2011

Issues With Youth Fitness Testing: A Review of Current Literature

Scott Steven Weinheimer

Issues With Youth Fitness Testing: A Review of Current Literature

Issues with Youth Fitness Testing: A Review of Current Literature

Submitted in Partial Fulfillment of the

Requirements for the Degree

Masters of Arts in Physical Education: Teaching/Coaching

Scott Steven Weinheimer

University of Northern Iowa

June, 2011

RESEARCH PROJECT APPROVAL PAGE

This paper by: Scott Steven Weinheimer

Entitled: Issues with Youth Fitness Testing: A Review of Current Literature

Has been approved as meeting the degree requirements for a Master of Arts Degree in Physical Education: Teaching/Coaching

06/02/11 _____

Date

_____

Dr. Trey Leech, Committee Chair

06/02/2011 _____

Date

_____

Dr. Fabio Fontana, Committee Member

Table of Contents

Introduction

As physical education makes its way through the dark ages of viewing student success in physical education classes as simply keeping them "busy, happy, good," (Placek, J., 1983) physical education teachers need to realize that they are on the front lines of America's health care crisis (Centers for Disease Control and Prevention, 2010). With the rise in lifestyle diseases in children, such as cardiovascular disease, certain types of cancers, and type two diabetes; it is becoming more obvious that the way people take care of themselves from an early age is vital to their longevity (CDC, 2011). Poor eating habits and a lack of physical activity have started to catch up with us as a nation and for the first time in history the current generation is not expected to outlive their parents (Pe4life, 2002). With the percentage of overweight and obese Americans at almost thirty percent and climbing, (CDC, 2010) the future does not look bright.

The problem may not only be physical, it could have financial implications as well. In 2006, health care costs related to obesity were estimated at as much as $147 billion (CDC, 2010). With obese patients paying as much as $1,429 higher than their average weight counter parts (CDC, 2010). Getting out in front of the problem early may be the key to treating this epidemic. Health care professionals have talked about the benefits of a healthy lifestyle for middle aged adults for years, pushing the importance of being physically fit and eating right. Health issues that only affected adults 30 years ago are becoming more prevalent in our youth, (Centers for Disease Control and Prevention, 2011) this has led to the realization that physical activity cannot be taken for granted and that early quality physical education is part of the answer to preventing many of these issues now and later in life.

All of these factors point to the fact that physical education professionals are a much more valuable asset than it has been thought in the past. Many administrators seem to value physical education less than core academic subjects. This can be shown by the number of physical education teaching positions and teaching time that has been cut to meet strict budget constraints, compared to

the number of positions and amount of time cut in core subject areas (Noffsinger, 2005). Schools across the country are cutting physical education in favor of academic time and funding. Some schools cannot legally cut physical education, because of state laws. However, some of these schools have removed physical education teachers from their teaching positions and require untrained classroom teachers to supervise physical education time (Noffsinger, 2005). In recent years, the focus of physical education has been shifted to developing life-long physical activity and ideas of overall wellness opposed to just a sports skill mastery focus or a physical activity focus. This means that a trained physical education professional is vital. One example of this shift is the emergence of organizations like PE4Life, which has a goal to "inspire and educate all students about the fundamental importance of lifetime physical activity and fitness" (PE4life, 2011).

While this is a step in the right direction, physical education can only become a respected profession if it can be proven that what is done in physical education does make a difference. Empirical data is the fastest way to validate a quality physical education program. If it can be shown that a physical education program is having positive physical effects on its students then it is easier to validate a program. One measure that is used to obtain this data is fitness testing. Being able to show what students can do when they start a P.E. program and the change in their ability when they finish a P.E. program is how it can be confirmed that a P.E. program does make a difference.

This review of the literature will look into previous literature that has been published on issues that can affect the results of fitness testing. First, an overview of fitness testing in the United States will occur. Then various social issues that can arise because of fitness testing and how they may affect the outcomes of fitness testing will be described. Lastly, problems associated with fitness testing procedures as well as possible ways to improve fitness testing procedures will be explained.

History of Fitness Testing in the United States

The history of physical fitness testing in the United States primarily began when a December 1953 article entitled *Muscular Fitness and Health* was published. The article, which was published in the journal of the American Association for Health, Physical Education, and Recreation, was coauthored by Hans Kraus and Bonnie Prudden and examined how the ease of the 20th century American lifestyle was causing Americans to lose muscle tone. Kraus and Prudden recommended that "Americans would have to engage in regular exercise to attain a state of physical fitness comparable to that of an earlier era" (Sturgeon, & Meer, 2010).

Because of the Kraus and Prudden article other researchers began to investigate American's physical fitness to see if the problem was really as bad as it seemed. Kraus and Prudden teamed with Dr. Sonja Webber and developed the "Kruse-Webber Test for Muscular Fitness." The study looked at the physical fitness of 4,400 public school students ages 6 – 16 from across the county and 3,000 European students ages 6 – 16. The study found that fifty six percent of American students failed at least one part of the Kraus-Webber test while only 8 percent of European students failed any part of the test (Sturgeon, & Meer, 2010). The test, which consisted of leg lifts, sit-ups, trunk lifts and toe touches showed that no matter what age or sex, European students seemed to be more physically fit. The findings were alarming enough that is was published in the *New York State Journal of Medicine*.

The United States government started to get involved in the health and physical fitness of the nation when the Kraus-Webber study made its way to President Dwight D. Eisenhower, who was "shocked" by this information and found the results "alarming" (p.41). President Eisenhower invited Kraus, Prudden, and Webber to the White House to present their findings to government leaders and the result of this presentation led to the organization of the President's Conference on the Fitness of American Youth, which was held in June 1956 at the United States Naval Academy. From this conference, the President's Council on Youth Fitness was established in July of 1956.

President Eisenhower explained that the Presidential Council's purpose would be to act as "a catalytic agency that would educate, stimulate, motivate, and encourage local communities and individual Americans to promote and adopt active lifestyles" (p.43). Eisenhower also recognized the need for the council citing the complaints of United States military recruiters about the poor fitness levels of the pool of draftees during the Second World War and the Korean conflict. It was reported that approximately 50% of the draftees showing up at draft boards where physically unfit (p.44). While he valued physical fitness, President Eisenhower, being a former Military leader, saw the military need for physical fitness as the most important reason to form the council.

Shortly after the formation of the President's Council, the American Alliance for Health, Physical Education, and Recreation (AAHPER/AAHPERD) came up with a test through the AAHPER Youth Fitness Project (YFP). The YFP pilot study of the fitness of children (ages 5 – 12) across the country, which indicated similar results to those found in the Kraus – Webber study (Plowman et al, 2006). The results of the study lead to the publishing of AAHPER Youth Fitness Test manual in 1958. The manual set fourth norms for the first national youth fitness test. The test included skill related activities such as pull-ups/flexed arm hang, straight leg sit-ups, shuttle run, standing broad jump, 50 yard dash, softball throw for distance, 600 yard run/walk, and three aquatic tests (Plowman et al, 2006). These activities varied greatly from the activities involved in the Kraus-Webber test which primarily focused on flexibility (Sturgeon, & Meer, 2010).

In 1964 the President's Council conducted a national fitness survey to help establish updated norms for youth aged 10 to 17 (Sturgeon, & Meer, 2010). This survey became the basis for the Presidential Award Program for the AAHPER YFT which began in 1966. This program awarded students who scored in or above the 85th percentile in all of the tests, were in good academic standing, and received a recommendation from the school principal (Plowman et al, 2006).

By the mid-1970s some physical fitness researches and physical education teachers started to question some of the activities tested and procedures used in the AAHPER YFT (Plowman et al, 2006). So a new AAHPER test was created which included several new test components including cardio respiratory function (1 mile/9 minute or 1.5 mile/12 minute), body composition (triceps or sum of triceps and sub scapular skin folds), and abdominal and low back-hamstring musculoskeletal function (bent knee, timed sit-ups; sit-and-reach) (Plowman et al, 2006). However AAHPER leaders decided to keep the original YFT and the Presidential awards system and create a new test called AAHPER Physical Fitness Test. The creation of a second AAHPER test caused some confusion and frustration so in 1986 the Presidents council decided to form their own test and award their Presidential Challenge awards to participants in their own test (Plowman et al, 2006). In 1986, the Presidential Challenge Physical Fitness test debuted and included pull-ups for boys/flexed arm hang for girls, 1 mile run, shuttle run, and sit-and-reach (Plowman et al, 2006). The current Presidential Challenge Physical Fitness test has changed over the years and now includes five components, the curl-ups (or partial curl-ups), the shuttle run, a one mile endurance run/walk, pull-ups (or right angle push-ups or flexed-arm hang) and v-sit reach (or sit and reach) (The President's Council on Fitness, Sports & Nutrition, 2011).

During the time period that the AAHPER YFT & PFT were causing some confusion and the President's Council decided to move away from AAHPER and form its own test, a group of researchers and professionals called for a fitness assessment that used different methods and offered a different prospective into fitness testing. This group created a fitness test using tracking software that AAHPER had used for years known as FITNESSGRAM. In 1987, when the FITNESSGRAM test debuted, it included the following items a 1 mile run, modified sit-up, sit and reach, pull up/flexed arm hang (either sex), body composition (4-12 grades) assessed by triceps and calf skin folds (or body mass index (BMI) if no skin folds available), and an optional shuttle run for grades K-3 (Plowman et al, 2006). While some of these tests where slightly different than those used in other fitness testing programs, one of the biggest

changes came with the fact that the FITNESSGRAM test would be criterion-referenced opposed to the norm-referenced used by other fitness testing programs.

The FITNESSGRAM test has evolved more than other popular fitness tests over the years. Because FITNESSGRAM is constantly revising their testing practices and has added and deleted certain aspects of the test depending on what scientific research has proven most effective (Plowman et al, 2006). Major additions or deletions since FITNESSGRAMs introduction in 1987 have happened in 1992, 1999, 2004, and 2005. The test has evolved into a series of options to test the four major areas, which are aerobic capacity, body composition, muscular strength and endurance, & flexibility (Plowman et al, 2006).

AAHPER known today as the American Alliance for Health, Physical Education, Recreation, and Dance (AAHERD) has consolidated it fitness testing into a programmed curriculum known as Physical Best (Plowman et al, 2006). The Presidential Challenge Physical Fitness test, which is still maintained by what has become known today as The President Council on Physical Fitness, Sport and Nutrition, and has expanded its programs to include different youth and adult programs and offers more incentives to those participating (The President's Council on Fitness, Sports & Nutrition, 2011). The FTINESSGRAM has also expanded to include the ACTIVITYGRAM which promotes lifelong physical activity (Plowman et al, 2006). The Presidential Challenge Physical Fitness test is the most commonly used fitness testing program in the United States (Keating, & Silverman, 2004). The FITNESSGRAM is the other major youth fitness testing program used in the United States today.

Social Issues Associated with Fitness Testing

Fitness testing has caused a variety of social issues to arise throughout the years including everything from issues such as the discomfort of students while they performing test items in front of peers to the validity of fitness testing to actually measure what they are trying to assess (Cale, & Harris, 2009). Several of these social issues have been researched and will be discussed in this section. The first such issue is whether or not the fitness testing is promoting healthy active lifestyles or just making students uncomfortable. The second issue pertains to how body composition testing is done and how this can affect students mentally and physically. Lastly, issues with the test components ability to evaluate what they are suppose to evaluate will be examined.

To many physical education teachers, fitness testing is a way to see if their program is effective and to help show the importance of healthy active lifestyles to student and parents. However research done by Cale and Harris (2009) has shown that fitness testing does little to motivate students to learn the skills needed to practice living a healthy active lifestyle (Cale, & Harris, 2009). Through an extensive research review, the researchers interviewed and surveyed fitness testing professionals and stakeholders from across the United Kingdom to gain an understanding of their ideas about fitness testing (Cale, & Harris, 2009). Among the findings was that, "fitness testing in P.E. may well represent a misdirected effort in the promotion of healthy lifestyles and physical activity and that P.E. time could therefore be better spent" (105, Cale, & Harris, 2009).

If the perspectives provided in Cale and Harris's article are true then maybe fitness testing itself is the cause of the social issues associated with fitness testing. Cale and Harris (2009) contest that since the fitness tests producers cannot agree on the best practices, then how can fitness testing be considered valid or reliable (p.93). They also point out how fitness test producers do not even agree on whether norms-based or criterion-based measurements are more appropriate. "Norms lead to

difficulties comparing children, and there is no scientific consensus on criterion-referenced ranges" (93, Cale, & Harris, 2009). They have found that while most fitness testing advocates say that testing motivates students, most students believe that fitness testing becomes boring and repetitive (Cale, & Harris, 2009). These findings lead one to believe that fitness testing used in a traditional manner could be its own worst enemy.

In the United States we have very similar issues. Looking at the two main fitness testing programs in the United States, the FITNESSGRAM and the Presidential Challenge Physical Fitness test, there are many of the same issues. FINESSGRAM uses Healthy Fitness Zones (HFZ) which are a set zone that students try and get their score into. These zones are criterion-referenced and based only on the score that each student should be able to do to achieve and maintain a healthy lifestyle (National Association for Sport and Physical Education, 2005). The Presidential Challenge Physical Fitness test awards are based on students who score in or above the 85th percentile of all students taking the test. This means that this norm-referenced awards system, honors students based on how well the rest of the student taking the test have done (National Association for Sport and Physical Education, 2005). The producers of these fitness tests also use different tests to measure the same aspect of fitness; students may do great on one test and poor on another test. Yet both tests are supposed to be measuring the same thing (Cale, & Harris, 2009).

Another major social issue associated with fitness testing stems from body composition testing. The issue involves skin fold body fat measurements, and the invasive procedures that must be used to get accurate readings (Malina, & Katzmarzyk, 1999). These invasive procedures have led to the Body Mass Index (BMI) becoming more widely used. This has been shown to be less accurate when determining an individual's body composition (Malina, & Katzmarzyk, 1999). Both of these methods can have negative effects on student's attitudes towards fitness testing when used inappropriately.

As stated before, skin fold measurements can be invasive and make students uncomfortable. This could be a reason that student have for not liking fitness testing. Skin fold measurements can also be inaccurate if the administrator is not skilled at taking the skin fold measurements (Malina, & Katzmarzyk, 1999). BMI can be a valuable alternative to skin fold measurements and is easier to figure than skinfold measurements, however BMI classifies students into categories that may not be accurate depending on their body type. This can lead to students getting the wrong impression about their body composition and physical appearance (Malina, & Katzmarzyk, 1999). This could cause students to become insecure about their bodies and even more uncomfortable about performing other fitness test in front others or at all.

The last social issue that will be discussed is the ability of the individual test components to evaluate what they are supposed to be evaluating. Cale and Harris (2009) describe fitness testing as being "plagued by severe limitations and the appropriateness, validity, and reliability of some fitness test and fitness test batteries for use with children" (p.92). A major problem they pointed out was with the multistage aerobic fitness test also known as the "beep test" or "PACER test", which they found "significantly underestimated children's peak VO2" (p.92). They say that this is caused because many of the tests used in popular fitness testing programs have been designed for use with adults and not children (Cale, & Harris, 2009). Cale and Harris show that by using child-specific formulas to calculate peak VO2 using a multistage aerobic fitness test, the percentage of error can be reduced, however the test still does not give an appropriate idea of the student's aerobic fitness level (Cale, & Harris, 2009). Despite all of this information the multistage aerobic fitness test or "beep test" is one of the most commonly used fitness test (Cale, & Harris, 2009).

Similar problems have been found with the sit-up/curl-up test, and the push-up and pull-up test (Cale, & Harris, 2009). The results of a study done by Saint Romain and Mahar, (2001) to find the

reliability of exchanging the push-up test for the modified pull-up test or vice versa, have shown that the agreement between the two tests to measure upper body strength is very low (Saint Romain, & Mahar, 2001). These two tests that are frequently used to measure upper body strength have been shown to get completely different results with the same students. To use the number of push-ups a student can do as a predictor for the number of pull-ups they can do seems to be almost impossible. Many physical educators only employee one of these two tests for upper body strength so the association between the two tests is not as important. Many physical educators say that they offer little practice leading up to the test. Romin and Mahar suggest "that practice from a fitness unit may help improve reliability of the push-up and modified pull-up scores" (p.78). Improving a student's ability to perform on fitness test could help student become more comfortable with fitness testing. Increased comfort and success could lead to more students enjoying physical fitness and physical education.

Fitness testing when used inappropriately can be detrimental to the success of a physical education program. Charles Corbin stated that "when institutional testing is done well, there are many potential benefits; however, it is not without its negatives" (S77, Corbin, 2010). This comment came after his work on the very detailed "Texas Youth Fitness Study", which looked into fitness testing practices across the state of Texas. One of their major conclusions was that fitness testing is a very dangerous "double edged sword" (S75, Corbin, 2010). This is because physical fitness testing can have great positive impacts and great negative impacts depending on how it is used. Designing a fitness testing program that is integrated into the design of your curriculum, so that you are teaching concepts that will have a direct effect on the results of the test is important so that students see that what they have done during class has had a positive effect on them (Corbin, 2010).

Fitness testing is the cause of many social issues and may be credited with many of the stereotypes associated with physical education today. However with the proper precautions and some

insight into things that may arise because of fitness testing many issues can be avoided. Proper planning and the incorporation on fitness testing into a physical education curriculum can do a lot to help make fitness testing more enjoyable for the student and the teacher. This could lead to physical education becoming more enjoyable as well, which may be a key to promoting a healthy active lifestyle.

Factors Affecting Results

To say that a physical education program is the cause of the changes seen in fitness testing scores is a very bold statement that cannot easily be proven. However with proper testing procedures it may be possible to support the theory that a physical education program can contribute to the positive changes seen in student fitness scores. There are many things that can influence scores and some of those things can be controlled. This section of the paper will first cover how the maturation of students can cause changes in fitness scores. Secondly, this section will cover how changes in the environment can affect the outcomes of testing based on influences that are not fitness related. Lastly, this section will cover how testing inconsistencies can affect results and how proper testing procedures and protocols can help prevent skewed test outcomes.

The natural growth and maturation of students can impact scores on fitness test. Maturation can cause a change in the results of most fitness test. When students get taller, shorter, heavier, or lighter their ability to move their body changes (Malina, 2004). They may be able to move more or less efficiently and may develop changes in their absolute and relative strength. As they grow they also develop a greater knowledge of how their body works and the best ways to use that to their advantage (Malina, 2004).

An example of the effects of maturation on fitness testing would be a student who runs the mile at the beginning of the school year, and then during the school year grows six inches taller. In the spring when fitness testing is done again, the student will most likely have a better mile time; however is it because of improved fitness or increased stride length? Another example would be the same student who does the maximum number of push-ups during testing in the fall; however in spring that student can barely meet the minimum number of push-ups. Did that student get weaker or did their body get heavier because of their growth spurt and their relative strength did not keep up? These are both

examples of how growth and maturation can affect the outcomes of fitness testing. There does not seem to be any solution to account for these situations, presented in the research, because growth and maturation are natural, expected and unavoidable (Malina, 2004).

Environmental changes can have just as much of an impact on the results of fitness testing however some of those effects can be avoided. When changes are seen in the results of fitness testing from a pretest to a post test, the circumstances in which the changes happened must be examined (Rothstein, 1984). Environmental changes can include everything from time of day to testing location. Time of day is important; however it may not be able to be avoided because of student class schedules and facility availability. Students' attitudes can be affected by the time of day and where P.E. falls into their schedule. If P.E. happens to be right after lunch they may be full and somewhat lethargic. If P.E. falls right before lunch they may be hungry and worried about food. These can both affect their attitudes towards the fitness testing. Fatigue is also important to consider, students at the end of the school day may be more fatigued than students at the beginning of the school day, which can affect their performance. The most important thing to consider with time of day is that you are doing your pre and post testing at the same time of the day so that the effect is potentially the same during both tests (Rothstein, 1984).

While it can be difficult to account for the effect that the time of day has on students, one issue that can be more controlled is the time of year effect. Doing fitness testing at the beginning and end of the year/semester is normal. However testing that happens at the end of the year/semester normally falls near some sort of school break or holiday. This means that students have either just returned from or are just getting ready to have an extended period away from school. If they are just returning from this break, it is important to give them time to get familiar with the testing procedures again and give them time to prepare mentally and physically for the test (Rothstein, 1984). If this is not possible,

because of time constrains it may be wise to perform you testing before the break, this way the students do not lose all that you have taught them before they see the results (Rothstein, 1984). However performing testing before a break or near a holiday can have its own effects on the results, because the student may be focused more on their impending vacation than the fitness testing. It may be wise to do testing before the end of the year/semester and give the students buffer time, so they can think about the testing and not the break (Rothstein, 1984).

The next type of environmental factor is more controllable than some of the others. Controlling your testing environment can have a big influence on testing scores. For example if you do the mile run on the track at the beginning of the year/semester and in the gym at the end of the year/semester, there will be implications on your scores. Things like this apply to all fitness tests and should be accounted for as much as possible. Providing the same setting for both the pre and post tests is vital to having valid results (Mahar, & Rowe, 2008). If you cannot recreate the same setting then expect to see a difference in your testing results (Mahar, & Rowe, 2008).

The last environmental factor that will be discussed in this section is the outside environmental influence. The main focus will be on non-physical education related physical activity and the effect that it can have on fitness testing. Research done in the United Kingdom shows that "the majority of young people's physical activity occurs outside school" (321, Cale, & Harris, 2006). This means that while physical education programs do all they can during school, students who get enough physical activity usually get it outside of school. However since most students do not get much physical activity outside of school and a push for physical education intervention programs is needed outside of the school setting (Cale, & Harris, 2006). However this research is concerned with the effects of physical education on the fitness scores of students. Students who participate in physical activity outside of school may well have higher fitness testing scores to begin with and therefore the results on both the pre and post test

will be affected (Cale, & Harris, 2006). This is an environmental factor that is not easily controlled. However, while we may not be able to account for the higher level or improved physical fitness of every student, improvement is the ultimate goal whether it comes from physical education or outside sources.

The last portion of this section will cover testing inconsistencies and how they can affect fitness testing results. Training to become a test administrator and properly administer the Presidential Challenge Physical Fitness test and the FITNESSGRAM fitness testing programs is fairly simple. The Presidential Challenge website describes the qualifications to be an instructor for their physical fitness testing program as "instructors are often P.E. or gym teachers—but not always. In fact, you don't have to be a certified teacher to administer any of our challenges. Any adult can do it as long as it's done in accordance with our guidelines, using discretion" (The President's Council on Fitness, Sports & Nutrition, 2011). They go on to explain how instructors work with coordinators who provide the instructors with the materials needed to conduct the testing (The President's Council on Fitness, Sports & Nutrition, 2011). This means that the only real requirement to administer the battery of test is that you are an adult and have the information. The FITNESSGRAM website explains their training process as "training opportunities for Fitnessgram are available with each Fitnessgram product purchase and do not require face-to-face training" (Human Kinetics, 2011). With these lenient stipulations for being a test administrator, the test producers are not providing a very scientific method of collecting the scores.

Administrator error can be a cause of discrepancies in fitness testing results. This means that the test administrator performed the test in a way that possibly affected the outcomes. This could lead to major issues in proving the effectiveness of you physical education programs. However during the "Texas Youth Fitness Study" researchers found that "Teacher administered, criterion-referenced health-related physical fitness test appear to be reliable and valid" (S29, Morrow Jr., et al, 2010). They did go on to say that "because reliabilities and validities can increase with training, it is important when

considering large-scale testing to conduct widespread training for those administering the test" (S29, Morrow Jr., et al, 2010). While the data that was collected during this study was good, they explain that during this study they performed training for their test administrators to make sure that everyone was testing the correct way (Morrow Jr., et al, 2010). Since training of this sort is not common, the majority of test administrators may not collect the data in the same way, which makes it difficult to compare results and repeat the same reliability results as found with Morrow Jr., et al (2010).

Another problem that can affect the validity of fitness testing results is the use of the different tests from pretest to posttest. An example of this would be using the mile run/walk to evaluate cardiovascular fitness during the pretest but the PACER test during the posttest. Both evaluate cardiovascular fitness and research has proven that the tests are very similar when determining cardiovascular fitness; however since they are different test they should not be interchanged (Beets, & Pitetti, 2006). Using the same test to evaluate for both pre and post test will help keep the results valid and in turn can help show that a P.E. program could be affecting student's fitness.

Many factors can affect the results of fitness testing and the more of those that can be accounted for, the more valid the test becomes and the more a P.E. program can promote itself as having affects on students physical fitness. It is impossible to account for every factor that could affect results, but the more emphasis that is put on eliminating those factors the better results you will see at the end.

Conclusion

This review of previous literature has looked at three main areas; 1) the history of fitness testing in the United States; 2) social issues associated with fitness testing; and 3) factors affecting results other than student fitness. Throughout this review a common theme has emerged, the idea that fitness testing does not seem to be the problem, yet the use of fitness testing is maybe where the problem lies.

Some research showed that in its current state the use of fitness testing was more of hindrance on physical education, than a promotion tool for healthy active lifestyles. The study described the beliefs that advocates have about fitness testing as "testing motivates young people, enhances learning and promotes positive attitudes" (95, Cale, & Harris, 2009) while students feelings toward fitness testing were reported as "repetitive and boring" (95, Cale, & Harris, 2009). These ideas contradict each other; however the fact is that if students do not enjoy fitness testing, then the idea that fitness testing motivates students may be wrong.

Many of these problems may have their roots historically. The fact that there are two major organizations that are promoting two entirely different fitness testing programs that supposedly measure the same thing is alarming. It is understandable that there will be disagreements when it comes to the science involved and that not everyone is going to agree on the best practices. However the science behind these tests should have evolved more than where it seems to be today and there should be more of a consensus of at least which test practices are better for students than others. This should not just be from a physiological stand point either. There should be psychological factors taken into account also. These should consider the age, gender, and the physical condition of the students that will be performing the test. However the science is not to that point and physical educators need some way to evaluate progress in their programs.

As stated before fitness testing does not seem to be the problem. The use of fitness testing is where the problem lies. This means that fitness testing administrators need to evaluate their practices and work to build fitness testing into their curriculum. They need to work to figure out the best testing programs for them and their students and stick with it so that they can evaluate progress longitudinally. They need to make sure that they work with other test administrators to have trainings so that every student is tested in the same manner. Fitness testing is a valuable tool if it is used appropriately. This is why consistency and conformity is so important when it comes to testing practices.

Fitness testing is not the enemy and it is not the problem. Fitness testing can be very beneficial to students, educators, schools, and communities if it used appropriately. However used unwisely fitness testing can create major problems for everyone involved and do the exact opposite of what the goal of physical education should be, to promoting healthy active lifestyles. Fitness testing does have major implications on the lives of everyone involved, it is important that these implications remain relevant and positive.

References

Beets, MW., & Pitetti, KH. (2006). Criterion-referenced reliability and equivalency between the pacer and 1-mile run/walk for high school student. *Journal of Physical Activity & Health*, *3*(S2), S21-S33.

Cale, L., & Harris, J. (2006). Interventions to promote young people's physical activity: issues, implication and recommendations for practice. *Health Education Journal*, *65*, 320-337.

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

Centers for Disease Control and Prevention. (2010, August 6). Vital signs: state-specific obesity prevalence among adults - United States, 2009. *MMWR Morbidity and Mortality Weekly Report*, *59*(30), 951-955.

Centers for Disease Control and Prevention. (2011, January 21). CDC grand rounds: childhood obesity in the United States. *MMWR Morbidity and Mortality Weekly Report*, *60*(2), 42-46.

Corbin, CB. (2010). Texas youth fitness study: a commentary. *Research Quarterly for Exercise and sport*, *81*(S3), S75-S78.

Human Kinetics, . (2011). *Fitnessgram*. Retrieved from http://www.fitnessgram.net/home/

Keating, XD., & Silverman, S. (2004). Teachers' use of fitness testing in school-based physical education programs. *Measurement in Physical Education and Exercise Science*, *8*(3), 145-165.

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

Malina, RM., & Katzmarzyk, PT. (1999). Validity of the body mass index as an indicator of the risk and presence of overweight in adolescents. *The American Journal of Clinical Nutrition*, *70*, 131S-136S

Malina, RM. (2004). *Growth, maturation, and physical activity*. Champaign, IL: United States: Human Kinetics.

Morrow Jr., JR., Martin, SB., & Jackson, AW. (2010). Reliability and validity of the Fitnessgram: quality of teacher-collected health-related fitness surveillance data. *Research Quarterly for Exercise and sport, 81*(S3), S24-S30.

National Association for Sport and Physical Education. (2005). *Physical education for lifelong fitness.* Champaign, IL: United States: Human Kinetics.

Sturgeon, J, & Meer, J. (2010, June). The first fifty years 1956-2006; the president's council on physical fitness and sports revisits its roots and charts future. *The President's Council on Physical Fitness and Sport; 50 years Promoting Health and Fitness*, 40-66.

Noffsinger, J. (2005, November 20). *Phys ed cuts may leave children's health behind.* Retrieved from http://www.foxnews.com

PE4life . (2011). *Pe4life: about us.* Retrieved from http://www.pe4life.org/about-us/

Pe4life. (2002, February 3). U.S. kids have shorter life expectancy than parents because of obesity. *News from Obesity , 2*(5), Retrieved from www.pe4life.com

Placek, J. (1983). Conceptions of success in teaching: Busy, happy and good? In T. Templin & J.Olson (Eds.), Teaching in physical education 46-56. Champaign, IL: Human Kinetics.

Plowman, SA., Sterling, CL., Corbin, CB., Meredith, MD., Welk, GJ., & Morrow Jr., JR. (2006). The history of Fitnessgram. *Journal of Physical Activity & Health, 3*, S5-S20.

Rothstein, AL. (1984). *Research design & statistics for physical education.* Prentice Hall.

Saint Romain, B., & Mahar, MT. (2001). Norm-referenced and criterion-referenced reliability of the push-up and modified pull-up. *Measurement in Physical Education and Exercise Science, 5*(2), 67-80.

The President's Council on Fitness, Sports & Nutrition. (2011). *Physical fitness test.* Retrieved from http://www.presidentschallenge.org/challenge/physical/index.shtml