fitness Test, and (2.2) Physical Fitness, Physical Education, and QDPE in High Schools.

2.1 Health-Related Physical Fitness Test

Research results have been reported concerning the selection of field tests in order to measure health-related physical fitness of children and adolescents. The research findings reviewed relative to this study were categorized and discussed according to the component of health-related physical fitness as follows:

1) Aerobic Capacity
2) Body Composition
3) Muscular Strength and Endurance
4) Muscular Flexibility.

2.1.1 Aerobic Capacity

In children and adolescents aerobic capacity can be obtained by measuring the maximal oxygen uptake (VO_2max), which is often expressed relative to body weight (ml/kg/min) (Astrand, 1976). VO_2max is usually predicted from performance in field tests in school settings instead of being measured directly in the laboratory.

Distance run tests are commonly used in children and adolescents. The use of these tests is based on the fact that for exhaustive exercise lasting longer than 2 minutes, energy is provided primarily through aerobic metabolism (Astrand & Rodahl, 1986) and that distance running performance is correlated with maximum aerobic power (Safrit, Hooper, Ehlert, Costa, & Patterson, 1988). Safrit (1990) summarized the reliability of distance run tests (600 yd, 1600 m, 9 min, and 12 min runs) in youth and concluded that the reliability of these tests varied (r=0.61 to 0.94). In general the reliability is high but mixed (Table 1). Lower reliability on distance runs in children may be due to variation in motivation, running skills, and pacing strategy (Cureton, 1982; McCormack, Cureton, Bullock, & Weyand, 1991; Saltarelli & Andres, 1993).

Table 1

Reliability of Aerobic Capacity Tests

Study	Test	Subject	Reliability r:interclass	R:intraclass
Buono, Roby, Micale, Sallis, & Shephard (1991)	1 mile run	15M & 15F (5th grade)	r=.91	
		15M & 15F (8th grade)	r=.93	
		15M & 15F (11th grade)	r=.98	
		45M & 45F (5, 8 & 11th grade)	r=.95	
Safrit & Wood (1987)	9 min run	27M (11 yr)	r=.54,	R=.69
		88M (12 yr)	r=.58,	R=.72
		104M (13 yr)	r=.43,	R=.61
		58M (14 yr)	r=.71,	R=.83
		44F (11 yr)	r=.82,	R=.90
		92F (12 yr)	r=.72,	R=.84
		85F (13 yr)	r=.58,	R=.73
		43F (14 yr)	r=.68,	R=.80
Doolittle & Bigbee (1968)	12 min run/walk	153M (9th grade)	r=.94	
Leger, Mercier, Gadoury, & Lambert (1988)	20 m shuttle run	139M & F (6-16 yr)	r=.89	
Liu, Plowman, & Looney (1992)	20 m shuttle run	20M & F (12-15 yr)	r=.93	
		12M (12-15 yr)	r=.91	
		8F (12-15 yr)	r=.87	
Mahoney (1992)	20 m shuttle run	12M (12 yr)	r=.73	
		8F (12 yr)	r=.88	

The validity of distance run tests has been evaluated by correlating distance run performance with VO_2max measured directly in the criterion tests (e.g., treadmill, bicycle). For youth, validity coefficients have been reported as moderately high (r=0.62 to 0.90), where runs of 600 yards to 2000 meters or 9 to 15 minutes and VO_2max measured in the criterion tests were correlated (Table 2).

Table 2

Validity of Aerobic Capacity Tests

Study	Test	Criterion Test	Subject	Validity (r)
Cureton, Boileau, Lohman, & Misner (1977)	600 yd run	treadmill	140M & 56F (7-12 yr)	-.62
Doolittle & Bigbee (1968)	600 yd run/walk	bicycle ergometer	9M (9th grade)	.62
Palgi & Gutin (1984)	600 m run		58M & F (10-14 yr)	-.73
	1000 m run		58M & F (10-14 yr)	-.75
Buono, Roby, Micale, Sallis, & Shephard (1991)	1 mile run	treadmill	15M & 15F (5th grade)	-.76
			15M & 15F (8th grade)	-.80
			15M & 15F (11th grade)	-.85
			45M & 45F (5,8,&11th grade)	-.73
Cureton, Boileau, Lohman, & Misner (1977)	1 mile run	treadmill	140M & 56F (7-12 yr)	-.66
Palgi & Gutin (1984)	2000 m run		58M & F (10-14 yr)	-.73
Van Mechelen, Hlobil, & Kemper (1986)	6 min run	treadmill	41M (12-14 yr)	.51
			41F (12-14 yr)	.45
			41M & 41F (12-14 yr)	.63
Jackson & Coleman (1976)	9 min run	treadmill	22M (4-6th grade)	.82
			25F (4-6th grade)	.71
Doolittle & Bigbee (1968)	12 min run	bicycle ergometer	9M (9th grade)	.90
Jackson & Coleman (1976)	12 min run	treadmill	22M (4-6th grade)	.82
			25F (4-6th grade)	.71
MacNaughton, Croft, Pennicott, & Long (1990)	15 min run	treadmill	42M & F (12 yr)	.88
			28M & F (13 yr)	.85
			44M & F (14 yr)	.67
			28M & F (15 yr)	.88
Armstrong, Williams, & Ringham (1988)	20 m shuttle run	treadmill	77M (11-14 yr)	.54
Boreham, Paliczka, & Nichls (1990)	20 m shuttle run	treadmill	23M (15.6 yr)	.64
			18F (15.4 yr)	.90
			23M & 18F (15.5 yr)	.87
Leger, Mercier, Gadoury, & Lambert (1988)	20 m shuttle run	retroextrapolating the O_2 recovery curve at time zero of recovery (Leger, Seliger, & Brassard, 1980)	188M & F (8-19 yr)	.71
Liu, Plowman, & Looney (1992)	20 m shuttle run	treadmill	22M (12-15 yr)	.65
			26F (12-15 yr)	.51
			22M & 26F (12-15 yr)	.69
Mahoney (1992)	20 m shuttle run	treadmill	10M (12 yr)	.83
			10F (12 yr)	.76
Van Mechelen, Hlobil, & Kemper (1986)	20 m shuttle run	treadmill	41M (12-14 yr)	.68
			41F (12-14 yr)	.69
			41M & 41F (12-14 yr)	.76

Disch, Frankiewicz, and Jackson (1975) investigated the construct validity of 10 distance run tests which ranged in distance from a 50 yd dash to the 2 mile run and 12 min run in college male students. They found that distance runs longer than 1 mile measured aerobic capacity, whereas the shorter distance runs of 1 mile or less tended to yield complex factor structures; they measured both speed (anaerobic capacity) and aerobic capacity. Correlations between distance runs of different distances and VO_2max (Baumgartner & Jackson, 1991; Jackson & Coleman, 1976; Krahenbuhl, Pangrazi, Burkett, Schneider, & Petersen, 1977; Safrit et al., 1988) suggest that if VO_2max is the primary determinant of distance runs, runs of 1 mile or longer should be used to assess VO_2max.

There are variables which have been found to affect distance running performance in youth and reduce the validity of distance run tests. The variables include environment, appropriate pacing, motivation, body fatness, running skills and economy, genetically determined biological characteristics such as muscle fiber types, and physiological variables affected by training such as skeletal muscle oxidative capacity and lactate threshold (Cureton, 1982; Cureton, Baumgartner, & McManis, 1991; Cureton, Boileau, Lohman, & Misner, 1977; Krahenbuhl, Morgan, & Pangrazi, 1989; McCormack et al., 1991).

Regarding such disadvantages in distance run tests, the 20 m shuttle run test was developed by Leger and Lambert (1982). This test is a progressive multistage maximal exercise test that closely simulates a graded, speed-incremented treadmill test used in the laboratory to directly measure VO_2max. Because the speed of running is controlled, variation in pacing has little influence on test outcome. Because a maximal effort is required only at the end of the test, motivation is probably less of a problem than with distance run tests in which a sustained near-maximal intensity is required (Leger & Lambert, 1982). It can be administered in a relatively small space either indoors or outdoors.

There are a few studies which have investigated the reliability of the 20 m shuttle run test and reported that the reliability of this test in youth was high ($r=0.73$ to 0.93) (Table 1). The range of validity coefficients is similar to that found using other distance run tests ($r=0.51$ to 0.90) (Table 2), indicating that the 20 m shuttle run test has moderate validity as a field test of VO_2max. In one study in which the 20 m shuttle run and a 6 min run were correlated with actual VO_2max in a same sample of 82 pupils of a high school, VO_2max was more highly correlated with the 20 m shuttle run than with the distance run (r=0.76 vs 0.63) (van Mechelen, Hlobil, & Kemper, 1986).

It has been shown that age affects the prediction of VO_2max in the 20 m shuttle run test (Leger, Mercier, Gadoury, & Lambert, 1988). The effect of gender is uncertain (Cureton, 1994). The age by gender interaction effect has not been investigated, although adjustment for this effect would likely improve the prediction because the gender difference in VO_2max changes with age (Krahenbuhl, Skinner, & Kohrt, 1985).

2.1.2 Body Composition

Direct assessment of body composition is impossible, making it necessary to estimate from indirect techniques. Although a number of indirect techniques such as underwater weighing have been developed to measure body composition, most are not practical for use in the school setting where large numbers of individuals must be measured in a short time period. A skinfold thickness measurement is commonly used to estimate body composition in school or other field settings. Compared to other techniques, it is an inexpensive and practical approach that can be applied in these settings and it has high reliability (r=0.89 to 0.98, R=0.94 to 0.99) (Table 3) and acceptable validity (r=-0.56 to -0.87) (Table 4) in youth.

Intra-examiner error (errors associated with the failure of the same investigator to obtain identical results upon repeated skinfold measurements) has been estimated to vary depending on the site, experience of the investigator, fatness of subject, and method of error estimation (Lohman, 1981). In general, reliability coefficients exceed 0.90 (Lohman, 1994). Experienced investigators have demonstrated repeatedly excellent test-retest reliability (r=0.90 to 0.96) (Shaw, 1986). Inter-examiner reliability has also been found to be good when investigators trained together (Morrow, Fridye, & Monaghen, 1986; Opplinger, Clark, & Kuta, 1992). Morrow et al. (1986) found only 4% of the total variation was associated with testers for skinfold measures when comparing three calipers, three testers, and three trials. Opplinger et al. (1992) also found similar results in training through clinics and concluded that testers could be adequately trained to provide accurate skinfold measurements. Standardization of training protocol is also essential (Lohman, Pollock, Slaughter, Brandon, & Boileau, 1984; Morrow et al., 1986; Opplinger et al., 1992). Shaw (1986) recommends the use of videotapes as a method of training in order to learn to take skinfold measurements properly. In her study, novice investigators who were trained with a video possessed higher reliability coefficients for both intra- and inter-examiner reliability than those trained with a regular reading manual. Sloan and Shapiro (1972) compared three standard calipers (i.e., the MNL, the Harpenden, and the Lange) in order to ascertain which gives the closest agreement between readings at four selected sites (i.e., thigh, abdomen, scapula, and arm) by two trained observers. Their results showed that the best agreement between observers was obtained with the Harpenden caliper and that the mean of the readings by the two observers at each site was not significantly different with any one caliper than with either of the others. Thus, calipers would influence the reliability of skinfold measurements. Lohman et al. (1984) and Morrow et al. (1986) also found calipers to be associated with 7.1 to 18.5% of the total variation depending on the site in skinfold measures.

Table 3

Reliability of Body Composition Tests

Study	Test	Subject	Reliability r:interclass	R:intraclass
DiNucci, McCune, & Shows (1990)	Skinfold sum of thigh, abdominal (male), suprailiac (female), pectoral (male), & triceps (female)	143M (college)	r=.98,	R=.99
		57F (college)	r=.93,	R=.96
		143M & 57F (college)	r=.98,	R=.99
Safrit & Wood (1987)	skinfold sum of triceps & subscapular	27M (11 yr)	r=.94,	R=.96
		88M (12 yr)	r=.97,	R=.98
		104M (13 yr)	r=.95,	R=.98
		58M (14 yr)	r=.97,	R=.98
		44F (11 yr)	r=.89,	R=.94
		92F (12 yr)	r=.97,	R=.98
		85F (13 yr)	r=.90,	R=.95
		43F (14 yr)	r=.93,	R=.96

The validity of skinfolds has been demonstrated consistently to be moderate to high (Table 4) when correlated between skinfolds and body fat estimated from laboratory techniques such as underwater weighing and potassium spectrometry. If body density alone is used to predict percent fat, the correlation (i.e., validity) with skinfolds is less than when multicomponent (i.e. density, water, and mineral measured in each child) model is used in prepubescent children (Slaughter, Lohman, Boileau, Horswill, Stillman, van Loan, & Bemben, 1988). It has been shown that the use of body density in children leads to an overestimate of body fatness in prepubescent children because of the lower density of the fat-free body.

Table 4

Validity of Body Composition Tests

Study	Test	Criterion Test	Subject	Validity (r)
Boileau, Wilmore, Lohman, Slaughter, & Riner (1981)	skinfold	body density from hydrostatic weighing	97M (8-11 yr)	-.77 Triceps -.64 Subscapular -.60 Midaxillary -.59 Supra-Iliac -.56 Abdomen
			86M (8-11 yr)	-.83 Triceps -.80 Subscapular -.82 Midaxillary -.77 Supra-Iliac -.81 Abdomen
Harsha, Frerichs, & Berenson (1978)	skinfold	body density from hydrostatic weighing	79 white M (6-16 yr) (12.1 +/- 2.0 yr)	-.76 Triceps -.79 Biceps -.75 Subscapular -.76 Supra-Iliac -.74 Subcostal -.82 Femoral -.79 Calf
			64 white F (6-16 yr) (12.5 +/- 1.4 yr)	-.75 Triceps -.77 Biceps -.80 Subscapular -.78 Supra-Iliac -.77 Subcostal -.80 Femoral -.73 Calf
			49 black M (6-16 yr) (13.1 +/- 1.6 yr)	-.82 Triceps -.76 Biceps -.75 Subscapular -.75 Supra-Iliac -.80 Subcostal -.84 Femoral -.77 Calf
			50 black F (6-16 yr) (12.5 +/- 1.8 yr)	-.82 Triceps -.82 Biceps -.87 Subscapular -.86 Supra-Iliac -.83 Subcostal -.79 Femoral -.83 Calf
			Total	-.81 Triceps -.79 Biceps -.76 Subscapular -.78 Supra-Iliac -.77 Subcostal -.84 Femoral -.80 Calf

Another aspect of validity is the establishment of the correct skinfold sites to best represent children. Logically, it would be important to include both the trunk and the extremities to see the pattern of fat distribution. Recent evidence suggests that chronic diseases are linked to both fatness and fat distribution and that a disproportionate development of abdominal fat may put an adolescent at additional risk for disease (Fox, Peters, Armstrong, Sharpe, & Bell, 1993; Williams, Going, Lohman, Harsha, Srinivasan, Webber, & Berenson, 1991, 1992). However, because of poor acceptance of the subscapular site, which is a commonly used site for the trunk, it has been replaced with the calf site (Going, 1988). For children, the combination of two or three extremity skinfolds is adequate for an estimate of percent body fat, however, as children become older (late adolescence), it is important to estimate truncal fat as well (Lohman, 1994). Slaughter et al. (1988) have demonstrated equations to estimate percent body fat of children and youth aged between 8 and 18 years from triceps and calf skinfolds. They are almost as valid as those from the triceps and subscapular skinfolds (r=0.88) and take into consideration the effects of maturation, race, and sex as well as the use of a multicomponent approach to body composition and account for the chemical immaturity of children. They are almost as valid as the sum of nine skinfold sites (Slaughter et al., 1988).

BMI (Body Mass Index) and waist to hip ratio are often used to assess body composition as an alternative method. They are very feasible for general use, since the procedures are straightforward and does not require any special equipment. BMI is designed to describe the proportion of fat in relation to body weight, regardless of the subject's height (Canada Fitness Survey, 1985). However, caution is warranted when comparing individuals to the normative data: subjects with high body weights but who are lean and muscular have relatively high BMI values, whereas thin, non-muscular subjects with low body weights have low BMI values (Canada Fitness Survey, 1985). The major disadvantage of the BMI is the lack of accuracy when compared to other methods such as skinfolds (Baumgartner & Jackson, 1991). Therefore, the combination of the BMI and skinfold measurements can provide a more comprehensive assessment of adiposity (Canada Fitness Survey, 1985). The waist to hip ratio provides a valid representation of the pattern of fat distribution (Canadian Standardized Test of Fitness [C.S.T.F.], 1986). It seems important to include this method for body composition assessment, since it is apparent that the patterning of adipose tissue distribution, independent of total body fat, alters the health risk of obesity (McArdle, Katch, & Katch, 1991). The trunk skinfold measurements assess subcutaneous fat in the trunk region very directly and are recommended to be added to the waist to hip ratio (C.S.T.F., 1986). Therefore, these are not the recommended procedures to be used alone for determining body composition. Consequently, it appears that skinfold measurements provide more valid information on body composition.

2.1.3 Muscular Strength and Endurance

In order to measure muscular strength and endurance in field tests, there are generally two sites used. They include the abdomen, trunk, and upper arm and shoulder girdle.

Abdomen

There has been considerable research done on the use of various versions of the sit-up test. In sit-ups, not only the abdominal but also the hip flexors are active which can have a negative impact on the low back. In electromyographical studies it has been shown that the abdominals are responsible for only the first 0-45^{o} of movement in the sit-ups, with the hip flexors being responsible for the rest (Flint, 1965: Jette, Sidney, & Cicutti, 1984: Ricci, Marchtti, & Figure, 1981). When the feet were held, less activity of the abdominals was evident than when the feet were not secured (Flint, 1965: Walters & Partridge, 1956). Since there is a high range of motion (66-75^{o}) (Robertson & Magnusdottir, 1987), there would be considerable work done by the hip flexors. Additionally, hyperextension of the trunk (lumber "hollowing") was noted in one study before the performer could come to the sitting position (Ricci et al., 1981). Furthermore, one of the serious drawbacks of any full sit-up exercise is the forced maximum flexion of the spine (Halpern & Bleck, 1979). Nachmson and Elfstrom (1970) reported that such a position in both flexed and straight knee full sit-ups actually caused intradiscal pressure according to the x-rays.

Although efforts have been made to develop a revised sit-up test which does not adversely affect the lower back, these problems still remain. The reliability coefficients are better for college students (r=0.83 to 0.85, R=.91) but even those for the younger students are generally acceptable (r=0.57 to 0.94, R=0.77 to 0.94) (Table 5).

Table 5

Reliability of Abdominal Strength and Endurance Tests

Study	Test	Subject	Reliability r:interclass	R:intraclass
Buxton (1957)	sit-up knees flexed feet held total N	53M & F (6-15 yr)	r=.94	
Cureton, Boileau, & Lohman (1975)	sit-up legs straight feet held N to max of 100	49M (8-11 yr)	r=.60	
DiNucci, McCune, & Shows (1990)	sit-up knees flexed feet held 1 min	143M (college) 57F (college) 143M & 57F (college)	r=.83, r=.85, r=.84,	R=.91 R=.91 R=.91
Klesius (1968)	sit-up	132M (10th grade)	r=.57 r=.68	
Safrit & Wood (1987)	sit-up knees flexed feet held 1 min	27M (11 yr) 88M (12 yr) 104M (13 yr) 58M (14 yr) 44F (11 yr) 92F (12 yr) 85F (13 yr) 43F (14 yr)	r=.62, r=.83, r=.79, r=.86, r=.64, r=.85, r=.89, r=.81,	R=.77 R=.91 R=.88 R=.93 R=.78 R=.92 R=.94 R=.89
Hyytiainen, Salminen, Suvitie, Wickstrom, & Pentti (1991)	curl-up knees bent at 90o feet free arms straight to knees 240s max hold	30M (35-44 yr)	r=.93	
Jette, Sidney, & Cicutti (1984)	curl-up knees bent at 140o feet free arms straight to knees cadence (20/min) max 100	43M & F (school children)	r=.88	
Robertson & Magnusdottir (1987)	curl-up knees bent feet free arms straight at the sides touch a flame (7.62 cm away) 1 min	15M (college) 15F (college)	r=.93 r=.97	
Vincent & Britten (1980)	curl-up knees bent feet free hands behind the head partner's fist under the back hold as long as possible	70M & 40F (7-12 yr) 138M & 22F (jr. high) 19M (college)	r=.62 r=.53 r=.71	

Despite the fact that the sit-up test is typically labeled as measuring both abdominal strength and endurance, only two studies have compared sit-up test performance with a criterion endurance test (Ball, 1993; DeWitt, 1944). The validity of various forms of the sit-ups as a test for abdominal strength and endurance ranges from low to moderate (r=0.04 to 0.65) when the sit-up test performance was correlated with the criterion test scores (Table 6). The poor validity and the wide variety of criterion tests that have been used by investigators point out the fact that no absolutely agreed upon criterion measure for abdominal strength and endurance exists, making statistical validation difficult (Plowman & Corbin, 1994).

Table 6

Validity of Abdominal Strength and Endurance Tests

Study	Test	Subject	Validity (r) & Criterion Tests Strength	 Endurance
Ball (1993)	sit-up knees bent feet free arms across the chest 1 min	144M (18-33 yr)	.39 1-RM trunk flexion	.18 60% 1-RM trunk flexion
DeWitt (1944)	sit-up knees bent feet free oblique max N	102M (college)	.04 1-RM abdominal lift dynamometer	.25 isometric sit-up position with feet held
	sit-up knees bent feet held oblique max N	102M (college)	.16	.37
	sit-up knees bent feet held oblique 2 min	102M (college)	.14	.26
Hall, Hetzler, Perrin, & Weltman (1992)	sit-up knees straight feet held hands behind the neck 1 min	23M (23.1 +/- 7.4 yr) 28F (22.2 +/- 4.6 yr)	-.18 (C) -.21 (E) .42 (C) .40 (E)	isokinetic dynamometer peak torque single effort concentric (C) and eccentric (E)
	sit-up knees bent feet held arms across the chest 1 min	23M 28F	-.25 (C) -.28 (E) -.27 (C) -.32 (E)	
Hall, Hetzler, Perrin, & Weltman (1992)	curl-up knees bent feet free touch a flame (7.62 cm away) 1 min	23M (23.1 +/- 7.4 yr) 28F (22.2 +/- 4.6 yr)	-.41 (C) -.38 (E) -.07 (C) -.08 (E)	isokinetic dynamometer peak torque single effort concentric (C) and eccentric (E)
Robertson & Magnusdottir (1987)	curl-up knees bent feet free arms straight on the sides touch a flame (7.62 cm away) 1 min	11M & 8F (college)	.65	sit-up knees bent feet held arms across the chest 1min
Vincent & Britten (1980)	curl-up knees bent feet free hands behind the head partner's fist under the back hold as long as possible	15M & 9F (7-12 yr) 138M & 22F (jr. high) 14M (college)	.30 .27 .39	sit-up knees bent 1min

Curl-ups have been recommended as an alternative to the sit-ups, based on extensive research. Robertson and Magnusdottir (1987) proposed a partial curl-up test which places the greatest demand on the abdominals and none of the hip flexors. Halpern and Bleck (1979) also found in their electromyographic study that the flexed-knee shoulder lift (i.e., curl-ups) provided maximum activity of external oblique and rectus abdominal muscles during the sit-up cycle compared to other four sit-ups. It requires a range of motion of less than 45°, which has been recommended for maximal involvement of the abdominals without hip flexor involvement (Flint, 1965; Ricci et al., 1981). Additionally, disc compression is not a major concern with the curl-ups (Plowman & Corbin, 1994). Halpern and Bleck (1979) found in x-rays that knee-flexed shoulder lift required minimum flexion of the lumber spine compared to other sit-ups.

Of the four studies which investigated the reliability of curl-ups (Hyytiainen, Salminen, Suvitie, Wickstrom, & Pentti, 1991; Jette et al., 1984; Robertson & Magnusdottir, 1987; Vincent & Britten, 1980) (Table 5), only the Robertson and Magnusdottir (1987) data are practical information, as the modified curl-up tests in the other studies have not been developed as extensively as the Robertson and Magnusdottir test. The test was highly reliable for adult men (r=0.93) and women (r=0.97), but the number of subjects was small and the age range was limited (22 to 39 years).

There is no study which has compared curl-up test performance with a criterion endurance test. The only study which specifically tested the curl-ups as an abdominal strength test (Hall, Helzler, Perrin, & Weltman, 1992) used isokinetic measures as criterion measures (Table 6). However, in this study the time was limited to 1 min. They found only weak relationships for the young male adults and almost no relationships for the young female adults. Additionally, all of these relationships were negative indicating that better curl-up test performance was associated with lower strength scores and vice versa. They speculated that the use of an isokinetic measure to validate a dynamic (isotonic) field test might have contributed to the poor results, and the results suggested that the use of the timed curl-up tests was not a valid method of estimating isokinetic abdominal muscular strength.

Two studies compared performances of full sit-ups and curl-ups (Robertson & Magnusdottir, 1987: Vincent & Britten, 1980) in an attempt to demonstrate concurrent validity (Table 6). However, the relationship between these two tests was found to be low, suggesting that the tests cannot be used interchangeably. Thus, the curl-ups are intended to utilize different muscles over a more restricted range of motion than the sit-ups. More validation work is needed for the curl-ups.

Trunk

Of the five anatomical and physiological areas (low back lumber, hamstring, and hip flexor flexibility as well as abdominal and trunk extensor strength and endurance) which have been identified as critical for the development and maintenance of low back function, only trunk extension strength and endurance has been shown to predict both first time and recurrent low back pain (Plowman & Corbin, 1994).

Retrospective studies have shown significant relationships between low back pain and back muscle fatigue associated with a lack of trunk extension strength and endurance, including three in which electromyographic records were able to distinguish between those who did and did not have low back pain (DeVries, 1968: Roy, DeLuca, & Casavant, 1989; Roy, DeLuca, Snyder-Mackler, Emley, Crenshaw, & Lyons, 1990). A prospective study by Biering-Sorensen (1984b) revealed that good isometric endurance of the back extensors measured by the 240 second over the table edge test might prevent first-time occurrence of low back pain in men but not in women. It was also found that in the maximal voluntary contraction at isometric backward extension of the trunk, trunk extensors tended to be weaker among those who experienced recurrence of low back pain compared to those without recurrence in the follow-up year.

The 240 second over the table edge test seems the only one which does not require sophisticated laboratory equipment for the measurement of back extensor strength and endurance. However, this test seems to be not practical for the school setting. This is because it does require a table and straps or person to hold the subject's lower body and is very time consuming. Therefore, the prone trunk extension lift has been suggested as being more practical (Plowman & Corbin, 1994).

Table 7

Reliability and Validity of Trunk Extension Tests

Study	Test	Subject	Reliability r.interclass R:intraclass	Validity (r) & Criterion Tests strength
Biering-Sorensen (1984b)	240s prone extension hold arms across the chest	449M (30-60 yr) 479F (30-60 yr)		.21 .25 backward static max voluntary extension contraction with strain gauge
Hyytiainen, Salminen, Suvitie, Wickstrom, & Pentti (1991)	240s prone extension hold arms across the chest	33M (35-44 yr)	r=.80, R=.74	

Little information is available on the reliability and validity of any version of trunk extension test. The reliability of prone trunk extension as a measure of muscular strength and endurance was found to be acceptable (r=0.74) in males (35-44 years) by Hyytiainen et al. (1991) (Table 7). This value may be artificially low since an upper limit of 240 seconds was placed on the results (Plowman & Corbin, 1994). Using the maximal voluntary contraction at isometric backward extension of the trunk, Biering-Sorensen (1984b) has reported the validity of prone trunk extension as a measure of muscular strength to be low (r=0.21, 0.25) in males and females (30-60 years), where time was limited to 240 seconds again (Table 7). More data are needed across ages and genders.

Upper Arm and Shoulder Girdle

There are various field tests which have been used as a measure of muscular strength and endurance of upper arms and shoulder girdle including chin-ups/pull-ups, flexed arm hang, and push-ups. However, these tests have been found to have some disadvantages. First of all, too many zero scores occur in these tests. Data from the National Children and Youth Fitness Study I (NCYFS I) (Ross, Dotson, Gilbert, & Ketz, 1985) reported that 10-30% of the boys from 10 to 14 years of age and over 60% of the girls from 10 to 18 years of age could not do one chin-up. The President's Council on Physical Fitness and Sports 1985 National School Population Fitness Survey (President's

Council on Physical Fitness and Sports, 1992) showed similar results: 40% of boys aged between 6 and 12 years could not do more than one pull-up and 25% could not do one; 70% of all girls aged between 6 and 17 years could not do more than one pull-up and 55% could not do any. Additionally, 45% of boys aged between 6 and 12 years and 55% of the girls aged between 6 and 17 years could not perform the flexed arm hang for more than 10 seconds. Pate, Burgess, Woods, Ross, and Baumgartner (1993) reported that pull-up and push-up tests yielded very high percentages of zero scores, particularly among girls, in the study with 9 and 10 year old children (38 boys and 56 girls). Sixty-six percent of boys and 91% of girls in pull-ups and 16% of boys and 57% of girls in push-ups had zero scores. In Canada CAHPER Fitness Performance II Test (Canadian Association for Health, Physical Education, and Recreation [CAHPER], 1980) showed that 10-35% of the boys from 6 to 11 years of age and 15-45% of the girls from 6 to 17 years of age could not perform the flexed arm hang for more than 10 seconds. Thus, it is questionable if such tests can discriminate between performers. Not surprisingly, however, the reliability of these tests are high (R=0.66 to 0.97) (Table 8) probably because so many subjects were unable to execute a single performance on either the test or retest days.

The validity of these tests as a measure of muscular strength and endurance of upper arm and shoulder girdle is low (r=0.02 to 0.47) (Table 9) when correlated with criterion tests. The criterion tests used were dynamic 1RM (one maximal repetition) tests of the latissimus dorsi pull, bench press, or arm curl for strength and total repetitions of a selected percent of 1RM for endurance. Thus, concurrent validity has not been established to support these tests as absolute measures of strength or muscle endurance. The difficulty may be in the selection of the criterion measures or in the inability to isolate specific muscle groups in both sets of measures. Additionally, it has been reported that body weight or body fatness may affect the performance in these tests (Pate et al., 1993; Berger & Medlin, 1969). Pate et al. (1993) found that the validity coefficients for strength improved considerably when the test results were expressed relative to body weight (per kg). However, most of the validity coefficients for endurance remained low even when adjusted by weight. In addition, chin-ups/pull-ups and flexed arm hang require facilities or equipment such as a horizontal bar. Furthermore, these tests do not seem to be particularly responsive to training (Rutherford & Corbin, 1994: Plowman & Corbin, 1994). Students need a realistic chance to be successful in testing and to improve with training in order to be motivated (Corbin, 1986).

Table 8

Reliability of Upper Arm and Shoulder Girdle Strength and Endurance Tests

Study	Test	Subject	Reliability r:interclass, R:intraclass	
Engleman & Morrow (1991)	pull-up overhand		mean of two trials,	single trial
		70M (3rd grade)	R=.95,	R=.91
		89M (4th grade)	R=.96,	R=.92
		83M (5th grade)	R=.91,	R=.83
		87F (3rd grade)	R=.95,	R=.91
		74F (4th grade)	R=.95,	R=.91
		67F (5th grade)	R=.96,	R=.92
		242M (3,4,&5th grade)	R=.94,	R=.88
		228F (3,4,&5th grade)	R=.95,	R=.91
Pate, Burgess, Woods, Ross, & Baumgartner (1993)	pull-up overhand	38M (9-10 yr)		R=.80
		56F (9-10 yr)		R=.66
		38M & 56F (9-10 yr)		R=.79
Pate, Burgess, Woods, Ross, & Baumgartner (1993)	push-up	38M (9-10 yr)		R=.83
		56F (9-10 yr)		R=.71
		38M & 56F (9-10 yr)		R=.85
DiNucci, McCune, & Shows (1990)	flexed arm hang overhand	143M (college)	mean of two trials,	single trial
		57F (college)	r=.93,	R=.96
		143M & 57F (college)	r=.92,	R=.96
			r=.94,	R=.97
Pate, Burgess, Woods, Ross, & Baumgartner (1993)	flexed arm hang overhand	38M (9-10 yr)		R=.90
		56F (9-10 yr)		R=.85
		38M & 56F (9-10 yr)		R=.88
Cotten (1990)	modified pull-up		mean of two trials,	single trial
		31M (5th grade)	R=.79,	R=.65
		33F (5th grade)	R=.83,	R=.71
		29M (6th grade)	R=.90,	R=.82
		33F (6th grade)	R=.95,	R=.90
Engleman & Morrow (1991)	modified pull-up		mean of two trials,	single trial
		70M (3rd grade)	R=.81,	R=.68
		89M (4th grade)	R=.91,	R=.83
		83M (5th grade)	R=.87,	R=.77
		87F (3rd grade)	R=.90,	R=.83
		74F (4th grade)	R=.87,	R=.77
		67F (5th grade)	R=.90,	R=.82
		242M (3,4,&5th grade)	R=.87,	R=.77
		228F (3,4,&5th grade)	R=.89,	R=.81
Kollath, Safrit, Zhu, & Gao (1991)	modified pull-up	61M (9th grade)	mean of two trials,	single trial
		44F (9th grade)	R=.95,	R=.91
			R=.84,	R=.72
Pate, Burgess, Woods, Ross, & Baumgartner (1993)	modified pull-up overhand	38M (9-10 yr)		R=.83
		56F (9-10 yr)		R=.81
		38M & 56F (9-10 yr)		R=.83
	underhand	38M (9-10 yr)		R=.85
		56F (9-10 yr)		R=.88
		38M & 56F (9-10 yr)		R=.87
Jackson, Fromme, Plitt, & Mercer (1994)			mean of two trials,	single trial
	90° push-up	40M (24.5 +/- 4.9 yr)	R=.96,	R=.99
	90° push-up knees on the floor	23F (24.7 +/- 5.0 yr)	R=.98,	R=.99
McManis & Wuest (1994)	90° push-up		two different days,	single day
		25M (3,4,&5th grade)	R=.90,	R=.82
		20F (3,4,&5th grade)	R=.91,	R=.84
		25M & 20F (3,4,&5th grade)	R=.91,	R=.83
		32M (9 & 10th grade)	R=.59,	R=.42
		23F (9 & 10th grade)	R=.94,	R=.88
		32M & 23F (9 & 10th grade)	R=.75,	R=.75

Table 9

Validity of Upper Arm and Shoulder Girdle Strength and Endurance Tests

Study	Test	Subject	Validity (r) & Criterion Test: Strength	Validity (r) & Criterion Test: Endurance
Pate, Burgess, Woods, Ross, & Baumgartner (1993)	pull-up overhand	38M (9-10 yr)	-.16	.25
		56F (9-10 yr)	.05	-.09
		38M & 56F (9-10 yr)	.11	.08
			latissimus pull-down, 1-RM	latissimus pull-down, 50% 1-RM, N
Pate, Burgess, Woods, Ross, & Baumgartner (1993)	push-up	38M (9-10 yr)	.36	.47
		56F (9-10 yr)	.02	-.14
		38M & 56F (9-10 yr)	.38	.17
			bench press, 1-RM	bench press, 50% 1-RM, N
Pate, Burgess, Woods, Ross, & Baumgartner (1993)	flexed arm hang overhand	38M (9-10 yr)	-.23	-.15
		56F (9-10 yr)	-.12	-.15
		38M & 56F (9-10 yr)	-.06	-.09
			arm curl, 1-RM	arm curl, 50% 1-RM, N
Pate, Burgess, Woods, Ross, & Baumgartner (1993)	modified pull-up overhand	38M (9-10 yr)	.09	.21
		56F (9-10 yr)	-.03	.00
		38M & 56F (9-10 yr)	.17	.17
	underhand	38M (9-10 yr)	.02	.05
		56F (9-10 yr)	-.09	.17
		38M & 56F (9-10 yr)	.12	.07
			latissimus pull-down, 1-RM	latissimus pull-down, 50% 1-RM, N
Jackson, Fromme, Plitt, & Mercer (1994)	90° push-up	40M (24.5 +/- 4.9 yr)	.30	.41
	90° push-up, knees on the floor	23F (24.7 +/- 5.0 yr)	.23	.40
			bench press, 1-RM	max.rep.at 45.5kg for M, 22.7kg for F

A modification of these tests has significantly reduced the occurrence of too many zero scores. One is the modified pull-up test. The NCYFS II (Ross, Pate, Delpy, Gold, & Svilar, 1987) showed that only 5% or less of children in the age groups between 6 and 9 years were unable to perform one modified pull-up. The median score for girls ranged from 6 to 9 as for boys from 6 to 10. Pate, Ross, Baumgartner, and Sparks (1987) explained that this was probably related to the fact that part of the body weight is supported. Engleman and Morrow (1991), however, found that the modified pull-up does not negate the effect of body composition (skinfolds) on upper body strength performance. The reliability of this test has been found to be generally acceptable (R=0.65 to 0.95) (Table 8), while the validity needs to be studied (Table 9). There is still, however, a disadvantage in this test: it requires equipment which must be adjusted for each individual tested.

The other modified test is a 90° push-up test, which has some practical advantages. The most important advantages are that it requires no equipment, and push-up scores seem to improve with training (Rutherford & Corbin, 1994; Plowman & Corbin, 1994). This will give the students motivation. Furthermore, only a few zero scores occur in this test. Massicotte (1990) reported that only 5% of both boys and girls over 8 years of age, and only 10% of both boys and girls aged between 6 and 8 years could not do one 90^{o} push-up in his Canadian youth national study. This range of zero scores is very similar to those in the modified pull-up and appear to be acceptable as a field test. The partially supported body weight and the limited angle of 90 degrees at the elbow may provide this better range of scores. The reliability of this test needs to be established specifically at all ages and for both genders. A limited number of studies have reported a generally high reliability (R=0.42 to 0.99) of this test (Table 8).

Little specific validation data are available for the 90^{o} push-up as a muscular strength and endurance measure (Table 9). Jackson, Fromme, Plitt, and Mercer (1994) studied the validity of the 90^{o} push-up test in college males and females, but the time was limited to 1 min. They used a 1 repetition maximum on the bench press as a test of strength and a maximum number of consecutive repetitions on the bench press with a weight load of 45.5kg for males and 22.7kg for females as a test of absolute endurance. r=0.30 and 0.23 for the strength as well as r=0.41 and 0.40 for the endurance were reported in males and females, respectively. They concluded that the scores were not highly related to measures of strength and endurance and were confounded with body weight. In terms of muscular endurance, the validity is higher than most other tests but not satisfactory. The magnitude of validity is almost the same as other tests in males, but females obtain much better values than in other tests, where lower values are typically obtained by females (Plowman & Corbin, 1994). More research is needed on the 90^{o} push-up, especially with elementary school aged children and adolescents.

2.1.4 Muscular Flexibility

Flexibility is the range of motion about a joint and specific to each joint

(Baumgartner & Jackson, 1991; Minkler & Patterson, 1994). Most commonly used tests evaluate flexibility of the lower back and hamstrings, which are large frequently used muscle groups. Flexibility of these joints are speculated to be related to low back pain.

Table 10

Reliability of Flexibility Tests

Study	Test	Subject	Reliability r:interclass,	R:intraclass
Buxton (1957)	stand & reach	50M & F (6-15 yr)	r=.95, .96	
DiNucci, McCune, & Shows (1990)	sit & reach	143M (college)	r=.92,	R=.96
		57F (college)	r=.95,	R=.97
		143M & 57F(college)	r=.94,	R=.97
Jackson & Baker (1986)	sit & reach	100F (13-15 yr)		R=.99
Jackson & Langford	sit & reach	52M (20-45 yr)	r=.99	
		52F (20-45 yr)	r=.99	
Safrit & Wood (1987)	sit & reach	27M (11 yr)	r=.94,	R=.97
		88M (12 yr)	r=.94,	R=.97
		104M (13 yr)	r=.94,	R=.97
		58M (14 yr)	r=.95,	R=.97
		44F (11 yr)	r=.87,	R=.93
		92F (12 yr)	r=.93,	R=.96
		85F (13 yr)	r=.88,	R=.93
		43F (14 yr)	r=.80,	R=.89
Wear (1963)	sit & reach	53M (college)	r=.94	
		100F (college)	r=.96	

The various forms of stand or sit and reach tests are intended to measure low back and hamstring flexibility. The reliability of these tests is consistently high (r=0.80 to 0.99) (Table 10). However, with one exception (Jackson & Langford, 1989), researchers have reported that the stand or sit and reach has been shown to be moderately correlated with hamstring flexibility, but not with low back flexibility when validated against criterion measures of hamstring and low back flexibility (Table 11). Thus, the stand or sit and reach test seems to be only a test of hamstring flexibility. Flexibility is highly specific, and a single test is unlikely to do an adequate job of measuring more than one kind of flexibility. If the measurement of low back flexibility is of interest, a new field test should be developed (Minkler & Patterson, 1994).

Table 11

Validity of Flexibility Tests

Study	Test	Subject	Validity (r) & Criterion Tests Hamstring	Low Back
Broer & Galles (1958)	stand & reach	100F (college)	.81 Leighton flexometer	
Mathews, Show, & Bohnen (1957)	stand & reach	66F (college)	.80 Leighton flexometer	
vanAdrichem & VanderKorst (1973)	stand & reach	84M (6-12 yr) 55M (13-18 yr) 60F (6-12 yr) 49F (13-18 yr)		.23 .14 .33 -.15 modified Schober
Jackson & Baker (1986)	sit & reach	100F (13-15 yr)	.64 straight leg raise with Leighton flexometer	.28 modified Schober
Jackson & Langford (1989)	sit & reach	52M (20-45 yr) 52F (20-45 yr)	.89 .70 straight leg raise	.90 .70 modified Schober

Wear (1963) suggested an additional limitation to the sit and reach test, which was that sit and reach flexibility was significantly related to excess of trunk and arm length over leg length in college-age males, although Mathews, Shaw, and Bohnen (1957) found no significant relationship between sit and reach performance and leg length and standing reach. Moreover, Broer and Galles (1958) also found a significant relationship between toe-touch performance and the trunk-plus-arm length to leg length ratio in an investigation of college women. Because of this potential bias due to limb length differences, a modified sit and reach test was proposed (Hoeger & Hopkins, 1992; Hoeger, Hopkins, Sherman, & Palmer, 1990). This test accounted for the bias by establishing a relative zero point for each person according to the distance between the finger tips and the box. In the studies with adolescents and adult women, it was found that there were significant differences in sit and reach performance, with subjects with proportionally longer arms than legs performing significantly better than others, and those with proportionally longer legs than arms performing the poorest (Hoeger & Hopkins, 1992; Hoeger et al., 1990). There were, however, no significant differences in modified sit and reach performance across those different groups. Reliability coefficients for the modified sit and reach have been reported as high (R=0.94 in Hoeger & Hopkins, 1992; R=0.94 in Hoeger et al., 1990; R=0.99 in Minkler & Petterson, 1994). Minkler and Petterson (1994) examined the criterion-related validity of this modified sit and reach test against criterion measure of hamstring and low back flexibility in college males and females. The test was moderately related to hamstring flexibility (r=0.66 in females and r=0.75 in males), but its relation to low back flexibility was quite low (r=0.25 and r=0.40 respectively). Therefore, they concluded that this test did no better job assessing flexibility than the widely used sit and reach test.

Whether low back and hamstring flexibility are really predictive of low back pain has recently been questioned. Biering-Sorensen (1984b) showed that when flexibility of low back was determined by the Schober test, a high degree of low back flexibility instead of a lack of flexibility was predictive of first-time low back pain among middle aged men. However, Battie, Bigos, Fisher, Spengler, Hansson, Nachemson, and Wortley (1990) showed that low back flexibility did not predict prospectively either first-time or recurrent low back pain. In addition, neither study indicated that hamstring flexibility was predictive of first-time low back pain, although it was predictive of recurrent low back problems in middle aged women (Biering-Sorensen, 1984b). No evidence is available on children.

Stretching both hamstrings simultaneously can result in "overstretching" the low back. Calliet (1988) suggested stretching one hamstring at a time by having the other leg flexed to protect the low back by avoiding excessive flexion of lumbosacral spine. Moreover, stretching one leg at a time eliminates the possibility of hyperextension of knee joints by avoiding stretch of the ligaments and joint capsule of the knee (Lindsey & Corbin, 1991). Although this test is recommended by FITNESSGRAM (Cooper Institute for Aerobic Research, 1992), neither specific reliability nor validity data are available on this version of sit and reach test. Correlation of 0.91 and 0.92 between left and right one-leg sit and reach and the two-leg sit and reach has been reported (Plowman & Corbin, 1994). More data are required for this test.

Another joint, which is measured in field testing, is the trunk. Little information is available for the test of trunk flexibility, although it has been recommended by FITNESSGRAM (Cooper Institute for Aerobic Research, 1992). Wear (1963) reported that prone back extension was a highly reliable (r=0.96) test of trunk flexibility when the subjects lay prone with their lower limbs secured and were asked to raise the chin as high off the mat as possible. However, no version of trunk extension has been validated against any criterion measure of trunk flexibility.

2.2 Physical Fitness, Physical Education, and QDPE in High Schools

2.2.1 Physical Activity, Exercise, and Physical Fitness

Caspersen, Powell, and Christenson (1985) proposed definitions to distinguish physical activity, exercise, and physical fitness, which are often confused with one another and sometimes used interchangeably in the literature. Physical activity is defined as any body movement produced by skeletal muscles that results in energy expenditure. Physical activity in daily life can be categorized into occupational, sports, conditioning, household, or other activities. Exercise is a subset of physical activity that is planned, structured, and repetitive and has as a final or an intermediate objective the improvement or maintenance of physical fitness. Physical fitness is a set of attributes that are either health- or skill-related as explained later in this chapter. The degree to which people have these attributes can be measured on specific tests.

Physical activity is generally viewed as having a favourable influence on the growth, biological maturation, and physical fitness of children and youth (Malina, 1994). It has been also suggested that enhanced physical activity during childhood has its merits in the prevention or lessening of adult health risk, even though there are no data to support or reject this hypothesis (Bar-Or, 1994).

One of the most well-documented effects of regular physical activity is a higher level of physical fitness (Blair, Kohl, & Gordon, 1992). This permits freedom from disease as well as a higher level of functional ability to achieve activities of daily life. "The relation between physical activity and fitness was probably known in antiquity" (Blair et al., 1992, p. 101). Physical fitness (product) is produced from physical activity (process) (Corbin, 1986).

2.2.2 Definition of Physical Fitness and Physical Fitness Test

The definition of physical fitness has evolved over the years and has led to so many different ways of fitness categorization (Caspersen et al., 1985; Corbin, 1991; Pate, 1983a). Among professional educators the term physical fitness, though familiar, may be so vague, ambiguous, and nonspecific as to be almost meaningless (Pate, 1988). However, there seems to be general agreement that physical fitness consists of a number of factors. Often the meaning or definition of physical fitness is based on the types of tests most commonly used to measure it. As tests of fitness have changed, the definition of physical fitness has evolved from a unidimensional to a multidimensional concept (Corbin, 1991).

The current body of knowledge in exercise science and society's view of physical fitness suggests that a definition of physical fitness should focus on the health-related aspects of fitness (Blair et al., 1983; Caspersen et al., 1985; Corbin, 1991: Gutin et al., 1992; Pate, 1983a, 1988), and most of the current youth fitness batteries (e.g., AAHPERD Physical Best, 1989; FITNESSGRAM, 1994) are based on the testing of health-related physical fitness (Corbin, 1991). Pate (1983a) presented two fundamental philosophies, which were health or health-related fitness (i.e., cardiorespiratory endurance, muscular strength and endurance, flexibility, and body composition) and motor or athletic fitness (i.e., speed, power, and agility as well as components of health-related fitness) philosophies. The health-related fitness includes only components that can prevent disease and/or promote health, and high levels of athletic qualities are not considered essential. It is quite responsive to training. The motor fitness is of particular importance only in the athletic situation and carry little significance in day-to-day life. It is heavily dependent on genetic factors. Figure 1 summarizes these two philosophies. These two philosophies have not always been clearly distinguished from one another by professionals or the public. The author concluded that motor fitness was important for athletes, but health-related physical fitness was important for the general populace.

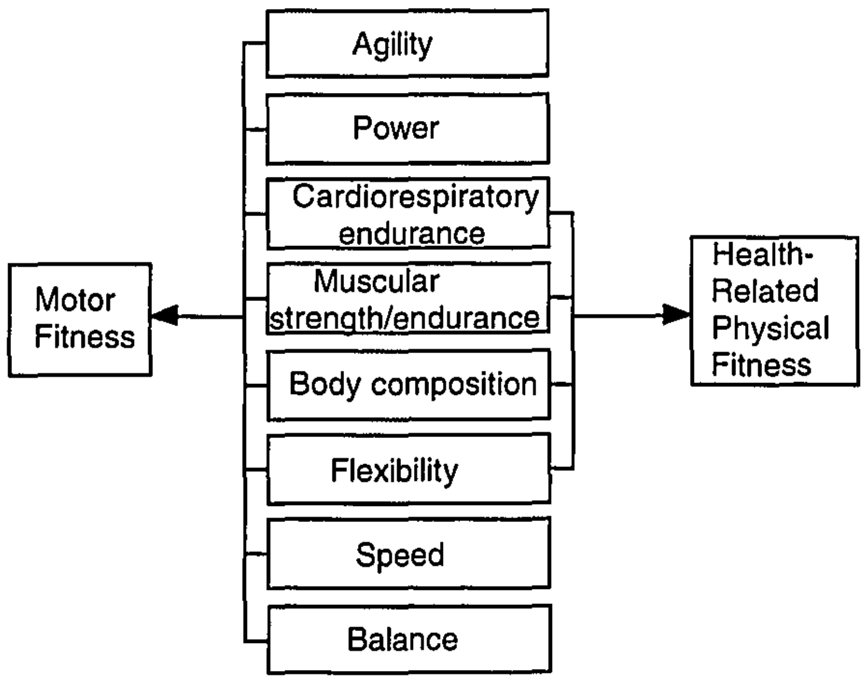


Figure 1. Components of Motor Fitness and Health-Related Physical Fitness.
Note. From "A new definition of youth fitness" by R.R. Pate, 1983, *The Physician and Sports medicine*, *11*(40), p. 78.

Corbin (1991) suggested a multidimensional hierarchical model of physical fitness (see Figure 2). Physical development is an integration of many "threads" in the multidimensional hierarchy. Two principal subdimensions of physical development are physical fitness and skill development. Physiological and health-related fitness are dimensions of physical fitness, while skill-related fitness is a dimension of skill development. Health-related physical fitness components are consistent with the definition by Pate (1983a) and Caspersen et al. (1985). Each component has subcomponents, for example, isometric and isotonic strength are subcomponents of strength. The health-related fitness is the dimension commonly considered as a definition of "physical fitness" for use in fitness testing. This is because physiological fitness testing (i.e., blood profiles) is somewhat invasive and may require equipment and expertise not generally available in non-medical settings. Skill-related fitness components can be measured in the field settings such as schools. However, skill-related fitness links more directly to skill development than to physical fitness as seen in the Figure 2. For this reason, such measures are not included in most current health-related physical fitness fieldtest batteries (Corbin, 1991).

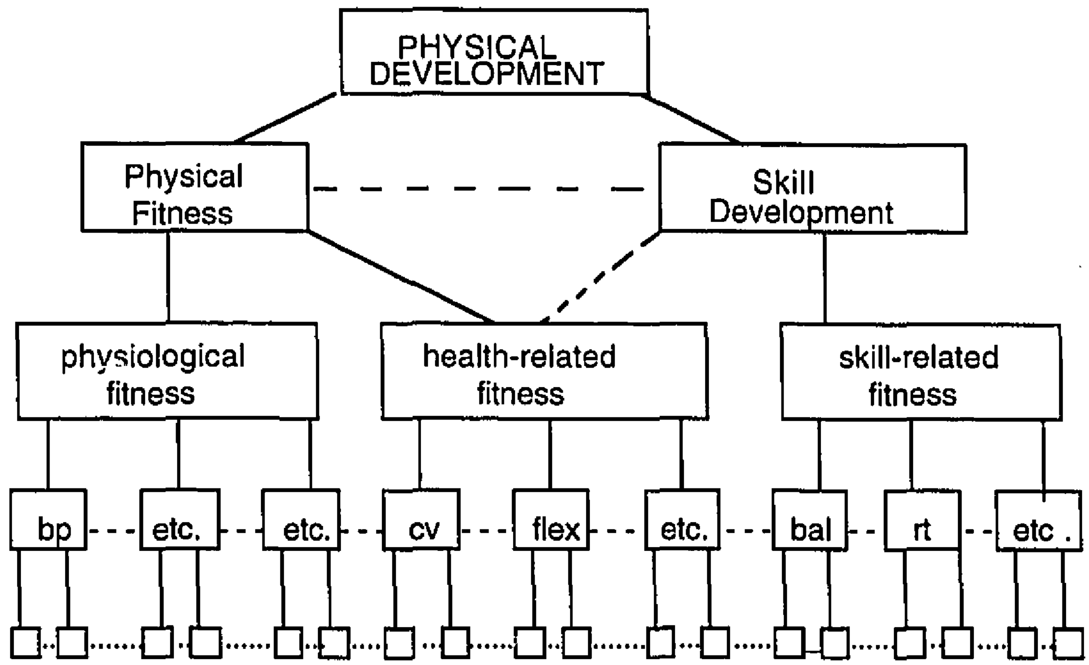


Figure 2. The Physical Domain: A Multidimensional Hierarchical Model.
Note. From "A multidimensional hierarchical model of physical Fitness: A basis for integration and collaboration" by C.B. Corbin, 1991, *QUEST*,*43*, p. 299.

2.2.3 Interpretation of Physical Fitness Test Scores

Interpretation of physical fitness test scores has been an issue addressed in promoting physical activities and lifetime physical fitness for children and youth (Corbin & Pangrazi, 1992; Cureton & Warren, 1990; Safrit, 1990). Traditionally, norm-referenced standards such as specific percentiles (e.g., 50th percentile) have been used in fitness testing to evaluate physical fitness of children and youth (Corbin & Pangrazi, 1992; Cureton & Warren, 1990). These standards make it possible to compare the performance of a child or a group of children to a known reference group (Corbin & Pangrazi, 1992). However, with the introduction of health-related physical fitness tests in the late 1970s and early 1980s, criterion-referenced standards have been suggested by researchers to be theoretically more sound than the normative standards (Blair et al., 1983; Cureton &

Warren, 1990): criterion-referenced standards would indicate levels of physical fitness required for good health, not maximal performance, and is irrespective of the levels of physical fitness of the reference group.

There has been a number of potential advantages as well as disadvantages addressed with the use of criterion-referenced standards. Cureton and Warren (1990) reviewed these issues. The one desirable characteristic of criterion-referenced standards is that it is explicitly linked to a criterion. They are standards that represent desired, specified levels of performance or health status on a criterion domain, behaviour, or attribute, that is, body composition, aerobic capacity, and neuromuscular function. Norm-referenced standards are not necessarily made to a criterion attribute or behaviour that is being evaluated.

A second advantage is that criterion-referenced standards represent an absolute, desired level of the criterion attribute or behaviour that is consistent with good health, while a normative standard such as the 50th percentile represents the current level of performance of children and youth compared to the "normal" population (Cureton & Warren, 1990). For example, when the normative population is relatively fit, and individual might score at the 15th percentile and yet still possess an adequate amount of fitness. This seems to be often the case that these standards are difficult to achieve and classify many children and youth as unfit (Blair, 1992; Corbin & Pangrazi, 1992). Fitness standards which appear unattainable may be more discouraging than encouraging to some children. "Learned helplessness" occurs when trying hard does not result in meeting fitness standards equal to those set by children who have better genetics or who have started to mature earlier (Corbin 1986, 1995). In order to maintain children's interest in activity and physical fitness their dreams need to have a chance to come true in their mind. "Repeated failure will make it clear that the dream will never come true, and the dream will begin to die" (Corbin, 1986, p. 83). Whitehead and Corbin (1991) found that intrinsic motivation of youth decreased when negative feedback was given for their physical fitness test performance. According to Deci and Ryan (1980), intrinsic motivation results from "the underlying need for competence and self-determination" (p. 42), and intrinsically motivated behaviours are "those that are performed in the absence of any apparent external contingency" (p. 42). Corbin (1986) suggested that as a result of poor intrinsic motivation "the physically rich are getting richer and the physically poor are getting poorer" (p. 82). Conversely, if the population was quite unfit, a score at the 90th percentile might still not be adequate in terms of health status. That is, youth will think that they are fit when they may not be. This could result in a decrease in exercise and future fitness erosion.

Third, it provides specific, individual, diagnostic information about whether status or performance is adequate by comparing individual's test scores with the standards. Consequently it becomes immediately apparent whether the individual need to modify his or her physical activity and/or diet (Cureton & Warren, 1990).

Fourth, the criterion-referenced scores in testing categorizes individuals into groups based on the standards, where those who meet or exceed the standards are referred to as masters and those who do not as nonmasters (Cureton & Warren, 1990). This is the principal purpose of the health-related physical fitness test. The degree to which the standard may be exceeded is often not important.

The principal criticisms of criterion-referenced standards are that they are arbitrary and that the consequences of misclassification can be severe (Cureton & Warren, 1990; Glass, 1978; Safrit, Baumgartner, Jackson, & Stamm, 1980). This results in false mastery, which incorrectly concludes that a person possess the minimal level of fitness, or false failure, which incorrectly classifies a person to be lacking the minimal level of fitness. The concern is that the false mastery may lead the person to a reduced level of physical activity and an increased risk of disease. The false failure may result in discouraging and preventing the person from making time and effort to participate in physical activities. Development of criterion-referenced standards usually requires empirical data, normative data, and expert judgment. It seems that the existing criterion-referenced standards have been derived exclusively from the normative data and experts' judgment (Corbin & Pangrazi, 1992; Cureton & Warren, 1990; Plowman, 1992). Because empirical data are mostly on adults, knowledge in children and youth, how much fitness in children and youth that is associated with minimal health risks in adulthood, is not available. The basis of the standards has not been revealed and, therefore, the extent to which they were established arbitrarily is unknown (Cureton & Warren, 1990).

Several steps can be taken to reduce the consequences of misclassification errors (Safrit et al., 1980). First, frequent retesting will increase the probability that a person will be properly classified within a short period of time. Second, more than one cut-off score can be established so that a serious misclassification is less likely. Third, just as the validity and reliability of a test should be determined, so should the validity and reliability of criterion-referenced standards. When well documented with a health rationale provided, criterion-referenced health standards are less arbitrary than norm-referenced criteria and become useful in efforts to improve public health (Corbin & Pangrazi, 1992).

Canada has not developed any test battery with criterion-referenced standards. Although at least five national youth health-related physical fitness tests currently have criterion-referenced standards in the United States of America (U.S.A.) (i.e., AAHPERD Physical Best [AAHPERD, 1988]; FITNESSGRAM [Cooper Institute for Aerobic Research, 1992]; Fit Youth Today [American Health and Fitness Foundation, 1986]; The South Carolina Fitness Test [Pate, 1983b], 1983; YMCA Youth Fitness Test [Franks, 1989]), a detailed description of the procedures used in their development has not been published nor have the standards been validated (Cureton & Warren, 1990; Plowman, 1992; Safrit, 1990). Therefore, these standards published for the recent version of fitness tests should be viewed as arbitrary until evidence of the validity of each standard is given. Safrit (1990) states that a properly developed test manual should include both criterion-referenced standards and tables of norms. It has been reported that standards based on sound research are only available for cardiovascular fitness (Cureton & Warren, 1990) and body composition (Williams, Going, Lohman, Harsha, Srinivasan, Webber, & Berenson, 1992) by researchers (Blair, 1992; Corbin & Pangrazi, 1992). Therefore, it was necessary in this study to select primarily fitness tests with which both criterion-referenced standards and tables of norms are available.

2.2.4 Physical Fitness of Children and Youth

With changes in the definition of physical fitness, fitness tests, and fitness standards used in the testing, recent reports in the U.S.A. started showing that fitness of North American children may not be as poor as the reports from 1950s made everyone believe for the past 40 years (Blair, 1992; Conger, Quinney, Gauthier, & Massicotte, 1982; Corbin & Pangrazi, 1992; Gauthier, Massicotte, Hermiston, & Macnab, 1983). There are two reasons for this. Firstly, many professionals have labeled generations of American children and youth as unfit without defining physical fitness and setting an acceptable standards for it (Blair, 1992; Corbin & Pangrazi, 1992). Until the mid 1970s, the motor fitness philosophy dominated physical education, and consequently fitness tests were developed to evaluate motor fitness (or skill-related fitness) components instead of health-related fitness components (Pate, 1983a). It was found that the significant decrease in fitness performance between 1975 and 1985 were only seen in items considered to be measures of athletic or skill-related fitness rather than health-related fitness (Corbin & Pangrazi, 1992). Four health-related fitness test items (i.e., sit-up, flexed-arm hang, pull-up, and sit-and-reach) generally showed improvement over the past 10 year period (Blair, 1992). In fact health-related fitness test performance (i.e., pull-up, flexed arm hang, toe-touch) in 1985 was superior to that in 1958 except for skinfold measurements (Corbin & Pangrazi, 1992). Secondly norm-referenced standards were used to evaluate children's fitness levels instead of criterion-referenced standards until recently (Corbin & Pangrazi, 1992). It was found that many more children are judged to be fit if criterion-referenced health standards were used instead of norm-referenced standards (Corbin & Pangrazi, 1992). However, because few health-related fitness variables have been studied over time, it is difficult to judge health-related fitness of youth with confidence (Corbin & Pangrazi, 1992). The sample is large but probably nonrepresentative of the U.S. population (Blair, 1992). In Canada, a comparison of the fitness performance of children and youth aged between 7 and 17 years was carried out on three of CAHPER Fitness-Performance Test items between 1966 and 1980 (Conger et al., 1982). Compared to 1966, health-related fitness test performances in the speed sit-ups and flexed arm hang were significantly superior in 1980 for boys and girls at all ages, and girls at age 7 and from 12 to 17 years, respectively. For skill-related fitness test performance, significant improvement was seen in the long jump for males and females at all ages in 1980. In 1983, Gauthier et al. compared the physical fitness levels of a representative sample of Canadian girls and boys aged 7 to 17 years with that of 1968. Physical work capacity adjusted to body weight increased by 7.3% for the boys and 20.2% for the girls. Results for all age groups of both sexes, with the exception of the 8, 12 and 13 (boys), and the 8 year old (girls), revealed a significant increase in performance. In terms of test standards, Canada has not developed any test with criterion-referenced standards. There is little data available on which to base a conclusion regarding changes in physical fitness in North American youth over the past several decades. The limited data available suggest that there has been no major change in health-related physical fitness of North American youth (Blair, 1992).

2.2.5 Role of Physical Education

Active participation in daily physical activities needs to be encouraged for individuals to enhance physical fitness and health. It is not likely to occur without the collaboration of many such as individuals, families, friends, school, community, and private agencies (Corbin, 1991; McKenzie, Alcaraz, & Sallis, 1994). The most obvious place to begin exercise programs to build lifetime fitness and related good health is in the school (Corbin, 1987a, 1987b, 1991, 1995; McKenzie et al., 1994; Sallis & McKenzie, 1991; Siedentop, 1992; Simons-Morton, O'Hara, Simons-Morton, & Parcel, 1987). Physical education is the logical primary target for intervention to increase participation in and appreciation of physical activity (Corbin, 1987b, 1991, 1995; Haywood, 1991; Nelson, 1991; Sallis & McKenzie, 1991; Siedentop, 1992) and can involve all children in the school, not just those interested in or otherwise encouraged to participate in organized sport after school (Simons-Morton et al., 1987). However, physical education classes cannot be held responsible for providing students all the physical activities needed to achieve desirable fitness levels and produce health benefits (Corbin, 1987b, 1991, 1995). First, children spend too little time in physical education classes and much time in these classes is not spent in activity. Second, there is not enough time to spend to develop all the fitness components in each class. Third, fitness is not the only objective of physical education; skill learning is often considered to be of equal or more importance. It is stimulus for movement and contributes to total fitness (Seefeldt & Vogel, 1987). A physical education program with health-related fitness alone is not complete: it lacks opportunities for the development of a wide range of motor-performance skills and experiences leading to satisfying participation in selected lifetime sports and physical activities (Jewett, Bain, & Ennis, 1995). There are other objectives as well such as cognitive, emotional, and social development and "higher order fitness objectives" discussed later (Corbin, 1987a). Therefore, collaboration within schools is important including exercise breaks in addition to physical education classes and recess (Corbin, 1987b), athletic (intramural) programs, interscholastic programs (Sallis & McKenzie, 1991), adventure education experiences, dance, and other movement experiences (Dyson, 1995).

2.2.6 Conceptual Physical Education Programs

In addition to increasing physical activity necessary to promote fitness development, physical education classes can also provide conceptual programs which enhance the various domains of learning. Physical education has the propensity to enhance the child's social, emotional, cognitive, and physical development (National Association for Sport and Physical Education [NASPE], 1992). Simultaneously, physical education is responsible for the child's knowledge in health-related topics such as nutrition, habitual behaviours (e.g., smoking, alcohol consumption), and stress management (Jewett et al., 1995). However, this role of physical education programs is often not recognized (Haywood, 1991). There have been studies reporting the attempt of the conceptual courses in school physical education programs (Bailey, 1985; Goldfine & Nahas, 1993; Phillipp et al., 1991; Rider et al., 1986; Slava, Laurie, & Corbin, 1984; Smith & Cestaro, 1992). "Higher order physical fitness objectives" such as establishing personal exercise programs, learning to test their own fitness, interpreting their own test results, and learning to solve their own fitness problems may be more important than temporary attainment of physical fitness (Corbin, 1987a, 1995). Physical fitness is transient and a lower level fitness objective. If people are to be fit for a lifetime they must move to higher order objectives (Corbin, 1987a). Rider et al.(1986) investigated the effects of fitness courses in high school physical education, which included cognitive, attitudinal, and physical fitness components. They found significantly improved knowledge of fitness concepts, physical fitness, and attitude toward participation in physical activity of high schools students. However, it is often commented that time during physical education classes are not sufficient enough to meet all fitness objectives (Corbin, 1987a, 1995). Health education programs within school in combination with physical education classes can increase physical activity levels and collaboration of classroom teachers and school nurses among others can encourage physical activity and other healthy life-style choices (Corbin, 1991; Corbin, Fox, & Whitehead, 1987). School can be particularly effective in teaching knowledge and attitudes and in working to alter other factors (e.g., beliefs, self-motivation, self-confidence, and a previous history of exercise) that predispose people to exercise (Corbin, 1991). However, when school health promotion programs have targeted physical activity without directly influencing physical education classes, the results have been much less encouraging (Sallis & McKenzie, 1991). School curriculum should be modified to ensure that enough time is spent on physical activities which are designed to build all components of fitness and that conceptual programs exists for the students to learn about physical activity, physical exercise, and healthy life-styles (Sallis & McKenzie, 1991). Furthermore, physical educators are challenged to collaborate with the public in developing and evaluating school physical education programs (Magnotta, 1991; McKenzie et al., 1994; Sallis & McKenzie, 1991). These recommendations have the support of other researchers (Fox, 1991; Haywood, 1991; Morris, 1991; Nelson, 1991; Smith & Cestaro, 1992; Wescott, 1992).

2.2.7 High School Physical Education

Adolescence appears to be a time when regular physical activity can produce marked changes in health related and motor fitness (Corbin, 1987a) and also the most appropriate time to introduce a conceptual curriculum that addresses health-related benefits of physic al activity (Goldfine & Nahas, 1993). If individuals' inactivity and disease risk profiles are to be altered, the younger years are the most opportune time to affect change before their lifestyles are firmly established (Goldfine & Nahas, 1993; Sallis & McKenzie, 1991). Furthermore, adolescents are more likely to comprehend health concepts compared to younger children. Younger children usually cannot conceptualize the abstract notion of the effects of activity on health, so their act cannot be stimulated through personal health concerns (Maddux, Roberts, Sledden, & Wright, 1986). The notion that a person should be fit to prevent health problems is too abstract to have great personal meaning to young children (Corbin, 1986). Adolescents, however, are intellectually capable of understanding health, fitness, and disease concepts and these knowledge raises students' awareness, contribute to their beliefs and attitudes, and increases their ability to make informed decisions about physical activity (Fox, 1991; Sallis & McKenzie, 1991). Franks (1984) reported that the high school physical education curriculum provides a transition for the student from childhood to life beyond the school years. However, Haywood (1991) claimed that high school physical education programs seemed to be the weakest link in the chain from initial positive experiences at elementary schools to adulthood. First, adolescents' participation in physical activity decreases when many schools no longer require yearly physical education to complete their graduation requirement (Haywood, 1991). Second, studies (Lacy & LaMaster, 1990; Li & Dunham, 1993) have shown that high school physical education classes did not produce adequate frequency, intensity, and duration of physical activity for students to achieve minimal fitness. Physical educators should plan fitness activities in the daily lesson incorporating exercise principles of intensity, duration, frequency, specificity, and progressive overload (Lacy & LaMaster, 1990). Third, adolescence is the time students begin to make their own choices. High school physical education programs are expected to foster knowledge and appreciation of the role of physical activity in health. When school physical education programs do not lead the students from their experiences in elementary school programs to high school programs, they spend more time in their lives for other interests rather than physical activities (Haywood, 1991). Although budgets are limited, high school physical education programs should provide various choices of activities so that adolescents can find one or a combination which suits their individual needs and interests (Bailey, 1985; Corbin, 1995; Fox, 1991; Franks, 1984; Lacy & LaMaster, 1990; Wescott, 1992). In this way their personal preference would be met, and a large drop in physical activity participation during the teen years would be improved. At this time, the personal health benefits should be emphasized rather than competitive team sports (Corbin, 1995; Fox, 1991). Consider new scheduling for high school physical education classes which allows more time to dress after class or activities which do not require strict dress codes (Corbin, 1987a). This is unlikely, but Wescott (1992) and Corbin (1987a) have provided challenging examples: monitor locker rooms to prevent locker room abuse and install hair dryers and private showers such as those used in modern fitness facilities. Physical educators must program long-term instructional activity classes where students work seriously towards mastery goals (Siedentop, 1994). Students should be helped to move toward responsibility such as exercising choice, which suggests that much of what physical educators do has to have elective features (Siedentop, 1992).

2.2.8 Quality Daily Physical Education

With the movement of increased time allotment for physical education as well as studies reporting positive effects of daily involvement in physical activities on children and youth, the need for quality daily physical education (QDPE) has been advocated by professionals in medicine, education, and related fields (Dahlgren, 1987; Hansen & McKenzie, 1988; Kay & Grant, 1992; Macfarlane, O'Neill, & Kay, 1992; Siedentop & Siedentop, 1985; Silver, 1983; Verabioff, 1986). Bailey (1973) reported the decline in fitness level with age of children and youth in Saskatchewan and justified the additional time for physical education throughout the school years. According to the Canadian Association for Health, Physical Education, Recreation, and Dance (CAHPERD), QDPE is more than daily instruction. Stereotypes of physical education as merely "play time" or a series of competitive team sports has been challenged. If implemented in schools, QDPE claims 1) maximum active participation, 2) wide range of movement experiences, 3) total fitness activities, 4) qualified and competent teachers, 5) adequate and appropriate equipment and facilities, 6) the principles of child growth and development as the problem base, and 7) opportunities to develop positive attitudes to activity and suitable competition (Hansen & McKenzie, 1988). In Canada, since 1970's it has led several provinces such as Saskatchewan, Manitoba, and New Brunswick to provincial-wide policy for daily physical education (Hansen & McKenzie, 1988; Verabioff, 1986). Studies which have investigated the effects of daily physical education programs on children and youth have shown conflicting results (Dragicevick et al., 1987; Johnson, 1969; Kemper et al., 1976; Phillipp et al, 1991; Pollatschek, 1989; Sinclair, 1983; Werner & Durham, 1988). However, in general, there is literature that demonstrates positive effects of QDPE on children and youth (Dahlgren, 1987; Hansen & McKenzie, 1988; Verabioff, 1986). Robbins (1987) surveyed selected schools with QDPE and reported the benefits of QDPE. They included 1) improved fitness, 2) increased knowledge and understanding of a healthy lifestyle, 3) improvements in psychosocial aspects of programming, 4) improved school morale, 5) increases in alertness, 6) attitudes toward study and academic achievement, and 7) improvements in the respect for others and property. Because physical education has been often viewed as a competitor for time and resources with other subject areas, the implementation of QDPE has been intermittent. Physical education needs to establish credibility as an essential subject in the school curriculum (Dahlgren, 1987; Hansen & McKenzie, 1988; Verabioff, 1986). Hansen and McKenzie (1988) identified seven categories of barriers existing for effective implementation of QDPE in Canada: 1) lack of involvement of publics, 2) lack of teacher commitment, 3) loss of academic time, 4) insufficient specialist, 5) quality of physical education as a subject, 6) lack of promotion of benefits of QDPE, and 7) lack of resources, facilities, equipment and safety.

2.2.9 QDPE in High School

There are few studies which have investigated effects of QDPE on health-related physical fitness of high school students. The earlier studies (Johnson, 1969; Kemper et al., 1976) examined the effects of the time differential of physical education instruction upon fitness and skill development. Johnson (1969) compared eighth grade boys and girls between 5-day-a-week to 2- and 3-day-a-week physical education classes. It was found that 5-day boys improved significantly better in push-ups, pull-ups, skinfold measurement, and another skill-related fitness component (i.e., standing broad jump), while 5-day girls improved in only a skill-related component (i.e., standing broad jump) after two consecutive years. Kemper et al. (1976) reported that after a whole school year 12 and 13 year old boys of 5-day physical education were significantly superior only in handgrip compared to those of 3-day physical education.

More recent studies have been conducted to examine the effects of time as well as content differential of physical education instruction. Dragicevick et al. (1987) conducted a year-long daily physical education program in a junior high school in New Zealand. Students received three 45 min normal physical education lessons and one 60 min sports period per week. In addition, a 10 min period of physical activities (i.e., aerobic game, an obstacle course, exercise from cards, and a 12 min run) followed with a 10 min warm-up period in ability groups each day. A series of three speakers and a forum on sport and fitness was also arranged for the first term to cover the broader aspects of physical education. Students in the control school received two 40 min periods of normal class physical education lessons and no additional whole-school activities other than one 60 min period of school sports per week. In order to measure physical fitness of students, five field tests, 1) sit and reach (flexibility), 2) standing jump (explosive power), 3) curls (abdominal strength and endurance), 4) shuttle run (speed and agility), and 5) 500 m run (aerobic endurance), were carried out. In addition, anthropometric measures (i.e., height, weight, mid-arm circumference, body mass index, and skinfold thickness) and the oxygen uptake at a heart rate of 170 beats per minute (VO_2170) on a bicycle ergometer were measured.

The results of this study showed that initial significant improvements at the pilot school in comparison with the control school were seen for females and males in the sit and reach, for males in the curls and shuttle run, and for females and males in the 500 m run. However, subsequent poor compliance with the program in the pilot school reduced the school difference in fitness, and the students of control school were scoring higher for females in the sit and reach and standing jump. Both schools showed a substantial loss in aerobic fitness over the summer holidays. There was no significant differences in the anthropometric measurements and the VO_2170. There was anecdotal evidence that the academic performance and social attitudes of the students in the pilot school were greatly improved in comparison with previous years. However, the performance in non-physical curriculum areas were not systematically analyzed in this study.

Phillipp, Piland, Seidenwurm, and Smith (1991) found no significant effects of daily physical education courses during two consecutive summer sessions on physical fitness of high school students when comparing with that of control groups. Both experimental and control groups undertook physical fitness tests (i.e., weight, sit and reach, sit-ups, skinfold measurement, and 1.5 mile run) and health behaviour tests (Teen Wellness Check, 1984). The daily experimental physical education course emphasized students' direct involvement in both team and individual physical fitness activities and also reinforced health related abilities and knowledge as well as individual responsibility through field experiences at an ice skating rink, an aerobics gym, a supermarket, a medical center and a hiking trail. The course was 4 hour long and lasted for 6 weeks during a summer session. Unlike the experimental group classes. the control group classes participated in the traditional summer school physical education curriculum consisting of a 6 week time period with daily sessions lasting 4 hours, which centered primarily on group related physical activities(i.e., team sport) with limited classroom instruction. The only statistically significant difference between the experimental and control groups in both years was for the 1.5 mile run, where the experimental group improved more. In Phillipp et al. (1991), this two year study was unable to verify statistically significant improvements due to participation in this experimental program.

A limited number of studies for this purpose have been conducted. In Canada studies have been conducted with elementary school students. Therefore, this study investigated the effect of QDPE on health-related physical fitness of high school students.

Chapter 3

Methodology

3.1 Subjects

The subjects in this study were 102 students from Argyle Academy and 96 students from Riverside Park Academy. These subjects were selected as the students in the seventh, eighth, and tenth grade intact physical education classes in each school. These two high schools are similar (e.g., size, geographical location, common feeder schools, and social economic status) and affiliated with the same school board (the Protestant School Board of Greater Montreal). These schools have different physical education programs, that is, emphasis on physical fitness as a component of the curriculum. Argyle Academy has a school curriculum which emphasizes fitness and physical education and has fifty minute daily physical education classes in the seventh and eighth grades (6 classes per 7 day cycle) but not in the tenth grade (3 classes per 7 day cycle). In comparison, Riverside Park Academy does not have daily physical education classes in any grade and has physical education classes only 3 classes per 7 day cycle.

During the 1984-85 academic year, Argyle Academy developed facilities, classes, and extra-curricular activities to promote physical fitness among its student enrollment. Specifically, a large-sized classroom was renovated to house fitness testing and training equipment such as cycle ergometers, a computerized treadmill, rowing machines, cross-country ski trainer, global gymnasium, heart rate monitors, skinfold calipers, and hand grip dynamometers. Supervision and access to the facilities was created for students. The scope of Argyle Academy's fifty minute daily physical education classes was increased to include a wide variety of activities designed to develop both interest and fitness (e.g., outdoor cross-country skiing, fitness room classes, golf, archery, and tennis). In addition, extra-curricular activities were designed in which participation was either encouraged or compulsory. For example, interschool athletics were encouraged but non-compulsory. On certain designated weekends outdoor education trips were compulsory (Peterson, 1987).

Finally, Argyle Academy has its own fitness test battery and evaluates students' physical fitness once a school year. The Argyle test battery is composed of the following tests which evaluates both major components of health-related and a specific component of skill-related physical fitness:

Body Composition

1) Body Mass Index
2) Sum of Three Skinfolds (triceps, Subscapular, and Iliac Crest)
3) Waist to Hip ratio
4) Sum of Two Trunk Skinfolds (Subscapular and Iliac Crest)

Muscular Strength/Endurance

5) Grip Strength
6) Sit-Ups
7) Push-Ups

Flexibility

8) Sit and Reach

Aerobic Capacity

9) The Leger Boucher 20 m Shuttle Run
(The 20min walk/run test was previously used until 1993)

Power

10) Standing Long Jump.

Riverside Park Academy also evaluates students' physical fitness once or twice a school year. The test battery evaluates muscular strength and endurance, muscular flexibility, and power, but no measures of body composition and aerobic capacity are included. The following tests are used:

Muscular Strength/Endurance

1) Push-Ups
2) Partial Curl-Ups

Flexibility

3) Sit and Reach

Power

4) Standing Long Jump.

A classification breakdown describing numbers, school group, grade, and gender of the subjects is presented in Table 12.

Table 12

Number of Subjects Classified by School, Grade, and Gender

School	Argyle			Riverside Park		
Grade	7	8	10	7	8	10
Males	13	10	29	13	14	19
Females	12	15	23	12	12	26
Total # of Subjects		102			96	

In both schools, students in the seventh and eighth grades participated in co-education classes, while those in the tenth grade participated in single-sex classes.

Only subjects who completed all the tests in the battery contributed to the data. Those who did not complete all the tests were often prevented due to injury and illness. In a few cases, female students refused the skinfold measurements. Eleven students were not able to complete all the tests.

3.2 Test Description and Instrumentation

A McGill test battery was developed to assess health-related physical fitness of high school students. The test battery was developed in order to discriminate differences in students' physical fitness and for future use in schools. Currently available test items were examined through a literature review (see Chapter 2). The most appropriate tests in this study for each fitness component were selected, with safety, reliability, validity, applicability (i.e., procedure, instrumentation), and availability of standards for the test being the crucial selection factors. As a result, this test battery is composed of test items recommended by C.S.T.F. (C.S.T.F., 1986), FITNESSGRAM (Cooper Institute for Aerobic Research, 1992), and Massicotte (1990).

Safety, reliability, and validity of tests are summarized in the Chapter 2. Applicability of tests was determined by available literature, expert consultation, physical educators and a pilot study. Since a combination of both criterion- and norm-referenced standards was recommended by experts, test items with both standards available were primarily selected. Except for tests of flexibility (i.e., back saver sit and reach, trunk lift), and waist to hip ratio, all others met all those criteria. Back saver sit and reach was chosen over the traditional sit and reach test because of its safety (Calliet, 1988; Lindsey & Corbin, 1991), although there are only limited data available on its reliability and validity, and only criterion-referenced standards are available for this test. Trunk lift was included in the test battery, since it is recommended to assess trunk flexibility in addition to flexibility of other joints measured by the back saver sit and reach (Cooper Institute for Aerobic Research, 1992). However, little information is available on reliability and validity of this test, and there are only criterion-referenced standards available for this test. For waist and hip ratio, there is only norm-referenced standards for youth aged 15 to 19 years available. This test was selected as a supplement to skinfold measurements, since it is important to assess the patterning of trunk adipose tissue distribution (McArdle et al., 1991). Height and weight were measured as anthropometric measurements and tables of norms are available.

The test battery, test procedures, and sources of standards for each test item are listed according to the component of fitness each test evaluates.

3.2.1 Aerobic Capacity

1) The Multistage 20 m Shuttle Run Test

In the gymnasium the 20 m course was measured by a measuring tape and marked with a masking tape and marker cones. The cassette tape was calibrated by using the one-minute test interval at the beginning of the tape. Students were lined-up behind the start line and given an instruction of the test. After a five second count down, on the sound signal students were required to start running back and forth on a 20 m course and touch the 20 m line by the time a sound signal was emitted from the prerecorded tape. If the students reached the other line before the beep, they had to wait for the beep before continuing. The frequency of the sound signals increased in such a way that running speed was increased by 0.5 km/hr each minute from a starting speed of 8.5 km/hr. A student was allowed to attempt to catch up with the pace until he/she missed two beeps. The test stopped when the subject was no longer able to follow the set pace. The last completed stage number was recorded. Students who had stopped the test moved to the outside of the testing area being careful not to interfere with others who were still running. They were advised to continue walking to cool down right after completing the test. (Leger, Mercier, Gadoury, & Lambert, 1988; Cooper Institute for Aerobic Research, 1992).

Both norm-referenced (Leger et al, 1988; Massicotte, 1990) and criterion-referenced standards (Cooper Institute for Aerobic Research, 1992) are available for this test.

3.2.2 Body Composition

2) Skinfold Measurements

All measurements were taken on the right side of the body. During the skinfold measurements, it was essential that the student relaxed the underlying musculature as much as possible. When the site of the skinfold had been determined, a fold of skin plus the underlying fat is grasped between the tester's thumb and forefinger with the back of the hand facing the tester. The John Bull caliper jaws were placed at right angle to the body surface one centimeter below the point where the skinfold was raised. While maintaining the pressure of the fingers on the skinfold, the trigger of the calipers was fully released and the measurement was taken. The measurement was noted when the indicator stabilized which was approximately two seconds after the full pressure of the caliper jaws was applied to the skinfold. The reading was recorded to the nearest 0.2 mm. The first set of all the skinfold measurements was completed before repeating the procedure to obtain a second skinfold measurement for each site. The mean of the two measures was recorded unless the difference between the first and second measure of that particular skinfold site was found to be greater than 0.4 mm. If so, a third measure of that skinfold site was taken and the closest two of the three scores were averaged. When the three measures were equidistant(e.g., 18.6, 19.2 & 19.8), the mean of all the three values was determined (C.S.T.F., 1986). Five skinfold sites (i.e., triceps, biceps, subscapular, iliac crest, and medial calf) were identified as described in the C.S.T.F. operations manual (C.S.T.F., 1986).

Norm-referenced standards for the sum of five skinfolds (Canada Fitness Survey, 1985) are available for this test.

3) Standing Height and Body Weight

a) Standing Height

A ruler or a tape was positioned vertically against a wall so as to ensure that it was perfectly straight and even with the floor. The student, without footwear, was asked to stand erect, place arms hanging by the sides, feet together, and the heels and back in contact with the wall, and look up straight ahead. The measurement was determined by employing a set square on the head and against the wall. The height was recorded to the nearest 0.5 cm (C.S.T.F., 1986).

b) Body Weight

A calibrated weight scale was placed on a flat hard surface. The student was without footwear and in light clothing (shorts and T-shirt). The weight was recorded to the nearest 0.1 kg (C.S.T.F., 1986).

Tables of norms (C.S.T.F., 1986) are available for the Standing Height and Body Weight.

4) Waist to Hip Ratio

Waist to hip ratio was determined by the following formula:

Waist Girth (cm)/ Hip Girth (cm).

a) Waist Girth

The student was asked to stand erect in a relaxed manner and place the arms hanging loosely at the sides. The tester used a cross-handed technique to position the tape horizontally at the level of noticeable waist narrowing. Tension was applied to the tape sufficient to maintain its position but not to cause indentation of the skin surface. The measurement was recorded to the nearest 0.1 cm (C.S.T.F., 1986).

b) Hip Girth

The student was asked to stand erect with feet together. The tape was positioned horizontally around the hips at the level of the greatest gluteal protuberance. The measurement was taken in the same manner as the waist girth measurement (C.S.T.F., 1986).

Norm-referenced standards for youth aged 15 to 19 years (C.S.T.F., 1986) are available for this test.

3.2.3 Muscular Strength and Endurance

Abdominal Strength and Endurance

5) Curl-Ups

Two 80 cm strips of masking tape were attached to the mat with the distance of 12 cm between them. Students were paired two by two; one who performed the curl-ups while the other helped the tester to count the number of curl-ups and placed his/her hands behind the head of student being tested. The student being tested lay in a supine position on the mat, knees bent at an angle of approximately 90°, feet flat on the floor, legs slightly apart, arms straight and parallel to the trunk with palms of hands resting on the mat. The fingers were stretched out, touching the outside of the first tape with the fingertips and the head was in contact with the partner's hand, resting on the mat. Keeping heels in contact with the mat, the student curled up slowly sliding palms of hands across the masking tapes on the mat until fingertips reached the outside of the second tape, then curled back down until the head touched the partner's hand. Movement was slow and controlled to the prerecorded cadence which was about 20 curl-ups per minute (one curl every 3 seconds). The student continued without pausing between curl-ups until he/she could no longer continue or had completed a maximum number of 75 curl-ups. The score was the number of correctly performed curl-ups. Count was made when the student's head returns to contact the partner's hand on the floor (Massicotte, 1990; Cooper Institute for Aerobic Research, 1992).

Both norm-referenced (Massicotte, 1990) and criterion-referenced standards (Cooper Institute for Aerobic Research, 1992) are available for this test.

Upper Arm and Shoulder Girdle Strength and Endurance

6) 90^{o} Push-Ups

The student being tested assumed a prone position on the mat with hands placed under the shoulders, fingers stretched out, legs straight, parallel and slightly apart, and toes tucked under. The student pushed up off the mat with the arms until arms were fully extended, keeping legs and back straight. The student then lowered the body using the arms until the elbows bent at 90^{o} angle and the upper arms were parallel to the floor, then returning to the straight-arm position. This movement was repeated as many times as possible. The rhythm was 20 push-ups per minute or one push-up every three seconds. The score was the number of push-ups completed successfully (Massicotte, 1990; Cooper Institute for Aerobic Research, 1992).

Both norm-referenced (Massicotte, 1990) and criterion-referenced standards (Cooper Institute for Aerobic Research, 1992) are available for this test.

3.2.4 Muscular Flexibility

7) Back Saver Sit and Reach

The sit and reach box used was approximately 12 inches high and had a measuring scale on top with the 9 inch mark even with the near edge of the box. The student removed his/her shoes and sat down facing the box with the zero end being nearest the student. One leg was fully extended with foot flat against the end of the box. The other knee is bent with the sole of the foot flat on the floor and 2-3 inches to the side of the straight knee. The arms were extended forward over the measuring scale with the hands placed one on top of the other. With palms down, the student reached directly forward with both hands along the scale four times and held the position of the fourth reach for at least one second. The extended knee was kept straight, and hands had to reach forward evenly. If necessary, the student was allowed to move the bent knee to the side as the body moved forward. After measuring one side, the student switched the position of the legs and reached again. The number of inches reached on each side was recorded (Cooper Institute for Aerobic Research, 1992).

Criterion-referenced standards (Cooper Institute for Aerobic Research, 1992) are available for this test.

8) Trunk Lift

The student being tested lay on a mat in a prone position. Toes were pointed and hands were placed under the thighs. The student lifted the upper body off the floor, in a very slow and controlled manner, to a maximum height of 12 inches. The position was held long enough to allow tester to place a ruler on the floor in front of the student and determine the distance of the student's chin from the floor. The ruler had a 12 inch mark with a tape and was placed at least as inch to the front of the student's chin and not directly under the chin. Once the measurement had been made the student returned to starting position in a controlled manner. Two trials were allowed and the highest score was recorded. The score was recorded to the nearest inch. Distances above 12 inches was recorded as 12 inches (Cooper Institute for Aerobic Research, 1992).

Criterion-referenced standards (Cooper Institute for Aerobic Research, 1992) are available for this test.

3.3 Pilot Study

Prior to conducting the data collection, testers conducted a pilot study to familiarize themselves with the assessment items. The pilot study was designed to identify problems with the test items and to clarify instructions given to students. It also allowed testers to refine the organization and efficiency of the testing process, thus minimizing the amount of class time taken to test students.

This pilot study lasted for one week in each school. Participants were 22 students in the eighth grade girls' class in Argyle Academy and 13 students in the seventh grade co-education class and 14 students in the eighth grade co-education class in Riverside Park Academy. These students did not participate in the study.

Since skinfold measurements require the tester to be adequately trained and experienced with the standardized procedures, the principle investigator in this study spent two weeks prior to the data collection testing and retesting school students. This group was composed of approximately 150 boys and girls from the seventh, eighth, and ninth grade classes who were not participants in the study. This tester administered all the skinfold measurements during the data collection.

3.4 Data Collection

Three physical education teachers in Argyle Academy and four physical education students from McGill University who were teaching assistants in either Argyle Academy or Riverside Park Academy assisted with data collection. These assistants received training with the test procedures prior to data collection. They were asked to adhere to the procedures in order to obtain accurate and reliable results. In addition, the data collection was periodically checked by an outside observer, a physical educator form McGill University, to ensure that the data collection was accurate and reliable.

All students were tested during the regular physical education classes over eight weeks in October, November and December. Each student was dressed in regular physical education uniform for the testing. Prior to testing, letters of consent were distributed and signed by parents and students. This study was approved by McGill University, the

Protestant School Board of Greater Montreal, and both Argyle Academy and Riverside Park Academy.

3.5 Experimental Design and Statistical Analysis

The dependent variables in this study are (1) sum of five skinfolds, (2) waist to hip ratio, (3) 20 m shuttle run, (4) mean of back saver sit and reach on the right and left, (5) trunk lift, (6) curl-ups, and (7) 90^{o} push-ups. Descriptive statistics (means and standard deviations) were calculated for these dependent variables. Hypothesis 1 to 4 for the effect of school group (Argyle and Riverside Park Academies), gender, and grade (grade 7, 8 and 10) on physical fitness performance variable scores were tested using a 2 x 2 x 3 factorial MANOVA (school x gender x grade). Univariate F tests were performed to identify significant differences among these factors for each dependent variable. This experimental design is presented in Table 13. The 0.05 level of significance was used for all statistical analyses in this study. Since the purpose of this study was to compare physical fitness of students of the same gender and grade between two different schools, comparisons of physical fitness between students of different genders and grades were not of interest. Following the significant F test results, Bonferroni simple planned contrasts were used to determine the location of significant differences.

Table 13

Experimental Design

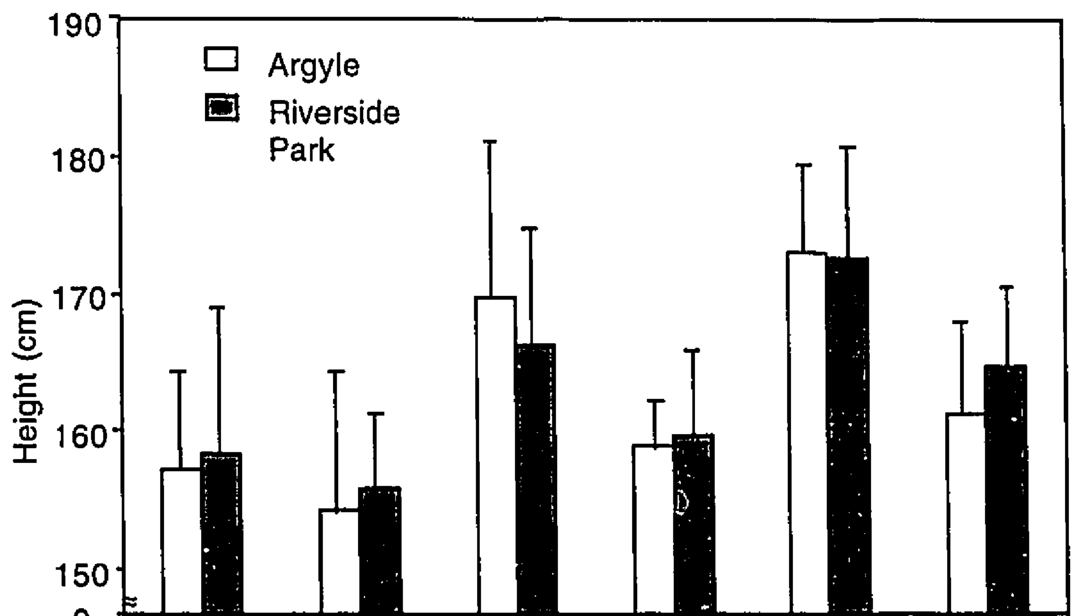
School Group	Argyle						Riverside Park					
Grade	7		8		10		7		8		10	
Gender	M	F	M	F	M	F	M	F	M	F	M	F
Sum of 5 Skinfolds												
Waist to Hip Ratio												
20m Shuttle Run												
Curl-Ups												
90° Push-Ups												
Back Saver Sit & Reach												
Trunk Lift												

Chapter 4

Results

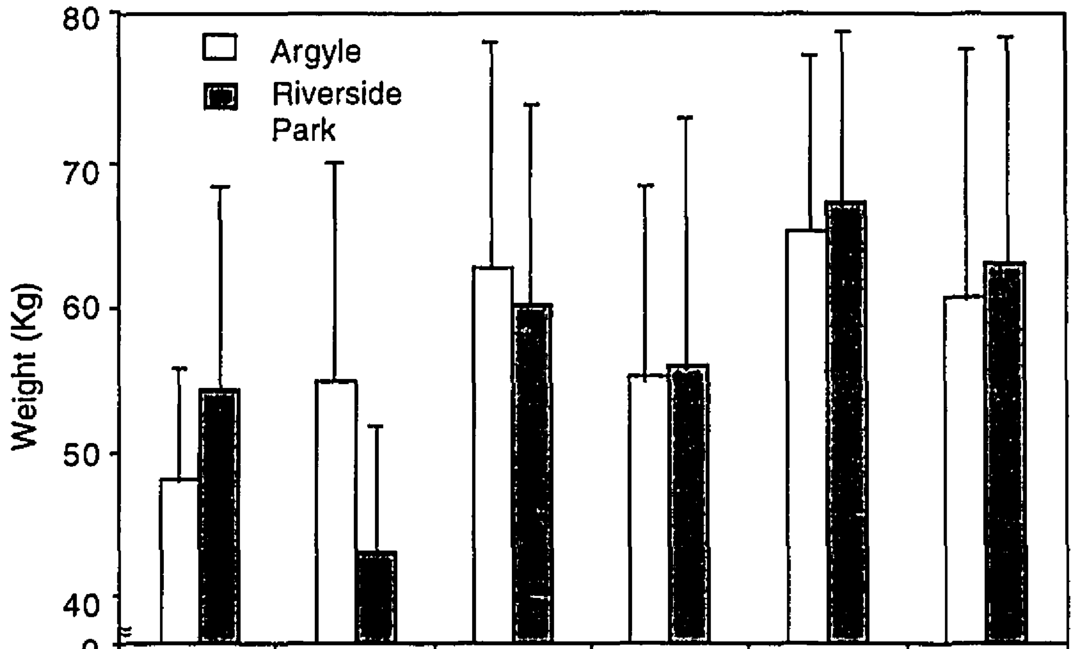
4.1 Descriptive Data for Health-Related Physical Fitness Performance Variables

Figure 3 through 11 present means and standard deviations for height, weight, body composition, aerobic capacity, muscular strength and endurance, and flexibility, respectively. These results are presented by test, school group, grade, and gender.



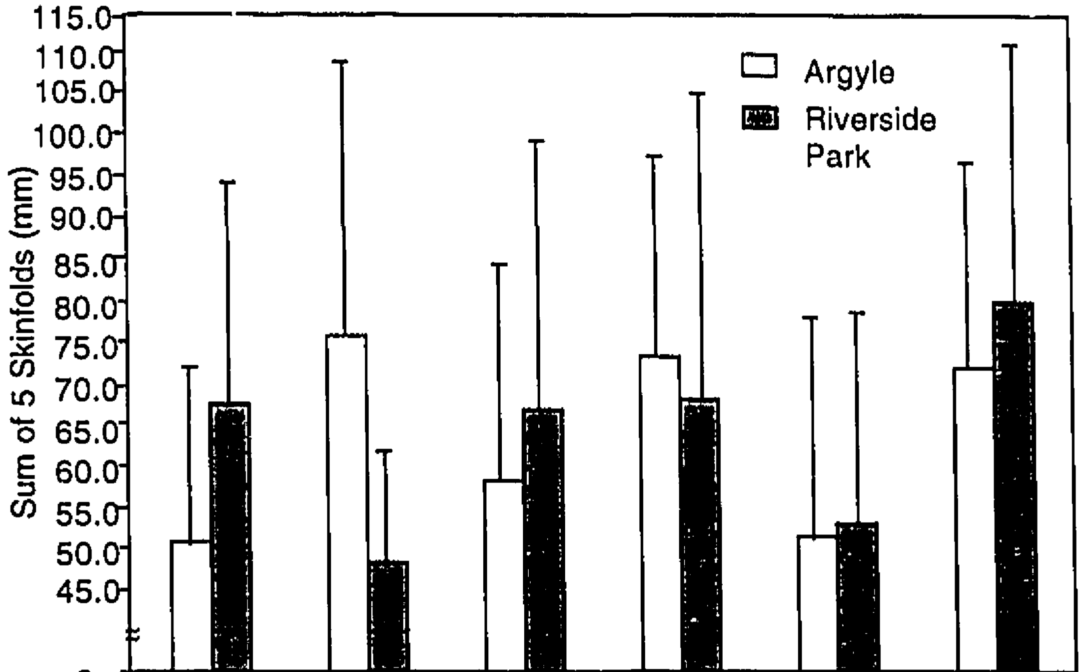
Gender	M	F	M	F	M	F
Argyle	157.0 ± 7.1	154.0 ± 9.4	169.4 ±11.4	158.9 ± 3.4	172.9 ± 6.2	161.2 ± 6.6
Riverside Park	158.5 ±10.4	155.8 ± 5.6	166.3 ± 8.4	159.8 ± 6.0	172.7 ± 7.6	164.3 ± 6.3
Grade	7		8		10	

Figure 3. Descriptive Data ($\bar{x}$ ± SD) for Height (cm)



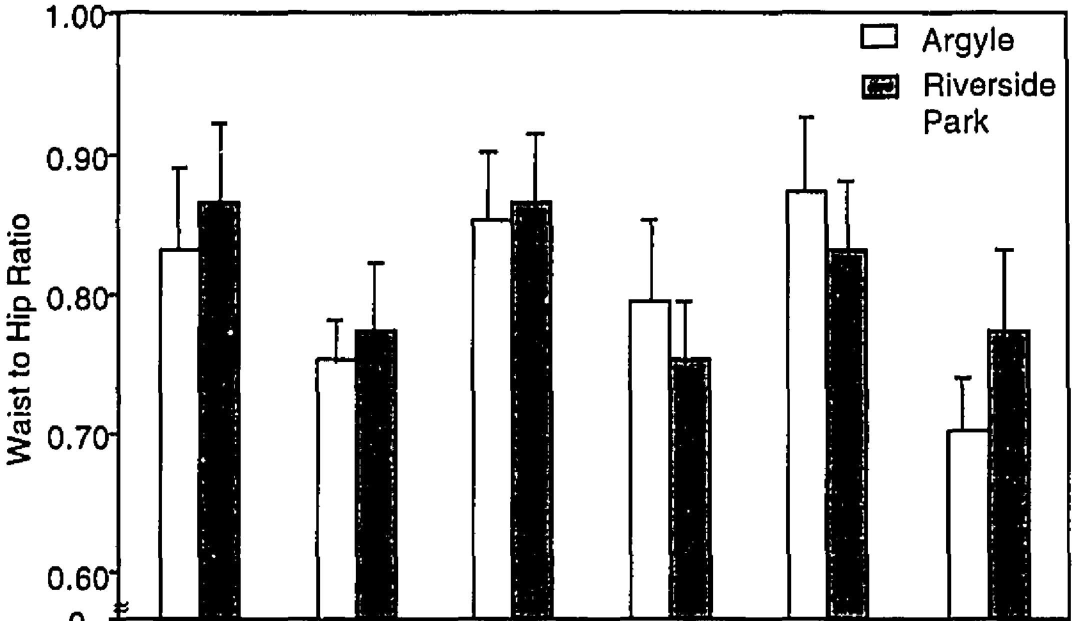
Gender	M	F	M	F	M	F
Argyle	47.7 ± 7.3	54.6 ± 14.3	62.6 ± 14.2	55.1 ± 12.0	65.1 ± 10.8	60.5 ± 15.6
Riverside Park	54.3 ±13.0	43.1 ± 8.3	60.2 ± 12.3	55.8 ± 15.8	67.3 ± 10.3	62.6 ± 14.4
Grade	7		8		10	

Figure 4. Descriptive Data ($\bar{x} \pm$ SD) for Weight (Kg)



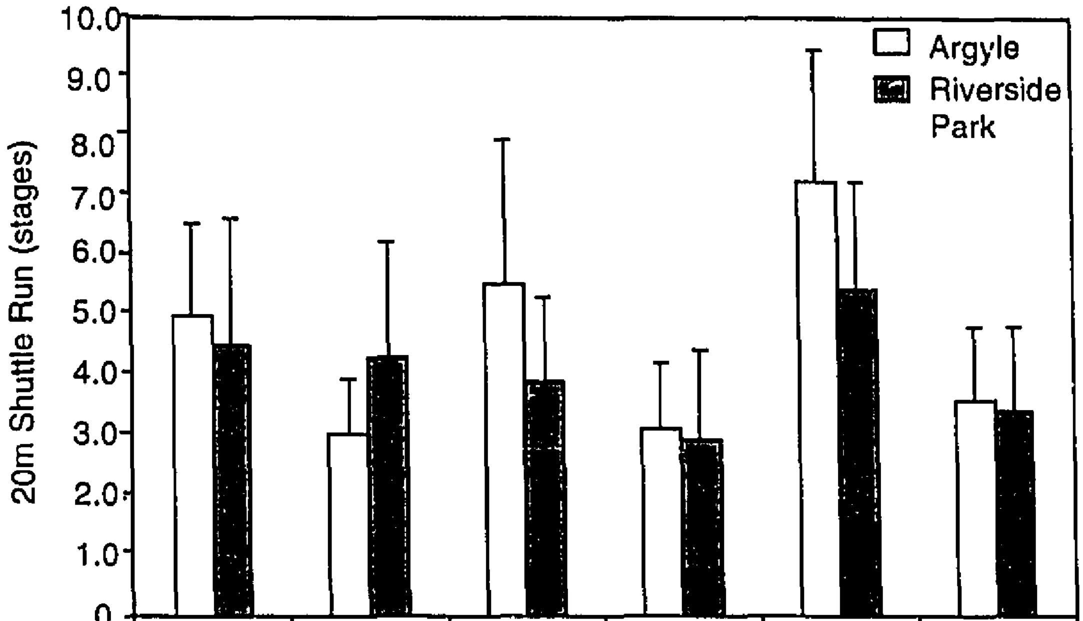
Gender	M	F	M	F	M	F
Argyle	50.8 + 21.3	75.4 ± 33.5	58.0 ± 26.0	73.3 ± 24.4	51.1 ± 27.3	71.7 ± 25.2
Riverside Park	66.3 ± 28.4	47.9 ± 14.4	66.6 ± 32.3	68.5 ± 36.8	53.6 ± 25.3	80.2 ± 31.0
Grade	7		8		10	

Figure 5. Descriptive Data ($\bar{x}$± SD) for Sum of 5 Skinfolds (mm)



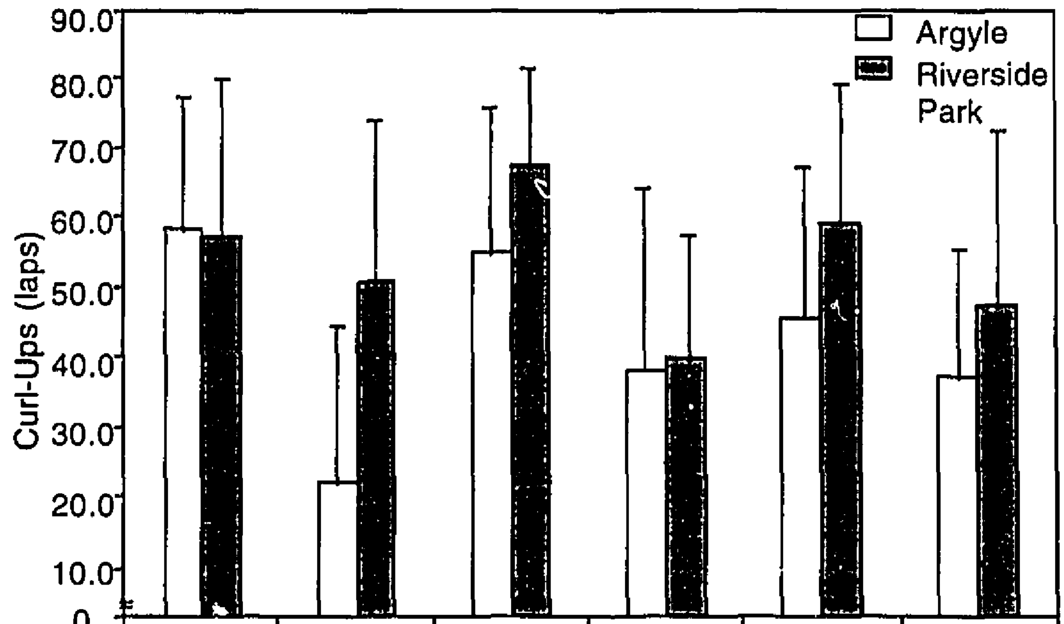
Gender	M	F	M	F	M	F
Argyle	0.83 ± 0.06	0.75 ± 0.03	0.85 ± 0.05	0.79 ± 0.06	0.87 ± 0.04	0.70 ± 0.04
Riverside Park	0.86 ± 0.06	0.77 ± 0.05	0.86 ± 0.05	0.75 ± 0.04	0.83 ± 0.05	0.77 ± 0.06
Grade	7		8		10	

Figure 6. Descriptive Data ($\bar{x}$ ±SD) for Waist to Hip Ratio



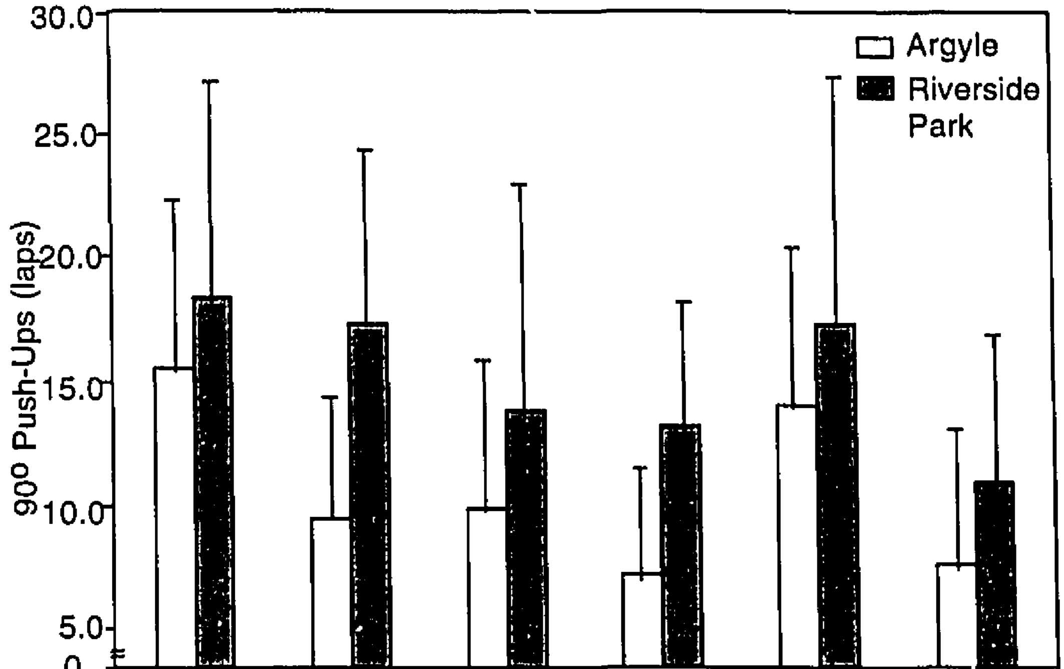
Gender	M	F	M	F	M	F
Argyle	5.0 ± 1.5	3.0 ± 1.2	5.5 ± 2.4	3.1 ± 1.1	7.2 ± 2.2	3.6 ± 1.2
Riverside Park	4.5 ± 2.1	4.3 ± 1.9	3.9 ± 1.7	2.9 ± 1.5	5.4 ± 1.8	3.3 ± 1.5
Grade	7		8		10	

Figure 7. Descriptive Data ($\bar{x}\pm SD$) for 20 m Shuttle Run (stages)



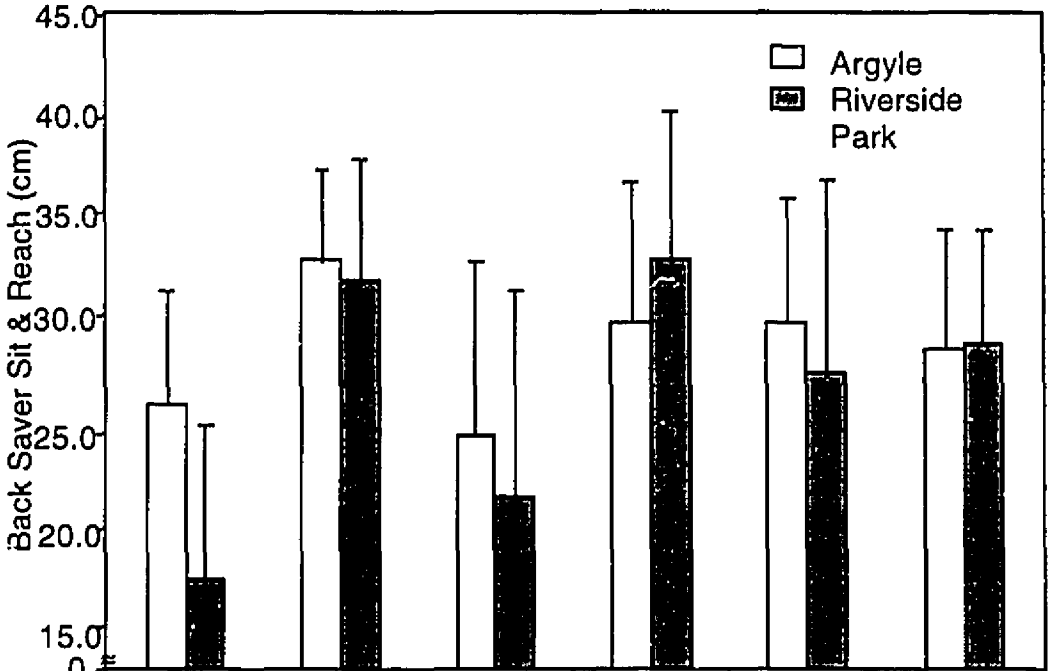
Gender	M	F	M	F	M	F
Argyle	58.2 ± 19.8	23.0 ±22.0	54.8 ± 20.7	38.0 ± 26.8	46.4 ± 21.4	37.8 ± 18.7
Riverside Park	58.1 ± 22.8	50.9 ±23.6	67.6 ± 14.1	40.0 ±18.1	59.5 ± 19.3	47.8 ± 24.3
Grade	7		8		10	

Figure 8. Descriptive Data ($\bar{x} \pm$ SD) for Curl-Ups (laps)



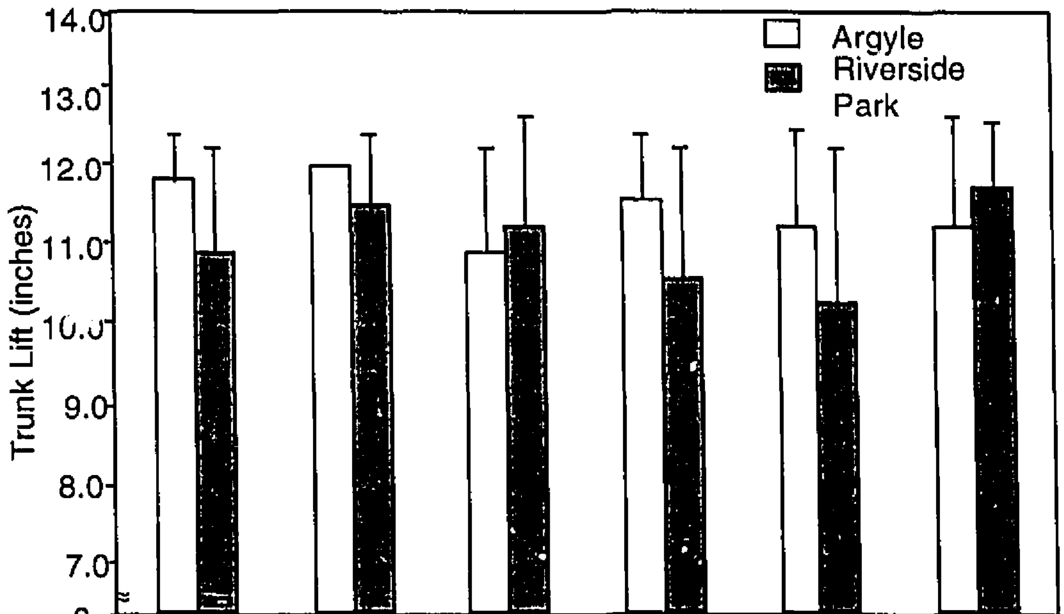
Gender	M	F	M	F	M	F
Argyle	15.2 ± 7.2	9.4 ± 5.0	9.9 ± 6.0	7.1 ± 4.4	13.9± 6.5	7.5 ± 5.8
Riverside Park	18.5 ± 8.7	17.3 ± 7.0	13.7 ± 9.2	13.0 ± 5.1	17.3± 10.1	10.8 ± 6.1
Grade	7		8		10	

Figure 9. Descriptive Data ($\bar{x}$± SD) for 90° Push-Ups (laps)



Gender	M	F	M	F	M	F
Argyle	25.6 ± 5.8	32.7 ± 4.7	24.3 ± 8.6	29.6 ± 6.9	29.7 ± 6.0	28.4 ± 5.8
Riverside Park	17.1 ± 7.7	31.8 ± 6.1	21.3 ± 10.1	32.8 ± 7.2	27.2 ± 9.5	28.6 ± 5.8
Grade	7		8		10	

Figure 10. Descriptive Data ($\bar{x} \pm SD$) for Back Saver Sit & Reach (cm)



Gender	M	F	M	F	M	F
Argyle	11.3 ± 0.6	12.0 ± 0.0	10.9 ± 1.4	11.6 ± 0.8	11.2 ± 1.3	11.3 ± 1.4
Riverside Park	10.9 ± 1.3	11.5 ± 0.9	11.3 ± 1.4	10.6 ± 1.6	10.3 ± 1.9	11.7 ± 0.9
Grade	7		8		10	

Figure 11. Descriptive Data ($\bar{x} \pm$ SD) for Trunk Lift (inches)

4.2 MANOVA Results

Hypothesis 1 states, "in comparison to students from Riverside Park Academy, students from Argyle Academy will demonstrate significantly better health-related physical fitness". The Multivariate and univariate test results for the main effect of school group are presented in Table 14.

Table 14

Multivariate and Univariate Test Results for The Effect of School Group

Source	df	F	P
Multivariate Analysis (Wilks' Lambda)	7, 180	8.975	0.000
Univariate Analysis			
Sum of 5 Skinfolds	1, 186	0.013	0.909
Waist to Hip Ratio	1, 186	1.312	0.254
20 m Shuttle Run	1, 186	4.305	0.039
Curl-Ups	1, 186	11.688	0.001
90^{o} Push-Ups	1, 186	19.282	0.000
Back Saver Sit & Reach	1, 186	3.303	0.071
Trunk Lift	1, 186	5.154	0.024

Results indicated a significant multivariate school group effect (F[7, 180]=8.975, P=0.000). Univariate analysis revealed that there was a significant school group effect on the 20 m shuttle run (F[1, 186]=4.305, P=0.039), trunk lift (F[1, 186]=5.154, P=0.024), curl-ups (F[1,186]=11.688, P=0.001), and 90^{o} push-ups (F[1, 186]=19.282, P=0.000).

Students from Argyle Academy performed significantly better in the 20 m shuttle run and trunk lift, while those from Riverside Park Academy performed significantly better in the curl-ups and 90^{o} push-ups.

There was no significant difference existing in the sum of five skinfolds, waist to hip ratio, and back saver sit and reach between the two schools.

Hypothesis 2 states, "there will be a significant interaction between school groups (Argyle and Riverside Park Academies) and genders for each health-related physical fitness performance variable". The multivariate and univariate test results for the interaction effect of school group by gender appear in Table 15.

Table 15

Multivariate and Univariate Test Results for The Interaction Effect of School Group and Gender

Source	df	F	P
Multivariate Analysis (Wilks' Lambda)	7, 180	3.273	0.003
Univariate Analysis			
Sum of 5 Skinfolds	1, 186	4.036	0.046
Waist to Hip Ratio	1, 186	2.358	0.126
20 m Shuttle Run	1, 186	9.032	0.003
Curl-Ups	1, 186	0.552	0.458
90^{o} Push-Ups	1, 186	1.072	0.302
Back Saver Sit & Reach	1, 186	6.710	0.010
Trunk Lift	1, 186	0.074	0.786

MANOVA results indicated a significant multivariate interaction effect of school group by gender (F[7, 180]=3.273, P=0.003). Univariate analysis indicated that there was a significant interaction effect of school group by gender on the sum of 5 skinfolds (F[1, 186]=4.036, P=0.046), 20 m shuttle run (F[1, 186]=9.032, P=0.003), and back saver sit and reach (F[1, 186]=6.710, P=0.010). There was no significant difference in the sum of 5 skinfolds for either girls or boys between Argyle and Riverside Park Academies.

Boys from Argyle Academy performed significantly better in the 20 m shuttle run than boys from Riverside Park Academy. There was no significant difference for girls between the two schools.

In the back saver sit and reach boys from Argyle Academy exhibited significantly better flexibility than boys from Riverside Park Academy. There was no significant difference for girls between the two schools.

Hypothesis 3 states, "there will be a significant interaction between school groups and grades (grade 7, 8, and 10) for each health-related physical fitness performance variable. The multivariate and univariate test results for the interaction effect of school group by grade are presented in Table 16.

Table 16

Multivariate and Univariate Test Results for The Interaction Effect of School Group by Grade

Source	df	F	P
Multivariate Analysis (Wilks' Lambda)	14, 360	1.454	0.126
Univariate Analysis			
Sum of 5 Skinfolds	2, 186	0.703	0.496
Waist to Hip Ratio	2, 186	2.565	0.080
20 m Shuttle Run	2, 186	3.019	0.051
Curl-Ups	2, 186	0.300	0.741
90° Push-Ups	2, 186	0.486	0.616
Back Saver Sit & Reach	2, 186	1.614	0.202
Trunk Lift	2, 186	0.637	0.530

MANOVA results indicated that there was no significant multivariate interaction effect of school group by grade (F[14, 360]=1.454, P=0.126). Univariate analysis also indicated no health-related physical fitness performance variable was significant at 0.05 level.

Hypothesis 4 states, "there will be a significant interaction among school groups, genders, and grades for each health-related physical fitness performance variable". The multivariate and univariate test results for the interaction among school group, gender, and grade appear in Table 17.

Table 17

Multivariate and Univariate Test Results for The Interaction Effect Among School Group, Gender, and Grade

Source	df	F	P
Multivariate Analysis (Wilks' Lambda)	14, 360	3.163	0.000
Univariate Analysis			
Sum of 5 Skinfolds	2, 186	3.217	0.042
Waist to Hip Ratio	2, 186	13.001	0.000
20 m Shuttle Run	2, 186	0.088	0.916
Curl-Ups	2, 186	3.045	0.050
90^{o} Push-Ups	2, 186	0.423	0.656
Back Saver Sit & Reach	2, 186	0.603	0.548
Trunk Lift	2, 186	4.553	0.012

MANOVA results indicated a significant multivariate interaction effect among school group, gender, and grade (F[14, 360]=3.163, P=0.000). Univariate analysis revealed that there was a significant interaction effect among school group, gender, and grade on the sum of 5 skinfolds (F[2, 186]=3.217, P=0.042), waist to hip ratio (F[2, 186]=13.001, P=0.000), trunk lift (F[2, 186]=4.553, P=0.012), and curl-up (F[2, 186]=3.045, P=0.050). For the sum of five skinfolds and trunk lift, there was no significant difference in either girls or boys in grade seven, eight, and ten between the two schools.

For the waist to hip ratio, boys in grade ten from Argyle Academy possessed significantly higher ratio than boys in grade ten from Riverside Park Academy. For the girls in grade ten, however, those from Riverside Park Academy exhibited significantly higher ratio than Argyle Academy. There was no significant difference among girls or boys in grade seven and eight between the two schools.

In grade seven, girls from Riverside Park Academy performed significantly better in curl-ups than girls from Argyle Academy. However, there was no significant difference in the boys in grade seven, and either boys or girls in grade eight and ten between the two schools.

Chapter 5

Discussion

In this chapter, the findings pertaining to the health-related physical fitness of high school students in two physical education programs are discussed in relation to previous studies that used high school students as subjects. The effects of QDPE at Argyle Academy in this study resulted in variation when compared to the regular 3-days-a-week physical education program at Riverside Park Academy. Recommendations for determining fitness levels of high school students are also provided.

Limited research has been carried out to examine the effects of QDPE on health-related physical fitness of high school students. To date the findings reported have not been consistent (Dragicevick et al., 1987; Johnson, 1969; Kemper et al., 1976; Phillipp et al., 1991). In this study results showed a significant difference in students' health-related physical fitness between Argyle Academy and Riverside Park Academy. The significant difference, however, resulted from the 20 m shuttle run, trunk lift, curl-ups, and 90^{o} push-ups, but not from the sum of five skinfolds, waist to hip ratio, and back saver sit and reach. Students from Argyle Academy performed significantly better in the 20 m shuttle run and trunk lift, while those from Riverside Park Academy performed significantly better in the curl-ups and 90^{o} push-ups. Therefore, it would seem that the QDPE at Argyle Academy showed its effects on only selected fitness components, aerobic capacity and trunk flexibility in this study.

5.1 Aerobic Capacity

Aerobic capacity was measured using the 20 m shuttle run (Leger et al., 1988; Cooper Institute for Aerobic Research, 1992) in this study. It appears that the students in this study generally possessed low aerobic capacity compared to norm- and criterion-referenced standards. Since the 20 m shuttle run has been reported to have high reliability (see Table 1) and moderate validity (See Table 2) as a field test, the test results are likely to show the actual aerobic capacity of the students.

Compared to the norms (Massicotte, 1990), the boys of Argyle Academy in grade seven, eight, and ten ranked in the 35th percentile, 33rd percentile, and 36th percentile, respectively, while the boys of Riverside Park Academy ranked in the 23rd percentile, 10th percentile, and 12th percentile, respectively. In comparison, the Argyle Academy girls ranked in the 35th percentile, 25th percentile, and 13th percentile, respectively, while the Riverside Park girls ranked in the 56th percentile, 10th percentile, and 12th percentile.

Compared to the criterion-referenced standards (Cooper Institute for Aerobic Research, 1992), the Argyle Academy boys in all the grades passed this test, while in the Riverside Park Academy the boys in grade seven passed this test and the boys in grade eight and ten did not reach the predetermined criterion. The standards of this test for the boys aged 13, 14, and 16 years were 35 to 74 laps, 41 to 80 laps, and 52 to 90 laps, respectively. In comparison, the girls in every grade of the two schools passed this test except those in grade ten of the Riverside Park Academy. The standards for the girls aged 13, 14, and 16 years were 15 to 42 laps, 18 to 44 laps, and 28 to 56 laps, respectively.

It was found that students from Argyle Academy performed significantly better than those from Riverside Park Academy on this test. In other studies (Dragicevick et al., 1987; Phillipp et al., 1991) effects of QDPE on aerobic capacity were also found when aerobic capacity was measured using distance running tests. Dragicevick et al. (1987) found an initial significant improvement in the 500 m run at the pilot school in comparison with the control school. However, the significant improvement in aerobic capacity diminished with subsequent poor compliance with the program and summer holidays. Phillipp et al. (1991) reported that a significant difference between experimental and control groups was only seen in the performance in 1.5 mile run after 2 consecutive summer sessions of QDPE. In their study, since the QDPE took place only during the 6 weeks of a summer session between the 2 summer sessions, it could be that the effects of QDPE have been affected by the activity levels of students during the rest of school year.

In this study, it was found that boys from Argyle Academy performed significantly better on the 20 m shuttle run than those from Riverside Park Academy. There was no significant difference for girls between the two schools. Therefore, the difference previously mentioned between the two schools in the 20 m shuttle run resulted from the difference in boys between the two schools and not in the girls. There was a 1.3 stage difference in boys between the two schools, while there was only a 0.2 stage difference in girls between the two schools. In the study by Dragicevick et al. (1987), a significant difference was seen both in boys and girls for the 500 m run performance when compared to those of a control group.

Greater fat mass of girls in Argyle Academy might have been hindering the effects of QDPE on aerobic capacity for girls. When the values of sum of five skinfolds are compared in girls between the two schools in this study, the Argyle Academy girls possessed thicker folds (7.9 mm more). In the boys between the two schools, for instance, the Argyle Academy boys possessed thinner folds (8.8 mm less) and performed significantly better on the 20 m shuttle run when compared with the Riverside Park Academy boys. However, it is likely that some other factors such as activity levels of the students both inside and outside of physical education classes contribute to differences in aerobic capacity.

5.2 Body Composition

Body composition was measured using the sum of five skinfolds (C.S.T.F., 1986) and waist to hip ratio (C.S.T.F., 1986). There were no significant differences between the two schools for body composition in this study. This is consistent with the results from other studies (Dragicevick et al., 1987; Kemper et al., 1976; Phillipp et al., 1991). Johnson (1969) is the only study that found a significant difference in triceps skinfold measures of the eighth grade boys between 5-day and 2- or 3-day-a-week physical education classes. In this study, although it was not statistically significant, students of Argyle Academy exhibited slightly better scores in both the sum of five skinfolds (0.5 mm less) and waist to hip ratio (0.008 less) compared to those of Riverside Park Academy.

The results of the sum of five skinfolds can be compared with the norms of

Canada Fitness Survey (1985), where the greater sum of skinfolds subjects posses, the higher percentile they ranks in. Compared to these norms, the Argyle Academy boys in grade seven, eight, and ten ranked in the 75th percentile, 84th percentile, and 78th percentile, respectively, while the Riverside Park Academy boys ranked in the 87th percentile, 88th percentile, and 82nd percentile, respectively. In comparison, the Argyle Academy girls ranked in the 82nd percentile, 74th percentile, and 76th percentile, whereas the Riverside Park Academy girls ranked in the 40th percentile, 66th percentile, and 85th percentile, respectively. Consequently, the students in this study generally possessed greater amount of body fat compared to the population of the norms. However, the norm-referenced standards have been criticized that they represents the current level of performance of the "normal" population instead of the absolute, desired level of the criterion attribute or behaviour that is consistent with good health (Cureton & Warren, 1990). This seems to be often the case that these standards are difficult to achieve and classify many children and youth as unfit (Blair, 1992; Corbin & Pangrazi, 1992). Therefore, one should avoid judging the fitness levels solely according to the norms. There is no criterion-referenced standards for the sum of five skinfolds available.

For the waist to hip ratio, it was found in the analysis of the interaction effect of school by gender by grade that boys in grade ten from Argyle Academy possessed significantly higher ratio than boys in grade ten from Riverside Park Academy. For the girls in the grade ten, however, those from Riverside Park Academy exhibited significantly higher ratio than those in Argyle Academy. Compared to the norms of C.S.T.F. (1986), boys in grade ten of Argyle Academy and Riverside Park Academy ranked in the 20th percentile and 50th percentile, respectively. In comparison, the grade tenth girls ranked in the 78th percentile and 30th percentile, respectively. Consequently, the difference is more in girls than boys in grade ten. There was no significant difference among girls or boys in grade seven and eight between the two schools. Norms are not available for youth aged under 15 years.

Wide ranges in amounts of subcutaneous fat exist in youth of all ages (Canada Fitness Survey, 1985). Adolescence is a time when growth takes part and "children vary considerably in growth and developmental characteristics during their school years" (Baumgartner & Jackson, 1991, p. 276). Therefore, it is difficult to see effects of QDPE programs on adolescents' body composition given the brief study period when they are not stable in their body composition. Canada Fitness Survey (1985) recommends monitoring the changes in skinfold thickness over the years. Since this study is cross sectional, it would be interesting to see if longitudinal studies show any effects of QDPE on body composition of adolescents.

5.3 Muscular Strength and Endurance

In order to assess muscular strength and endurance of abdominal and both upper arm and shoulder girdle in this study, curl-ups and 90° push-ups (Cooper Institute for Aerobic Research, 1992) were used, respectively. Compared to the norm referenced standard of curl-ups (Massicotte, 1990), the boys in grade seven, eight, and ten of Argyle Academy ranked in the 79th percentile, 70th percentile, and 59th percentile, respectively. In comparison, the boys of Riverside Park Academy ranked in the 79th percentile, 80th percentile, and 74th percentile, respectively. For the girls, the Argyle Academy girls ranked in the 35th percentile, 68th percentile, and 67th percentile, while the Riverside Park Academy girls ranked in the 87th percentile, 70th percentile, and 78th percentile, respectively. In the 90° push-ups, compared to the norms (Massicotte, 1990), the Argyle Academy boys in grade seven, eight, and ten ranked in the 31st percentile, 10th percentile, and 12th percentile, respectively. In comparison, the Riverside Park Academy boys ranked in the 41st percentile, 16th percentile, and 15th percentile, respectively. For the girls, the Argyle Academy girls ranked in the 42nd percentile, 30th percentile, and 24th percentile, whereas the Riverside Park Academy girls ranked in the 73rd percentile, 55th percentile, and 37th percentile, respectively.

Compared to the criterion-referenced standards of curl-ups (Cooper Institute for Aerobic Research, 1992), the boys in any grade of the two schools passed the test by exceeding the upper end of the standards when the scores required for the boys aged 13, 14, and 16 years were 21 to 40 laps, 24 to 45 laps, 24 to 47 laps, respectively. In comparison, the girls in all the grades of the two schools also passed the test. They exceeded the upper end of the standards except those in grade seven of Argyle Academy when the scores required for the girls aged 13, 14, and 16 years were 18 to 32 laps, 18 to 32 laps, and 18 to 35 laps, respectively. For the 90° push-ups, however, the boys in grade seven of the two schools and grade eight of Riverside Park Academy passed the test but the rest of boys failed when compared to the criterion-referenced standards (Cooper Institute for Aerobic Research, 1992). The required scores for the boys aged 13, 14, and 16 years were 12 to 15 laps, 14 to 30 laps, and 18 to 35 laps, respectively. The boys in grade seven of the two school exceeded the required score, while the grade eight Riverside Park Academy boys scored the lower end of the standard. In comparison, the girls in all the grades of the two schools passed the test when the standards were 7 to 15 laps for the girls aged 13, 14, and 16 years. The girls in grade seven of Riverside Park Academy exceeded the upper end of the standards, while the rest of the girls scored between the two ends of the standards. Therefore, the students in both schools possessed the optimal muscular strength and endurance fitness.

It is surprising that the students from Riverside Park Academy, which has a regular 3-day-a-week physical education program, performed significantly better in the curl-ups and 90° push-ups. In any grade and gender, students from Riverside Park Academy scored better than those from Argyle Academy in these tests. However, students receiving QDPE in the study by Dragicevick et al. (1987) as well as those who received 5-day-a-week physical education program in the study by Johnson (1969) and Kemper et al., (1976) performed significantly better than those in a control group in muscular strength and endurance measures such as push-ups, pull-ups, handgrip, and curl-ups. During the data collection period of this study students in Riverside Park Academy were engaging in gymnastics. This might have positively affected the muscular strength and endurance of students in the school. As well, Riverside Park Academy had been using the curl-ups and 90° push-ups in the test battery. Therefore, the students of Riverside Park Academy had been tested on these tests periodically and encouraged to achieve high scores. Consequently, it is possible that they were used to and already trained for these tests. Furthermore, there were anecdotal comments that students of Argyle Academy were frustrated because they could not join their regular physical education classes while they were participating in the tests. During the data collection period when not being tested, students were playing volleyball games in their classes. It was observed by testers that some students were eager to finish testing fast and go back to their volleyball games. This might have negatively affected their motivation in their testing.

In further analysis, an interaction effect of school group by gender by grade was noted in curl-ups but not in 90° push-ups. In grade seven, girls from Riverside Park Academy performed significantly better (50.9 laps) than those from Argyle Academy (23 laps) in curl-ups. The differences in curl-ups of the boys in grade seven and either boys or girls in grade eight and ten were not statistically significant between the two schools. There were girls in grade seven of Riverside Park Academy who exhibited outstanding performance in these muscular strength and endurance test items. According to one of physical education teachers, these girls played in a intramural sport program and possessed higher fitness than other students. This might help explain the result.

5.4 Muscular Flexibility

In this study trunk flexibility was measured using the trunk lift (Cooper Institute for Aerobic Research, 1992). When compared to the criterion-referenced standards of this test (Cooper Institute for Aerobic Research, 1992), both genders of any grade in the two schools met the standard of 9 to 12 inches. The effects of QDPE on trunk flexibility was not examined in other studies with high school students. Results of this study showed that students from Argyle Academy exhibited significantly better trunk flexibility than those from Riverside Park Academy. It is interesting to note that there was no significant difference between the two schools in the back saver sit and reach test, which is considered to be a test of hamstrings and lower back flexibility, although students of Argyle Academy performed better in this test (2 cm more) compared to those of Riverside Park Academy.

For the back saver sit and reach, it was found in the analysis of the interaction effect of school group by gender that boys from Argyle Academy exhibited significantly better flexibility (4.7 cm more) than those from Riverside Park Academy. The girls of Riverside Park Academy scored 0.8 cm higher than those of Argyle Academy, although the difference was not of statistical significance. Dragicevick et al. (1987) reported a significant difference in the sit and reach performance of both boys and girls between experimental and control groups after seven months of QDPE. They also reported that after that point, where the summer holiday started, this significant difference was lost. Using the 50th percentile scores from the Canada Fitness Survey (1985), trunk flexion (sit and reach) scores show an increase in adolescent boys until the age of 17 years, while adolescent girls show leveling off after 15 years of age. Girls consistently outscore boys and the difference between the genders are ranging from 3 to 7 cm between the age of 12 and 18 years. Much of the change with age, that is evident between the genders, may be related to growth itself (Canada Fitness Survey, 1985). Trunk flexion is lowest in both boys and girls just prior to the growth spurt (Canada Fitness Survey, 1985). Compared to the criterion-referenced standards of FITNESSGRAM (Cooper Institute for Aerobic Research, 1992) for the back saver sit and reach, the boys of Riverside Park Academy passed the required score of 20.32 cm (8 inches) by 1.6 cm, while those of Argyle Academy exceeded the required score by 6.2 cm. In comparison, for the girls both schools possessed high scores (30.2 cm and 31.0 cm for Argyle Academy and Riverside Park Academy respectively), which were similar to the criterion-referenced scores (25.4 cm and 30.5 cm for those aged under 14 years and above 15 years respectively). It could be possible that girls in Argyle Academy showed a ceiling effect: since they already possessed high sit and reach performance, the QDPE could not improve their performance. The effects of QDPE program on the muscular flexibility warrants further investigation.

5.5 Interaction Effect of School Group by Grade

There was no significant interaction between school groups and grades for each health-related physical fitness performance variable. Consequently, the QDPE program in this study seems to have affected the health-related physical fitness of students in the same manner regardless the grade difference. In the study of Werner and Durham (1988) with upper elementary school children, an interaction effect of grade (4, 5, & 6) by treatment (daily vs. 2-days-a-week physical education) was found on the endurance run and sit-ups. They speculated that the effect of classroom teacher, the motivation of the students, the stability of physiological body changes versus the onset of puberty all contributed to the grade by treatment interaction effect.

5.6 Health-Related Physical Fitness Test Items

In this study selection of optimal health-related physical fitness tests was another important issue. Of the seven tests used, some were well-established already in terms of reliability, validity, and procedures, while others seemed to require more research (see Chapter 2). The 20 m shuttle run, sum of five skinfolds, waist to hip ratio, and trunk lift were generally straightforward and easy to administer, although skinfold measurements required the investigator to gain adequate technique and experiences before data collection. For curl-ups and 90^{o} push-ups, however, there were some problems observed.

The 20 m shuttle run was found to be practical for the use in schools as suggested by Leger and and Lambert (1982). Since it requires a relatively small space, it was not problematic to administer in the gymnasium. This makes it possible to carry out testing regardless the weather or season. Also, this test was straightforward to administer and carry out with a prerecorded tape which explained the procedure and controlled the running pace during testing. The 20 m shuttle run has been reported to be one of the most reliable and valid tests of aerobic capacity in children and adolescents (see Chapter 2). Most importantly, students generally liked this test and appeared to be motivated during testing. It may be an advantage compared to other distance running tests. Therefore, this test is recommended for future use in schools.

The use of sum of five skinfolds in the school settings is controversial. It has been shown that excess body fat of children and adolescents is associated with risk factors for cardiovascular disease (Williams et al., 1992). Childhood obesity is a serious health problem in North America, and skinfold tests are often included in health-related physical fitness tests (Baumgartner & Jackson, 1991). The skinfold measurements seem the best method available in order to assess body composition in field testing today (Baumgartner & Jackson, 1991). It is feasible in the field compared to underwater weighing, although underwater weighing is more accurate. Compared to Body Mass Index (BMI), skinfold measurements are more accurate and give more direct information, although BMI is much easier to administer. It has been reported that testers can be trained in order to take skinfold measurements accurately and consistently (Morrow et al., 1986; Opplinger et al., 1992; Shaw, 1986). In this study, the skinfold measurements were taken by a single tester who had spent time in training before the measurements. The skinfold caliper used was accurate and belonged to the McGill University, since there was no reasonable caliper available in the high schools. Although it is an inexpensive approach compared to other techniques such as underwater weighing, skinfold calipers are expensive and not affordable in many schools. Also training for skinfold measurements is time consuming and requires following a standardized procedure. Few teachers in schools are adequately trained. Furthermore, skinfold measurements is very time consuming, and it does not seem practical to administer during physical education classes. In addition, this test seems unpopular among students, especially among girls. Therefore, the skinfold measurements deserve further investigation in schools.

Waist to hip ratio was very straightforward to administer and did not require much time. Since it is important to assess body fat distribution in addition to total body fat, it is recommended to include waist to hip ratio in body composition assessment (C.S.T.F., 1986). However, this test needs test standards for children aged under 15 years.

It seemed difficult for some students to execute curl-ups instead of sit-ups. In some cases it was caused by the differences in body size: for students with long arms, it required less effort for them to reach the second line on the mat than those with shorter arms when curling up. Moreover, sliding of the body was often observed. As a result, students' finger tips were not placed correctly at the first line on the mat. Consequently, students were asked to interrupt their performance and correct their position. This caused frustration in individuals in their attempts to execute the curl-ups properly and disturbed their concentration and motivation. However, this test has been suggested by researchers because it requires greater involvement of the abdominal and is safer for the back compared to other tests (Flint, 1965; Halpern & Bleck, 1979; Ricci et al., 1981; Robertson & Magnusdottir, 1987; Plowman & Corbin, 1994). Therefore, this test is recommended for future use in high schools. However, this test warrants further inquiry.

In the 90° push-ups, the 90° angle at the elbow was not easily maintained. As well, the pace of the push-ups recommended (1 push-up per 3 seconds) was thought to be too slow by some students. These points were also reported in the study of McManis and Wuest (1994). Strict adherence to test procedures as well as practice before testing might help solve these problems, and this test is recommended.

In terms of the administration, the back saver sit and reach was straightforward. However, problems were recognized when students with either longer or shorter extremities were tested: those with longer legs as well as shorter arms could not reach the scale on the box as easily as other students when they bent forward. This limitation of the test has been discussed in Hoeger et al. (1990) and Hoeger and Hopkins (1992). Since the back saver sit and reach avoids overstretching of the low back and hyperextension of knee joints, it is recommended over the traditional sit and reach.

Trunk lift was straightforward and expedient to complete. Little information is available on reliability and validity for this test (see Chapter 2). Also, standards need to be developed based on sound research (see Chapter 3). Therefore, this test warrants further investigation to be included as part of a test battery.

5.7 Student Appreciation of Fitness Tests

Argyle Academy in this study has its own physical fitness test battery which includes body composition measurements. The Argyle test battery is composed of the following: 1) body mass index, 2) sum of three skinfolds (triceps, subscapular, & Iliac crest), 3) waist to hip ratio, 5) grip strength, 6) sit-ups, 7) push-ups, 8) sit and reach, 9) the Leger Boucher 20 m shuttle run, and 10) standing long jump. In comparison, Riverside Park Academy's test battery has four items and does not measure body composition and aerobic capacity. Their test battery includes 1) push-ups, 2) partial curl-ups, 3) sit and reach, and 4) standing long jump. There were differences noticed in the students' reaction towards fitness testing between the two schools: students of Argyle Academy seemed more accustomed to fitness testing than those of Riverside Park Academy. Some students from Argyle Academy showed negative reactions towards some test items such as skinfold measurements. As a result, some girls refused skinfold measurements, and some students did not seem motivated in some tests such as curl-ups and 90° push-ups. In those cases cautions were taken to secure students' maximal motivation by discussion with the students, testers, and physical educators and verbal encouragement during performance. However, the students were never forced to participate in the tests. In comparison, at times students at Riverside Park Academy had mixed reactions to some tests. The students experienced some test items for the first time; for example, the skinfold measurements and 20 m shuttle run. Since the tests were novel, they appeared to show less negative reactions compared to the students of Argyle Academy. Initially some students were tentative to have the skinfold caliper touch them, but many of them were curious about the new tests and eager to participate. Students of Riverside Park Academy appeared more motivated in their testing.

A comment by one of the teachers of Argyle Academy was "fitness tests are not popular things". Fitness tests should not be something negative. Appreciation of fitness tests has a very important impact on students' motivation and their future decisions about physical fitness (Corbin, 1986, 1995). "Students must move to higher order objectives such as establishing personal exercise programs, learning to test their own fitness, interpreting their own test results, and learning to solve their own fitness problems" (Corbin, 1987a, p. 51). Feedback for the test results must be given to students in a encouraging way so that their motivation to do physical activities will remain high (Corbin, 1987a; Whitehead & Corbin, 1991). Corbin (1995) has suggested that testing needs to be a positive experience for students in order to motivate them so that they will want to be physically active in their future. If testing is a negative experience, fitness tests should be replaced with more enjoyable physical activity.

This preliminary study has revealed additional information and questions regarding QDPE and fitness testing in high schools. These questions deserve further research.

Chapter 6

Summary, Conclusions, and Future Directions

6.1 Summary and Conclusions

The purpose of this study was to compare health-related physical fitness of secondary school students between two schools which have different physical education programs in terms of the emphasis on fitness as a component of the school curriculum. Argyle Academy has a school curriculum which emphasizes fitness and physical education and has fifty minute daily physical education classes in the seventh, eighth, and eleventh grades (6 days per 7 day cycle) and less in the ninth and tenth grade (3 days per 7 day cycle). In comparison, Riverside Park Academy does not have daily physical education classes in any grade (3 days per 7 day cycle). Subjects in the study were 102 students from Argyle Academy and 96 students from Riverside Park Academy. These subjects were selected as the students in the seventh, eighth, and tenth grade intact physical education classes in each school. Subjects were tested with the McGill test battery during the regular physical education classes over eight weeks in October, November, and December. The test battery consisted of height, weight, 20 m shuttle run, sum of five skinfolds, waist to hip ratio, curl-ups, 90^{o} push-ups, back saver sit and reach, and trunk lift.

The experimental design included three factors, school group, gender, and grade. Statistical differences were examined using a 2 x 2 x 3 factorial MANOVA, followed by univariate F tests. Following the significant F test results, Bonferroni simple planned contrasts were used to determine the location of significant differences.

MANOVA results indicated that there was a significant school effect. Subsequent univariate analysis revealed that 20 m shuttle run, curl-ups, 90^{o} push-ups, and trunk lift were significant at the 0.05 level. Students from Argyle Academy performed significantly better in the 20 m shuttle run and trunk lift, while those from Riverside Park Academy performed significantly better in the curl-ups and 90^{o} push ups.

There was a significant school by gender interaction effect for the following variables: sum of five skinfolds, 20 m shuttle run, and back saver sit and reach. Bonferroni contrasts indicated that the significant F test result for the sum of five skinfolds resulted from significant differences between different genders. For the 20 m shuttle run and back saver sit and reach, boys for Argyle Academy performed significantly better than those from Riverside Park Academy. There was no significant school by grade interaction effect for any of health-related physical fitness test item at the 0.05 level.

There was a significant school by gender by grade interaction effect for the following variables: sum of five skinfolds, waist to hip ratio, curl-ups, and trunk lift. Bonferroni contrasts revealed that the significant F test results for the sum of five skinfolds and trunk lift resulted from the significant differences among different genders and grades. For the waist to hip ratio, boys in grade ten from Argyle Academy possessed significantly higher ratio than those from Riverside Park Academy. For girls in grade ten, however, those from Riverside Park Academy exhibited significantly higher ratio than those from Argyle Academy. On the curl-ups, girls in grade seven from Riverside Park Academy performed significantly better than those from Argyle Academy.

From this study it is problematic to make broad based conclusions regarding fitness testing in high school physical education programs due to the small sample size. Within the delimitations and limitations of this study, the following conclusions seem justified:

1) Compared to students from Riverside Park Academy, those from Argyle Academy obtained significantly better scores for the 20 m shuttle run and trunk lift. Students from Riverside Park Academy obtained significantly better scores for the curl-ups and 90^{o} push-ups.

2) There was a significant school by gender interaction effect for the sum of five skinfolds, 20 m shuttle run, and back saver sit and reach.

3) There was no significant school by grade interaction effect for any of health-related physical fitness test item at the 0.05 level.

4) There was a significant school by gender by grade interaction effect for the sum of five skinfolds, waist to hip ratio, curl-ups, and trunk lift.

6.2 Future Directions

This study was conducted with the intention of serving as one of the preliminary studies on the effects of QDPE on health-related physical fitness of high school students. Investigation of the effects of QDPE in terms of the frequency of physical education was the major purpose of this study. Through this study, additional important aspects of the study of QDPE were noted for further investigation.

First, in order to determine health-related physical fitness of high school students, the following tests are recommended to be utilized: the 20 m shuttle run, sum of five skinfolds, waist to hip ratio, curl-ups, 90^{o} push-ups, and back saver sit and reach. However, the sum of five skinfolds, curl-ups, 90^{o} push-ups, and back saver sit and reach require further research. The waist to hip ratio needs test standards for children aged under 15 years. The trunk lift warrants further investigation to be included as part of a test battery.

Second, this study did not investigate several factors in physical education such as teacher instruction, content of classes, intensity of activities, exercise time (duration) in classes, and physical activities outside of classes. Further research is necessary to examine effects of these factors in addition to the frequency of physical education on health-related physical fitness in order to clarify the "quality" of QDPE program.

Third, studies of QDPE including this study are commonly cross sectional. In order to examine the longer-term effects of QDPE and the course of the effects on health-related physical fitness of high school students, there is a need of longitudinal studies of QDPE program.

Fourth, studies of QDPE have been often carried out on elementary school students. Therefore, further studies of QDPE with high school students seem warranted. In schools, especially in high schools, other subjects such as math, English, and science are considered to be more important than physical education (Bailey, 1973; Dahlgren, 1987; O'Sullivan et al., 1994; Verabioff, 1986). Physical education becomes an elective subject in the schools, and fitness levels of students begin to decline during the school years (Bailey, 1973; Dahlgren, 1987). Dahlgren (1987) proposed that no school subject could be as important as physical education. She pointed out the reasons for this comment as 1) constantly rising health care costs, 2) constantly increasing amounts of leisure time available, 3) positive effects of regular physical activity, and 4) increased quality of life from regular physical activities. Physical education in high schools awaits future research in order to establish credibility as an essential subject in the school curriculum.

Last, as suggested by researchers (Bailey, 1985; Corbin, 1987a, 1991; Goldfine & Nahas, 1993; Haywood, 1991; Phillipp et al., 1991; Rider et al., 1986; Slava et al., 1984; Smith & Cestaro, 1992), conceptual physical education courses can teach students about important concepts such as physical activity, fitness, and healthy lifestyles. It has been also suggested that adolescence is the most appropriate time to introduce these conceptual courses (Goldfine & Nahas, 1993). Therefore, it would be meaningful to implement conceptual courses into these QDPE programs in each grade and investigate their effects.

References

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

American Health and Fitness Foundation. (1986). *Fit Youth Today*. Austin, TX: Author.

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

Astrand, P.O. (1976). The child in Sport and physical activity-physiology. In Albinson, J.G. & Andrew, G.M. (Eds.), *Child in sport and physical activity*, (pp.19-33). Baltimore: University Park Press.

Astrand, P.O., & Rodahl, K. (1986). *Textbook of work physiology* (3rd ed.), New York: McGraw-Hill.

Bailey, D.A. (1973). Exercise, Fitness and Physical Education for the Growing Child- A Concern. *Canadian Journal of Public Health, 64*, 421-430.

Bailey, K. (1985). Health related fitness course. *British Journal of Physical Education*, 16(6), 218-220.

Ball, T.E. (1993). The predictability of muscular strength and endurance from calisthenics. *Research Quarterly for Exercise and Sport, 64*(Suppl.), A-39. (Abstract).

Bar-Or, O. (1994). Childhood and adolescent physical activity and fitness and adult risk profile. In Bouchard, C., Shephard, R.J., Stephens, T., & Associates (Eds.), *Physical activity, fitness, and health: International proceedings and consensus statement*, (pp.931-942). Champaign, IL: Human Kinetics.

Battie, M.C., Bigos, S.J., Fisher, L.D., Spengler, D.M., Hansson, T.H., Nachemson, A.L., & Wortley, M.D. (1990). The role of spinal flexibility in back pain complaints within industry: A prospective study. *Spine, 15*(8), 768-773.

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

Berger, R.A., & Medlin, R.L. (1969). Evaluation of Berger's 1-RM chin test for junior high males. *Research Quarterly, 40*, 460-463.

Biering-Sorensen, F. (1984a). A one-year prospective study of low back trouble in a general population. *Danish Medical Bulletin, 31*, 362-375.

Biering-Sorensen, F. (1984b). Physical measurements as risk indicator for low-back trouble over a one-year period. *Spine, 9*, 106-119.

Blair, S.N. (1992). Are American children and youth fit? The need for better data. *Research Quarterly for Exercise and Sport,* 63(2), 120-123.

Blair, S.N., Falls, H.B., & Pate, R.R. (1983). A new physical fitness test. *The Physician and Sports medicine, 11*(4), 87-95.

Blair, S.N., Kohl, H.W., & Gordon, N.F. (1992). How much physical activity is good for health? *Annual Reviews of Public Health, 13*, 99-126.

Boileau, R.A., Wilmore, J.H., Lohman, T.G., Slaughter, M.H., & Riner, W.F. (1981). Estimation of body density from skinfold thicknesses, body circumferences and skeletal widths in boys aged 8 to 11 years: Comparison of two samples. *Human Biology, 53*(4), 575-592.

Boreham, C.A.G., Paliczka, V.J., & Nichols, A.K. (1990). A comparison of the PWC170 and 20-MST tests of aerobic fitness in adolescent schoolchildren. *Journal of Sports Medicine and Physical Fitness, 30*, 19-23.

Broer, M.R., & Galles, N.R.G. (1958). Importance of relationship between various body measurements in performance of the toe-touch test. *Research Quarterly, 29*, 253-263.

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

Buxton, D. (1957). Extension of the Kraus-Weber test. *Research Quarterly, 28*, 210-217.

Calliet, R. (1988). *Low back pain syndrome* (4th ed.). Philadelphia: F.A. Davis.

Canadian Association for Health, Physical Education, and Recreation (CAHPER). (1980). *CAHPER Fitness Performance II Test Manual.* Ottawa: Author.

Canada Fitness Survey. (1985). *Physical fitness of Canadian Youth.* Ottawa: Government of Canada Fitness and Amateur Sport.

Canadian Standardized Test of Fitness (C.S.T.F.). (1986). *Canadian Standardized Test of Fitness (C.S.T.F.) Operations manual* (3rd ed.). Ottawa: Minister of State, Fitness and Amateur Sport.

Caspersen, C.J., Powell, K.E., & Christensen, G.M. (1985). Physical activity, exercise, and physical fitness: Definitions and distinctions for health-related research. *Public Health Reports, 100*(2), 12-131.

Conger, P.R., Quinney, H.A., Gauthier, R., & Massicotte, D. (1982). A comparison of the CAHPER Fitness-Performance Test, 1966-1980. *Canadian Association of Health, Physical Education, and Recreation Journal, 49*(1), 6-11.

Cooper Institute for Aerobic Research. (1992). *The Prudential FITNESSGRAM Technical Reference Manual.* Dallas, TX: Author.

Corbin, C.B. (1986). Fitness is for children. Developing lifetime fitness. *Journal of Physical Education, Recreation, and Dance, 57*, 82-84.

Corbin, C.B. (1987a). Physical fitness in the K-12 curriculum. Some defensible solutions to perennial problems. *Journal of Physical Education, Recreation, and Dance, 58*(7), 49-54.

Corbin, C.B. (1987b). Youth fitness, exercise and health: There is much to be done. *Research Quarterly for Exercise and Sport, 58*(4), 308-314.

Corbin, C.B. (1991). A multidimensional hierarchical model of physical Fitness: A basis for integration and collaboration. *QUEST, 43*, 296-306.

Corbin, C.B. (1995). Health benefits of physical activity. Paper presented at the American Alliance Health, Physical Education, Recreation, and Dance, Portland, Oregon.

Corbin, C., Fox, K., & Whitehead, J. (1987). Fitness for a life time. In Biddle, S. (Ed.), *Foundations of Health Related Fitness in Physical Education.* London: Ling Publishing House.

Corbin, C.B., & Pangrazi, R.P. (1992). Are American children and youth fit? *Research Quarterly for Exercise and Sport, 63*(2), 96-106.

Cotten, D.J. (1990). An analysis of the NCYFS II modified pull-up test. *Research Quarterly for Exercise and Sport, 61*, 272-274.

Cureton, K.J. (1982). Distance running performance tests in children: What do they mean? *Journal of Physical Education, Recreation, and Dance, 53*(10), 64-66.

Cureton, K.J. (1994). Aerobic capacity. In Marrow, J.R., Falls, H.B., & Kohl, H.W. (Eds.), *The Prudential FITNESSGRAM Technical Reference Manual*, (pp.33-55). Dallas, TX: The Cooper Institute for Aerobic Research.

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

Cureton, K.J., Boileau, R.A., & Lohman, T.G. (1975). Relationship between body composition measures and AAHPER test performances in young boys. *Research Quarterly, 46*, 218-229.

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

Cureton, K.J., & Warren, G.L. (1990). Criterion-referenced standards for youth health-related fitness tests: A tutorial. *Research Quarterly for Exercise and Sport, 61*(1), 7-9.

Dahlgren, W.J. (1987). A case for daily, high-quality physical education throughout the school years. *Canadian Association of Health, Physical Education, and Recreation Journal, 53*(6), 39-43.

Deci, E.L. & Ryan, R.M. (1980). The empirical exploration of intrinsic motivational processes. In Berkowitz, L. (Ed.), *Advances in Experimental Social Psychology, 13*, (pp.39-80). New York, NY: Academic Press.

DeVries, H.A. (1968). EMG fatigue curves in postural muscles. A possible etiology for idiopathic low back pain. *American Journal of Physical Medicine, 47*(4), 175-181.

DeWitt, R.J. (1944). A study of the sit-up type of test, as a means of measuring strength and endurance of the abdominal muscles. *Research Quarterly, 15*, 60-63.

DiNucci, J., McCune, D., & Shows, D. (1990). Reliability of a modification of the health-related physical fitness test for use with physical education majors. *Research Quarterly for Exercise and Sport, 61*, 20-25.

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

Doolittle, T.L., & Bigbee, R. (1968). The twelve-minute run-walk: A test of cardiorespiratory fitness of adolescent boys. *Research Quarterly, 39*, 491-495.

Dragicevick, A.R., Hill, P.M., Hopkins, W.G., & Walker, N.P. (1987). The effects of a year of physical education on physical fitness in two Auckland schools. *New Zealand Journal of Health, Physical Education, and Recreation, 20*(1), 7-11.

Dyson, B. P. (1995). Physical culture as an alternative physical education curricula model. Paper presented at the Focus on Teacher Education: Visions and Ventures across the Career Span Conference, Morgantown, West Virginia.

Engelman, M.E., & Morrow, J.R., Jr. (1991). Reliability and skinfold correlates for traditional and modified pull-ups in children grades 3-5. *Research Quarterly for Exercise and Sport, 2*, 88-91.

Flint, M.M. (1965). Abdominal muscle involvement during the performance of various forms of sit-up exercise. *American Journal of Physical Medicine, 44*, 224-234.

Fox, K. (1991). Motivating children for physical activity: Towards a healthier future. *Journal of Physical Education, Recreation, and Dance, 62*(7), 34-38.

Fox, K., Peters, D., Armstrong, N., Sharpe, P., & Bell, M. (1993). Abdominal fat deposition in 11 year old children. *International Journal of Obesity, 17*, 11-16.

Franks, B.D. (1984). Physical fitness in secondary education. *Journal of Physical Education, Recreation, and Dance, 55*(9), 41-43.

Franks, B.D. (1989). *YMCA Youth Fitness Test Manual.* Champaign, IL: Human Kinetics.

Gauthier, R., Massicotte, D., Hermiston, R., & Macnab, R. (1983). The physical work capacity of Canadian children, aged 7 to 17, in 1983. A comparison with 1968. *Canadian Association of Health, Physical Education, and Recreation Journal, 50*(2), 4-9.

Glass, G.V. (1978). Standards and criteria. *Journal of Educational Measurement, 15*(4), 237-261.

Goldfine, B.D., & Nahas, M.V. (1993). Incorporating health-fitness concepts in secondary physical education curricula. *Journal of School Health, 63*(3), 142-146.

Going, S. (1988). Physical Best-Body composition in the assessment of youth fitness. *Journal of Physical Education, Recreation, and Dance, 59*(7), 32-36.

Gutin, B., Manos, T., & Strong, W. (1992). Defining heath and fitness: First step toward establishing children's fitness standards. *Research Quarterly for Exercise and Sport, 63*(2), 128-132.

Hall, G., Hetzler, R.K., Perrin, D., & Weltman, A. (1992). Relationship of timed sit-up tests to isokinetic abdominal strength. *Research Quarterly for Exercise and Sport, 63*, 80-84.

Halpern, A.A., & Bleck, E.B. (1979). Sit-up exercises: An electromyographic study. *Clinical Orthopaedics and Related Research, 145*, 172-178.

Hansen, H., & McKenzie, L. (1988). Needs, benefits, barriers, change strategies, politics. Quality, daily physical education- is it possible? *Canadian Association of Health, Physical Education, and Recreation Journal, 54*(2), 29-38.

Harsha, D., Frerichs, R.R., & Berenson, G.S. (1978). Densitometry and anthropometry of black and white children. *Human Biology, 50*(3), 261-280.

Haywood, K.M. (1991). The role of physical education in the development of active lifestyles. *Research Quarterly for Exercise and Sport, 62*(2), 151-156.

Hoeger, W.W.K., & Hopkins, D.R. (1992). A comparison of the sit and reach and the modified sit and reach in the measurement of flexibility in women. *Research Quarterly for Exercise and Sport, 63*(2), 191-195.

Hoeger, W.W., Hopkins, D.R., Button, S, & Palmer, T.A. (1990). Comparing the sit and reach with the modified sit and reach in measuring flexibility in adolescents. *Pediatric Exercise Science, 2*, 156-162.

Hyytiainen, K., Salminen, J.J., Suvitie, T., Wickstrom, & Pentti, J. (1991). Reproducibility of nine tests to measure spinal mobility and trunk muscle strength. *Scandinavian Journal of Rehabilitation Medicine, 23*, 3-10.

Jackson, A.W., & Baker, A.A. (1986). The relationship of the sit and reach test to criterion measures of hamstring and back flexibility in young females. *Research Quarterly for Exercise and Sport, 57*, 183-186.

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

Jackson, A.W., Fromme, C., Plitt, H., & Mercier, J. (1994). Reliability and validity of a 1-minute push-up test for young adults. *Research Quarterly for Exercise and Sport, 65*(Suppl.), A57-A58. (Abstract).

Jackson, A.W., & Langford, N.J. (1989). The criterion-related validity of the sit and reach test: Replication and extension of previous findings. *Research Quarterly for Exercise and Sport, 60*, 384-387.

Jette, M., Sidney, K., & Cicutti. (1984). A critical analysis of sit-ups: A case for the partial curl-up as a test of abdominal muscular endurance. *Canadian Association of Health, Physical Education, and Recreation Journal, 51*, 4-9.

Jewett, A.E., Bain, L.L., & Ennis, C.D. (1995). *The curriculum process in physical education.* Dubuque, IA: Wm. C. Brown.

Johnson, L.C. (1969). Effects of 5-day-a week vs. 2-and 3-day-a week physical education class on fitness, skill, adipose tissue and growth. *Research Quarterly, 40*(1), 93-98.

Kay, S., & Grant, B.C. (1992). What do the children think? *New Zealand Journal of Health, Physical Education , and Recreation, 25*(1), 19-21.

Kemper, H.C.G., Verschuur, R., Ras, K.G.A., Snel, J., Splinter, P.G., & Tavecchio, L.W.C. (1976). Effects of 5-versus 3-lessons-a-week physical education program upon the physical development of 12 and 13 year old school boys. *Journal of Sports Medicine, 16*, 319-326.

Klesius, S.E. (1968). Reliability of the AAHPERD Youth Fitness Test items and relative efficiency of the performance measures. *Research Quarterly, 39*, 801-811.

Kollath, J.A., Safrit, M.J., Zhu, W., & Gao, L.G. (1991). Measurement errors in modified pull-ups testing. *Research Quarterly for Exercise and Sport, 62*, 432-435.

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

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

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

Lacy, A.C., & LaMaster, K.J. (1990). Analysis of fitness activities in junior high school physical education. *Physical Educator, 47*(3), 176-179.

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

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

Leger, L.A., Seliger, V. & Brassard, L. (1980). Backward extrapolation of VO_2max values from O_2 recovery curve. *Medicine and Science in Sport and Exercise, 12*, 24-27.

Li, X.J., & Dunham, P. Jr. (1993). Fitness load and exercise time in secondary physical education classes. *Journal of Teaching in Physical Education, 12*, 180-187.

Lindsey, R., & Corbin, C. (1991). Questionable exercises: Some safer alternatives. *Canadian Association of Health, Physical Education, and Recreation Journal, 57*(4), 6-13.

Liu, N.Y.-S., Plowman, S.A., & Looney, M.A. (1992). The reliability and validity of the 20-meter shuttle test in American students 12 to 15 years old. *Research Quarterly for Exercise and Sport, 63*(4), 360-365.

Lohman, T.G. (1981). Skinfolds and body density and their relationship to body fatness: A review. *Human Biology, 53*, 181-225.

Lohman, T.G. (1994). Body composition. In Morrow, J.R., Falls, H.B., & Kohl, H.W.III.(Eds.), *The Prudential FITNESSGRAM Technical Reference Manual*, (pp.57-72). Dallas, TX: The Cooper Institute for Aerobic Research.

Lohman, T.G., Pollock, M.L., Slaughter, M.S., Brandon, J., & Boileau, R.A. (1984). Methodological factors and the prediction of body fat. *Medicine and Science in Sports and Exercise, 16*, 92-96.

Macfarlane, D., O'Neill, C., & Kay, S. (1992). Are daily fitness programmes worth the effort? *New Zealand Journal of Health, Physical Education, and Recreation, 25*(1), 16-18.

MacNaughton, L., Croft, R., Pennicott, J., & Long, T. (1990). The 5 and 15 minute runs as predictors of aerobic capacity in high school students. *Journal of Sports Medicine and Physical Fitness, 30*(1), 24-28.

Maddux, J.E., Roberts, M.C., Sledden, E.A., & Wright, L. (1986). Developmental issues in child health psychology. *American Psychologist, 41*(1), 25-34.

Magnotta, J.R. (1991). Promoting fitness enhancement programs. *Journal of Physical Education, Recreation , and Dance, 62*(5), 14-15.

Mahoney, C. (1992). 20-MST and PWC170 validity in non-caucasian children in the UK. *British Journal of Sports Medicine, 26*(1), 45-47.

Malina, R.M. (1994). Physical activity: Relationship to growth, maturation, and physical fitness. In Bouchard, C., Shephard, R.J., Stephens, T., & Associates (Eds.), *Physical activity, fitness, and health: International proceedings and consensus statement* (pp.918-930). Champaign, IL: Human Kinetics.

Massicotte, D. (1990). *Partial curl-ups, push-ups, and multistage 20 meter shuttle run, national norms for 6 to 17 year olds*. Project\# 240-0010-88/89. Ottawa: Canadian Association for Health, Physical Education, and Recreation (CAHPER) and Fitness and Amateur Sport Canada.

Mathews, D.K., Shaw, V., & Bohnen, M. (1957). Hip flexibility of college women as related to length of body segments. *Research Quarterly, 28*, 352-356.

McArdle, W.D., Katch, F.I., & Katch, V.L. (1991). *Exercise Physiology: Energy, Nutrition, and Human Performance (3rd ed.)*. Malvern, PA: Lea & Febiger.

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

McKenzie, T.L., Alcaraz, J.E., & Sallis, J.F. (1994). Assessing Children's Liking for Activity Units in an Elementary School Physical Education Curriculam. *Journal of Teaching in Physical Education, 13*, 206-215.

McManis, B.G., & Wuest, D.A. (1994). Stability reliability of the modified push-up in children. *Research Quarterly for Exercise and Sport, 65*(Suppl.), A58-A59. (Abstract).

Minkler, S., & Patterson, P. (1994). The validity of the modified sit-and-reach test in college-age students. *Research Quarterly for Exercise and Sport, 65*(2), 189-192.

Morris, H.H. (1991). The role of school physical education in public health. *Research Quarterly for Exercise and Sport, 62*(2), 143-147.

Morrow, J.R.Jr., Fridye, T., & Monaghen, S.D. (1986). Generalizability of the AAHPERD health related skinfold test. *Research Quarterly for Exercise and Sport, 57*(3), 187-195.

Nachemson, A., & Elfstrom, G. (1970). *Intervital dynamic pressure measurements in lumbar discs: A study of common movements, maneuvers and exercises.* Stockholm: Almqvist & Wiksell.

National Association for Sport and Physical Education (NASPE). (1992). *The Physically Educated Person.* Reston, VA: American Alliance of Health, Physical Education, Recreation, and Dance.

Nelson, M.A. (1991). The role of physical education and children's activity in the public health. *Research Quarterly for Exercise and Sport, 62*(2), 148-150.

Oppliger, R.A., Clark, R.R., & Kuta, J.M. (1992). Efficacy of skinfold training clinics: A comparison between clinic trained and experienced testers. *Research Quarterly for Exercise and Sport, 63*(4), 438-443.

O'Sullivan, M., Siedentop, D., & Tannehill, D. (1994). Breaking out: Condependency of high school physical education. *Journal of Teaching in Physical Education, 13*(4), 421-428.

Palgi, Y., & Gutin, B. (1984). Running performance in children. *Medicine and Science in Sports and Exercise, 16*, 158. (abstract).

Pate, R.R. (1983a). A new definition of youth fitness. *The Physician and Sports medicine, 11*(40), 77-83.

Pate, R. (Ed.). (1983b). *South Carolina physical fitness test manual (2nd Ed.)*.Columbia: South Carolina Association for Health, Physical Education, Recreation, and Dance.

Pate, R.R. (1988). The evolving definition of physical fitness. *QUEST, 40*, 174-179.

Pate, R.R., Burgess, M.L., Woods, J.A., Ross, J.G., & Baumgartner, T. (1993). Validity of field tests of upper body muscular strength. *Research Quarterly for Exercise and Sport, 64*, 17-24.

Pate, R.R., Ross, J.G., Baumgartner, T.A., & Sparks, E. (1987). The modified pull-up test. *Journal of physical Education, Recreation and Dance, 58*(9), 71-73.

Peterson, J.C. (1987). *Effect of gender on fitness performance variables (Argyle Test Battery) in high school students.* Unpublished Master's Thesis: McGill University.

Phillipp, A., Piland, N.F., Seidenwurm, J., & Smith, H.L. (1991). Improving physical fitness in high school students: Implications from an experimental course. *Journal of Teaching in Physical Education, 13*(4), 58-73.

Plowman, S.A. (1992). Criterion referenced standards for neuromuscular physical fitness tests: An analysis. *Pediatric Exercise Science, 4*, 10-19.

Plowman, S.A. (1992). Physical activity, physical fitness, and low back pain. In Holloszy, J.O. (Ed.), *Exercise and Sport Sciences Reviews, 20*,(pp.221-242). Baltimore: Williams & Wilkins.

Plowman, S.A., & Corbin, C.B. (1994). Muscular strength, endurance, and flexibility. In Morrow, J.R., Falls, H.B., & Kohl, H.W. (Eds.), *The Prudential FITNESSGRAM Technical Reference Manual*. (pp.73-100). Dallas, TX: The Cooper Institute for Aerobic Research.

Pollatschek, J.L. (1989). The Linwood project - quality daily physical education (QDPE). *British Journal of Physical Education, 20*(4), 197-201.

President's Council on Physical Fitness and Sports. (1992). *Normative Data from The 1985 School Population Fitness Survey for Use with The President's Challenge Youth Physical Fitness Program.* Washington, DC: Author.

Pollatschek, J.L. (1989). The Linwood project - quality daily physical education (QDPE). *British Journal of Physical Education, 20*(4), 197-201.

Ricci, B., Marchetti, M., & Figure, F. (1981). Biomechanics of sit-up exercises. *Medicine and Science in Sports and Exercise, 13*, 54-59.

Rider, R.A., Imwold, C.H., & Johnson, D. (1986). Effects of Florida's personal course on cognitive, attitudinal and physical fitness measures of secondary students: A pilot study. *Perceptual and Motor Skills, 62*, 548-550.

Robbins, S.G. (1987). A survey of selected schools with quality daily physical education programmes. *Canadian Association of Health, Physical Education, and Recreation Journal, 53*(6), 10-13.

Robertson, L.D., & Magnusdottir, H. (1987). Evaluation of criteria associated with abdominal fitness testing. *Research Quarterly for Exercise and Sport, 58,* 355-359.

Ross, J.G., Dotson, C.O., Gilbert, G.G., & Ketz, S.J. (1985). New standards for fitness measurement. *Journal of Physical Education, Recreation, and Dance, 56*(1), 62-66.

Ross, J.G., Pate, R.R., Delpy, L.A., Gold, R.S., & Svilar, M. (1987). New health-related fitness norms. *Journal of Physical Education, Recreation, and Dance, 5 8*(9), 66-70.

Roy, S.H., DeLuca, C.J., & Casavant, D.A. (1989). Lumber muscle fatigue and chronic lower back pain. *Spine, 14,* 992-1001.

Roy, S.H., DeLuca, C.J., Snyder-Mackler, L., Emley, M.S., Crenshaw, R.L., & Lyons, J.P. (1990). Fatigue, recovery, and low back pain in varsity rowers. *Medicine and Science in Sports and Exercise,* 22, 463-469.

Rutherford, W.J., & Corbin, C.B. (1994). Validation of criterion-referenced standards for tests of arm and shoulder girdle strength and endurance. *Research Quarterly for Exercise and Sport, 65*(2), 110-119.

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

Safrit, M.J., Baumgartner, T.A., Jackson, A.S., & Stamm, C.L. (1980). Issues in setting motor performance standards. *QUEST, 32*(2), 152-162.

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

Safrit, M.J., & Wood, T.M. (1987). The test battery reliability of the health-related physical fitness test. *Research Quarterly for Exercise and Sport, 58,* 160-167.

Sallis, J.F., & McKenzie, T.L. (1991). Physical education's role in public health. *Research Quarterly of Exercise and Sport, 62*(2), 124-137.

Saltarelli, W.A., & Andres, F.F. (1993). Teaching steady pacing to students-a practical method. *Journal of Physical Education, Recreation and Dance, 68*(8), 67-70.

Seefeldt, V., & Vogel, P. (1987). Children and fitness: A public health perspective. A response. *Research Quarterly for Exercise and Sport, 58*(4), 331-333.

Shaw, V.W. (1986). The accuracy of two training methods of skinfold assessment. *Research Quarterly for Exercise and Sport, 57*, 85-90.

Siedentop, D. (1992). Thinking differently about secondary physical education. *Journal o of Physical Education, Recreation, and Dance, 63*(7), 69-72.

Siedentop, D. (Ed.). (1994). *Sport education: Quality PE through positive sport experiences.* Champaign, IL: Human Kinetics.

Siedentop, D., & Siedentop, B. (1985). Daily physical education in Australia. *Journal of Physical Education, Recreation, and Dance, 56*(2), 41-43.

Silver, J. (1983). Daily physical education: A report of a study tour. *New Zealand Journal of Health, Physical Education, and Recreation, 16*, 12-14.

Simons-Morton, B.G., O'Hara, N.M., Simons-Morton, D.G., & Parcel, G.S. (1987). Children and fitness: A public health perspective. *Research Quarterly for Exercise and Sport, 58*(4), 295-302.

Sinclair, G.J. (1983). A daily physical education pilot project. *Canadian Association of Health, Physical Education, and Recreation Journal, 49*, 22-26.

Slaughter, M.S., Lohman, T.G., Boileau, R.A., Horswill, C.A., Stillman, R.J., van Loan, M.D., & Bemben, D.A. (1988). Skinfold equations for estimation of body fatness in children and youth. *Human Biology, 60*, 709-723.

Slava, S., Laurie, D.R., & Corbin, C.B. (1984). Long-term effects of a conceptual physical education program. *Research Quarterly for Exercise and Sport, 55*(2), 161-168.

Sloan, A.W., & Shapiro, M. (1972). A comparison of skinfold measurements with three standard calipers. *Human Biology, 44*, 29-36.

Smith, T.K., & Cestaro, N. (1992). Saving future generations - the role of physical education. *Journal of Physical Education, Recreation, and Dance, 63*(8), 75-79.

VanAdrichem, J.A.M. & vanderKorst, J.K.(1973). Assessment of the flexibility of the lumbar spine. *Scandinavian Journal of Rheumatology, 2*, 87-91.

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

Verabioff, L.J. (1986). Can we justify daily physical education? *Canadian Association of Health, Physical Education, and Recreation Journal, 52*(2), 8-11.

Vincent, W.J., & Britten, S.D. (1980). Evaluation of the curl-up - a substitute for the bent knee situp. *Journal of Physical Education and Recreation, 51*(2), 74-75.

Walters, C.E., & Partridge, M.J. (1956). Electromyographic study of the differential action of abdominal muscles during exercise. *American Journal of Physical Medicine, 36*, 259-267.

Wear, C.L. (1963). Relationship of flexibility measurements to length of body segments. *Research Quarterly, 34*, 234- 238.

Werner, P. & Durham, R. (1988). Health related fitness benefits in upper elementary school children in a daily physical education program. *Physical Educator*, *45*(2), 89-93.

Westcott, W.L. (1992). High school physical education: A fitness professional's perspective. *QUEST*, *44*, 342-351.

Whitehead, J.R., & Corbin, C.B. (1991). Youth fitness testing: The effect of percentile-based evaluative feedback on intrinsic motivation. *Research Quarterly for Exercise and Sport*, *62*(2), 225-231.

Williams, D.P., Going, S.B., Lohman, T.G., Harsha, D.W., Srinivasan, S.R., Webber, L.S., & Berenson, G.S. (1991). Truncal fat patterning and its relation to cardiovascular disease (CVD) risk factors in children and adolescents. *Medicine and Science in Sports and Exercise*, *23*, S51. (Abstract).

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