TRENDS IN TEXAS YOUTH FITNESS TESTING:

A LONGITUDINAL LOOK AT STATEWIDE REPORTING

A Dissertation

by

MEGAN CORINNE HAMILTON GIEBERT

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Kelly L. Wilson
Committee Members,	Susan E. Ward
	E. Lisako J. McKyer
	Jean A. Madsen
Head of Department,	Richard Kreider

December 2016

Major Subject: Health Education

ABSTRACT

Since 2007, Texas has mandated physical fitness assessment and reporting requirements for students in public school districts. Results from these yearly assessments are summarized in aggregate form as publicly accessible data, available for monitoring, analysis, and evaluation. The purpose of this research was to examine longitudinal trends in reporting to inform on the reach, impact, and value of routine statewide fitness data collection. This research manifested itself in two broad studies, each encompassing data from students in Grades 4, 6, 7, and 9.

The first study used seven years of data to examine reporting frequency over time and to identify predictors of school district reporting compliance. Reporting rates among districts were high each year, but several district characteristics strongly correlated with exemplary reporting frequency. The strongest predictors of reporting frequency were: teacher turnover rate, fitness performance, economic status, and among older grades, attendance rate. The second study used three years of data to measure the longitudinal relationship of achievement rates on two common fitness tests, which were proxies for aerobic capacity and body composition, among various student subgroups. Some correlation between these achievement rates was expected, but a very high correlation was detected among Grade 6 girls and persisted over the time period. Measuring how the correlations endure over time offers valuable insight about the unique contributions that each test can offer to students, highlighting how the mathematical calculations of the tests can impact the students attempting them.

Together, the findings from these studies have implications for students who are required to take annual fitness tests; for parents who help interpret students' performance; for teachers who train to administer tests and report the results; and for researchers who develop and evaluate tests. Although considered beneficial and informative, statewide fitness testing is a large endeavor that relies on comprehensive student participation and dedicated adherence from school and district staff. These results can help health professionals determine how to optimize statewide fitness testing and use it to further a culture of school-based physical activity and health.

ACKNOWLEDGEMENTS

I would like to thank my committee members for their tremendous assistance with completing this work. Dr. Wilson, my committee chair, helped me select this area of study, locating it at the intersection of school health and physical activity; since then, I have appreciated her patience and guidance during every meeting, presentation, and feedback session. Dr. Ward and Dr. McKyer consistently aided with answering research questions, offering analytical insights, and organizing my thoughts, while Dr. Madsen provided interpretations that broadened the research beyond just health specialists.

Additionally, I would like to thank my immediate department supervisors, Bruce Hanik and Saurabh Singh, for their support in my academic endeavors. My colleagues and peers in the Department of Health & Kinesiology also helped make this project possible, listening to insights and negating moments of doubt. I have been grateful for their encouragement.

Finally, I thank my family; in particular, Matt, Mom, and Dad. At every mile you believed in me.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of four members: Committee chair, Dr. Kelly Wilson, of the Department of Health and Kinesiology; Dr. Susan Ward, of the Department of Health and Kinesiology; Dr. Lisako McKyer, of the Department of Health and Kinesiology and the School of Public Health; and Dr. Jean Madsen, of the Department of Educational Administration and Human Resource Development. All work for the dissertation was completed independently by the student.

Funding Sources

There are no outside funding contributions to acknowledge related to the research and compilation of this document.

NOMENCLATURE

AP	Academic Performance
BMI	Body Mass Index
HFZ	Healthy Fitness Zone
NIZ	Needs Improvement Zone
PA	Physical Activity
PE	Physical Education
PFAI	Physical Fitness Assessment Initiative
PACER	Progressive Aerobic Cardiovascular Endurance Run
PEIMS	Public Education Information Management System
SY	School Year
TEA	Texas Education Agency

TABLE OF CONTENTS

Page

LIST OF TABLES

1.0 INTRODUCTION

The 2008 Physical Activity Guidelines for Americans recommend 60 minutes of physical activity (PA) per day for children and adolescents, which should consist of aerobic, muscle- or bone-strengthening skills, activities, and playtime (United States Department of Health and Human Services [USDHHS], 2008). The benefits of PA for youth are well-established and highly promoted (Centers for Disease Control and Prevention [CDC], 2013; World Health Organization [WHO], 2016), but PA levels among American children fall below the recommended amounts (National Physical Activity Plan [NPAP], 2014).

Schools are a strategic setting for increasing youth PA amounts through a variety of school, community, and policy efforts (CDC, 2013). The CDC and the Association for Supervision and Curriculum Development (ASCD) identify Physical Education (PE) and PA as core components of the Whole School, Whole Community, Whole Child Model (Lewallen, Hunt, Potts-Datema, Zaza, & Giles, 2015), which portrays the role of health in any student's education. Schools also have a direct incentive to foster and provide PA due to substantial research suggesting a positive correlation between PA amounts and student academic performance (AP) (CDC, 2010; Rasberry et al., 2011; Singh, Uijtdewilligen, Twisk, van Mechelen, & Chinapaw, 2012).

School-based PE is an institutionalized form of youth PA promotion. When implemented well, PE can have long-reaching, positive effects on youth health (CDC, 2013; Lewallen et al., 2015; NPAP, 2016). According to the Society of Health and

Physical Educators (SHAPE America), components of high-quality PE include well-developed curricula, appropriate instruction from credentialed teachers, supportive school policies and environment, and student assessment. One type of formative student assessment is fitness testing (SHAPE America, 2015).

Fitness testing can indicate whether or not students are obtaining PA's health benefits and identify areas of fitness in which students can improve. Debate exists surrounding the value of fitness testing (Cale, Harris, & Chen, 2007; Rowland, 1995), but academic consensus regards fitness testing as a constructive part of PE classes if used to a) formatively evaluate students (Wiersma & Sherman, 2008), b) educate students about health and fitness concepts (SHAPE America, 2015) and c) offer personally relevant feedback to students (Meredith & Welk, 2013). The CDC recommends assessments as "an ongoing, vital part of the physical education program" (CDC, 2013, p. 13) assuming that testing is administered in accordance with national or state PE standards. Youth fitness assessment can also be completed on a statewide scale if thoroughly planned in advance (Morrow & Ede, 2009).

Morrow and Ede (2009) name three primary reasons for the increased interest in school-based fitness assessment: the rise of childhood obesity, the connections between PA and AP, and the long-term outcomes of youth fitness and PA habits at an early age. Coupled with consistent national support and gradual institutionalization, fitness testing is popular in schools despite a general decrease in PE requirements (CDC, 2015c). In a nationwide 2014 survey, 63% of American schools administered fitness tests at least once during their academic school years. After assessment, 72.0% of those schools compared fitness test results with national, state, or local criterion-referenced standards (CDC, 2015b), which adheres to recommended assessment implementation procedures (Meredith & Welk, 2013).

The state of Texas has implemented many policies to normalize school-based youth fitness testing, analysis, and reporting. As of 2012, Texas was one of only 14 states to require all its public school districts to annually complete fitness testing (Presidential Youth Fitness Program [PYFP], 2014). Required statewide fitness testing began in 2007 after the passage of Texas Senate Bill (SB) 530 modified the Texas Education Code to require fitness assessment for all students in Grades 3 to 12. After assessments are conducted, these fitness results must be annually reported to the Texas Education Agency (TEA) so that data can be publicly available for research purposes (§ 530, 2007). The process of aggregating and reporting these results is called the Physical Fitness Assessment Initiative (PFAI).

Texas's PFAI procedures have created many avenues of research, analysis, and discussion. The most comprehensive report using PFAI data appeared in 2010, when *Research Quarterly for Exercise and Sport* published a special supplement issue covering the initiative's inaugural year and identifying future directions of the project (Cooper et al., 2010). Since then, scholars have also used PFAI data to examine the PA-AP relationship in fuller detail (Van Dusen, Kelter, Kohl, Ranjit, & Perry, 2011) and to document health-related fitness by geographic and legislative region (Janak et al., 2014). Furthering the efforts to monitor fitness, the TEA annually partners with The Cooper Institute to report summaries of fitness trends as a way to establish baselines for further research (Allums-Featherston, Bai, & Welk, 2014b). However, very few longitudinal analyses of PFAI data have been conducted since the PFAI's inception.

This dissertation research seeks to utilize PFAI data to answer two broad questions about statewide youth fitness assessment. The first question relates to the reach and scope of reporting compliance among Texas school districts: what are the characteristics of Texas school districts who consistently report, fully complying with PFAI data protocol? Although many PFAI reports have been gathered, it is not clear which school districts reported every year and where improvements in reporting compliance may be made. Answering this question could reveal predictors of reporting compliance and confirm that PFAI efforts are uniformly reaching all Texas school districts.

The second question concerns the types of tests used in fitness testing: what is the relationship over time between the two most popular fitness tests used among Texas students? These two tests are the Body Mass Index (BMI), which approximates body composition, and the Progressive Aerobic Cardiovascular Endurance Run (PACER), which estimates aerobic capacity. Both tests are from the battery FITNESSGRAM®. Although the tests are expected to correlate to some extent because of their interacting health effects, the strength of the relationship over time has not yet been studied among Texas youth. Body composition and aerobic capacity are two distinct health indicators, despite their interactions. Determining how these two components relate can provide insight on how to refine test components to most comprehensively reflect student health.

In order to answer these questions, two studies were developed and conducted, both using the publicly available collection of PFAI data. The remaining sections of this document detail the context, background, procedures, and results of these efforts. Section Two provides a brief history of fitness testing in the United States and chronicles the changes in philosophies behind testing as a practice. It also describes the political environment that helped institutionalize statewide fitness testing in Texas, prompting the appearance of PFAI data and related scholarship, and outlines the aims of Studies A and B. Section Three addresses the first research question; the purpose of Study A was to examine reporting frequencies of Texas public school districts and determine common characteristics among districts that demonstrate exemplary reporting compliance. Section Four describes Study B, which offered context to the BMI and PACER fitness tests and sought to measure the relationship over time between the scores on these tests according to student subgroup. Finally, Section Five summarizes the conclusions from Studies A and B and offers recommendations for ensuring that statewide PFAI procedures maximally serve students and schools.

2.0 LITERATURE REVIEW

2.1 Introduction

Physical fitness is defined as "the ability to carry out daily tasks with vigor and alertness, without undue fatigue, and with ample energy to enjoy leisure-time pursuits and respond to emergencies" (Centers for Disease Control and Prevention [CDC], 2015a). Physical fitness testing measures the extent of this fitness ability by assessing five broad health-related fitness indicators: cardiorespiratory endurance, muscular strength, muscular endurance, flexibility, and body composition. National fitness batteries are constructed to guide the assessment process. The purpose of this literature review was to provide a brief history of national fitness testing in youth populations, to explain the paradigms that shape the current fitness test batteries, and to contextualize the prominence that fitness testing has in Texas school districts.

2.2 National Youth Fitness Testing History and Philosophy

Youth fitness testing has been common in the United States since the 1950s; this decade witnessed the first national fitness assessment conducted in American public schools. Interest in fitness testing has only expanded since the 1950s, but many questions regarding what fitness tests should consist of have been proposed, debated, tested, refuted, and evaluated. The history of youth fitness testing has been satisfactorily covered in previous literature (see Morrow, Zhu, Franks, Meredith, & Spain, 2009; Pate, Welk, & McIver, 2013), but a review of major concepts will minimize confusion in terminology and philosophies surrounding the current fitness testing practices.

The catalyst for national interest in youth fitness can be traced to a 1953 publication by Kraus and Hirschland titled "Muscular Fitness and Health," which suggested that contemporary American youth were largely unfit for military service (Kraus & Hirschland, 1953). This finding prompted executive action. Being in the early Cold War years, fitness at the time was considered necessary for national security. The Eisenhower administration formed the Presidential Council on Youth Fitness as a way to assess, monitor, and begin to improve this youth fitness status. This council collaborated with the American Association of Health, Physical Education, and Recreation (AAHPER) – today known as the Society of Health and Physical Educators (SHAPE America) – to create a national Youth Fitness Test (YFT). The YFT was first pilot-tested in 1957 with a sample of 8,500 American students and would eventually become the President's Challenge (Morrow et al., 2009; President's Council on Fitness, Sports, and Nutrition [PCFSN], 2016).

Because the original YFT was meant to gauge military preparedness, these fitness assessments were grounded in motor skills such as throwing, sprinting, and endurance (Morrow et al., 2009). In 1965, an awards component was attached to the testing process, known as the Presidential Physical Fitness Award, which honored the students who achieved the highest scores. This percentile-based award tended to go only to the most athletically gifted students (Pate et al., 2013; PCFSN, 2016). Because the fitness assessment program rewarded motor ability, the unintentional message this paradigm conveyed was that fitness was an exclusive achievement. Fitness, it seemed, could belong only to athletes.

Despite these concerns about athletic exclusivity, consistent national support and advocacy from AAHPER helped the YFT grow into an institutionalized norm for students until the mid-1970s, when "mounting dissatisfaction" (Plowman et al., 2006, p. S7) with the YFT test protocol was too great to ignore. Several states began developing their own test batteries, starting as variations of YFT tests but incrementally adapting to meet the needs of specific states (Morrow et al., 2009). The rationale behind fitness testing gradually shifted from a skills-based concept to a health-based one; measuring fitness abilities was considered useful because of the health benefits children could enjoy by achieving and maintaining fitness. By the 1980s, the state-specific changes birthed a new paradigm centered in health-related fitness (Morrow et al., 2009; Plowman et al., 2006).

In 1975, the organization AAHPER – contemporarily known as the American Association of Health, Physical Education, Recreation and Dance (AAHPERD) – was actively considering changes to the YFT (Plowman et al., 2006). It attempted to accommodate the health-related paradigm by offering a second test battery, the Health-Related Physical Fitness Test (HRPFT), in conjunction with its YFT (Morrow et al., 2009). Due to its well-established reputation, the YFT remained the more popular test option of the two. In 1982, a second professional agency emerged with a promising new test battery. The Cooper Institute's FITNESSGRAM® test battery began to gain traction as a reasonable alternative to both the YFT and HRPFT. A full history of FITNESSGRAM is documented elsewhere (see Plowman et al., 2006), but its origins are key to understanding the present-day support for fitness testing in Texas.

Throughout the 1980s, attempts to merge the test batteries were thorough yet unsuccessful (Morrow et al., 2009; Plowman et al., 2006). By the end of the 1980s, there were three prominent national youth fitness tests: the YFT, created by AAHPERD and receiving presidential endorsement; the Physical Best (PB), another of AAHPERD's health-related fitness tests that was supported by United States Department of Health and Human Services (USDHHS); and the FITNESSGRAM, from The Cooper Institute. Each battery was situated somewhere on the continuum between health- and skill-related fitness (Pate et al., 2013), offered slightly different components (Morrow et al., 2009), and had its own merits. The YFT was the most visible of the three programs and contained a prominent recognition component. Simultaneously, AAHPERD's PB battery incorporated more health-related tests, offered a more inclusive evaluation system, and contained helpful public health surveillance data. Finally, FITNESSGRAM had developed a positive reputation as a health-related assessment tool and became popular in schools due to its well-received style of fitness feedback (Pate et al., 2013).

Schools and their parent districts were left to choose among the three national batteries. Statewide batteries were also available in some areas. Ultimately a compromise between national and professional organizations was reached in 2003, which minimized confusion for the districts and schools attempting to implement youth fitness testing. In that year, FITNESSGRAM became the battery of choice for the Presidential Youth Fitness Program (PYFP), part of the PCFSN's President's Challenge program (PCFSN, 2016). Because the President's Challenge program had expanded to include adults, the PYFP became the term for the youth-based component of fitness assessment. In 2007, Mood, Jackson, and Morrow summarized the twenty-first century's status quo: "Currently, AAHPERD does not publish a youth fitness test, FITNESSGRAM works jointly with AAHPERD on youth fitness assessment and educational programs, and the [PCFSN] maintains both performance-based and health-fitness test batteries with its President's Challenge program" (Mood, Jackson, & Morrow, 2007, p. 222).

Three characteristics of FITNESSGRAM helped elevate it as a nationally renowned test battery. First, the test battery's developers prioritize using the most up-to-date software and technology, which eases test implementation and reporting. This technological commitment makes the tests more feasible to complete and summarize across large groups, streamlining both the assessment process as well as results communication (Plowman et al., 2006). Second, FITNESSGRAM seeks to recognize all students who meet a minimum level of fitness for their gender and age group. These minimum levels, called criterion-referenced standards (CRS), are benchmarks that represent the points at which students will start seeing health benefits from the level of fitness achieved; health in this context is defined as "minimal disease risk" and "the ability to carry on with tasks of daily life" (Plowman et al., 2006, S12). FITNESSGRAM was not the first assessment tool to use the CRS format, but by fully embracing them, it distinguished itself from other tests by attenuating the stigma of athleticism. Finally, further advertising the philosophy that fitness is achievable for everyone, FITNESSGRAM offers a secondary test battery called ACTIVITYGRAM® that allows students to track their daily physical activity (PA) amounts.

ACTIVITYGRAM is a three-day PA recall, which enables students to assess their PA levels in addition to their fitness levels (Plowman et al., 2006). Although meant to be a natural extension of FITNESSGRAM, ACTIVITYGRAM is not exclusively recommended as part of the PYFP, which continues to promote fitness assessment over PA. Instead, PA tracking appears in the Presidential Active Lifestyle Awards (PALA) program, another component of the President's Challenge (PCFSN, 2016).

The emerging interest in PA tracking tools such as ACTIVITYGRAM is leading to a fieldwide discussion about the relevance of fitness assessment in schools when compared with PA assessment. Catalyzing this discussion was an editorial from Rowland (1995), written in response to an article by Corbin, Pangrazi and Welk that described the rise of a new lifestyle-health based fitness model (1994). Stemming from a basic understanding of the new model, Rowland asked: to what extent does fitness assessment resemble the lifestyle behaviors of active, healthy children? Children, the editorial implied, stay active by riding bikes, swimming, chasing, joining team sports, and romping on playgrounds. If children can be physically active without being physically fit, then is fitness assessment necessary in schools? Corbin, Pangrazi, and Welk responded to the Rowland editorial, proposing that fitness testing is compatible with the new activity-based model because fitness testing can both educate students about physical health and spark interest in PA. Rather than abandoning fitness testing altogether, the authors clarified, the field should abandon using fitness testing in irresponsible ways, such as grading children on their performances or forcing children to engage in PA as punishment (ie. having to run extra laps if tardy) (Corbin, Pangrazi, &

Welk, 1995). These authors' succinct guidelines defended fitness assessment as a personally relevant teaching tool for students to use. Despite these clarifying remarks, Rowland's questions continue to be entertained. In 2007, Cale, Harris, and Chen articulated the most thorough argument against fitness testing, but fitness testing advocates continue to debate these points. As evidenced by the commentaries published in response to Cale, Harris, and Chen, some advocates grow weary of the debate (2007).

Although the question continues to be explored, a truce-like consensus in the field occurred in 2008 when *Measurement in Physical Education and Exercise Science* published a special issue on the benefits of fitness testing and the best practices involved in it (Liu, 2008). This issue covered multiple professional perspectives regarding youth fitness assessments: the pedagogical benefits of testing (Welk, 2008; Silverman, Keating, & Phillips, 2008), guidelines for maximizing reliability and validity of testing (Mahar & Rowe, 2008), and tips for promoting student learning and enjoyment (Wiersma & Sherman, 2008). These articles indicated professional support for fitness testing but also discouraged inappropriate fitness testing trends, so that students perceive testing as a fun, useful, and informative experience.

2.3 Youth Fitness Testing in Texas

The use of fitness assessments varies by state, but statewide physical fitness testing – mandated through statewide policy or education code amendments – is becoming popular in public schools. As of 2010, 17 states mandated or were considering mandating statewide fitness testing for public school students (Morrow & Ede, 2009). Texas is one of the leaders in this approach. Not only does Texas regularly assess its students' fitness levels, but it is also one of only 27% of states that reports their fitness results to an education agency. Of these 27%, Texas is also among the 50% that publicly report the aggregated findings (Lee, Nihiser, Fulton, Borgogna, & Zavacky, 2013).

One contributor to Texas's advanced approach to fitness testing is that Texas's fitness batteries have historically subscribed to the health-related fitness paradigm. The 1976 Texas Physical Fitness Motor Ability Test (TPFMAT) drew a distinction between motor ability and physical fitness – one of the first tests of its kind to do so (Morrow et al., 2009). Another test developed in Texas was the Fit Youth Today battery, appearing in 1986 and using the more health-centered CRS methods (Morrow et al., 2009; Pate, 1989). These were the first glimmers of paradigmatic unity across the state and may have led to Texas prioritizing fitness testing as a health initiative for students.

Another influential factor in Texas's comprehensive fitness testing is The Cooper Institute. This Dallas-based agency was founded in 1970 and generates data and research on the effects of fitness and exercise on Americans' health. Their work seeks to support and influence the promotion of aerobic physical activity, including for youth, by empirically supporting the benefits of lifelong PA and achievable fitness. The Cooper Institute was responsible for developing FITNESSGRAM, the battery currently used in the PYFP (The Cooper Institute, 2014b). The Cooper Institute is not singlehandedly responsible for the culture of youth fitness assessment in Texas, but the agency was almost certainly a driving factor behind its promotion. For example, Morrow, Martin, Welk, Zhu, and Meredith (2010) highlighted four legislative bills passed between 2001 and 2007, which all impact school health initiatives in Texas public schools. Each was authored by Texas State Senator Jane Nelson, who represents the Denton and Tarrant county areas, close to where The Cooper Institute is headquartered. The Cooper Institute's advocacy on behalf of school health played a considerable factor in the fitness-focused nature of Texas classrooms. The fact the bills passed reflects an evolving culture surrounding youth fitness in Texas. These senate bills helped propel Texas into being a leader in youth fitness according to the national standards of school-based physical activity and fitness. In addition, evaluations of the bills' effectiveness are ongoing and apparent (Barroso et al., 2009; Kelder et al., 2009).

Texas Senate Bill (SB) 530, signed into law in 2007, was the first legislative piece to directly address fitness testing. This statute introduced required fitness assessment and monitoring protocol for Texas public school students, specifically outlined in Subchapter C. According to this legislation, school districts became required to "assess the physical fitness of students enrolled in grades 3 through 12" (§ 530, 2007) on the following health-related fitness components: aerobic capacity, body composition, and muscular strength, endurance, and flexibility. All public school students in these grades, unless exempt due to disability or conditions that render the assessment unfeasible, were compelled to participate in this assessment. This legislative act also declared that school districts must report their findings annually in aggregate form to the Texas Education Agency (TEA) (§ 530, 2007).

The passing of SB 530 marked the beginning of a statewide Physical Fitness Assessment Initiative (PFAI). The PFAI was established to meet four objectives related to fitness in schools required by SB 530: assessment, measurement selection, result compilation, and associative comparison. All school districts were and continue to be expected to participate in the PFAI on a yearly basis. As of May 2016, seven years of data were archived, spanning from the 2007-2008 school year to 2013-2014. These records are available electronically and to the public, accessible on the TEA's website (TEA, 2016). The Texas PFAI "represents the most comprehensive statewide youth fitness initiative pursued to date" (Cooper et al., 2010, p. iii). The combined legislative and financial efforts created publicly accessible archives of fitness data, available for longitudinal analysis and comparisons. In short, as of 2007, Texas public schools found themselves at the heart of the fitness revolution.

Texas's PFAI data has been analyzed in previously published studies. The 2007 statute permitted agencies and entities to use PFAI data to research associations between fitness scores and other pertinent variables, specifically academic achievement, attendance, obesity, discipline, and school meal programs. Districts were allowed to collaborate with other agencies to complete these types of analyses (§ 530, 2007). Studies regarding fitness scores were conducted in Texas prior to the PFAI (Santiago, Roper, Disch, & Morales, 2013), but SB 530 led to an influx of research concerning Texas youth fitness.

2.3.1 Texas Youth Fitness Study and Texas Youth Evaluation Project

The largest published study using PFAI data was a collaboration called the Texas Youth Fitness Study (TYFS), which aligned with the objectives of SB 530. Ultimately, this study's lofty goals were to reduce childhood health issues through the promotion of physical activity and physical fitness in schools (Morrow et al., 2010). Bolstering the TYFS was a parallel project called the Texas Youth Evaluation Project (TYEP), which "promoted awareness across the state and helped build schools' capacity to continue tracking and promoting youth physical fitness" (Cooper, 2010, S79). While the former project yielded objective, systematically analyzed findings related to Texas youth fitness, the latter gathered resources and funding to assist schools in completing these fitness tests annually and complying with the new statewide policies. A detailed, grant-supported report on TYFS and TYEP was published in *Research Quarterly for Exercise Science*'s 2010 supplementary issue, which demonstrated the many ways that PFAI data can be used and interpreted. Feiden (2011), writing for the agency that helped financially support the TYFS, offers a concise summary of major findings from this endeavor.

The TYFS is an ongoing interpretation PFAI results, but as of 2016 no peer-reviewed publication of results has appeared that summarizes findings from more recent school years. However, The Cooper Institute and the TEA released updated findings in 2014, which described statewide trends in fitness achievement from the 2010-2011 to 2013-2014 school years (Allums-Featherston et al., 2014b).

2.3.2 Associations with Academic Performance

Academic performance (AP) is a popular concept to study because many schools are incentivized to improve student AP. Rather than cutting PE time to make room for instructional time, the CDC recommends expanding PA opportunities at schools to achieve both health and academic benefits (CDC, 2010). Using PFAI data, scholars have been able to partner with Texas school districts to match student fitness results with AP. Published in the *RQES* supplementary issue, Welk, Jackson et al. (2010) described the positive associations between fitness and AP in 2007-2008 PFAI data, including relationships between fitness and school attendance. A later contribution from The Cooper Institute summarized PA-AP associations observed in PFAI data for the 2011-2012 to 2013-2014 school years. Although this work remains unpublished, it set a firm succession of baseline findings regarding student fitness (Allums-Featherston et al., 2014a).

Other notable efforts to compare PFAI data with AP data come from Van Dusen et al. (2011), who found statistically significant positive associations between PA and AP in the 2007-2008 school year. Sampling from 99 districts and obtaining 13 usable records, this analysis constitutes a sample size of over 250,000 students in grades 3-11 across Texas (Van Dusen et al., 2011). Janak et al. (2014) also studied the PA-AP relationship, merging FITNESSGRAM scores with standardized test scores from the 2008-2009 school year. Using all available PFAI data, this study contained findings from over 2.5 million Texas students (Janak et al., 2014).

2.3.3 Future Areas of Study with PFAI Data

Although SB 530 stipulates that fitness scores should be used to measure associations with disciplinary problems, school meal programs, and obesity, research involving these relationships remains underdeveloped. School meal programs are often used as a proxy for socioeconomic status in PA-AP studies, but further research on fitness performance and the nutritional quality of school meal programs is desired.

Disciplinary problems are potentially harder to compare without violating student privacy. Because the Body Mass Index (BMI) can be used as either a fitness test or an obesity indicator, comparing the two is often fruitless. Nevertheless, SB 530 invites research to use these constructs to gauge the effects of fitness achievement, assessment, and monitoring in schools.

2.4 Conclusion

This literature review aimed to provide context to Texas's youth fitness assessment initiative. By tracing the national history of fitness assessment, describing evolutions in assessment philosophy, and naming key influences in Texas PFAI data collection and compilation, this review highlighted how PFAI data has already been utilized and proposed ways in which the data can be studied in future scholarship.

2.5 Specific Aims of Dissertation

The remainder of this section briefly describes the specific aims of Studies A and B. At the beginning of this record of study, no published work had endeavored to examine all seven years of publically accessible PFAI data at once. Therefore, the overarching purpose of Studies A and B was to summarize reporting trends among PFAI data over time.

2.5.1 Aim One

The aim of Study A was to examine the reporting compliance over time of Texas public school districts. Data from all seven years of PFAI reports were compiled to measure how frequently each district reported and to identify district characteristics that predicted reporting compliance.

The hypotheses for Study A were:

a. Null: There are no variables that predict frequency of district reporting.

b. Alt: There are some variables that predict frequency of district reporting.

2.5.2 Aim Two

The aim of Study B was to determine longitudinal associations between two popular fitness tests, which indicated student performance on aerobic capacity and body composition. Study B compared these tests' correlations over time within various student subgroups.

The hypotheses for Study B were:

a. Null: There is no variation by student subgroup in the longitudinal relationship between aerobic capacity and body composition.

b. Alt: There is some variation by student subgroup in the longitudinal relationship between aerobic capacity and body composition.

These studies are fully presented in Sections Three and Four of this document. Section Five summarizes general conclusions from this scholarship and reflects upon future directions for Texas youth fitness testing.

3.0 STUDY A: PREDICTORS OF FITNESS ASSESSMENT REPORTING COMPLIANCE IN TEXAS SCHOOL DISTRICTS

3.1 Introduction

Health-related fitness assessment is a prominent component of youth physical education (PE) programs and can provide feedback regarding students' achievement of both immediate and long-term health benefits (SHAPE America, 2015). In 2007, Texas passed Senate Bill (SB) 530, which mandated that all public school districts "annually shall assess the physical fitness of students enrolled in grades 3 through 12" (§ 530, 2007). This legislation also required all school districts to report their fitness assessment outcomes in aggregate form to the Texas Education Agency (TEA). When SB 530 passed, it marked the beginning of comprehensive fitness testing protocol in Texas school districts.

The statewide process of assessment, reporting, monitoring, and analysis is called the Physical Fitness Assessment Initiative (PFAI). All Texas public school districts are required to report PFAI data; however, no penalty exists for districts that do not comply with this policy. Peer-reviewed interpretations of PFAI reporting rates indicated that the majority of districts comply with reporting requirements: from 2007-2008 to 2009-2010, an average of 2.7 million Texas students were assessed each year, encompassing about 85% of all Texas public school districts annually (Cooper et al., 2010). Descriptions of PFAI efforts from 2010-2011 to 2013-2014, although not peer-reviewed, also documented high reporting rates (The Cooper Institute, 2014a).

Although these summaries suggest that PFAI reporting from 2007-2008 to 2013-2014 was high on a year-to-year basis, it is not clear which districts demonstrated consistent reporting compliance. Several questions have yet to be answered regarding reporting frequency: how many districts reported consistently over the seven-year period, and how many did not? What common features may exist among districts that fully adhered to SB 530's reporting requirements? What district characteristics serve as predictors of reporting frequency?

To explore the longitudinal reach and scope of Texas's youth fitness assessment initiatives, the purpose of this study was to examine PFAI data to measure districts' reporting frequency over time. Variables related to district size, district demographics, and district performance outcomes were analyzed to determine what characteristics, if any, predict district reporting frequency and indicate high compliance with the statewide initiative.

3.2 Methods

Three publicly available databases were compiled to complete this study. All records were retrieved from the TEA in January 2016:

1. Physical Fitness Assessment Initiative (PFAI)
2. Texas Education Agency Directory (AskTED)
3. Public Education Information Management System (PEIMS)

3.2.1 PFAI Data

The PFAI data summarized fitness performance by school district and described the fitness outcomes of Texas students enrolled in Grades 3 to 12. Any student who attended school on their classroom or school's day of fitness testing would have been included in aggregated PFAI reports for the specific school year. Fitness testing for these students was mandatory due to SB 530's statewide fitness assessment requirements, although exemptions were permitted for students with disabilities that prevented them from safely completing the fitness test (§ 530, 2007). The fitness tests that students completed were from FITNESSGRAM®, a nationally recognized test battery developed by The Cooper Institute, a nonprofit research institute located in north Texas. This battery assessed four broad areas of fitness – aerobic capacity, muscular strength and endurance, flexibility, and body composition – using six fitness tests (Plowman & Meredith, 2013).

When students attempted each fitness test, their individual scores fell within one of two categories: a Healthy Fitness Zone (HFZ), which suggested the student was fit enough to experience health-related benefits appropriate for their gender and age group; or a Needs Improvement Zone (NIZ), which suggested the student should focus on improving that area of fitness in order to maximize health benefits. Any Texas student could earn between zero and six HFZs. These HFZ standards are regularly evaluated and adjusted if needed to reflect contemporary research (Plowman & Meredith, 2013). The HFZ standards used in this reporting period corresponded to FITNESSGRAM versions 8 and 9 (TEA, 2016).

After fitness assessments were completed at the classroom or school levels, results were aggregated by district, grade, and gender and submitted to the TEA for annual PFAI reporting. Any time a student was in a subgroup with fewer than six students, the data for that subgroup was reported as “< 5” or a blank cell to reduce the possibility of student identification. Once aggregated, the PFAI data quantified district-level fitness performance based on the percentages of students passing all six fitness tests. Reporting data in this manner intended to represent fitness achievement at a district level, rather than a student level.

The PFAI database was most significant to the study because it allowed the compilation of fitness testing reporting frequency by district. It was also used to create a variable describing the district’s overall fitness by grade; this fitness variable was analyzed as a predictor of reporting frequency.

3.2.2 AskTED Data

The second source of TEA data used was from AskTED, an online directory of Texas public school district information. This dataset listed names, district IDs, addresses, regions, counties, size, and district type. AskTED records for the 2013-2014 school year were selected, as they were thought to most thoroughly match the most recently available PFAI data. The most pertinent information from AskTED data was its description of district type. Districts were either classified as independent school districts (ISDs), common school districts (CSDs), or charter districts (CDs). Due to the relatively small number of the latter two types, CSDs and CDs were combined into non-ISDs for comparative analysis with ISDs.

3.2.3 PEIMS Data

The third source of TEA data used was PEIMS data. These reports are collected annually and include extensive information about district size, demographics and performance. Reports from the 2013-2014 school year were used. Ten variables from PEIMS data were selected to be analyzed as predictors of reporting frequency. These variables fit into three broad categories: district size, which included number of campuses, enrollment, and community size; district demographics, which included racial/ethnic percentages and a proxy for socioeconomic status (SES); and district performance, which included attendance rate, teacher turnover rate, and standardized test scores.

3.2.4 Merging Datasets

The PFAI data was reported as seven distinct files; it was consolidated into one dataset before merged with AskTED and PEIMS data. When consolidating PFAI data, two distinct reporting methods were detected: early years of PFAI data included aggregated subgroups by district, while later years subdivided each district's subgroup according to school. To ensure consistency for this study, school data reports were aggregated to the district level to maintain analysis of district-level trends.

Additionally, PFAI data was consolidated by gender. Although PFAI data separated results into binary gender subgroups, student gender was not expected to influence district-level reporting characteristics; larger groups were also more desired to represent the overall district profile.

Due to the size of PFAI datasets, four grades were chosen to measure district reporting trends: Grade 4, Grade 6, Grade 7, and Grade 9. Multiple grades were included to ensure that districts were uniformly adhering to SB 530 reporting requirements across all grades. These four grades were chosen because they spanned multiple levels of elementary, middle, and high schools and represented a wide range of Texas youth who were required to complete PFAI testing.

Once this consolidation was complete, PFAI data was merged with AskTED directory information and PEIMS characteristics. The common variable that linked all three data sources was District IDs. In order to be included in the present analysis, a district needed to exist in all three databases. Districts that were not listed on all three datasets were excluded.

3.2.5 Analytical Procedures

Descriptive statistics were analyzed to measure the number of times each district reported per year and to calculate each district's reporting frequency within the seven-year period. Comparative *t*-tests were conducted to compare the reporting frequency of ISDs to non-ISDs. Pearson *r* correlations were calculated between all predictors to measure relationships with reporting frequency and with each other.

Three types of linear regression analyses were conducted to measure the extent to which the predictors determined reporting frequency. First, a multiple linear regression (MLR) was completed using all eleven predictors and the entire sample for each grade. To check model predictive capacity, R^2 values were examined. Then, to examine internal replicability of the datasets, three randomly selected subsets from each grade were constructed and compared. This process shed light on sampling error and potential for sample variation. Finally, two-fold cross-validation was used to ensure the inclusion of the best available predictors and determine if a model with fewer variables would provide more predictive value. The two-step process included a training set comprised of one-third of each dataset and a test set comprising the remaining two-thirds. All statistical analyses were conducted using Statistical Package for the Social Sciences (SPSS), Versions 22 and 23.

3.2.6 Human Subjects Research Statement

Because all datasets were publicly available for download from the TEA, no human participants were contacted in order to access the data. Texas A&M University's IRB declared this study exempt from review in December 2015.

3.3 Results

3.3.1 Sample Size

In the final school year (SY) of analysis, there were 1,227 districts in Texas. However, when combining the demographic information from the two independent sources (AskTED and PEIMS), some districts did not successfully match; therefore, the highest possible number of eligible districts was 1,195.

3.3.2 Total Amount of PFAI Reporting by Year

When paired with PFAI data, the number of total districts who ever contributed PFAI data averaged 1,098 districts, varying slightly by grade. These districts reported at least once in all seven years, meaning they completed fitness testing and submitted PFAI results. Number of reporting districts decreased slightly as grade increased; the highest amount of reporting occurred in Grade 4 at 1,125 districts, while the lowest amount occurred in Grade 9 at 1,052 districts.

General trends in reporting existed across all four grades. The SY that received the most reporting was SY0910, with an average of 93.4% of eligible districts submitting

PFAI data. Reporting dropped noticeably from SY1011 to SY1112, from 92.4% to 82.6%. The only SY with outlying reporting was SY1314, which PFAI administration labeled "preliminary" (TEA, 2016). This SY averaged 25.4% of districts reporting, accounting for an average of 278.7 districts. Because of this incomplete dataset, reporting frequencies were calculated for both a total of seven SYs as well as a total of six SYs (see Table 3.1).

Table 3.1
Total Number and Percentage of Districts Reporting by Year

	Grade 4		Grade 6		Grade 7		Grade 9		Average	
	n	%	n	%	n	%	n	%	n	%
SY0708	989	87.9	988	88.5	979	89.2	935	88.6	972.7	88.6
SY0809	1045	92.9	1032	92.4	1022	93.2	980	92.9	1019.7	92.9
SY0910	1048	93.2	1044	93.5	1029	93.3	982	93.4	1025.7	93.4
SY1011	1034	91.9	1029	92.1	1018	93.0	978	92.4	1014.7	92.4
SY1112	932	82.8	924	82.7	903	82.6	869	82.6	907.0	82.6
SY1213	901	80.1	888	79.5	881	78.7	828	79.6	874.5	79.6
SY1314	285	25.3	279	25.0	281	25.7	270	25.4	278.7	25.4
Total n	1125		1117		1097		1052		1097.7	

Total potential number of districts: 1195
Total actual number of districts: 1227

3.3.3 Average Reporting Frequency

For all grades, reporting frequency averaged 5.5 years out of the seven potential SYs of reporting. Districts most commonly reported six out of seven years; in Grade 4, 47.1% of the sample reported six out of seven years. Only 18.8% of districts reported all seven years. These means and percentages were similar in Grades 6, 7, and 9. To determine how the outlying SY1314 dataset affected averages, reporting frequency without that year was also examined. Without SY1314, reporting frequency in Grade 4 was 5.28 years out of six years, with 62.2% of districts reporting all six SYs; these values were similar for other grades (see Table 3.2a-b).

Table 3.2a
Average Reporting Frequency, Including SY1314

	Grade 4		Grade 6		Grade 7		Grade 9		Average	
	Mean = 5.54 SD = 1.30		Mean = 5.53 SD = 1.31		Mean = 5.57 SD = 1.28		Mean = 5.55 SD = 1.29		Mean = 5.54 SD = 1.30	
	n	%	N	%	n	%	n	%	n	%
1	20	1.8	25	2.2	21	1.9	23	2.2	22.2	2.0
2	33	2.9	25	2.2	24	2.2	24	2.3	26.5	2.4
3	40	3.6	39	3.5	38	3.5	21	3.0	34.5	3.4
4	92	8.1	102	9.1	88	8.0	82	7.8	91.0	8.2
5	193	17.2	184	17.2	188	17.1	196	18.6	190.2	17.3
6	536	47.6	530	47.4	528	48.1	498	47.3	523.0	47.6
7	212	18.8	212	19.0	210	19.1	197	18.7	207.7	18.9

Table 3.2b
Average Reporting Frequency, Excluding SY1314

	Grade 4		Grade 6		Grade 7		Grade 9		Average	
	Mean = 5.28 SD = 0.034		Mean = 5.28 SD = 0.035		Mean = 5.31 SD = 0.034		Mean = 5.29 SD = 0.035		Mean = 5.29 SD = 0.034	
	n	%	n	%	n	%	n	%	n	%
1	22	2.0	26	2.3	22	2.0	24	2.3	23.5	2.1
2	32	2.8	24	2.1	22	2.0	25	2.4	25.7	2.3
3	44	3.9	43	3.8	42	3.8	33	3.1	40.5	3.6
4	104	9.2	110	9.8	97	8.8	95	9.0	101.5	9.2
5	223	19.8	216	19.3	226	20.6	231	22.0	224.0	15.9
6	700	62.2	697	62.4	687	62.6	644	61.2	682.0	62.1

Table 3.3
Mean Differences in Reporting Frequency by District Type

	ISD			Non-ISD				
Grade	n	Mean	SD	n	Mean	SD	*t*	*p*
Grade 4	1013	5.69	1.13	112	4.16	1.88	8.399	<0.01
Grade 6	1010	5.68	1.12	107	4.13	1.99	7.923	<0.01
Grade 7	1004	5.68	1.14	93	4.33	1.91	6.715	<0.01
Grade 9	967	5.64	1.17	85	4.48	1.97	5.360	<0.01
Average	998.5	5.67	1.14	99.2	4.27	1.93	7.099	<0.01

3.3.4 Reporting Frequency Differences by District Type

Approximately 10% of reporting districts for each grade were considered non-ISDs. On average, ISDs reported more frequently than non-ISDs. Uneven group sizes led to some heterogeneity in variance, but even when equality of variance with not assumed, *t*-test values indicated statistically significant differences in mean years of reporting between ISDs and non-ISDs (see Table 3.3). A closer examination of non-ISD reporting frequencies detected a discrepancy in reporting trends in SY0708, the first SY of PFAI reporting. In SY0708, reporting among non-ISDs was lower than the overall reporting trends.

3.3.5 Correlations Between Reporting Frequency and District Characteristics

Many statistically significant correlations were discovered between reporting frequency and the predictor variables (see Table 3.4). All unspecified Pearson *r* correlations were from Grade 6 data, unless otherwise noted to reflect a grade-specific correlational difference.

Table 3.4
Correlations Between Reporting Frequency and Variables

Variable	Grade 4	Grade 6	Grade 7	Grade 9	Total <0.05
Campuses	0.194**	0.194**	0.199**	0.195**	4
Enrollment	0.198**	0.198**	0.201**	0.197**	4
Community Type	-0.073*	-0.072*	-0.009	0.021	2
Percentage Black	-0.127**	-0.120**	-0.078**	-0.056	3
Percentage White	0.068*	0.058	0.028	0.034	1
Percentage Hispanic	-0.004	-0.004	0.006	-0.016	0
Low-SES	-0.071*	-0.051	-0.059	-0.073*	2
Fitness	0.267**	0.345**	0.310**	0.250**	4
Turnover Rate	-0.249**	-0.268**	-0.236**	-0.212**	4
STAAR Scores	0.105**	0.117**	0.113**	0.087**	4
Attendance Rate	0.043	0.009	-0.010	-0.022	0
Total p<0.05 per Grade	9	7	6	6	

**=$p<0.01$; *=$p<0.05$

Variables related to district size that persistently correlated with reporting frequency, regardless of grade, were number of campuses ($r = 0.194$, $p<0.01$) and enrollment ($r = 0.198$, $p<0.01$). Community type was negatively correlated with reporting frequency in lower grades only ($r = -0.072$, $p<0.01$).

District demographic variables sometimes correlated with reporting frequency, depending on grade. In three out of four grades, the district-wide percentage of Black students was statistically significantly related to reporting frequency ($r = -0.120$,

$p<0.01$). District percentage of White students was statistically significant in Grade 4 only ($r = 0.068$, $p<0.05$). District percentage of Hispanic students was never statistically significantly related to reporting frequency. District percentage of low-SES students formed a negative relationship with reporting frequency but was statistically significant only in Grades 4 and 9 ($r = -0.071$, $p<0.05$; $r = -0.073$, $p<0.05$).

In all four grades, all district performance variables except for attendance rate correlated with reporting frequency. Fitness performance and reporting frequency were positively correlated and statistically significant; the association was strongest in Grade 6 ($r = 0.345$, $p<0.01$). District standardized test scores had a positive relationship with reporting frequency ($r = 0.117$, $p<0.01$), while teacher turnover rate had a negative relationship ($r = -0.268$, $p<0.01$).

3.3.6 Regression Analyses

Regression analyses posited reporting frequency as the outcome variable and the numerous district characteristics as predictors. Three types of regression tests were conducted.

The first linear regression, the MLR, used all 11 predictors and explained the most variance in reporting frequency. The combined inclusion of all predictors was statistically significantly related to reporting frequency. This model produced R^2 values averaging 0.177, with some variation by grade, explaining approximately 18% of the reporting frequency variance. Explanatory power was highest in Grade 6 ($R^2 = 0.212$) and lowest in Grade 6 ($R^2 = 0.137$). These were the highest R^2 values achieved at any point during the regression analyses.

The most influential predictors in the MLR were teacher turnover rate, fitness performance, and percentage of low-SES students. The latter predictor approached statistical significance for all grades except Grade 6, in which it was considered statistically significant. Attendance rate was a statistically significant predictor only in upper grades.

To test for internal replicability of results, three randomly generated subsets were selected and analyzed. Each subset comprised of one-third of the data for each grade. Grade 6 demonstrated the most stable values across its three samples (R^2 = 0.204, 0.187, 0.209), while Grade 9 demonstrated the least stable values (R^2 = 0.098, 0.233, 0.151).

The third subset in each internal replicability test was used as a training set in twofold cross-validation. The remaining data, comprising two-thirds of each grade's sample, was treated as a test set in a linear regression run with three to five of the strongest predictors. Predictors were ruled "in" or "out" depending on statistical significance of the predictor's Beta coefficient in the training set or if they approached statistical significance in the MLR. Based on training set findings, three predictors – teacher turnover rate, fitness performance, and percentage of low-SES students – were added to all grades for regression analysis in the test set. Attendance rate was added to Grades 7 and 9. Community type was added to Grade 7 because its beta weight was considered statistically significant in the training set.

In all grades, test sets produced R^2 values smaller than that of the original regression; values ranged from 0.106 to 0.171. The original MLR model remained the most predictive for Grade 6 and least predictive in Grade 9 (see Table 3.5a-d).

Table 3.5a

Grade 4 Beta Coefficients for Regression Analyses

Predictors	MLR R^2 = 0.173 Adj. R^2 = 0.165		IR1 R^2 = 0.167 Adj. R^2 = 0.140		IR2 R^2 = 0.160 Adj. R^2 = 0.135		IR3 [Training] R^2 = 0.225 Adj. R^2 = 0.199		Test R^2 = 0.116 Adj. R^2 = 0.112	
	β	*p*	β	*p*	β	*p*	β	*p*	β	*p*
Fitness Performance	0.289	0.001*	0.214	0.001*	0.277	0.001*	0.342	0.001*	0.249	0.001*
Turnover Rate	-0.204	0.001*	-0.257	0.001*	-0.218	0.001*	-0.171	0.005*	-0.245	0.001*
Low-SES	0.083	0.069	0.113	0.171	0.094	0.229	0.122	0.133	0.089	0.018*
Attendance Rate	-0.004	0.881	0.001	0.999	-0.070	0.210	0.005	0.924		
Community Size	-0.018	0.633	-0.146	0.032	0.068	0.268	-0.098	0.138		
Campuses	0.021	0.879	0.019	0.934	0.023	0.955	0.182	0.440		
Enrollment	0.202	0.139	0.240	0.309	0.171	0.672	0.047	0.841		
Percentage Black	0.016	0.885	-0.071	0.777	-0.047	0.817	-0.184	0.350		
Percentage Hispanic	0.173	0.395	-0.176	0.698	-0.061	0.879	-0.135	0.706		
Percentage White	0.171	0.421	-0.249	0.598	0.003	0.994	-0.263	0.480		
Academic Scores	-0.025	0.540	-0.059	0.464	-0.046	0.530	0.076	0.279		

*=$p<.05$

Displaying results for Grade 4 only

Table 3.5b
Grade 6 Beta Coefficients for Regression Analyses

Predictors	MLR $R^2 = 0.212$ Adj. $R^2 = 0.204$		IR1 $R^2 = 0.204$ Adj. $R^2 = 0.177$		IR2 $R^2 = 0.187$ Adj. $R^2 = 0.161$		IR3 [Training] $R^2 = 0.209$ Adj. $R^2 = 0.184$		Test $R^2 = 0.171$ Adj. $R^2 = 0.168$	
	β	*p*	β	*p*	β	*p*	β	*p*	β	*p*
Fitness Performance	0.340	0.001*	0.243	0.001*	0.271	0.001*	0.347	0.001*	0.330	0.001*
Turnover Rate	-0.196	0.001*	-0.233	0.001*	-0.188	0.001*	-0.227	0.001*	-0.243	0.001*
Low-SES	0.101	0.019*	0.138	0.099	0.107	0.149	0.089	0.271	0.127	0.001*
Attendance Rate	-0.035	0.208	0.015	0.787	-0.058	0.249	0.028	0.589		
Community Size	-0.037	0.290	-0.037	0.576	-0.010	0.875	0.026	0.671		
Campuses	0.010	0.940	0.079	0.700	-0.075	0.771	-0.206	0.515		
Enrollment	0.200	0.125	0.135	0.519	0.310	0.231	0.390	0.222		
Percentage Black	-0.020	0.860	-0.016	0.937	0.016	0.941	-0.011	0.957		
Percentage Hispanic	0.075	0.729	0.034	0.922	0.055	0.896	0.069	0.878		
Percentage White	0.067	0.763	0.066	0.854	0.107	0.808	0.127	0.783		
Academic Scores	0.002	0.949	0.040	0.595	0.053	0.430	-0.059	0.409		

*= $p<.05$
Displaying results for Grade 6 only

Table 3.5c

Grade 7 Beta Coefficients for Regression Analyses

Predictors	MLR R^2 = 0.187 Adj. R^2 = 0.179		IR1 R^2 = 0.211 Adj. R^2 = 0.186		IR2 R^2 = 0.255 Adj. R^2 = 0.232		IR3 [Training] R^2 = 0.201 Adj. R^2 = 0.176		Test R^2 = 0.164 Adj. R^2 = 0.158	
	β	*p*	β	*p*	β	*p*	β	*p*	β	*p*
Fitness Performance	0.348	0.001*	0.309	0.001*	0.358	0.001*	0.371	0.001*	0.329	0.001*
Turnover Rate	-0.197	0.001*	-0.186	0.002*	-0.273	0.001*	-0.103	0.069	-0.264	0.001*
Low-SES	0.089	0.050	0.074	0.312	0.166	0.036*	0.079	0.361	0.113	0.003*
Attendance Rate	-0.062	0.028*	-0.099	0.087	0.077	0.150	-0.077	0.125	0.35	0.088
Community Size	0.03	0.399	-0.02	0.765	-0.035	0.586	0.174	0.006*	0.038	0.288
Campuses	0.029	0.824	0.188	0.440	0.104	0.755	-0.164	0.659		
Enrollment	0.173	0.197	0.063	0.795	0.091	0.786	0.296	0.431		
Percentage Black	0.106	0.352	0.056	0.793	0.291	0.130	-0.092	0.617		
Percentage Hispanic	0.301	0.173	0.177	0.605	0.689	0.061	-0.126	0.749		
Percentage White	0.279	0.218	0.206	0.570	0.686	0.067	-0.259	0.511		
Academic Scores	-0.01	0.797	0.001	0.993	-0.037	0.621	0.04	0.602		

*=*p*<.05

Displaying results for Grade 7 only

Table 3.5d

Grade 9 Beta Coefficients for Regression Analyses

Predictors	MLR R^2 = 0.137 Adj. R^2 = 0.127		IR1 R^2 = 0.098 Adj. R^2 = 0.067		IR2 R^2 = 0.233 Adj. R^2 = 0.207		IR3 [Training] R^2 = 0.151 Adj. R^2 = 0.125		Test R^2 = 0.106 Adj. R^2 = 0.100	
	β	*p*	β	*p*	β	*p*	β	*p*	β	*p*
Fitness Performance	0.283	0.001*	0.179	0.005*	0.354	0.001*	0.314	0.001*	0.253	0.001*
Turnover Rate	-0.154	0.001*	-0.15	0.018*	-0.167	0.007*	-0.146	0.013*	-0.192	0.001*
Low-SES	0.084	0.093	0.031	0.723	0.107	0.215	0.183	0.035*	0.053	0.208
Attendance Rate	-0.092	0.003*	-0.131	0.022*	-0.111	0.064	-0.144	0.027*	0.038	0.046*
Community Size	0.050	0.184	0.096	0.178	0.034	0.608	0.08	0.247		
Campuses	0.019	0.894	0.215	0.584	0.353	0.123	0.007	0.975		
Enrollment	0.182	0.202	-0.034	0.930	-0.093	0.682	0.228	0.301		
Percentage Black	0.133	0.290	0.186	0.435	0.705	0.014	0.225	0.399		
Percentage Hispanic	0.317	0.214	0.494	0.312	1.702	0.007	0.487	0.460		
Percentage White	0.335	0.195	0.586	0.244	1.782	0.006	0.518	0.442		
Academic Scores	0.000	0.994	-0.021	0.790	0.033	0.673	0.063	0.478		

*=p<.05

Displaying results for Grade 9 only

3.4 Discussion

These results confirm that annual reporting during the PFAI was, overall, very high. Except for SY1314, approximately 80-90% of districts reported each year, reaching an average of 1,097 districts. These reporting rates match what Welk, Meredith, Ihmels, and Seeger (2010) approximated during the first year of PFAI reporting. Percentages may be slightly lower in the 2010 calculations because districts who never reported within the entire seven-year period were included. Nevertheless, confirming that at least 80% of districts reported at least once over the seven years exemplifies the reach of PFAI reporting initiatives.

One seemingly prominent finding in this study is the apparently low reporting in SY1314, when only 25% of the districts reported compared to high reporting in all other SYs. According to the TEA, SY1314 data is described as “preliminary” and contains data only using a specific electronic submission system (TEA, 2016). The Cooper Institute’s Texas Youth Fitness Project shared that approximately 6,100 schools submitted PFAI data in SY1314, a number comparable to previous reporting years (The Cooper Institute, 2014a). It can be inferred that SY1314 was a strong year for reporting, even if not indicated in the publicly available electronic results. It is not known, however, if such a delay in public availability is typical.

A more subtle decrease in reporting occurred between SY1011 and SY1112; reporting rates fell from approximately 90% of the districts to 80%. Even though PFAI procedures navigated formatting changes between SY2010 and SY2011 with little effect on reporting rates, reporting rates dropped one SY later. The reasons for this drop are not documented but could include changes in technology, the launch of new FITNESSGRAM standards (TEA, 2016), or TEA administrative turnover. After the SY1112 drop, the number of districts submitting PFAI reports stabilized at the new reporting rate (The Cooper Institute, 2014a).

The reporting frequency by district was comparable among all four grades sampled; the reporting trends within each year matched. However, only 1,052 districts submitted data for Grade 9 students, compared to 1,125 submitted for Grade 4. This finding – occurring at the district level over a longitudinal period – expands on what Welk, Meredith et al. (2010) noticed at the school level their analysis from the SY0708 period: PFAI reports came from elementary schools more often than middle or high schools. This difference in reporting could reflect the less stringent statewide PE requirements for middle and high school students (Martin, Ede, Morrow, & Jackson, 2010; Morrow, Martin, Welk, Zhu, & Meredith, 2010). Additionally, Martin et al. (2010) reported more challenges when working with secondary school students, who proved relatively difficult to motivate to complete fitness testing. The present findings confirm what others previously indicated: reporting decreases as grade increases, even though all students in Grades 3 through 12 are expected to participate (§ 530, 2007).

In general, ISDs reported more frequently than non-ISDs. The greatest discrepancy between reporting rates was in SY0708, suggesting a delayed effect in achieving reporting compliance for charter and common districts. The wording of SB 530 could imply only public school districts need to report fitness scores; perhaps non-ISDs assumed they could opt out of this legislation and later opted in. Another possibility is that ISDs had more awareness of or access to PFAI reporting resources. Another consideration is methodological; the ISD group size is larger than non-ISDs, limiting the statistical meaning of the disparity.

Overall, districts submitted reports an average of 5.5 out of seven years – or 5.3 out of six years when omitting SY1314 – and this finding confirms that Texas school districts are reporting consistently, adhering to SB 530's requirements and treating fitness testing as an institutionalized norm. Even so, predictors of high reporting frequency were detected in districts by grade, and most were related to district performance.

Fitness performance highly correlated with reporting frequency and served as one of its strongest predictors. A methodological caveat exists regarding this variable; in this study, it was calculated as a composite score that represented each district's highest-achieving fitness scores, not factoring in lower scores from other reporting years. If a district reported more often, higher scores could replace lower-achieving fitness scores from previous years. On one hand, it is theoretically reasonable for a district to report all seven years and never have its students perform in an exemplary fashion. On the other hand, perhaps only more fit districts are regularly submitting reports, while less fit districts do not bother. Another concern with the fitness variable is that despite the satisfactory reliability and validity scores of the FITNESSGRAM tests (Morrow, Martin, & Jackson, 2010), many measurement and reporting mistakes were committed during fitness testing administration (Martin et al., 2010), making this study's fitness variable even more susceptible to error.

District percentage of low-SES students correlated occasionally with reporting frequency and emerged as a marginal predictor in all grades and a statistically significant predictor in Grade 6. Previous evidence supports a positive connection between fitness and SES (Coe, Peterson, Blair, Schutten, & Peddie, 2013; Welk, Meredith et al., 2010), but connecting reporting frequency to SES is a new finding. It is reasonable to speculate that priorities in lower-reporting districts may not include consistent PFAI reporting; meeting basic testing requirements and student health needs may be more pressing district tasks. Youth fitness outcomes, though, are lower in nonwhite and low-SES populations (Welk, Meredith et al., 2010). Based on this study's findings, the districts most likely in need of routine fitness monitoring appear to be reporting less frequently than districts with high-SES students.

Another noteworthy predictor was teacher turnover rate. As turnover rate increased, reporting frequency decreased. Statewide assessment requires a strong delegation of responsibility considering the extensive training, cost, and reporting systems needed to facilitate it (Morrow & Ede, 2009). If teacher turnover rate is high, PFAI reporting responsibilities could be lost from person to person, providing a barrier to reporting compliance. The negative association between teacher turnover rate and reporting frequency suggests a need to transcend this barrier.

A final district performance variable, attendance, had no statistically significant correlation with reporting frequency, but it did serve as a predictor of reporting frequency for Grades 7 and 9. Statistically, attendance appears to function as a suppressor variable (Thompson, 2008), but it is unclear what this status means interpretively. In later years of PFAI data, attendance increases with aerobic capacity performance (Allums-Featherston et al., 2014a), but a causal relationship between the two is indeterminate. Because attendance becomes more autonomous as students age, students in older grades who excel at fitness testing may choose to attend class on the testing day, while others may stay home or opt out of taking PE during the semester of testing. This phenomenon could occur with minimal effect on district reporting frequency, but it may statistically inflate the power of the already-limited fitness variable. Nevertheless, fitness testing is mandatory for students in all grades, so investigating the practical effects of attendance is useful for increasing reporting compliance.

3.5 Limitations

This secondary data analysis required merging three separate datasets, making it prone to reporting errors inherent in any of the three. After compiling and merging the data, the total potential districts in this study was 1,195, but the total number of Texas school districts in SY1314 was 1,227 (Williams, 2015). Approximately 30 districts were lost while merging demographic datasets. In addition, some PFAI data was not used because older SYs occasionally contained reports from districts not extant or active in SY1314. These issues were minor, as the final reports included an average of 1,097 districts across the four grades examined, a generous size consistent with other PFAI-based analyses and representative of Texas school districts.

Fitness performance as a variable was limited because the variable's calculation did not factor in lower scores, meaning it may have favored higher-reporting districts.

Fitness, like other percentage-based variables, may also be naturally biased towards larger districts. The fitness variable was meant to be a snapshot look at a district's best fitness performance during the reporting period; further exploration of why fitness and reporting frequency correlate so strongly is needed to articulate the specifics of this relationship.

Due to the high reporting overall, the dependent variable of reporting frequency was positively skewed, which may have affected validity during the regression analyses. Datasets may also be prone to sampling variance, as indicated by the inconsistent R^2 values produced during the internal replicability tests, especially in higher grades.

Finally, the aim of this study was to measure district reporting compliance over time. These findings do not represent student-level reporting compliance or fitness performance. The results are only applicable at the district level.

3.6 Conclusion

Previous studies have reported high Texas PFAI reporting compliance on a year-to-year basis, but this study confirmed these claims longitudinally and also examined reporting according to district characteristics. Regarding district type, ISDs were more likely to fully comply than non-ISDs, but both district types began reporting at similar rates after the inaugural reporting year. Based on these successes, Texas's comprehensive approach to youth fitness assessment can serve as a model for other states to follow if similar monitoring is deemed appropriate.

Despite this confirmation, this study also found that the districts most likely to demonstrate exemplary reporting compliance achieved higher fitness scores, had lower teacher turnover rates, served more economically privileged students, and in higher grades, recorded higher attendance rates. Knowing that these predictors exist can guide researchers and practitioners towards refinement of PFAI protocol. Ultimately, the goals of initiatives such as PFAI include reaching all Texas school districts and so that they benefit all Texas public school students.

4.0 STUDY B: CORRELATIONS OVER TIME BETWEEN AEROBIC CAPACITY AND BODY COMPOSITION

4.1 Introduction

Youth fitness assessments are a recommended part of physical education (PE) courses, which are a key ingredient in coordinated school physical activity programs (CDC, 2013). School-based youth fitness testing in Texas became mandatory in 2007 with the passing of Texas Senate Bill (SB) 530, which required all districts to assess and report fitness achievement rates to the Texas Education Agency (TEA) (Cooper et al., 2010). The Physical Fitness Assessment Initiative (PFAI) is the statewide process of compiling student fitness scores and summarizing aggregate performance across all Texas school districts (TEA, 2016).

The five components of health-related fitness generally assessed in fitness testing are cardiorespiratory endurance, muscular strength, muscular endurance, flexibility, and body composition. FITNESSGRAM®, a well-known fitness battery developed in Texas, contains tests for all five fitness components (Plowman & Meredith, 2013) and is the assessment tool most often chosen for youth fitness assessment initiatives across the nation, including Texas's PFAI (TEA, 2016). These tests have high reliability and validity (Morrow, Martin, & Jackson, 2010) and the tests' criterion-based standards are routinely evaluated to reflect the most current evidence-based findings about the amount of fitness ability children need to have in order to maximize health benefits (Plowman & Meredith, 2013).

Body composition is important to measure because of the health risks associated with having an overly high percent body fat (Friedemann et al., 2012; Going et al., 2011). One of FITNESSGRAM's tests is the Body Mass Index (BMI), which serves as a reasonable approximation of body composition even though it does not directly measure percent body fat (Laurson, Eisenmann, & Welk, 2011). Measuring BMI in youth is considered a valid surveillance strategy (Must & Anderson, 2006), but BMI measurement in schools remains controversial (MacLean et al., 2010; Nihiser et al., 2007; Ruggieri & Bass, 2015).

One argument against measuring youth body composition is relates to the debate between "fitness" and "fatness." In popular terms, being "fit" is typically considered having high cardiovascular fitness or athletic ability, while being "fat" refers to having a high BMI or a large amount of body fat. Some evidence indicates that youth cardiovascular fitness can have beneficial effects independently of fatness (Kwon, Burns, & Janz, 2010), potentially negating the need for BMI measurements; others argue the interaction between the two concepts necessitates measuring both regularly (Eisenmann, Welk, Wickel, & Blair, 2007).

The most popular cardiovascular fitness test included in FITNESSGRAM is the Progressive Aerobic Cardiovascular Endurance Run (PACER) (Plowman & Meredith, 2013). In Versions 8.6 and 9.1 of FITNESSGRAM, a student's ability to achieve the PACER's Healthy Fitness Zone (HFZ) was dependent on the student's individual BMI. This relative standard of HFZ achievement was meant to reflect the interaction between fitness and fatness. The results were described as a "more valid" reflection of student healthiness but were "more awkward" for teachers to administer and explain to children (Plowman & Meredith, 2013).

These HFZ standards were used during the 2011-2014 years of the Texas PFAI, and HFZ achievement rates of BMI and PACER tests have been previously summarized (The Cooper Institute, 2014a). The HFZ achievement rates of PACER and BMI tests in Texas youth are expected to be correlated to some extent based on the method of HFZ achievement calculations. What remains unclear, however, is how stable the relationship between these rates was over time. It is also not known which groups of students, if any, would be most affected by the BMI-dependent PACER results.

The purpose of this study is to compare BMI and PACER HFZ achievement rates during the 2011-2014 period of PFAI data collection. Two primary elements were examined: 1) the mean HFZ achievement rates for both tests and 2) the correlations between the two achievement rates over time.

4.2 Methods

4.2.1 PFAI Data

This study was completed using the publicly available database for Texas's statewide PFAI data. All PFAI data was publicly available in aggregate form at the TEA's website. Datasets were accessed in January 2016.

At this study's onset, seven years of PFAI data were available but only three years of data could be used to complete the study. These years spanned the 2011-2012 to 2013-2014 school years. This data contained HFZ achievement rates of each specific test and used the same criterion-referenced standards, making them valid to compare to one another (TEA, 2016). The three datasets included in this study are referred to as Years 1, 2, and 3.

Students who contributed PFAI data were enrolled in Texas public schools and were present on the time and day of their district, school, or classroom's annual fitness testing. After these students completed the testing, all PFAI data was aggregated by district, school, grade, and gender. In unusual cases, PFAI data was aggregated at the classroom level; to maintain analytical consistency, the reports in these cases were averaged into school-level composite scores. This reporting style ensured protection of student identities while still permitting comparisons by grade, gender, and school characteristics.

The unit of analysis for this study was at the school level, but each school included subgroups divided by grade and gender. In order to offer a detailed look at which student groups may be affected by the BMI-dependent PACER calculation, four grades and two genders were selected for analysis, allowing each school to have up to eight subgroups. Grades 4, 6, 7 and 9 were selected because they encompassed students enrolled in elementary, middle, and high schools, spanned a wide range of developmental stages, and represented a large sample of Texas students tested. These subgroup characteristics were useful to identify which groups of students, if any, had the strongest relationship between BMI and PACER HFZ achievement rates.

4.2.2 Inclusion Criteria

Because a key variable was change over time, schools had to report at least two out of three years to be included. Schools reporting only one year were not included.

Each school's subgroups could also only be included if the subgroup contained more than five attempts on both the BMI and PACER tests. This requirement was necessary because PFAI data concealed details in reporting results among subgroups with fewer than five students or five attempts, using "N/A," "< 5," or percentages to limit identifying information. As a result, records from school subgroups were omitted if the subgroup had a) fewer than five total students tested, b) fewer than five BMI test attempts, or c) fewer than five PACER test attempts.

4.2.3 PEIMS Data

After compiling all PFAI data, information about each school's overarching district was merged with the data. These district characteristics came from the Public Education Information Management System (PEIMS), which is publicly available from the TEA. Four PEIMS characteristics were selected: percentage of economically disadvantaged students (used as a proxy for socioeconomic status [SES]), standardized test scores (used as a proxy for academic performance [AP]), student attendance rate, and teacher turnover rate. These four variables were averaged by school year and applied to each unit of analysis. Performance quartiles were formed to simplify analytical procedures.

4.2.4 Analysis

Three types of change over time were explored: changes in BMI scores, changes in PACER scores, and changes in the correlations between BMI and PACER. Comparative means testing was used to summarize changes in both BMI and PACER scores over time and identify which subgroups may have experienced changes.

Differences were described by grade, gender, and combined grade and gender; differences related to district characteristics were also explored. For the third type of change over time, bivariate Pearson correlations were calculated between BMI and PACER scores for each year. This analysis added a longitudinal component to a classic correlational comparison, articulating the relationship between BMI and PACER over time. Differences in correlational change were also described by grade, gender, and combined grade and gender, to determine if these changes were occurring uniformly in school subgroups. All statistical analyses were conducted using Statistical Package for Social Sciences (SPSS), Versions 22 and 23.

4.2.5 Human Subjects Research Statement

Both PFAI data and PEIMS data were publicly available for download from the TEA. Texas A&M University's IRB declared this study exempt from review in December 2015.

4.3 Results

4.3.1 Sample Size

After applying all inclusion criteria, a total of 2,331 schools were included in the analysis. This amount represented 531 Texas school districts and was subdivided into 5,987 school subgroups. Overall, these results were from 2.36 million students, tested at any point over the three years. Year 2, the middle year, contained the most students, at 1.12 million (see Table 4.1).

Table 4.1
Total Sample Size

Year	Districts	Schools	School Subgroups*	Students
Year 1	519	2,246	5,745	591,313
Year 2	522	2,248	5,722	1,123,294
Year 3	173	1,145	2,982	645,968
*Total***	*531*	*2,331*	*5,987*	*2,360,575*

*Subgroups defined as school-grade-gender unit of analysis
**Total numbers signify those who reported two out of the three analyzed years

4.3.2 Total Means over Time

Overall, the HFZ achievement rate on BMI tests averaged 51.75%, with minimal variation from year to year. BMI HFZ achievement rates by year were: 51.89%, 52.51%, and 50.86%. In each year, BMI achievement rates were normally distributed. These mean differences were considered statistically significant ($F=16.916$, $p<0.001$).

The overall HFZ achievement rate on PACER tests averaged 79.26%, with even less variation from year to year than the BMI achievement rates. PACER HFZ achievement rates by year were 79.35%, 79.62%, and 78.83%. These mean differences neared statistical significance ($F=2.985$, $p=0.051$).

Although HFZ achievement rates on both tests varied only slightly from year to year, the rates varied considerably by subgroup grade and gender. BMI HFZ achievement rates were approximately 10% higher for girls than for boys. BMI scores also improved gradually as students aged; higher percentages of older students achieved their respective BMI HFZs. When combining grade and gender, Grade 9 girls obtained the highest BMI achievement rates, averaging 56.67%, while Grade 6 boys obtained the lowest rates, averaging 47.80% (see Table 4.2).

Like the BMI rates, the PACER HFZ achievement rates changed minimally from year to year but varied considerably by grade and gender subgroups. When comparing results by gender, PACER scores were higher for boys than for girls in all grades except Grade 9, when achievement rates among both genders were close to identical. Grade 4 boys obtained the highest PACER achievement rates, averaging 90.60%, and Grade 7 girls obtained the lowest rates, averaging 63.84%. Age-related decline in PACER HFZ achievement rates varied by gender. Between Grades 4 and 6, girls' PACER achievement rate dropped dramatically and then stabilized, while boys' PACER achievement rate fell more gradually from Grades 6 to 9 (see Table 4.3).

Table 4.2
BMI Achievement Rate over Time, by Subgroup

		Year 1	Year 2	Year 3	Average	Range
Grade 4 Girls	Mean	54.25	54.54	53.02	**53.93**	1.52
	n	1463	1462	759		
	SD	12.16	12.44	12.90	12.50	
Grade 4 Boys	Mean	49.93	51.03	48.55	**49.84**	2.49
	n	1467	1458	761		
	SD	12.36	12.49	12.96	12.60	
Grade 6 Girls	Mean	52.10	52.76	52.07	**52.31**	0.69
	n	605	600	334		
	SD	11.76	12.73	12.20	12.23	
Grade 6 Boys	Mean	48.25	48.29	46.87	**47.80**	1.41
	n	599	592	327	lowest rate	
	SD	11.49	11.45	11.98	11.64	
Grade 7 Girls	Mean	54.40	52.80	51.18	**52.80**	3.22
	n	473	474	243		most change
	SD	11.68	12.32	12.66	12.22	
Grade 7 Boys	Mean	49.82	52.10	49.37	**50.40**	2.74
	n	465	465	239		
	SD	11.20	11.94	12.37	11.84	
Grade 9 Girls	Mean	56.73	56.65	56.63	**56.67**	0.09
	n	330	326	156	highest rate	least change
	SD	12.79	13.31	12.37	12.82	
Grade 9 Boys	Mean	51.14	53.16	53.32	**52.54**	2.18
	n	343	345	163		
	SD	12.81	12.97	12.33	12.70	
Total	*Mean*	*51.90*	*52.51*	*50.86*	***51.75****	*1.65*
	n	*5745*	*5722*	*2982*		
	SD	*12.30*	*12.59*	*12.87*	*12.59*	

* = $p<0.05$ difference in means over time
Bold signifies averages by subgroup; italics signifies total averages

Table 4.3
PACER Achievement Rate over Time, by Subgroup

Subgroup		Year 1	Year 2	Year 3	Average	Range
Grade 4 Girls	Mean	85.49	85.39	85.18	**85.35**	0.31
	n	1463	1462	759		least varied
	SD	8.88	8.88	9.07	8.94	
Grade 4 Boys	Mean	90.59	91.04	90.17	**90.60**	0.87
	n	1467	1458	761	highest rate	
	SD	6.86	6.53	6.88	6.76	
Grade 6 Girls	Mean	64.45	64.92	64.67	**64.68**	0.47
	n	605	600	334		
	SD	10.53	11.85	12.19	11.53	
Grade 6 Boys	Mean	79.12	79.93	79.04	**79.36**	0.89
	n	599	592	327		
	SD	8.49	8.81	9.68	8.99	
Grade 7 Girls	Mean	65.11	64.83	61.59	**63.84**	3.52
	n	473	474	243	lowest rate	most varied
	SD	10.88	12.94	13.24	12.36	
Grade 7 Boys	Mean	76.69	77.74	76.16	**76.86**	1.58
	n	465	465	239		
	SD	9.81	10.48	10.23	10.17	
Grade 9 Girls	Mean	67.78	67.36	65.78	**66.97**	1.99
	n	330	326	156		
	SD	13.37	14.67	14.51	14.18	
Grade 9 Boys	Mean	66.17	66.34	67.12	**66.54**	0.95
	n	343	345	163		
	SD	14.25	15.79	13.94	14.66	
Total	*Mean*	*79.35*	*79.62*	*78.83*	***79.27***	*0.78*
	n	*5745*	*5722*	*2982*		
	SD	*13.85*	*14.34*	*14.51*	*14.23*	

Bold signifies averages by subgroup; italics signifies total averages

No major year-to-year changes were observed in any particular district characteristics; for example, schools in districts with the highest attendance rates consistently observed HFZ achievement for BMI around 54% and for PACER around 80%, with no outlying performance in any year. Discrepancies in HFZ achievement rates, however, did occur across district characteristic quartiles.

Both HFZ achievement rates were lower among schools in low-SES districts. In the highest SES quartile, BMI achievement averaged 58% and PACER averaged 84%; in the lowest SES quartile, these rates fell to 45% and 76%, respectively. In all three years of reporting, BMI achievement rate and SES were statistically significantly correlated (average r = -0.218, p<0.05), while PACER achievement rate and SES were statistically significantly correlated in only two out of the three years (average r = -0.109, p<0.05).

Similar patterns occurred in district's reported AP. In the quartile of highest AP, BMI achievement averaged 54% and PACER averaged 80%; in the quartile with the lowest AP, these rates fell to 47% and 76%, respectively. Both BMI and PACER achievement rates were statistically significantly correlated with standardized scores in all three years of reporting, with higher Pearson r values occurring with BMI (average r = 0.182, p<0.05) than with PACER (average r = 0.112, p<0.05).

Similar variability occurred in attendance rates; as attendance percentile increased, BMI and PACER achievement increased as well. Attendance rate was statistically significantly correlated with BMI in two out of three years (average r = 0.120, p<0.05); and correlated with PACER in one out of three years (r = 0.087, p<0.05).

Table 4.4

Correlations Between District Characteristics and Achievement Rates

	SES	AP	Attendance	Turnover	BMI Year 1	BMI Year 2	BMI Year 3	PACER Year 1	PACER Year 2	PACER Year 3
SES	1.0	-0.587*	-0.075*	0.124*	-0.214*	-0.229*	-0.211*	-0.130*	-0.088*	-0.061
AP	-0.587*	1.0	0.287*	-0.234*	0.220*	0.199*	0.127*	0.146*	0.104*	0.086*
Attendance	-0.075*	0.287*	1.0	-0.202*	0.141*	0.100*	-0.058	0.087*	0.032	-0.054
Turnover	0.124*	-0.234*	-0.202*	1.0	0.005	0.006	-0.049	-0.047	-0.001	-0.024
BMI Year 1	-0.214*	0.220*	0.141*	0.005	1.0	0.236*	0.243*	0.325*	0.082*	0.123*
BMI Year 2	-0.229*	0.199*	0.100*	0.006	0.236*	1.0	0.219*	0.061*	0.328*	0.076*
BMI Year 3	-0.211*	0.127*	-0.058	-0.049	0.243*	0.219*	1.0	0.062	0.042	0.349*
PACER Year 1	-0.130*	0.146*	0.087*	-0.047	0.325*	0.061*	0.062	1.0	0.593*	0.600*
PACER Year 2	-0.088*	0.104*	0.032	-0.001	0.082*	0.328*	0.042	0.593*	1.0	0.606*
PACER Year 3	-0.061	0.086*	-0.054	-0.024	0.123*	0.076*	0.349*	0.600*	0.606*	1.0

*= $p < 0.05$

As teacher turnover percentile increased, BMI and PACER scores decreased, but the variation by quartile was minimal relative to other district characteristics. Neither BMI nor PACER achievement rates were statistically significantly correlated with turnover rate in any of the years (see Table 4.4).

4.3.3 BMI and PACER Achievement Rate Correlations

Overall, PACER achievement rates formed a moderate positive correlation with BMI achievement rates (r = 0.344, p<.01) across all three years. Each year, these correlations were similar, with negligible change over time detected (r = 0.332, 0.358, 0.344).

When the sample was divided into grade-by-gender subgroups, the correlations between BMI and PACER achievement rates grew stronger in all years. Grade 6 girls had the highest correlation between BMI and PACER rates, at r = 0.800. Grade 4 boys had the lowest correlation at r = 0.456. Grade 9 girls had the most variation in correlation from year to year (r = 0.511, 0.499, 0.614) (see Table 4.5).

Correlations between BMI and PACER achievement rates experienced some variation when classified into district characteristic quartiles. The district characteristic containing the most noticeable differences in correlation strength was attendance rate; districts with highest attendance rates had stronger positive correlations between BMI and PACER achievement rates (r = 0.392, 0.396, 0.404).

Table 4.5
Correlations Between BMI and PACER Achievement Rates over Time

Subgroup	Year 1	Year 2	Year 3	Average	Range	Findings
Grade 4 Girls	0.455	0.505	0.494	0.484	0.05	
Grade 4 Boys	0.449	0.458	0.462	**0.456**	**0.013**	Lowest *r*
Grade 6 Girls	0.784	0.795	0.821	**0.800**	0.037	Highest *r*
Grade 6 Boys	0.686	0.638	0.645	0.656	0.048	
Grade 7 Girls	0.783	0.686	0.754	0.741	0.097	Second-highest r
Grade 7 Boys	0.637	0.649	0.675	0.653	0.038	
Grade 9 Girls	0.511	0.499	0.614	0.541	**0.115**	Most varied *r*
Grade 9 Boys	0.524	0.499	0.562	0.528	0.063	
Total	*0.332*	*0.358*	*0.344*	*0.344*	0.026	

4.4 Discussion

The sample size in this study was smaller than previously reported PFAI summaries, which were compiled in the Texas Youth Fitness Project (TYFP). In this report's findings, the TYFP tallied roughly 9.6 million students tested within the three years (Allums-Featherston et al., 2014b). Comparatively, only 2.3 million students were represented in these results due to stricter inclusion criteria, removal of one-time reporters, and omission of smaller subgroups. Nevertheless, this study showed many similarities to TYFP reports that cross-validate the statewide findings.

Average HFZ achievement rates on both BMI and PACER tests mirrored what Allums-Featherston et al. (2014b) reported at the conclusion of 2011-2014 TYFP. In their report, achievement rates on aerobic capacity tests ranged from 63 to 93% in boys and 67 to 89% in girls. BMI achievement ranges were smaller, from 47 to 59% in boys and 52 to 68% in girls. Aerobic capacity HFZ achievement in both genders also steadily declined with age. The differences in HFZ achievement rates by grade, gender, and grade-by-gender subgroups matched statewide trends: girls performed better on body composition, boys performed better in aerobic capacity, and all aerobic capacity achievement decreased with age (Allums-Featherston et al., 2014b).

Differing from this study, statewide findings reported that over time, both BMI and aerobic capacity achievement increased steadily during the TYFP assessment period (Allums-Featherston et al., 2014b). The improvements in body composition and aerobic capacity documented elsewhere were not prominent in this study's findings, for unclear reasons. In this study, only BMI achievement rates were statistically significantly different over time. Not only did these rates fail to consistently improve, but interpretively, the changes appeared negligible considering the realistic value of the achievement rate. Since this study consisted only of schools that used BMI and PACER data, perhaps the effects of other fitness assessments improved the overall averages, or the larger sample allowed for a greater range of improvement. Nevertheless, the observed trends do not fully reflect statewide trends.

Average HFZ achievement rates varied according to district characteristics as well, with lower-SES districts reporting lower achievement rates and higher-AP districts reporting higher achievement rates on both FITNESSGRAM tests. These findings are not surprising considering the well-documented evidence of relationships between American childhood obesity rates and SES (CDC, 2014), and between youth fitness and AP (CDC, 2010). However, the moderating effects of district characteristics must be considered when examining and interpreting fitness findings in order to reduce fitness disparities across the state of Texas, which had not been documented in this portion of PFAI data. These patterns in PFAI data were last analyzed in 2010, using a previous version of FITNESSGRAM (Welk, Meredith et al., 2010).

A prominent finding in this study related to correlations between BMI and PACER achievement over time, particularly in how they varied by subgroup. Because a student's ability to achieve a passing score on the PACER test was, in these FITNESSGRAM versions, dependent on the student's BMI, some score correlation was expected. These relationships strengthened considerably when the study was partitioned by subgroup, with Grade 6 girls experiencing the strongest associations between the achievement rates. Statewide trends documented a drop in both girls' and boys' aerobic capacity achievement as students age, while BMI achievement remained comparable throughout all grades (Allums-Featherston et al., 2014b). It is unusual for only the girls, then, to experience extreme correlation between the two rates. The underlying cause of this high correlation between BMI and PACER achievement has yet to be fully articulated, especially in girls of middle school age. Fortunately, a 2015 update to FITNESSGRAM removed this BMI-dependent PACER calculation, potentially remedying this issue. Future research should continue to monitor this relationship to confirm all subgroups are equally affected by the independently-assessed constructs.

4.5 Limitations

Due to changes in FITNESSGRAM versions and reporting styles, this data represents only three years of publicly available PFAI reporting. Of this portion, schools and their subgroups were included only if they had at least five students tested and at least two out of three years reported. Although these methods strengthened the integrity of the analysis, smaller schools or schools that may have reported less often are not represented. Additionally, this study involved merging two secondary datasets, increasing the risk of random error inherent in either dataset, which may limit internal validity of study results.

4.6 Conclusion

The results of this study helped to confirm findings from previous PFAI results, documenting the changes in HFZ achievement rates in Texas. The study's more novel findings demonstrated other considerations with how the relationship between BMI and PACER achievement can vary, and in which student groups this correlation can occur. Because the interactions between the two fitness components can disproportionately affect middle school girls, studying why these discrepancies are observed can help strengthen the overall quality of fitness assessments and the assessment experience. Ultimately, evaluating fitness assessment standards will lead to improved fitness testing that is effective, educational, and enjoyable for all students.

5.0 CONCLUSION

The overarching purpose of this research was to examine trends in statewide youth fitness assessment reporting using publicly available data from the Texas Physical Fitness Assessment Initiative (PFAI). This data analysis aimed to provide evidence of the reach and impact of statewide youth fitness testing and ultimately improve the value of this statewide initiative. Two studies were developed to complete this purpose.

The first study sought to examine characteristics of school districts with exemplary reporting compliance. Its key findings confirmed previous reports that assessment and reporting requirements were annually upheld with high district-level reporting and documented impressive district-level reporting frequency over time. Nevertheless, predictors of reporting compliance emerged: districts more likely to report PFAI data had higher fitness performance among students, lower teacher turnover rate, higher socioeconomic status, and – among older grades only – higher attendance rate. These results are concerning because they suggest that the districts in most need of fitness improvement are not fully complying with the statewide assessment and reporting protocol. In other words, fitness assessment requirements may not be reaching the students in most need of fitness improvement; these students may also be missing the high-quality physical education (PE) programs intended to provide such assessment.

Based on this study, a more thorough investigation of the barriers to reporting occurring in lower-reporting districts is necessary, particularly in low-performing, low-SES, and high-turnover districts. Strategies to help increase reporting compliance can include aiding districts in establishing avenues of regular PFAI reporting that can withstand personnel changes and turnover. Additionally, districts should be flexible in allowing schools time for fitness testing, but the process should not overly tax school schedules, resources, teacher time, or student learning. The first step to overcoming these barriers, however, is to more fully articulate them.

The second study examined the trends in and the relationship between the two most common fitness tests in Texas, the Body Mass Index (BMI) and the Progressive Aerobic Cardiovascular Endurance Run (PACER). Respectively, these tests are proxies for "fatness" and "fitness" in common terms, which are interacting but independent constructs. The most novel finding from this study concerned the correlations between BMI and PACER achievement rates in Grade 6 girls. This relationship was expected to be moderately positive, but the unusually strong correlation may indicate that fitness and fatness are harder to distinguish for this subgroup. The evidence presented here helped justify the change to PACER calculations so that body composition and aerobic capacity can be measured and taught to students as independent constructs.

Other findings in the second study both supported and contradicted what previous reports have summarized. The Cooper Institute documented gradual improvement in aerobic capacity and body composition achievement over time (2014a), but this study did not replicate these results – likely due to the smaller amount of PFAI data used. Nevertheless, the study did match these observations: BMI achievement was slightly higher in girls, PACER achievement was higher in boys, and PACER achievement decreased as students aged, regardless of gender. Recommendations from the findings regarding fitness achievement changes suggest a need for physical activity (PA) programs, interventions, and services in these student populations to attenuate the observed trends.

Overall, the findings from these two studies generate many questions regarding the previously-debated theory and rationale of youth fitness assessment, which may shape the future of statewide, school-based fitness assessment policies and protocol. Although fitness testing advocates have spent over 20 years refuting the suggestion that "the horse is dead," (Rowland, 1995, p. 117) the question of statewide fitness testing still remains relevant in a time where the American education system is overwhelmed with standardized testing. Requiring fitness assessment, reporting, and monitoring for all school districts may be an ill-timed addition to the already saturated testing environment in Texas schools. Ideally, fitness testing is conducted so students learn from it and use the process to improve their health. The statewide policies currently framing the PFAI do not gauge student comprehension about fitness or measure teachers' ability to instruct effectively; based on the reported PFAI data alone, it is unclear if fitness assessment is working as an educational tool or as a diagnostic test for future PE or PA initiatives.

Complicating matters further, fitness test versions changed three times within the seven years of examined data, and an even newer version launched in early 2016. At best, these updates are adjustments for maximal accuracy and ease of implementation, making fitness testing more relevant and feasible in schools. At worst, they necessitate renewed training protocol for each version and new technological adjustments to ensure reporting compliance; they also obstruct validly measuring changes over time.

From one perspective, the state of Texas should be commended for developing school health policies related to PE, PA, and fitness promotion. Fitness assessment in schools can have beneficial effects: it provides students a chance to learn about their fitness and physical health; it can correct misconceptions about body image and healthiness; its standards can be adjusted to reflect the best impressions of health; and it can monitor statewide fitness trends (Morrow & Ede, 2009; Plowman & Meredith, 2013). Legislative frameworks for these objectives can help shape a culture of health and fitness among all students and schools. However, from another perspective, the completed studies demonstrated issues with assessing at the statewide level. Certain district characteristics may predict reporting compliance, and changes to fitness test standards can unintentionally affect certain student subgroups over others. Finally, based on the available PFAI data, there is neither a guarantee that learning is taking place during fitness assessments, nor evidence that students had a safe, enjoyable, and educational fitness testing experience. Fitness assessment without education is a disservice to students.

The problems observed with statewide fitness testing may not result from the assessment battery itself, but from how it is being used. FITNESSGRAM® is the most well-known and well-established fitness assessment battery available, but it was initially developed to provide individualized feedback to students about their health and fitness. In Texas, however, the test battery is being used for required statewide reporting and monitoring on an annual basis, aggregating results of millions of students in multiple grades. Fitness testing in Texas is caught between these two approaches: the former demands customizable, easy-to-interpret results shared with students, parents, and teachers; the latter demands standardized methods of reporting to be disseminated statewide and needs constant technological upkeep. The current PFAI requirements may limit how freely schools and districts can operate and could narrow the opportunities to actually educate students using fitness assessment procedures.

One broad solution is to allow more flexibility in statewide reporting requirements, which manifests itself in a number of ways. Perhaps assessing fewer grades, or formally reporting the results once every two years, can promote fitness testing as a benefit, not a burden, to school faculty and staff. In addition, embracing the emerging paradigm of PA assessment alongside fitness assessment could create customizable educational opportunities for students. Assessing PA, rather than fitness, may be more useful to monitor because fitness is a health outcome, while PA is a measurable behavioral choice (PYFP, 2014). Nationwide initiatives are following beginning to conform to this paradigm, including the Presidential Active Lifestyle Award (PALA) that appeared in 2008 (PCFSN, 2016). Because fitness and PA are both essential to student health, perhaps there is room for measuring both in Texas classrooms, or allowing teachers choose which one to assess. Since 1999, the FITNESSGRAM battery has included an ACTIVITYGRAM® component that fulfills this purpose (Plowman et al., 2006). Allowing for ACTIVITYGRAM reporting may further improve the assessment climate by letting Texas districts, schools, and classrooms choose what will best serve their students.

Finally, Texas policies should reflect that assessment is only the first step to changes in Texas youth fitness and PA. Fitness testing belongs as part of PE programs, but testing should not constitute all PE efforts. Rather, fitness testing should be a springboard to other health and PA-promoting activities, programs, and services in Texas school districts. Based on the findings from the completed studies, these endeavors should focus on aerobic PA opportunities in older grades, on body-image education that both measures BMI and clarifies what it means, and on building capacity for fitness and PA in districts with high teacher turnover rate and in low-SES populations. The combined effects of PE, PA, and fitness can improve both student physical health and also academic achievement (CDC, 2010; PYFP, 2014).

Overall, the results of this research are meant to fulfill one purpose, central to the tenets of fitness assessment philosophy: they are meant to evaluate. The first study provided feedback on district-level compliance with statewide reporting requirements over the first seven years of PFAI initiatives, highlighting both accomplishments and areas of improvement. The second study validated previous reports of Texas youth fitness achievement and offered new information on relationships between fitness performance indicators over time, demonstrating why changes to fitness measurements are beneficial. Both studies offered perspectives on statewide youth fitness testing in Texas and can now be used to create further strategies for future success. The results are meant to unite fitness assessment advocates and opponents alike in our mutual goal: healthy and happy Texas students.

REFERENCES

Allums-Featherston, K., Bai, Y., & Welk, G. J. (2014a). *Associations between health-related fitness and academic achievement (2011 to 2014): A report from the Texas youth fitness project* [PDF document]. Dallas, TX: The Cooper Institute. Retrieved from http://tea.texas.gov/Texas_Schools/Safe_and_Healthy_Schools/Physical_Fitness_Assessment_Initiative

Allums-Featherston, K., Bai, Y., & Welk, G. (2014b). *Levels of health-related physical fitness in Texas school children (2011 to 2014): A report from the Texas youth fitness project* [PDF document]. Dallas, TX: The Cooper Institute. Retrieved from http://tea.texas.gov/Texas_Schools/Safe_and_Healthy_Schools/Physical_Fitness_Assessment_Initiative

Barroso, C. S., Kelder, S. H., Springer, A. E., Smith, C. L., Ranjit, N., Ledingham, C., Hoelscher, D. M. (2009). Senate bill 42: Implementation and impact on physical activity in middle schools. *Journal of Adolescent Health, 45*, S82-S90. DOI: 10.1016/j.jadohealth.2009.06.017

Cale, L., Harris, J., & Chen, M. (2007). More than 10 years after “the horse is dead”: Surely it must be time to “dismount”?! *Pediatric Exercise Science, 19,* 115-131.

Centers for Disease Control and Prevention. (2010). *The association between school-based physical activity, including physical education, and academic performance* [PDF document]. Retrieved from http://www.cdc.gov/

healthyyouth/health_and_academics/

Centers for Disease Control and Prevention. (2013). *Comprehensive school physical activity programs: A guide for schools* [PDF document]. Retrieved from http://www.cdc.gov/healthyschools/physicalactivity/ cspap.htm

Centers for Disease Control and Prevention (2014). *Childhood obesity facts.* Retrieved from http://www.cdc.gov/obesity/data/childhood.html

Centers for Disease Control and Prevention. (2015a). *Full report: Results from the school health policies and practices study 2014* [PDF document]. Retrieved from http://www.cdc.gov/healthyyouth/data/shpps/results.htm

Centers for Disease Control and Prevention. (2015b). *Glossary of terms.* Retrieved from http://www.cdc.gov/physicalactivity/basics/glossary/

Centers for Disease Control and Prevention. (2015c). *School health policies and practices study: Trends over time 2000-2014* [PDF document]. Retrieved from http://www.cdc.gov/healthyyouth/data/shpps/results.htm

Coe, D. P., Peterson, T., Blair, C., Schutten, M. C., & Peddie, H. (2013). Physical fitness, academic achievement, and socioeconomic status in school-aged youth. *Journal of School Health, 83*(7), 500-507.

Cooper, K. H. (2010). Reflections on the Texas youth evaluation project and implications for the future. *Research Quarterly for Exercise and Sport, 81*(Suppl. 3), S79-S83. DOI: 10.1080/02701367.2010.10599697

Cooper, K. H., Everett, D., Meredith, M. D., Kloster, J., Rathbone, M., & Read, K. (2010). Preface: Texas statewide assessment of youth fitness. *Research*

Quarterly for Exercise and Sport, 81(Suppl. 3), ii-iv. DOI:10.1080/02701367.2010.10599687

Corbin, C. B., Pangrazi, R. P., & Welk, G. J. (1994). Toward an understanding of appropriate physical activity levels for youth. *PCPFS Research Digest, 1*(8), 1-8.

Corbin, C. B., Pangrazi, R. P., & Welk, G. J. (1995). A response to "the horse is dead; let's dismount". *Pediatric Exercise Science, 7*, 347-351.

Eisenmann, J. C., Welk, G. J., Wickel, E. E., & Blair, S. N. (2007). Combined influence of cardiorespiratory fitness and body mass index on cardiovascular disease risk factors among 8-18 year old youth: The aerobics center longitudinal study. *International Journal of Pediatric Obesity, 2*, 66-72. DOI: 10.1080/17477160601133713

Feiden, K. (2011). *The Texas youth fitness study: Looking at school policies as they relate to physical fitness and academic variables.* Princeton, NJ: Robert Wood Johnson Foundation.

Friedemann, C., Heneghan, C., Mahtani, K., Thompson, M., Perera, R., & Ward, A. M. (2012). Cardiovascular disease risk in healthy children and its association with body mass index: Systematic review and meta-analysis. *The BMJ, 345,* DOI: 10.1136/bmj.e4759

Going, S. B., Lohman, T. G., Cussler, E. C., Williams, D. P., Morrison, J. A., & Horn, P. S. (2011). Percent body fat and chronic disease risk factors in U.S. children and youth. *American Journal of Preventive Medicine, 41*(4S2), S77-S86. doi:

10.1016/j.amepre.2011.07.006.

Janak, J. C., Gabriel, K. P., Oluyomi, A. O., Perez, A., Kohl, H. W., & Kelder, S. H. (2014). The association between physical fitness and academic achievement in Texas state house legislative districts: An ecologic study. *Journal of School Health, 84*(8), 533-542. DOI: 10.1111/josh.12176

Kelder, S. H., Springer, A. S., Barroso, C. S., Smith, C. L., Sanchez, E., Ranjit, N., & Hoelscher, D. M. (2009). Implementation of Texas senate bill 19 to increase physical activity in elementary schools. *Journal of Public Health Policy, 30*(Suppl. 1), S221-S247. DOI: 10.1057/jphp.2008.64

Kraus, H., & Hirschland, R. P. (1953). Muscular fitness and health. *Journal of the American Association for Health, Physical Education, and Recreation, 24*(10), 17-19. DOI: 10.1080/23267232.1953.10627704

Kwon, S., Burns, T. L., & Janz, K. (2010). Associations of cardiorespiratory fitness and fatness with cardiovascular risk factors among adolescents: The NHANES 1999-2002. *Journal of Physical Activity and Health, 7*, 746-753.

Laurson, K. R., Eisenmann, J. C., & Welk, G. J. (2011). Body mass index standards based on agreement with health-related body fat. *American Journal of Preventive Medicine, 41*(4S2), S100-S105. DOI: 10.1016/j.amepre.2011.07.004

Lee, S. M., Nihiser, A. J., Fulton, J. E., Borgogna, B., & Zavacky, F. (2013). Physical education and physical activity. In *Results from the School Health Policies and Practices Study 2012* (33-47). Atlanta, GA: United States Department of Health and Human Services.

Lewallen, T. C., Hunt, H., Potts-Datema, W., Zaza, S., & Giles, W. (2015). The whole school, whole community, whole child model: A new approach for improving educational attainment and healthy development for students. *Journal of School Health, 85*(11), 729-739. DOI: 10.1111/josh.12310

Liu, Y. (2008). Youth fitness testing: If the "horse" is not dead, what should we do? *Measurement in Physical Education and Exercise Science, 12*(3), 123-125. DOI: 10.1080/10913670802219134

MacLean, L. M., Meyer, M., Walsh, A., Clinton, K., Ashley, L., Donovan, S., & Edwards, N. (2010). Stigma and BMI screening in schools, or "Mom, I hate it when they weigh me". In J. A. O'Dea & M. Eriksen (Eds.), *Childhood obesity prevention: International research, controversies, and interventions* (17-30). Oxford: Oxford University Press.

Mahar, M. T., & Rowe, D. A. (2008). Practical guidelines for valid and reliable youth fitness testing. *Measurement in Physical Education and Exercise Science, 12*(3), 126-145. DOI: 10.1080/10913670802216106

Martin, S. B., Ede, A., Morrow, J. R., & Jackson, A. W. (2010). Statewide physical fitness testing: Perspectives from the gym. *Research Quarterly for Exercise and Sport, 81*(Suppl. 3), S31-S41. DOI: 10.1080/02701367.2010.10599692

Meredith, M., & Welk, G. J. (eds.). (2013). *FITNESSGRAM/ACTIVITYGRAM test administration manual* (updated 4th ed.). Champaign, IL: Human Kinetics.

Mood, D. P., Jackson, A. W., & Morrow, J. R. (2007). Measurement of physical fitness and physical activity: Fifty years of change. *Measurement in Physical*

Education and Exercise Science, 11(4), 217-227. DOI: 10.1080/10913670701585502

Morrow, J. R., & Ede, A. (2009). Statewide physical fitness testing: A BIG waist or a BIG waste? *Research Quarterly for Exercise and Sport, 80*(4), 696-701. DOI: 10.1080/02701367.2009.10599610

Morrow, J. R., Martin, S. B., & Jackson, A. W. (2010). Reliability and validity of the FITNESSGRAM. *Research Quarterly for Exercise and Sport, 81*(Suppl. 3), S24-S30. DOI: 10.1080/02701367.2010.10599691

Morrow, J. R., Martin, S. B., Welk, G. J., Zhu, W., & Meredith, M. D. (2010). Overview of the Texas youth fitness study. *Research Quarterly for Exercise and Sport, 81*(Suppl. 3), S1-S5. DOI: 10.1080/02701367.2010.10599688

Morrow, J. R., Zhu, W., Franks, B. D., Meredith, M. D., & Spain, C. (2009). 1958-2008: 50 years of youth fitness tests in the United States. *Research Quarterly for Exercise and Sport, 80*(1), 1-11. DOI: 10.1080/02701367.2009.10599524

Must, A., & Anderson, S. E. (2006). Body mass index in children and adolescents: Considerations for population-based applications. *International Journal of Obesity, 30*, 590-594. DOI:10.1038/sj.ijo.0803300

National Physical Activity Plan. (2014). *The United States report card on physical activity for children and youth* [PDF document]. Retrieved from http://www.physicalactivityplan.org/projects/reportcard.html

National Physical Activity Plan. (2016). *United States national physical activity plan* [PDF document]. Retrieved from http://www.physicalactivityplan.org/

Nihiser, A. J., Lee, S. M., Wechsler, H., McKenna, M., Odom, E., Reinold, C., … & Grummer-Strawn, L. (2007). Body mass index measurement in schools. *Journal of School Health, 77*(10), 651-671. DOI: 10.1111/j.1746-1561.2007.00249.x

Pate, R. R. (1989). The case for large-scale physical fitness testing in American youth. *Pediatric Exercise Science, 1*, 290-294.

Pate, R. R., Welk, G. J., & McIver, K. L. (2013). Large-scale youth physical fitness testing in the United States: A 25-year retrospective review. *Pediatric Exercise Science, 25*, 515-523.

Plowman, S. A., & Meredith, M. D. (Eds.). (2013). *FITNESSGRAM/ACTIVITYGRAM reference guide (4th edition)* [PDF document]. Dallas, TX: The Cooper Institute.

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

President's Council on Fitness, Sports, and Nutrition. (2016). *Our history*. Retrieved from http://www.fitness.gov/about-pcfsn/our-history/

Presidential Youth Fitness Program. (2014). *Monitoring student fitness levels* [PDF document]. Retrieved from http://www.cdc.gov/healthyschools/physicalactivity/facts.htm

Rasberry, C. N., Lee, S. M., Robin, L., Laris, B. A., Russell, L. A., Coyle, K. K., & Nihiser, A. J. (2011). The association between school-based physical activity, including physical education, and academic performance: A systematic review of the literature. *Preventive Medicine, 52*(Suppl.), S10-S20. doi:http://dx.doi.org/10.1016/j.ypmed.2011.01.027

Rowland, T. W. (1995). The horse is dead; let's dismount. *Pediatric Exercise Science, 7*, 117-120.

Ruggieri, D. G., & Bass, S. B. (2015). A comprehensive review of school-based body mass index screening programs and their implications for school health: Do the controversies accurately reflect the research? *Journal of School Health, 85*(1), 61-72. doi: 10.1111/josh.12222.

Santiago, J. A., Roper, E. A., Disch, J. G., & Morales, J. (2013). The relationship among aerobic capacity, body composition, and academic achievement in fourth and fifth grade Hispanic students. *The Physical Educator, 70,* 89-105.

Silverman, S., Keating, X. D., & Phillips, S. R. (2008). A lasting impression: A pedagogical perspective on youth fitness testing. *Measurement in Physical Education and Exercise Science*, *12*(3), 146-166. DOI: 10.1080/10913670802216122

Singh, A., Uijtdewilligen, L., Twisk, J. W., van Mechelen, W., & Chinapaw, M. J. (2012). Physical activity and performance at school: A systematic review of the literature including a methodological quality assessment. *Archives of Pediatrics & Adolescent Medicine, 166*(1), 49-55.

Society of Health and Physical Educators. (2015). *The essential components of physical education* [PDF document]. Retrieved from http://www.shapeamerica.org

Texas Education Agency. (2016). *Physical fitness assessment initiative*. Retrieved from http://tea.texas.gov/Texas_Schools/Safe_and_Healthy_Schools/Physical_Fitness_Assessment_Initiative

Texas Senate Bill § 530. (2007). Retrieved from http://www.capitol.state.tx.us

The Cooper Institute. (2014a). *Texas youth fitness project (SY2011 to SY2014) executive summary* [PDF document]. Retrieved from http://tea.texas.gov/Texas_Schools/ Safe_and_Healthy_Schools/Physical_Fitness_Assessment_Initiative

The Cooper Institute. (2014b). *The Cooper Institute: Who we are.* Retrieved from http://www.cooperinstitute.org/about/

Thompson, B. (2008). *Foundations of behavioral statistics: An insight-based approach.* New York, NY: The Guilford Press.

United States Department of Health and Human Services. (2008). *2008 physical activity guidelines for Americans* [PDF document]. Retrieved from http://health.gov/paguidelines/

Van Dusen, D. P., Kelder, S. H., Kohl, H. W., Ranjit, N., & Perry, C. L. (2011). Associations of physical fitness and academic performance among schoolchildren. *Journal of School Health, 81*(12), 733-740.

Welk, G. J. (2008). The role of physical activity assessments for school-based physical activity promotion. *Measurement in Physical Education and Exercise Science, 12*(3), 184-206. DOI: 10.1080/10913670802216130

Welk, G. J., Jackson, A. W., Morrow, J. R., Haskell, W. H., Meredith, M. D., & Cooper, K. H. (2010). The association of health-related fitness with indicators of academic performance in Texas schools. *Research Quarterly for Exercise and Sport, 81*(Suppl. 3), S16-S23. DOI: 10.1080/02701367.2010.10599690

Welk, G. J., Meredith, M. D., Ihmels, M., & Seeger, C. (2010). Distribution of health-related physical fitness in Texas youth. *Research Quarterly for Exercise and Sport, 81*(Suppl. 3), S6-S15. DOI: 10.1080/02701367.2010.10599689

Wiersma, L. D., & Sherman, C. P. (2008). The responsible use of youth fitness testing to enhance student motivation, enjoyment, and performance. *Measurement in Physical Education and Exercise Science, 12*(3), 167-183. DOI: 10.1080/10913670802216148

Williams, M. (2015). *Pocket edition: 2013-14 Texas public school districts* [PDF document]. Retrieved from http://tea.texas.gov/communications/pocket-edition/

World Health Organization. (2016). *Physical activity.* Retrieved from http://www.who.int/mediacentre/factsheets/fs385/en/