The Relationship Between Academic Achievement and Physical Fitness

LeaAnn Tyson Martin and Gordon R. Chalmers

Abstract

Over the past 50 years research has typically demonstrated either no, or a weak relationship between academic performance and physical performance. Nevertheless, the fact that any positive relationship has been found has lead to the "healthy children learn better" concept currently being promoted in schools, and used as a rationale to justify physical education programs. The conclusions of previous studies have been based on the statistical significance of the results, and have not incorporated the practical significance. The authors investigated the relationship between academic performance, as measured by scores on the Iowa Tests of Basic Skills, and physical fitness, as measured by performance on the President's Challenge. Subjects were 5847 Seattle School District students in grades three, five, six and eight. The correlation between mean physical and academic percentile scores was 0.19, p <.05. While statistically significant, this correlation indicated that only 3.6% of the variance in academic performance could be explained by physical fitness. The low level indicated that the relationship between academic performance and physical fitness is of little practical importance. Accordingly, physical education programs should not be advocated as a means to promote academic achievement in students.

It has been suggested that the benefits gained from, or performance in, physical education are related to academic achievement (Almond & McGeorge, 1998; Black, 1995). Some have advocated that not necessarily physical education, but physical activity and fitness are related to cognitive performance and achievement (Dustman, Emmerson, & Shearer, 1994; Etnier et al., 1997). Thus, the concept that "healthy children learn better" is being promoted in schools and elsewhere (Symons, Cinelli, James, & Groff, 1997).

Investigations into the relationship between academic achievement and physical fitness have produced mixed results. Weber (1953) correlated fitness, using the Iowa Physical Proficiency Profile (including sit-ups, pull-ups, running), to entrance exam scores and grade point averages for 246 male college students. Fitness level had a significant positive relationship with grade point average (r = .41), but did not relate to performance scores on entrance exams. Hart and Shay (1964) examined mathematics and verbal SAT scores and the Physical Fitness Index in 60 college women. When the relationships between verbal scores and mathematics scores and fitness index were examined, the r values were .068 and .146, respectively, although neither was significant at the .05 level. A battery of fitness tests (e.g., flexed arm hang, curl-ups, and step test) were administered to 827 female freshmen and subjects were placed in one of three categories of fitness: high, fair or poor (Arnett, 1968). Arnett (1968) found significant differences in grade point average between the groups, with participants with higher fitness levels having higher GPAs.

Studies on school-aged children, using a variety of academic and fitness measures, have also resulted in inconsistent findings. Clarke and Jarman (1961), examining 217 boys (aged 9, 12 and 15), found that there was a consistent, and for some fitness measures, a significant tendency for the students in the high fitness group to have higher means on both standard achievement tests and grade point average.

More recent studies have used standardized achievement and fitness tests as measures. A study involving 1,767 students in second, fourth and sixth grades examined the relationship between performance on the Georgia Criterion Referenced Test for Reading, Mathematics and Career Education and performance on a variety of physical fitness tests from the Minnesota Performance Test, the American Alliance for Health, Physical Education, Recreation and Dance (AAHPERD) Health Related Physical Fitness Tests, and the Texas Physical Fitness-Motor Ability Test (Harris & Jones, 1982). For the boys and girls, multiple regression analysis demonstrated a low, but significant, relationship between reading and mathematics ability and the combination of eight motor performance measures examined, five of which were fitness measures.

Winn (1993) studied 302 fourth and fifth grade children and examined the relationship between scores on the California Test of Basic Skills (reading, mathematics) and performance on the AAHPERD President's Challenge. Using national norms, total fitness and total academic achievement scores were determined. The overall correlation between the total scores was .213. When each test item was correlated with scores in each of reading, mathematics and language, the correlations ranged from .043 to .462, although none of the correlations were significant at the .05 level.

An examination of 7,961 youngsters from 7 to 15 years of age in Australia was conducted by Dwyer, Sallis, Blizzard, Lazarus and Dean (2001). School ratings of scholastic ability were compared with performance on a variety of fitness measures including sit-ups, push-ups, and a 1.6 kilometer run. Across the age groups, there were significant, but weak, correlations (ranging from .1 to .27) between fitness (cardiorespiratory endurance, muscular force and power) and academic performance.

Most recently, the California Department of Education (2002; 2005) reported the results of two studies that examined the relationship between scores on achievement tests and the Fitnessgram. In the first study, performance on the Stanford Achievement Tests and scores on the Fitnessgram for 884,715 students in grades 5, 7, and 9 were investigated. A composite score, ranging from zero to six, was created for physical fitness, in which a student obtained one point for each of the six test items for which the student was determined to be in the "healthy zone." In each of the three grades, higher levels of fitness were related to higher academic achievement. The relationship was stronger for math achievement and fitness, especially at higher fitness levels. This study has yet to be published. As a result, no statistical measures are available. Nevertheless, the results were cited by professional sources, such as the National Association for Sport and Physical Education (no date) and the PE Central web site (no date) as evidence that there is a direct relationship between physical fitness levels and academic achievement.

In the latter study (California Department of Education, 2005), performance on the California Standards Tests and the Fitnessgram for 1,036,386 students in grades 5, 7 and 9 were compared. Again in this study, students were awarded a composite score, representing the number of fitness test battery items in which they were in the "healthy zone." Results were similar to the 2002 study, with higher fitness scores associated with higher scores in English-language arts and mathematics ($p < .05$). In this study (California Department of Education, 2005), however, only means were reported; thus, no standard deviations were given for the groups compared, nor were effect size measures made to quantify the practical significance of the differences observed between groups.

In summary, research examining the relationship between academic achievement and physical fitness has produced mixed results. Of these, one study has been published only as a press release in which no statistical analysis was reported and a second study had incomplete statistical information to effectively interpret the results (California Department of Education, 2002; 2005). In the remaining investigations the interpretation of the results focused on whether a statistically significant finding was observed. A number of statistical researchers, however, have emphasized that the correct interpretation of research results requires that not only the statistical significance of the data be considered, but also the practical significance of the findings (e.g., Speed, 1998; Sterne & Smith, 2001; Thomas, Salazar & Landers, 1991; Vincent, 1999). This is particularly important in studies such as the present one, and the ones discussed above, which typically involve very large sample sizes of hundreds to hundreds of thousands of subjects. Due to the effect of sample size on the calculation of statistical significance, with large sample sizes it is possible to calculate statistical significance on a result that has no practical significance (Vincent, 1999).

As evidenced by the history of investigations, the importance of understanding the relationship between physical fitness and academic performance in children and youth is relevant, and increased by recent evidence from studies conducted on animals and elderly humans that increased physical activity results in improved cognitive function (Colcombe et al., 2004; McAuley, Kramer, & Colcombe, 2004; Rhodes et al., 2003). The purpose of this study was to investigate the relationship between school academic performance and physical fitness, using the Iowa Tests of Basic Skills and the President's Challenge. Further, a unique contribution of the current study to the existing literature is that the meaningfulness of the results is discussed. Thus, data are not simply interpreted based on the statistical significance found.

Methods

Subjects

Subjects in the study were students in grades three, five, six and eight in 46 elementary schools, 9 middle schools, and 5 "alternative" K-8 or K-12 schools in the Seattle School District during the 2000-2001 academic year. Subjects were administered a fitness test battery and an academic achievement test during the school year, and only those students with complete sets of scores for both tests were included in the study.

The President's Challenge Youth Physical Fitness Program

The President's Challenge Youth Fitness test is a norm referenced assessment. Although the President's Challenge has a test battery that includes 11 separate tests, some of the individual tests are only one of two or three options. For example, abdominal strength and endurance can be measured with "curl-ups" or "partial curl-ups." The participants in this study participated in curl-ups, the one-mile run/walk, the V-sit reach, the shuttle run, and the best of the pull-up or flexed arm hang. Test items and administration are standardized, and descriptions of each test item are available from the President's Council on Physical Fitness and Sports (http://www.fitness.gov). Each student's raw score on each test item was converted to a performance percentile as per norms available from the President's Council on Physical Fitness and Sports (1992). Physical education teachers at each of the schools administered the tests during the academic year on a voluntary basis, students participated on a voluntary basis, and teachers submitted scores to the district administration on a voluntary basis. Teachers administered the tests at any time during the academic year they deemed appropriate for their academic program and curriculum.

Iowa Tests of Basic Skills

The Iowa Tests of Basic Skills (ITBS) were used to measure academic achievement. This is a frequently used achievement test and has been validated by its developers (Riverside Publishing, Houghton Mifflin Company, see www.river/pub.com). The ITBS assessment is norm referenced and consists of multiple forms (forms A, K, L and

M), subjects were administered Form M. Third and sixth graders are administered the "basic skills" version, while fifth and eighth graders are given the "integrated writing" version. The test battery and administration are standardized. The test was administered by classroom teachers to all students in the classroom during the first three weeks of March during the academic year.

Data Collection

ITBS and President's Challenge test scores were received by the Office of Research, Evaluation and Assessment of the Seattle School District. Scores were then matched according to student number, and the student number was replaced by a code number. Permission to use, and access to, this pre-existing data set was given to the researchers by the Office of Research, Evaluation and Assessment of the Seattle School District, and use of this data followed guidelines established by the Western Washington University human subjects review committee.

Statistical Analysis

For each subject, the percentile scores for each of the three academic tests (reading, language arts, and mathematics) were averaged, to give an average academic percentile score. Similarly, the percentile scores for each of the five fitness tests were averaged, to give an average physical percentile score. The relationship between average academic and physical percentile scores was plotted, a Pearson coefficient of correlation, and the significance of the correlation were calculated (Vincent, 1999). Statistical analysis was performed using Excel 2002 (Microsoft Corp., Redmond, Washington) and a p-value <.05 was considered significant.

Results

The relationship between physical fitness, measured by an average score on the President's Challenge test battery, and academic performance, determined by the Iowa Tests of Basic Skills, in Seattle school children was examined. Data from 5847 students in grades three (n = 2049), five (n = 2169), six (n = 948) and eight (n = 681) were analyzed. The Pearson coefficient of correlation between mean fitness and academic percentile scores was 0.19, p <.05 (Figure 1). A Pearson coefficient of correlation is calculated by fitting a straight line of best fit to the relationship between two variables to estimate the strength of the relationship. In some cases the line that best describes the relationship between two variables is not a straight line, and if so using a straight line to describe the relationship results in an underestimation of the strength of the relationship. To determine if the low relationship between the variables was due, in part, to the improper use of a straight line to describe the relationship, various curved lines (logarithmic, exponential, and power) were also fitted to the data. Each of these curved lines was found to not fit the data as well as a straight trend line. Hence, the relationship was described by the Pearson coefficient of correlation.

Discussion

The distribution of percentile scores observed for both the fitness and the academic measures displayed a trend toward more scores being greater than the fiftieth percentile than below. This is not surprising given that the individual student raw scores were being compared to national norms to produce the individual percentile scores, and a single school district is unlikely to mirror the national norm distribution.

While some may argue about the validity of physical fitness assessments, their use is common practice in physical education, and measures of fitness are often required in state and local curricula. The most widely utilized fitness assessments in the United States are the President's Challenge test battery and the Cooper Fitnessgram. While these two tests differ in some aspects, they are both considered far better than fitness testing methods used in previous decades in most of the prior studies reviewed in the introduction. For this study, teachers administered the tests and students

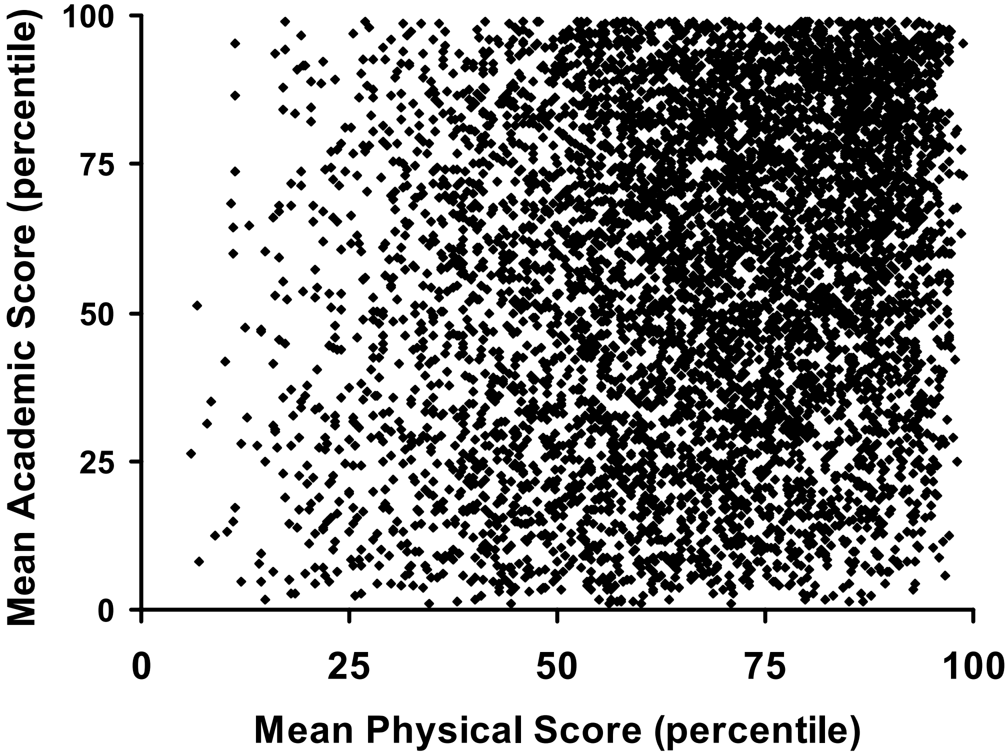

Figure 1. The relationship between the mean academic percentile score and mean physical percentile score for 5847, third, fifth, sixth and eighth grade students. The academic score is the average student performance on Iowa Tests of Basic Skills. The physical score is the average student performance on the five tests in the President's Challenge fitness test. Correlation, $r = 0.19$, $p < .05$.

participated on a voluntary basis. While this might have limited the number or type of subjects, a sample size of nearly 6000 subjects provides ample statistical power. Further, the measures on the subjects were distributed across the entire range of possible scores, indicating that subjects with all possible combinations of fitness and academic abilities were included in the sample.

Similarly, the Iowa Tests of Basic Skills are currently widely used in schools in the United States, indicating that they are considered by many school districts as being among the best methods available to assess academic performance. Standardized testing of academic subjects has been a common practice to assess achievement for decades, and more recently, has become a criterion for graduation from public schools in some states. Thus, the measures used in the current study are generally accepted as valid, useful appraisals.

The correlation of .19 between academic and fitness scores found in the present study is comparable to the mid point of the range of correlations reported by most of the previous comparable investigations (Dwyer et al., 2001; Hart & Shay, 1964; Winn, 1993). The strength of the relationship between academic and fitness scores may be greater than the value determined in the present study, due to inherent error present in any measurement. A careful interpretation of a correlation is essential for determining if such data provide supportive evidence for the idea that "…students achieve best when they are physically fit" (California Department of Education, 2002). A correlation of .19 is described in the field of exercise science as "small" (Hopkins, 2001). In education statistical texts, a correlation between 0 and .4 is described as being "Of little practical importance except in unusual circumstances: perhaps of theoretical value" (Fraenkel & Wallen, 2003, p. 259), and a correlation of .20-.35 indicates "…only a slight relationship…may be valuable to explore interconnection of variables, but of little value in prediction studies" (Creswell, 2002, p. 372). The coefficient of determination ($r2$) expresses the proportion of variability in one variable that can be explained by a second variable. For the current data, a coefficient of determination indicates that only 3.7% of the variability in academic performance can be explained by the variance in physical performance. Almost all (i.e. 96.3%) of the variance in academic performance is unexplained by physical fitness.

Of course, the interpretation of any study examining the relationship between two measures must also be tempered by the fact that a causal link cannot be concluded, no matter what the relationship strength. For those in the profession of health promotion or physical education, causation would be a beneficial result and would no doubt enhance the argument for "healthy children learn better." However, no inference can be made by these results or others (e.g., California Department of Education, 2005) that one result produces the other.

Based on the above interpretation guidelines, it is up to the reader to assess if a correlation of 0.19 between physical fitness and academic performance, using measures of these variables that are currently widely accepted, provides evidence that there is a meaningful relationship between physical fitness and academic performance in the students examined. Hence, is it reasonable to propose that school physical education is important for the role it can play in promoting academic achievement via physical fitness? The authors question the conclusions of other papers (for example, California Department of Education, 2002; 2005) that there is solid evidence that healthy children learn better. The present data, and two similar data sets over the last 20 years showing a similar relationship strength (Dwyer et al., 2001; Winn, 1993), do not provide evidence that school physical education, producing increased student fitness levels, can be promoted for more than a trivial positive effect on academic achievement.

Quality school physical education programs have many positive effects, including motor skill and physical fitness development of students (National Association for Sport and Physical Education, 2002). Given the recently escalating problem of obesity and associated diseases such as diabetes in children (National Center for Chronic Disease Prevention and Health Promotion, 2004a, 2004b; Ogden, Flegal, Carroll & Johnson, 2002), it is increasingly important that schools provide physical activity and health education to students. Ironically, school physical education programs are often marginalized or the target of budget cuts because of their lack of inclusion in the No Child Left Behind Act and other high stakes testing. Retaining and enhancing quality physical education programs for the learning of motor skills and health benefits may be justified. Rationalizing school physical education as a means to also improve academic achievement of students, however, does not appear to be justified.

REFERENCES

Almond, L., & McGeorge, S. (1998). Physical activity and academic performance. *British Journal of Physical Education, 29*(2), 8-12.

Arnett, C. (1968). Interrelationships between selected physical variables and academic achievement of college women. *Research Quarterly, 39*, 227-230.

Black, S. (1995). Just do it. *Executive Educator, 17*(4), 33-36.

California Department of Education. News release (2002, December, 10). www.cahperd.org/images/pdf_docs/CDE_News_Release.pdf

California Department of Education. (2005). *A study of the relationship between physical fitness and academic achievement in California using 2004 test results*. Sacramento, CA: Author.

Clarke, H., & Jarman, B. O. (1961). Scholastic achievement of boys 9, 12, and 15 years of age as related to various strength and growth measures. *Research Quarterly, 32,* 155-162.

Colcombe, S. J., Kramer, A. F., Erickson, K. I., Scalf, P., McAuley, E., Cohen, N. J., et al. (2004). Cardiovascular fitness, cortical plasticity, and aging. *Proceedings of the National Academy of Sciences, USA, 101,* 3316-3321.

Creswell, J. W. (2002). *Educational research. Planning, conducting, and evaluating quantitative and qualitative research*. Upper Saddle River, NJ: Merrill Prentice Hall.

Dustman, R. E., Emmerson, R., & Shearer, D. (1994). Physical activity, age and cognitive function. *Journal of Aging and Physical Activity, 2,* 143-181.

Dwyer, T., Sallis, J. F., Blizzard, L., Lazarus, R., & Dean, K. (2001). Relation of academic performance to physical activity and fitness in children. *Pediatric Exercise Science, 13*, 225-237.

Etnier, J. L., Salazar, W., Landers, D. M., Petruzzello, S. J., Han, M., & Nowell, P. (1997). The influence of physical fitness and exercise upon cognitive functioning: A meta-analysis. *Journal of Sport and Exercise Psychology, 19,* 249-277.

Fraenkel, J. R., & Wallen, N. E. (2003). *How to design and evaluate research in education* (5th ed.). Boston: McGraw Hill.

Harris, D. I., & Jones, M. A. (1982). Reading, math and motor performance. *Journal of Physical Education, Recreation and Dance, 53*(9), 21-22.

Hart, M. E., & Shay, C. T. (1964). Relationship between physical fitness and academic success. *Research Quarterly, 35,* 443-445.

Hopkins, W. G. (2001). *New view of statistics: Effect magnitudes*. Retrieved July 10, 2004, from http://www.sportsci.org/resource/stats/effectmag.html

McAuley, E., Kramer, A. F., & Colcombe, S. J. (2004). Cardiovascular fitness and neurocognitive function in older adults: A brief review. *Brain, Behavior, and Immunity, 18,* 214-220.

National Association for Sport and Physical Education. (2002). *2001 Shape of the nation report.* Reston, VA: Author.

National Association for Sport and Physical Education. (no date). *Physical Education is Critically Important* [Brochure]. Reston, VA: Author.

National Center for Chronic Disease Prevention and Health Promotion. (2004a). *Diabetes public health resource. Data and trends*. Retrieved July 15, 2004, from http://www.cdc.gov/diabetes/statistics/prev/national/figbyage/htm

National Center for Chronic Disease Prevention and Health Promotion. (2004b). *Nutrition and physical activity. Overweight and obesity.*. Retrieved July 15, 2004, from http://www.cdc.gov/nccdphp/dnpa/obesity/trend/maps/

Ogden, C. L., Flegal, K. M., Carroll, M. D., & Johnson, C. L. (2002). Prevalence and trends in overweight among U.S. children and adolescents, 1999-2000. *Journal of the American Medical Association, 288*, 1728-1732.

PE Central (no date). *Physical Education Research In Action*. Retrieved September 30, 2005, from http://www.pecentral.com/research/index.asp

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: U.S. Government Printing Office.

Rhodes, J. S., van Praag, H., Jeffrey, S., Girard, I., Mitchell, G. S., Garland, T. Jr., et al. (2003). Exercise increases hippocampal neurogenesis to high levels but does not improve spatial learning in mice bred for increased voluntary wheel running. *Behavioral Neuroscience, 117,* 1006-1016.

Speed, H. (1998). Power, effect size and the misinterpretation of statistical inference in exercise science, In: *Australian Conference of Science and Medicine in Sport*, Adelaide 13-16 October, Canberra, Sports Medicine Australia, p.215. Available: URL: http://www.ausport.gov.au/fulltext/1998/acsm/smabs215.htm

Sterne, J. A. C., & Smith, G. D. (2001). Sifting the evidence what's wrong with significance tests? *British Medical Journal*, 322, 226-231.

Symons, C. W., Cinelli, B., James, T. C., & Groff, P. (1997). Bridging student health risks and academic achievement through comprehensive school health programs. *Journal of School Health, 76,* 220-227.

Thomas, J. R., Salazar, W., & Landers, D. M. (1991). What is missing in p is less than .05? Effect size. *Research Quarterly for Exercise and Sport*, 62(3), 344-348.

Vincent, W. J. (1999). *Statistics in kinesiology.* Champaign, IL: Human Kinetics.

Weber, J. R. (1953). Relationship of physical fitness to success in college and to personality. *Research Quarterly, 24,* 471-474.

Winn, K. L. (1993). *A study of the relationship between physical fitness levels and the academic achievement of fourth and fifth grade students.* Unpublished master's thesis, Western Washington University, Bellingham, WA.

Drs. LeaAnn Tyson Martin and Gordon R Chalmers are faculty members at Western Washington University, Bellingham, Washington.