The Physical Educator Vol. 81 • pp. 404–428 • 2024
https://doi.org/10.18666/TPE-2024-V81-I4-11867

PHYSICAL FITNESS

Physical Fitness Testing Concerns: Bullying and Inclusion

Melissa Bittner, David N. Daum, Tonya Moore, Debra Patterson, Dianne Wilson-Graham, and Patricia Suppe

Abstract

In February 2020, California Governor Gavin Newsom's proposed budget recommended suspending the state-mandated physical fitness test (FitnessGram®) for three years due to concerns over bullying and test discrimination against students who identify as gender non-binary and students with disabilities. The purpose of this study was to gain an understanding of California physical educators' perceptions of the state-mandated physical fitness assessment. Snowball sampling was used to recruit 1,082 participants who completed a survey with quantitative and open-ended questions. Analysis of the open-ended questions indicated the two main themes were "how to improve fitness testing" and "challenges in fitness testing." It is imperative to consider the needs of all students, including those who identify as gender non-binary or

Melissa Bittner, Associate Professor, California State University; David N. Daum, Associate Professor of Kinesiology, San Jose State University; Tonya Moore, Physical Education and Health Coordinator, Los Angeles County Office of Education; Debra Patterson, Professor, CSU; Dianne Wilson-Graham, Executive Director, CA Physical Education-Health Project; Patricia Suppe, Retired Physical Education Teacher, Loma Vista Middle School and District Coordinator Supervisor of Student Teachers CSU. The authors would like to acknowledge Dr. Barry Lavay, emeritus professor, CSU, for additions to previous versions of this manuscript. Please send author correspondence to Melissa.Bittner@csulb.edu

have a disability, and ensure physical fitness testing is conducted in a safe and welcoming environment.

Introduction

The FitnessGram® was developed by The Cooper Institute and is used to identify student progress related to the Healthy Fitness Zone®. It is estimated that FitnessGram® is used by approximately 67,000 schools and more than 22 million students each year (Cooper Institute, 2017). National organizations and programs, such as the Presidential Youth Fitness Program (PYFP) and SHAPE America (formerly American Alliance for Health, Physical Education, Recreation and Dance), have also adopted FitnessGram® as their official health-related physical fitness test (SHAPE America, 2017). The FitnessGram® includes a variety of standardized health-related physical fitness test items designed to assess (a) aerobic capacity; (b) muscle strength, endurance, flexibility; and (c) body composition. The scores that students obtain from the tests are evaluated against norm- and criterion-referenced standards that comprise Healthy Fitness Zones®. These fitness zones are based on potential health risks and considers various age and gender differences (Cooper Institute, 2017).

In California, the California Department of Education has designated the FitnessGram® as its state-mandated physical fitness test. The *California Education Code Section 60800* is the law that dictates the physical performance test approved by the state board of education be administered during the months of February through May to students in fifth, seventh, and ninth grades (California Department of Education, 2021). School districts must submit the results of physical fitness testing to the California Department of Education, at least once every two years. In the state of California, 10th-grade students are required to take four years of physical education but can opt out of two years by passing five out of six of the fitness subtests (California Department of Education, 2012).

In February 2020, California Governor Gavin Newsom's proposed 2020-2021 budget recommended suspending the state-mandated physical fitness test for three years due to concerns related to bullying of students, not accommodating for students who identify as gender non-binary, and not meeting the needs of students with

disabilities. Newsom's office outlined the plan as follows: "During the period of suspension, the department shall consult with experts and other stakeholders, including but not limited to, individuals with expertise in fitness, adapted physical education, gender identity, and students with disabilities, in order to provide recommendations regarding the purpose and administration of the physical performance test. The department may contract with a research entity to conduct a study regarding the physical performance testing of students" (Luery, 2020). The California Department of Education solicited membership and an assessment committee was formed in Fall 2021. Though the suspension did not occur, and the program continues, these are issues that should still be addressed.

Bullying

The Centers for Disease Control and Prevention defines bullying as "any unwanted aggressive behavior(s) by another youth or group of youths who are not siblings or current dating partners that involves an observed or perceived power imbalance and is repeated multiple times or is highly likely to be repeated" (Gladden et al., 2014). Bullying can happen in many forms, including physical, verbal, social, or cyber (Lavay et al., 2016). Bullying may be more prevalent in physical education because, when a student performs an activity, the results are public. For example, if a student fails a math test, their peers do not know. However, if a student shoots an air ball while playing basketball or strikes out when swinging with a long-handled implement, the results are seen by peers. Bullying can occur more frequently in physical education locker rooms and large class sizes where teacher supervision is limited (Lavay et al., 2016). Furthermore, Hurley (2010) indicated that 18.3% of respondents had experienced physical bullying in physical education, 23.7% had experienced verbal bullying, and 20.4% experienced social bullying. Those who experienced frequent bullying in physical education did not intend to take the class in the future.

A concern related to physical fitness testing is body shaming, which is when someone mocks or criticizes a person for a supposed body imperfection (Kulkarni, 2021). There is evidence that weight-based testing during physical activity is significantly associated with reduced sports enjoyment, reduced perceived activity compared with peers, and reduced mild-intensity physical activity (Faith et al.,

2002). Weight-based testing may affect aspects of mental, emotional, and physical health. During physical fitness testing, obtaining height and weight scores to determine body mass index may lead to bullying. For example, conducting body mass index testing (i.e., height, weight) in a public manner where others can view results may lead to body shaming from peers.

Gender Non-Binary Students

There is a growing awareness of gender identities other than the binary options of male and female, including transgender, non-binary, intersex, and gender expansive. Gender non-binary is defined as an umbrella term for all genders that go beyond male and female only (Thorne et al., 2019). It is important to recognize that not all non-binary people identify as transgender and not all transgender people identify as non-binary (Trans Student Educational Resources, 2020). In California, The Gender Recognition Act (2017) requires government-issued identification to provide the option for a non-binary gender category (the letter "x"). Specifically, birth certificates, driver's licenses, identity cards, and gender-change court orders are required to have a third category of gender non-binary. Individuals in California have the right to a third gender option on all official documents and public entities (like schools) must comply with gender recognition (e.g., identity, pronouns, gender markers, names) for unofficial records as well. To date, 21 states feature legislation that prohibits discrimination based on gender identity in either employment, housing, and/or public accommodations (Freedom for All Americans, 2022).

Consequently, the Gender Recognition Act (2017) has a direct impact on FitnessGram® testing and reporting procedures because schools use government-issued identification. The norms used to interpret the performance data from the FitnessGram® are based on binary gender categories (i.e., male, female), leading some to question whether students who identify as gender non-binary may be placed in an awkward or uncomfortable position if required to declare a binary gender identity (Woodworth, 2021). Additionally, if score reporting only has a binary option, it may be an illegal practice under The Gender Recognition Act as FitnessGram® does not allow for a third gender option. In response to the Gender Recognition Act and similar movements across the United States, the FitnessGram®

Scientific Advisory Board released a position statement titled "Gender Uses in FitnessGram®." This report identified that when reporting data, teachers need to use a child's sex assigned at birth to obtain the most accurate FitnessGram® results (California Physical Fitness Test Reference Guide, 2020). The report recommends that for students who identify as non-binary, the "Status" field should be left blank. Since FitnessGram® is intended to be used as an assessment tool, leaving a score blank does not assist adapted or general physical educators with the information needed to provide effective programming or aid in pedagogically appropriate practices such as goal setting.

Students With Disabilities

According to the National Center for Education Statistics (NCES), 14% of students in public schools have identified disabilities and are served under the Individuals with Disabilities Education Act (IDEA), reflecting 7.3 million children (NCES, 2021). Not all children with disabilities will require modifications or accommodations to equitably access physical fitness testing, but it is estimated that upwards of 29% of students with disabilities (approximately 2.1 million children) could benefit from inclusive assessment practices.

Although the FitnessGram® manual includes a section for addressing "strategies for testing students with disabilities" (Cooper Institute, 2017), specific test items and Healthy Fitness Zone standards for students with disabilities have not been provided. In California, students who are physically unable to complete any of the FitnessGram® tests are to be given as much of the test as conditions permit (*Education Code* Section 60800 and the *California Code of Regulations*, Title 5, Section 1041). The California Department of Education policy is to leave the score blank if a student is unable to perform a standardized fitness assessment skill, which is not an inclusive practice. Figure 1 indicates the PFT options in California, however, these options are not inclusive for all students. For example, a wheelchair user may struggle to find appropriate options within many of the testing areas, a student with an arm amputation is not accommodated within the upper body strength and endurance testing options, and students with high needs of support may not be able to perform any test items in aerobic capacity.

Figure 1
Current Approved PFT Items in California

Aerobic Capacity

- PACER (Progressive Aerobic Cardiovascular Endurance Run)
- One-Mile Run
- Walk Test (only for ages 13 or older)

Abdominal Strength and Endurance

- Curl-Up

Upper Body Strength and Endurance

- Push-Up
- Modified Pull-Up
- Flexed-Arm Hang

Body Composition

- Skinfold Measurements
- Body Mass Index
- Bioelectric Impedance Analyzer

Trunk Extensor Strength and Flexibility

- Trunk Lift

Flexibility

- Back-Saver Sit and Reach
- Shoulder Stretch

The Brockport Physical Fitness Test (BPFT; Winnick & Short, 1999) is an assessment that examines physical health in children with disabilities from ages 10 to 17. The test and its standards were specifically designed and researched for children with various disabilities, including cerebral palsy, congenital anomalies and amputations, intellectual disability, spinal cord injuries, and visual impairments. The need for this specific physical fitness test arose when it was noted that the standard physical test, FitnessGram®, did not meet the needs of people with disabilities. The BPFT should be used as an accompaniment to FitnessGram®, with students completing the FitnessGram® areas first and only using a modified BPFT item if a student is unsuccessful or determined best by an Individualized

Education Program (IEP) team decision (Winnick & Porretta, 2022). However, some states (i.e., California) have not adopted BPFT as an assessment test option. While FitnessGram® and BPFT are both published by Human Kinetics, they are separate manuals and assessments. This sends an inconsistent message on inclusion and if students with disabilities must complete their assessment from a different manual and battery of tests.

Importance of Physical Fitness Testing

Regardless of the issues that Governor Newsom identified related to fitness testing, school-based physical fitness testing is a means of educating students about their health (Pate et al., 2012; Silverman et al., 2008) and may motivate youth to engage in healthful levels of activity for a lifetime. It is important to monitor the epidemiology of youth fitness because of the very robust evidence regarding the beneficial health aspects of fitness (Keating et al., 2016). In the absence of school-level data, disparities in youth fitness across school districts, communities, and states may go unnoticed and, therefore, unaddressed. In addition, physical fitness testing data has led to significant amounts of funding for physical education programs. Other benefits of fitness testing include identifying strengths/needs for programming, student motivation through goal setting, monitoring improvement, and advocacy by promoting the importance of physical education (Cooper Institute, 2017). The purpose of this survey was to understand California physical educators' perceptions of the mandated physical fitness test related to bullying, gender identity, and meeting the needs of students with disabilities.

Method

Participants and Recruitment

There were 1,082 physical educators recruited by email through the California Association for Health, Physical Education, Recreation, and Dance (CAHPERD) listserv and CAHPERD social media (i.e., Facebook, Twitter, Instagram). Inclusion criteria, determined by survey questions, required participants to be over 18 years of age and a current K-12 physical education or adapted physical education teacher in California. All data collection procedures were

approved by a university Institutional Review Board for the protection of human subjects.

The survey was emailed (or hyperlinked if accessed through social media) to potential participants using Qualtrics, an online survey platform. Two weeks after the initial email, the survey was sent a second time to all the participants through a follow-up email and social media blast Participants anonymously completed the informed consent and online survey at a location of their choice.

Procedures

A 14-question quantitative and qualitative survey was developed for this study. The online survey was designed to focus on issues of bullying, gender identity, and students with disabilities during the administration of the FitnessGram®. The survey was developed by the six members of the CAHPERD Assessment Task Force. Three of the members were university faculty with specializations in physical education and/or adapted physical education, one worked in a county office of education with LGBTQIA+ expertise, one worked for a state-funded professional development organization, and one was a retired K-12 physical educator. All members of the taskforce had experience using and/or training teachers on the administration of the FitnessGram®, working with students with disabilities, and extensive experience working with K-12 students. Content validity was established after the initial revisions through sending the survey to three experts with experience researching assessment in physical education for feedback on content relevance and question structure. After receiving feedback, the survey was again examined and revised until all changes were unanimously agreed.

The final survey consisted of two sections: demographics (seven questions) and questions regarding bullying, gender identity, and students with disability (seven questions). The demographic section pertained to participant characteristics (e.g., gender, highest degree, and ethnicity). Questions related to bullying, gender identity, and students with disabilities consisted of three quantitative questions and four open-ended questions. Three of the four open-ended questions related to bullying, gender identity, and challenges related to administering the FitnessGram®. The final open-ended question invited participants to share any additional thoughts related to the topics on the survey.

Data Analysis

Descriptive statistics were reported for the demographic variables. Chi-square Test of Independence analyses were used to evaluate the relationship between the individual demographic characteristics and the level response to a question using a 5-point Likert Scale. A 5-point Likert scale was used for four of the questions. Choices for the four questions were 1 = never, 2 = rarely, 3 = neutral, 4 = often, 5 = very often. For statistical analysis, choices 1 and 2 were combined as were choices 4 and 5 to yield never/rarely, neutral, and often/very often. The rationale for combining responses was that respondents may interpret the subjective choices "often" and "very often" or "rarely" and "never" differently. By consolidating the responses, it was possible to look at general trends. A probability level of $p \leq .05$ was used to determine statistical significance for all analyses using the Statistical Package for Social Sciences (SPSS) v.25 (IBM Inc., 2017). Where there was an association found, a Goodman's follow-up test was run, and p values were adjusted using a Bonferroni correction to consider multiple comparisons.

The six authors reviewed the open-ended question responses with frequency counts of the categories reported. Each of the three open-ended questions was coded by teams of two; each team had one university faculty member experienced with qualitative data analysis. Measures of trustworthiness were enacted by the second author. The results from the open-ended questions were thematically analyzed using Braun and Clarke's (2006) method for qualitative analysis. This entailed the research team (a) familiarizing themselves with the data through reading and re-reading of data; (b) generating initial codes and organizing the codes into meaningful groups; (c) searching for themes through sorting the codes into potential themes; (d) refining of the themes through further review and analysis; and (e) defining and naming themes.

Results and Discussion

The following section includes results and discussion in relation to Governor Newsom's concerns during physical education assessment on bullying, students with disabilities, and students who identify as gender non-binary. Demographics are briefly discussed, and then the data are discussed in relation to two themes with sub-

themes. The two main themes are "how to improve fitness testing" and "challenges in fitness testing."

Demographics

There were 1,082 physical educators in California that completed the anonymous and confidential survey. Of the respondents, 58% (n = 625) identified as female, 41% (n = 445) identified as male, and 1% (n = 7) preferred to self-describe. The majority (68%) of participants were white (n = 813), followed by 15% Latinx, Hispanic, or Spanish (n = 176), and other identified ethnicities had less than 5% total representation. Fifty-eight percent (n = 621) of participants had obtained a master's degree, with 41% (n = 435) having a bachelor's degree. Participants identified as general physical education (GPE) 76% (n = 832), adapted physical education (APE) 6% (n = 63), or both GPE and APE 13% (n = 140).

How to Improve Fitness Testing

When responding to the question of how to eliminate or decrease bullying, over 60% of participants (n = 522) identified teaching practices or testing administration practices that lead to privacy. There were six sub-themes that emerged from the data: testing individually or in small groups, testing for privacy, teaching students pro-social behaviors, teaching the "why" of fitness testing, addressing bullying, and accommodating gender identity.

Testing Individually or in Small Groups

The most prevalent recommendation to allow for privacy was to test students one-on-one, in pairs, or small groups. Many participants responded with just "small groups" or a similarly brief response with no explanation. However, some participants expanded their rationale behind the benefit of testing individually or in small groups because this allows the teacher to "assess students in small groups or individual to limit the amount of mass exposure to the total class" (participant 138, female, 20+ years teaching, Adapted Physical Education [APE]). Additional benefits of small group testing will "help them [physical educators] organize, administer, and monitor better keeping a safer environment. The process would go more smoothly as well" (participant 137; 4-10 years; female; APE). For some participants, grouping depended on the test being admin-

istered; for example, they generally tested "students individually or in groups of no more than 2-4 at a time (except pacer which I do half a class at a time running" (participant 966; female; 20+ years; elementary). Some participants identified that they were able to test in small groups because of the size of their school or class. However, testing in small groups can create logistical issues for other physical educators. For example, one participant discussed that when a teacher is administering a test to a small group it is "hard to maintain order when you have to constantly look up to see if the other 60-90 kids are still on task" (participant 86, male, 20+ years, elementary).

Sentiments from the survey participants echoed best practice recommendations from Cooper Institute (2017). These tips included testing in small groups and having the other students rotate in and out when not being assessed to ensure the whole class is engaged in physical activities. Spreading the assessment skills over a period (e.g., test one or two items per week) can also help with administering in smaller groups. Another tip included developing and training a cadre of volunteers who can assist on testing days. While these practices may better allow for testing in small groups, it does not always protect student privacy.

Testing for Privacy

There were two connected concepts within testing for privacy (i.e., keeping test scores private from other students and setting up the testing environment). Several participants discussed having conversations with students about how they discuss the results of the test within class, for example, one participant said, "that students only have a right to comment on their own scores" (participant 966; female; 20+ years; elementary), while another educator admitted that while he had a strict rule to not allow discussion of results in class, "I can't control what they do after class, though" (participant 99; male; 0-3 years; high school). One participant went further and said they "don't acknowledge or reward students who are doing well in the tests" (participant 410; female; 11-19 years; middle school). Several participants discussed not making the fitness test a competition. There was consensus among participants that public posting or discussion of results was an inappropriate practice.

The second concept related to privacy was setting up the testing environment so students cannot view the results of others. Privacy

in the testing environment was discussed mostly in the context of taking height and weight. For example, one participant said, "When taking weight of students we make sure no other students can see" (participant 1060; female; 4-10 years; middle school). Testing in a private setting was very important for one participant because "even if it takes longer, it's worth it for student wellbeing" (participant 501; female; 0-3 years; middle school). Similarly, one participant recommended that educators need to "be clear that your [student] results are NO ONE's business. Be discreet while testing without shouting names. Be encouraging to ALL. Keep height and weight private. Record all the answers yourself.... without a student helper" (participant 557; female; 20+ years; middle school). Privacy of results was generally discussed as a fundamental right for the students and "ALL students should have the option to test in private if they are embarrassed" (participant 1060; female; 4-10 years; middle school).

For some tests, setting up the learning space so students cannot view the results of other students is challenging because of the public nature of being physically active in front of other students, as participant 269 (female; 11-19 years; elementary) discussed:

> The test is entirely public to the rest of the class. It is extremely easy to see if a student stops early for a component of the assessment, or if they are behind in say the 1-mile test while the rest of the students wait. Another is the BMI portion is very transparent for students based off of visually seeing their body type. The weight might not be made public but unfortunately students are easily able to identify other students who weigh more

Similarly, another educator compared the public nature of physical performance to public speaking and how "being called on to read aloud in class builds character and improves communication skills, performing and playing with and in front of peers builds the same character and skills" (participant 551; male; 20+ years; middle school).

Privacy sentiments are echoed by the Society for Health and Physical Education (SHAPE) America's position statement on *Appropriate and Inappropriate Practices Related to Fitness Testing* (2017). It concludes that individual student results should be made

available to students and their parents/guardians and should not be shared publicly. Teachers and schools should keep data confidential at all times, including teaching students that fitness data are personal; scores should never be posted or called out in class.

Teach Students Pro-Social Behaviors

Some participants (n = 213) discussed creating a positive and bully-free zone. For these educators it was about building the culture that "everyone is positive in the understanding that everyone has strengths and weaknesses" (participant 84; female; 20+ years; middle school) and to "create an environment that supports everyone and also giving the students an understanding that everyone is different and that is ok" (participant 610; female 11-19 years; high school). Several participants discussed teaching students about positive social behaviors, such as frontloading students "about respect and personal responsibility well before testing begins" (participant 597; male; 4-10 years; elementary and middle school) and emphasizing "personal development and acceptance of differences" (participant 739; female; 20+ years; middle school). One elementary educator (participant 445; female; 20+ years; elementary) went into detail about how they review appropriate social behaviors and allow students to pair up with whomever they chose. "This may allow for a little cheating, but I prefer minimal dishonesty to anyone being paired with someone who will be with someone discouraging or unkind. They are partners and are there to encourage, support, and are accountable for the other person's score... They love it and it is fun. I have kids who ask to do test over three and four times because they don't feel they gave their best."

Having students try their best was a common sentiment. There were several approaches identified by participants, such as downplaying the importance of the test, having students cheer on each other, having students challenge each other, and telling success stories. For example, one educator likes to tell students how they "failed every one of the tests until I was in high school… [to] help them understand that everybody is different and some of the tests will be hard for some and easy for others" (participant 277; male; 4-10 years; elementary and middle school).

Teach the "Why" of Fitness Testing

While the survey did not specifically ask participants about the importance of fitness testing, there were participants who used fitness testing to teach students the value of fitness testing and why it is important to the student. For example, participant 608 (female; 4-10 years; high school) discussed how students should understand health and fitness and how "it [PFT] relates to everyday life beyond their general education years. They should know the reasons for taking a fitnessgram test and why its important to know your healthy fitness zone." Using the test results for self-improvement and goal setting was stressed by many participants. The test is "about self-improvement and not comparing themselves with others. [The] best approach is communication and individual goal setting" (participant 642; female; 20+ years; high school). The test is just "one tool in a tool bag of tools in developing healthy choices and fitness levels."

Doing the test more than once in a school year was also a strategy employed by some. For example, this educator (participant 215; male; 4-10 years; elementary) said, "I prepare students for Fitnessgram since the beginning of the year so they look forward to see their levels of fitness." By employing the fitness test multiple times a year, this allows educators to utilize "meaningful practices like goal setting, personal fitness plans etc. to make it meaningful and shift the focus to personal growth" (participant 279; male; 0-3 years; high school). Using strategies such as goal setting allows "students to celebrate growth over time (regardless of official pass/fail)" (participant 83; female; 20+ years; high school). This educator further discussed not communicating the healthy fitness zones to their students "so that the pressure is eliminated, and they can focus on growth over time."

Addressing Bullying

A few participants (n = 54) identified they had minimal or no problem with bullying. For example, participant 369 (male, 20+ years, middle school) said, "I don't allow bullying period." Or that they "have never seen bullying, or even heard of it during PFT in 17 years" (participant 229; female; 20+ years; elementary). Similarly, another participant (116; female; 11-19 years; elementary) said:

> In all the years I have taught PE and administered the FitnessGram each year, I have never heard of a student at

> our school being bullied during any portion of the test. That doesn't mean it hasn't happened and not brought to my attention, but we do our best to explain the tests purpose, importance and individuality of performance. We also make sure to keep testing a no scoring private and provide as much privacy as we can for those that want it.

An outlier response was defiant in that bullying is even a problem. The participant said, "Just test the kids, there really isn't a bullying problem. Stop raising marshmallows" (participant 395; female; 20+ years, middle school). Some respondents ($n = 27$) that identified bullying was not a problem did not expand on why they believed this. Conversely, when addressing bullying another participant (205; male; 11-19 years; high school) identified that the culture of a class matters in the prevalence of bullying and that if the teacher already has good management, the problems will be minimal: "If the teacher struggles with management, the problems will be plentiful. At our site, whenever there is a teacher who isn't great in a daily basis... there have been problems."

This response is similar to results from Woodward and Schneider (2021), who conducted a critical evaluation of the case for pausing California's school-based fitness testing with results not supporting the assertion that physical fitness testing placed students at elevated risk for bullying and/or weight-based teasing as compared to other school settings. However, there was a dearth of evidence as only three research studies were indicated on the topic specific to bullying and physical fitness testing; more rigorous research is warranted.

Accommodating for Gender Identity

Regarding receiving professional learning in the new laws around gender identity (i.e., Gender Recognition Act), 77% ($n = 820$) of participants reported that they had not received any professional development. Experienced teachers were significantly more likely to have received professional learning around gender identity χ^2 (3, 1063) = 21.02, $p < .001$ compared to new teachers (within the first four years of teaching). Most teachers indicated "neutral" (i.e., 3 on the 5-point Likert scale) when asked if they know how best to support students who identify as gender nonbinary ($M = 2.89$; $SD = 1.05$). Physical education teachers in California are not stating with

confidence that they know how to support students who identify as gender nonbinary; thus, professional development in this area may be warranted.

Gender identity is an issue of privacy, comfort, and personal choice. Like many other issues within education, issues of gender identity are politicized and loaded with personal and religious beliefs. Some participants felt no accommodation is necessary because "a male is a male, a female is a female. One's 'feelings' do not determine if you have a vagina or penis" (participant 1000; male; 20+ years; high school). Another participant (357; female; 20+ years high school) who felt similarly said, "In my professional opinion, gender identity should have no impact on the PFT. PFT scoring should NOT be based on identity but rather anatomy. Penis = scoring as a male and vagina = scoring as a female. What the student identifies as is irrelevant to the PFT." This type of thinking can cause harm to any transgender, non-binary, or gender-fluid students and indicates physical education teachers in California need additional training related to Senate Bill 179 Gender Recognition.

On the other hand, there were participants who felt that accommodating for gender identity was needed. Some suggestions related to creating a third category of "non-binary entry for gender for students, score and norms for students in this category" (participant 202; male; 4-10 years; middle school) or to have just "one score sheet that includes all genders" (participant 378; female; 4-10 years; middle school). Another suggestion (participant 192; female; 20+ years; high school) was to allow transgender students to "choose which zones apply to them. Results are discussed individually as well as strategies on how to improve." Another participant (240; female; 20+ years; elementary) felt that it might take reimagining how results are presented:

> Have one set of criteria. Then set up three columns, (approaching, meeting, and exceeding) competency. All students use the same scale. In addition, let us use our Physical Education standards—they cover all of the knowledge that students need to have. Put our content on the standardized tests—then and only then will admin and parents care about learning in Physical Education.

Currently, scoring is presented in a binary way; there is a male and female set of criteria. As one participant (269; female; 4-10 years; elementary) noted, "How can we have biological standards without addressing gender identity?" This is important to note because teachers can accidentally "out" students by having them identify publicly if they are male or female.

The concerns surrounding gender non-binary issues are not central to FitnessGram®. The International Olympics Committee and National Collegiate Athletics Association (NCAA) are also navigating transgender participation in a manner that preserves opportunity for transgender athletes while balancing fairness, inclusion, and safety for all who compete (NCAA, 2022). Both the Olympics and NCAA policies currently call for gender non-binary participation for each sport to be determined by the policy for the national governing body of that sport.

Challenges in Fitness Testing

Participants were asked what challenges they faced when administering the FitnessGram®. Most participants identified challenges (n = 908). The sub-themes include supporting structures (e.g., administrators), time and class size, the need for professional learning, and the test needs to change. It is important to note that, across these themes, challenges related to fitness testing were barriers to having accurate data. For example, one participant (949; female; 20+ years; elementary) said due to equipment issues, "I don't take it seriously—the scores are inaccurate, student's form is inaccurate. It's ridiculous." However, there was a small subset (n = 48) that identified they did not have any issues. The participants who stated they had no challenges simply put "none" or "I don't have any challenges in administering the fitnessgram."

Supporting Structure

Support came out as a multifaceted issue. There were issues such as administrative support, support from parents, and support from other educators. One participant (269; female; 20+ years; high school) mentioned there was "not enough importance given *[to the FitnessGram®]* by *[the]* district." Without support from the district and school administrators, there is likely minimal support related to budget because as one teacher noted, "We don't have enough testing

equipment to allow an efficient testing procedure. I'm sure that accuracy *[of the test results]* is impacted due to this" (participant 116; female; 11-19 years; middle school). Administrative support is critical to assist with large class sizes, scheduling of the FitnessGram® testing window, and supervision assistance. A participant (462; male; 11-19 years; elementary) identified there was "no support for supervision of the students who are not being tested." Administrative and parent support were seen as important and connected to student participation. For example, one participant (326; female; 0-3 years; high school) indicated that "students just refuse; we need more parent and admin support for PFT!" Another participant (258; female; 4-10 years; elementary and high school) noted the issue was a lack of communication "regarding scheduling/calendar for subtest administration." Additionally, there was a desire for "all teachers administering the test in the same way and technique" (participant 25; female; 20+ years; all levels).

Students With Disabilities

Related to not being supported, participants were asked, "How often did student Individualized Education Plan (IEP) and Section 504 plans give direction for accommodations or modification with the physical fitness test?" Sixty-two percent responded rarely or never (n = 674). General physical education teachers were significantly more likely to indicate IEP and Section 504 plans never/rarely give direction for accommodations or modifications for the physical fitness test χ^2 (4, 1031) = 9.55, p = .01. Students with disabilities should have appropriate modifications during physical fitness assessment. The policy of leaving the "Status" field blank (i.e., California Department of Education) if a student with a disability is not able to complete one of the recommended modifications does not indicate what the student can do and compromises meeting that student's needs.

Time and Class Size

Additional factors related to challenges in fitness testing were time and size of classes. Limited time is attributed to two factors, the first is the testing window, which is the time frame that testing must be completed, and the second is the amount of time for the class period. Frequently, time and class size were listed together because they are inherently linked. Large classes take more time to admin-

ister the test. Good testing takes time "especially when you have accommodations for students" (participant 115; female; 20+ years; elementary and middle school) and to "administer testing privately" (participant 15; female; 20+ years; middle school).

The challenge of physical education dealing with larger class sizes is well-established in the literature (Kirkam-King et al., 2017; National Association for Sport and Physical Education, 2006). The participants in this investigation frequently cited large class sizes as an issue for FitnessGram® testing because it made it hard to supervise, meet the needs of special education students, and be able to assess small groups while still teaching the whole class. Several teachers identified that having too many students was impacting the accuracy of scoring; for example, participant 94 (male; 20+ years; middle school) said there are "too many students to accurately test while keeping the rest of class safe and engaged in an activity. Impossible to monitor both the students testing and the students not testing." Another participant (279; male; 0-3 years; high school) identified that their class sizes were big but other than that "I do not face challenges. I fitness test three times a year and honestly, my students love it. I worked so hard to create an environment where students feel safe to fitness test in."

The Need for Professional Learning

Professional learning has played a central role in teachers' growth for many decades (Collinson & Ono, 2001). Professional learning for teachers in physical education can take on various forms (e.g., workshops, training, seminars, university coursework, journals, and social media). The need for professional learning regarding PFT fell into two categories. The first is regarding areas of participant need and the other is based on inappropriate testing practices or not understanding the test. This survey indicated that for the FitnessGram® to meet the needs of all students, changes may be proposed related to test items, administered protocols, and procedures on scoring and reporting. If this occurs, 88% of teachers indicated it would be 'very important/important' (n = 922) to have new professional learning programs. The most frequently cited area of professional learning was related to student engagement and motivation. Teachers had issues with "students' willingness to participate fully" (participant 238; female; 11-19 years; high school), "earnest effort to perform at

best ability" (participant 282; male; 20+ years; high school), and that "students don't care about their time/score and don't meet their potential" (participant 330; female; 11-19 years; elementary). Teaching accommodations and modifications, as well as motivating students with disabilities, were also frequently identified by participants as areas where professional learning is needed.

It cannot be understated that it is important that physical educators understand the science behind the tests. Teacher comments included, "kids are not motivated because it [the fitnessgram] is outdated. Why are we testing them on things that adults don't get tested on" (participant 314; male; 11-19 years; middle school), "I don't believe students with higher BMIs should be running the mile in a lower time" (participant 374; female; 20+ years; middle school), and "tests do not accurately assess the energy system assessed" (participant 455; male; 20+ years; middle school). It was clear from participant responses that professional learning needs to focus on the "why" of the tests; not just how to communicate it to students, but also for the physical educators. Aligning each test with what component of fitness it is testing, how it is testing that area, and why this test is being used is clearly needed. Professional learning related to BMI and cardiovascular fitness is a necessity.

The Test Needs to Change

Some participants (n = 78) simply want the test to go away or parts of the test to be eliminated. Several teachers felt the test is "outdated for today's youth and are not attainable within the culture today" (participant 537; female; 20+ years; middle school). Most frequently the BMI (n = 32) was identified as the part of the test that needed to change. The criticism tended to present in two forms. The first is that height and weight measurements cause issues of test administration and privacy, and the second is an issue of bullying and mental health. For example, participant 227 (male; 20+ years; elementary) said, "Find a new system to test Body Composition that is less intrusive, one that doesn't stress the student."

It is important to note that BMI, like the other test items, is not a tool for diagnosing health risks among students, but it is simply a screening tool (SHAPE, 2017). If a student's BMI score signals a warning (too high or too low), the school informs the student's family, with a suggestion to follow up with the family's physician for fur-

ther evaluation. To date, there is not enough evidence for scientists to conclude whether school-based BMI measurements are effective at preventing or reducing childhood obesity or whether they cause harm, by either increasing the stigma attached to obesity or increasing pressures to engage in unsafe weight control behaviors (Centers for Disease Control and Prevention [CDC], 2017). To minimize the risk of potential negative consequences, schools that measure students' heights and weights can follow the 10 safeguards outlined by the CDC (e.g., appropriate expertise and privacy).

Another issue was how the testing norms are presented and whether the categories of healthy, or not, are potentially harmful. As one participant indicated (1047; male; 11-19 years; middle school), "make it more of a personal assessment that is used to measure growth over a child's development from fifth, seventh, and ninth grade." It is important to note that the FitnessGram® results can be used in this way, but clearly it is not, perhaps due to issues identified earlier, such as class size, time, and support.

Limitations and Future Research

Within the limitations of this investigation, this survey was only open to physical education teachers in California. It is feasible that nationally, physical education teachers may feel differently about fitness testing in relation to bullying, students who identify as gender non-binary, and students with disabilities. Though California is currently the only state with a Gender Recognition Act *and* identifies FitnessGram® as its state-mandated physical fitness assessment, it behooves other states to be aware of these issues. These topics reflect contemporary concerns within the educational landscape and are an important discussion for all educators. Thus, it is critical to address such factors to ensure the equity of physical education for all students. As physical education professionals, we want our students to have a positive experience in physical education, including during fitness testing, and feel empowered to participate in lifetime fitness. Thus, we should take an introspective lens to consider what we are doing about these concerns during physical fitness testing.

Future research should include methods to ensure physical fitness testing is equitable, diverse, and inclusive. There are many considerations and strategies for physical educators that may help to support all students' ability to thrive and succeed in the educational

environment. Using gender-neutral language (i.e., they) when referencing gender identity or sexual orientation can create an inclusive and safe environment (Cooks-Campbell, 2021). It is important not to make assumptions about how a student identifies based on appearance and to be as culturally sensitive and inclusive as possible when discussing gender and relationships. It is also important to increase students' access to an inclusive curriculum with sensitivity to pronouns (e.g., they/them). Generally, avoid dividing the class by gender, having single-gender teams or activities, and try to mix traditional gender roles and eliminate "boys" versus "girls." Be mindful of students who are not public about their gender identity or sexual orientation and ensure gender identity or sexual orientation is never revealed or discussed with anyone without the student's explicit consent. This is especially true when educators communicate with students, families, and other educators about physical fitness test scores. This study reveals that few physical education teachers are receiving professional learning in the new laws around gender identity (i.e., Gender Recognition Act) as 77% (n = 820) of participants reported that they had not received any professional training. More research is needed on this topic as we work to move toward more inclusive pedagogy.

An option to create a more inclusive physical fitness test may be to ensure equal access and training in the *Brockport Physical Fitness Test* (BPFT; 1999, 2014). Several states (e.g., Vermont and Illinois) use the BPFT as an accompaniment to the FitnessGram®. The BPFT has 27 unique test items specific to various disabling conditions. Physical education teachers are receiving very little guidance and direction in fitness testing for students with disabilities as 62% (n = 674) responded never/rarely with regard to direction for accommodations or modifications for the physical fitness test.

Conclusion

This study adds to the literature base that physical fitness testing is often implemented in a way that is inconsistent with recommended best practices, which may be the underlying reason for widespread concern about the impact of physical fitness testing on students' well-being (Woodward & Schneider, 2021). As identified in earlier themes, professional learning is needed to address issues, such as ensuring student privacy, addressing bullying, managing large class

sizes, addressing the "why" of fitness testing with students, incorporating goal setting, working with overweight/obese students, meeting the needs of students with disabilities, and ensuring accuracy of the test. For example, Vermont has an online module for professional development on FitnessGram® and BPFT, with explanations of each skill and short videos to accompany each. When teachers complete the modules, they are sent a certificate of completion.

As was noted by Woodworth and Schneider (2021), many of the concerns expressed by California's Governor Gavin Newsom are not empirically supported. This is not to say that a problem does not exist, but rather that further inquiry into these concerns is required. Until the criticisms of FitnessGram® are addressed, it will remain difficult to conclusively regard physical fitness testing as an integral part of population level health promotion (Suminski et al., 2019).

Though it comes with challenges, fitness testing is worthwhile because it provides students with information about their health-related fitness and alerts them to potential health risks to which they might not otherwise have access (SHAPE, 2017). Fitness testing can be used to help students learn to set goals, develop an appropriate training program, and monitor progress. Fitness testing helps add legitimacy to the field of physical education through the implementation of standardized assessment. Ultimately, we want all our students to have a positive learning experience during fitness testing, which may lead to lifetime fitness and living a healthy lifestyle.

References

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3(*2), 77–101. doi: 10.1191/1478088706qp063oa

California Department of Education. (2012). *CalEd Facts, Physical Fitness Testing*. California Department of Education. https://www.cde.ca.gov/ta/tg/pf/cefpft.asp

California Physical Fitness Test Reference Guide. (2020). https://pftdata.org/files/Reference_Guide.pdf

Centers for Disease Control and Prevention. (2017). *Body mass index (BMI) measurement in Schools*. Centers for Disease Control and Prevention.

Collinson, V., & Ono, Y. (2001). The professional development of teachers in the United States and Japan. *European Journal of Teacher Education, 24*(2), 223–248. https://doi.org/10.1080/02619760120095615

Cooks-Campbell, A. (2021). *What gender inclusive means, and why your use of pronouns matters*. https://www.betterup.com/blog/gender-inclusive

Faith, M. S., Leone, M. A., Ayers, T. S., Heo, M., & Pietrobelli, A. (2002). Weight criticism during physical activity, coping skills, and reported physical activity in children. *Pediatrics*, *110*(2), e23–e23. https://doi.org/10.1542/peds.110.2.e23

Freedom for All Americans. (2022). https://freedomforallamericans.org/states/

Gender Identity Act of 2017. (2017). Senate Bill 179, Chapter 853 *et seq*. https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201720180SB179

Gladden, R. M., Vivolo-Kantor, A. M., Hamburger, M. E., & Lumpkin, C. D. (2014). *Bullying surveillance among youths: Uniform definitions for public health and recommended data elements*. Version 1.0. In Centers for Disease Control and Prevention.

Hurley, V., & Mandigo, J. (2010). Bullying in physical education: Its prevalence and impact on the intention to continue secondary school physical education. *PHENex Journal*, *2*(3).

Keating, X. D., & Lambdin, D. (2016). Effects of a three-tiered intervention model on physical activity and fitness levels of elementary school children. *The Journal of Primary Prevention, 37*(4), 313–327. https://doi.org/10.1007/s10935-016-0430-y

Kirkham-King, M., Brusseau, T. A., Hannon, J. C., Castelli, D. M., Hilton, K., & Burns, R. D. (2017). Elementary physical education: A focus on fitness activities and smaller class sizes are associated with higher levels of physical activity. *Preventive Medicine Reports, 8,* 135–139.

Kulkarni, S. (202). *Perfectly Imperfect: "Body Shaming."* https://www.psychline.in/blog/perfectly-imperfect-body-shaming/

Lavay, B., French, R., & Henderson. H. (2016). *Positive behavior management in physical activity settings* (3rd ed.). Human Kinetics.

Luery, M. (2020). *Newsom wants to suspend PE testing: 3 things to know*. https://www.kcra.com/article/newsom-wants-suspend-pe-testing-3-things-to-know/30786294#

National Association for Sport and Physical Education. (2006). *Teaching large class sizes in physical education: Guidelines and strategies* [Guidance document].

National Collegiate Athletic Association. (2022). *Board of Governors updates transgender participation policy*. https://www.ncaa.org/news/2022/1/19/media-center-board-of-governors-updates-transgender-participation-policy.aspx

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.

Society for Health and Physical Education. (2017). Appropriate and inappropriate practices related to fitness testing. *Journal of Physical Education, Recreation & Dance, 88*(6), 3-9. doi:10.1080/07303084.2017.1331636

Suminski, R. R., Blair, R. I., Lessard, L., Peterson, M., & Killingsworth, R. (2019). Physical education teachers' and principals' perspectives on the use of FitnessGram. *SAGE Open Medicine*, *7*, 2050312119831515. https://doi.org/10.1177/2050312119831515

The Cooper Institute. (2017). *Fitnessgram by the Cooper Institute*. https://fitnessgram.net/

Thorne, N., Yip, A. K. T., Bouman, W. P., Marshall, E., & Arcelus, J. (2019). The terminology of identities between, outside and beyond the gender binary: A systematic review. *International Journal of Transgenderism, 20*(2-3), 138–154.

Trans Student Educational Resources. (2020, April 15). *TSER*. Trans Student Educational Resources (TSER). https://transstudent.org/about/definitions/

Winnick, J. P., & Porretta, D. L. (2022). *Adapted physical education and sport* (7th ed.). Human Kinetics.

Winnick, J. P., & Short, F. X. (1999). *The Brockport physical fitness test manual*. Human Kinetics.

Woodworth, A., & Schneider, M. (2021). Critical evaluation of the case for pausing California's school-based fitness testing. *Health Behavior and Policy Review, 8,* 168–183.