Historical Analyses of Fitness Testing of College Students in China

Xiaolu Liu [1], Xiaofen D. Keating (corresponding author)[2], Rulan Shangguan[3]

[1]*Institution: Department of Curriculum and Instruction, The University of Texas at Austin, US, Email: Xiaolu.liu@utexas.edu, Phone number: 737-781-0932. Mailing address: 7131 Wood Hollow Dr. Apt.151, Austin, TX, 78731*

[2]*Institution: Department of Curriculum and Instruction, The University of Texas at Austin, US, Email: xk93@austin.utexas.edu, Phone number: 512-232-3565, Mailing address: Department of Curriculum and Instruction, The University of Texas at Austin, 1912 Speedway Stop D5000, Austin, TX, 78712*

[3]*Institution: Department of Curriculum and Instruction, The University of Texas at Austin, US, Email: Rulan@utexas.edu, Phone number: 405-385-1464, Mailing address: 3543 Greystone Dr. Austin, TX, 78731*

Abstract

This study examined changes in China's college student fitness test batteries since its inception in 1954. Using the constant content comparison method, the testing components, testing items and related cut-off values, testing methods, testing results utility, and testing material distribution were examined to identify the salient trends. The results presented a notable shift of testing components from skill-related fitness to health-related fitness. The total number of testing items has been reduced, resulting in fewer optional testing items. New test methods were implemented since 2014 and the cut-off values for aerobic fitness decreased, while others remained the same or slightly increased over time. The test results were included in student grade point average for partially assessing students' academic achievement. There is a need to improve the quality of the test website, and more attention should be given to the possible use of new technology in college student health-related fitness testing.

Keywords: Physical fitness assessment, test battery, higher education, college students

Introduction

Youth physical fitness tests have been widely used in school-based physical education programs throughout the world for various purposes (Harris & Cale, 2006). It has greatly impacted student short- and long-term fitness and physical activity (PA) behaviors (Mercier & Silverman, 2014), attitudes toward physical education (Pasek, Michalowska-Sawczyn, & Nowak-Zaleska, 2014), physical education program assessment (Csányi et al., 2015), and sport participation (Cale & Harris, 2009). To date, numerous researchers and practitioners have suggested that physical fitness testing may play a key role as an educational tool in fitness and PA promotion if implemented appropriately (Cvejic, Pejovic, & Ostijic, 2013). As a result, many different types of youth fitness tests (e.g., FitnessGram, EuroFit, etc.) have been developed and revised for more than half of a century throughout the world (Castro-Piñero et al., 2010; Morrow, Zhu, Franks, Meredith, & Spain, 2009). Moreover, health-related fitness has been regarded as a focus in current fitness test batteries (Castro-Piñero et al., 2010; Vanheist, Beghin, Czaplicki, & Ulmer, 2014).

While debate on the role of youth fitness testing in educational settings continues (Silverman, Keating, & Phillips, 2008), new technologies have emerged and dramatically changed educational practices (Collins & Halverson, 2010). Many previously identified fitness testing related problems such as lack of validity and reliability (Currell & Jeukendrup, 2008), students' limited knowledge about the need for youth fitness testing (Hopple & Graham, 1995), time-consuming for test implementation in large classes (Keating & Silverman, 2009; Ruiz et al., 2011), and the embarrassment caused by performing in front of the class (Garrett & Wrench, 2008) may be eliminated by employing technologies. For instance, websites have been used to distribute testing materials and report testing results (De Villiers & van Staden, 2011). This may have helped increase the awareness of youth fitness testing among teachers, students, and parents. Noticeably, a radio-frequency identification (RFID)-based autoscoring system was developed for the Progressive Aerobic Cardiovascular Endurance Run (PACER) test (Youm, Jeon, Park, & Zhu, 2015), which is widely known as one of the commonly implemented shuttle runs for aerobic fitness testing (Castro-Piñero et al., 2010). As suggested by the authors, RFID-based autoscoring system could significantly reduce the needed testing time and increase the accuracy of measuring aerobic fitness.

College students are a unique subgroup of young adults, who are being trained to be future professionals (Kulavic, Hultquist, & Mclester, 2013). Changes in physical fitness of college students may directly affect student academic performance (Keating, Castelli, & Ayers, 2013; Scott, De Souza, Koehler, & Murray-Kolb, 2017). However, few researchers draw upon important changes of physical fitness test practice for college students (Chen, 2015). Unlike that in many other countries, Chinese college students are required to take part in fitness testing each year they are pursuing a bachelor's degree, even though no mandated physical education for juniors and seniors (Keating, Huang, Deng, & Qu, 2003). It is apparent that fitness testing is an important educational practice in higher education settings in China (Hong, 2014). To date, surprisingly, limited research has been available on changes in China's college student fitness testing practices.

Through a historical perspective on the development and evolution of college student fitness testing in China, it is critical to examine fitness testing practices in Chinese higher education settings as it affects millions of young adults' health. The data from this study will provide a basis for further research related to assessing and monitoring college student fitness in the Chinese population. It could also lend a foundation for comparing and contrasting youth fitness testing practices across nations to shed new light on how fitness testing could be implemented to promote a healthy lifestyle among college students, which is currently an understudied topic. It is hoped that the current study will help enhance college student health-related fitness by improving the design and implementation of the China National Physical Fitness Test (CNPFT). Therefore, the purpose of the study was to investigate changes of the CNPFT for Chinese college students since its inception in 1954. Instead of examining student fitness levels, the study focused on how fitness testing batteries have been designed to measure and assess college student fitness in China, focusing on testing components, methods, cut-off values, and the website designed for youth fitness testing.

Methods

Data Resources

The officially published CNPFT documents were the data sources for the study. The earliest fitness test used in Chinese schools was the former Russian *System of Labor and Defense* in 1954 (Keating et al., 2003), which was adopted by the country's military and workplaces at that time. According to the China Student Health Network (2008), with the development of society and the break-up with the former Russia, China developed its own test battery named as *China's National Youth Fitness Testing* in 1975. As it evolved, the test was revised and renamed the CNPFT in 2007. After seven years of experimenting with the CNPFT, it was revised and become the official CNPFT in 2014.

Data Collection

The testing components, items included in each component, testing methods, and cut-off values used for each test item were coded in Chinese first and then translated into English by the first investigator. Besides collecting data related to the testing components, items and cut-off values for teach test item, the test award program, the use of test results, and information included in the website for CNPFT was also examined focusing on its information richness and interactivity, which are the important features for assessing the quality of websites (Lu, Kim, Dou, & Kumar, 2014). The collected data were checked for accuracy by the second and third authors who were professionally trained in both China and the US.

It is necessary to note that the first fitness test battery, *The System of Labor and Defense* developed in 1954, did not include specific tests for college students and simply included a group of '18 and older'. In 1975, *The National Standard for Sports Exercise* had no specific group for college students, and the adult group (over 19 years old) was the equivalent to the college student group. The third version of the test program was titled as the CNPFT in 2007, which included college student group and used the same testing items for high school students. Two groups (i.e., 1st two-year group vs. the 2nd two-year group) were added to the CNPFT in 2014. Because the one developed in 2014 was based on that used in 2007, only slight differences existed in the tests used in 2007 and 2014 (see Table 1).

Data Analyses

The comparative content analysis method (Charmaz, 2000) was used to analyze the data. Specifically, college students' fitness tests implemented in 1954, 1975, 2007 and 2014 were coded to identify differences in testing batteries. Testing methods used in each testing component were investigated in terms of feasibility and accuracy. The percentages of change in each cut-off value for the same test items over time were computed. Unfortunately, the assessment standards used in 1954 could not be found in the literature. As a result, the differences in cut-off values for each college student group and sex were calculated using the data in 1975, 2007 and 2014. As noted earlier, the fitness test website data were analyzed focusing on its appearance, functionalities, and content of texts, video, photos, etc. No website design changes could be identified due to the lack of data concerning a previous China's fitness test website, however.

Results

Testing Components and Items

Fitness test components. Both skill- and health-related test components were included from 1954 to 2014 (see Table 1). However, health-related tests such as body component, cardiovascular endurance muscular strength and endurance, and flexibility dominated the CNPFT since 2007. Comparing to the fitness test batteries developed in 1954 and 1975, body composition and flexibility tests were added to the CNPFT since 2007, placing a strong focus on health-related fitness. Life skill related tests such as rifle shooting and team sport skills such as basketball and soccer skills were deleted in 2007. Fitness knowledge has never been a testing component since 1954, however.

Test items. There was a smallest number of total test items in 2014, compared to that in 1954, 1975, and 2007, respectively, indicating the reduction in optional testing items. Pull-ups was the only test item used for males in all test batteries, as well as sit-ups for females. The 50-meter shuttle run was used by both males and females, and the 1000-meter and 800-meter run for males and females, respectively have been used in all tests. The standing long jump has been kept since 1975, for both sexes. Only the endurance running items were used for aerobic fitness testing. Other types of aerobic activities such as swimming, jump roping, and skiing were used in earlier tests, but have not been employed since 2007. Vo2max was a newly added item for the aerobic fitness test component in 2007 (see Table 1).

Table 1. *Physical fitness testing components and items included in youth fitness test for college 3students in 1954, 1975, 2007 and 2014*

Testing Component	Testing Items	1954	1975	2007	2014	
		18 or older	19 or older	College students	1st two years	2nd two years
Body composition	BMI			Yes	Yes	Yes
Cardiovascular endurance (Aerobic fitness)	Vo_2 max			Yes	Yes	Yes
	60-meter shuttle run run	Yes				
	100-meter shuttle run	Yes				
	50-meter shuttle run		Yes	Yes	Yes	Yes
	1500-meter run (male)	Yes	Yes			
	3000-meter run (male)	Yes				
	1000-meter run (male)		Yes	Yes	Yes	Yes
	800-meter run (female)	Yes	Yes	Yes	Yes	Yes
	Step test			Yes		
	200-meter swimming	Yes	Yes			
	Marching 10 km (female)/ 5 km (female)	Yes				
	Ski run	Yes	Yes			
Muscular strength and endurance	Rope climb	Yes				
	Pull-up (male)	Yes	Yes	Yes	Yes	Yes
	Push-up (female)	Yes				
	Sit-up (female)		Yes	Yes	Yes	Yes
	Handgrip			Yes		
	Grenade throw	Yes				
Flexibility	Sit and reach			Yes	Yes	Yes
Power	High jump	Yes				
	Long jump	Yes	Yes			
	Standing long jump		Yes	Yes	Yes	Yes
	Medicine ball throw) (2kg		Yes	Yes		
	Shot put 5kg (male)/ 4kg (female)		Yes			
Life and sport skill related component	Team sports: basketball, volleyball, soccer	Yes	Yes	Yes		
	Gymnastics skills	Yes	Yes			
	Shooting (pneumatic rifle)	Yes				

Test Implementation Methods

The test implementation methods remained almost the same for about 60 years -- field testing students with minimum required equipment to control costs for testing all college students. For instance, push-ups and sit-ups use self-weight to assess muscular strength and endurance in a class setting for freshmen and sophomores while juniors and seniors were tested outside their class time because physical education is only required for the first two years in higher education in China (Keating et al., 2003). As noted earlier, the newly added Vo2 max test for aerobic fitness was used to replace 50-meter shuttle run and/or 1000-meter or 800-meter run. This test implementation was in a lab setting. However, the required equipment was quite expensive (i.e., about US$4,000 per unit). Noticeably, the accuracy of measuring cardiovascular endurance could be significantly improved and the amount of time on testing could be greatly reduced even if the equipment is limited because it only takes a few seconds for a student to complete the test. More complicated technology such as sensors and computers has not been widely used as testing methods.

Cut-off Values

The cut-off values for the tests used in 1954 were not available and were excluded in the data analyses. There were three sets of cut-off values for all test items: excellent, good, and pass. Due to changes in testing items over the years, only 50-meter shuttle run, 1000-meter (male)/800-meter (female) run, pull-up for males, standing long jump, and sit and reach were used more than twice and, therefore, changes in the assessment criteria could be computed.

The cut-off values for the test of 50-meter shuttle run for "excellent and pass" were lowered down for both male and females since 1954 (see Figure 1). The endurance test (i.e., 1000-meter run for male) had a slightly lower cut-off value for passing the test (see Figure 2). However, the standards for "excellent and good" became the highest ones in 2014 (see Figure 2). The requirement for "pass" the endurance test (i.e., 800-meter run) for females was lower than that in the previous years. With respect to pull-up (male) and sit-up (female), the lowest standards were used in 1975 (see Figure 3). The cut-off values for "excellent and good" were the highest in 2014, except that for "excellent" for male and "pass" for female, respectively (see Figure 3). The standards for both standing long jump and sit and reach were increased each time when the test battery was revised (see Figure 4). The changes in standing long jump were less than 10% in general. Greater changes were associated with the sit and reach test cut-off values (see Figure 5). Furthermore, freshmen and sophomores had different cut-off vales than those for juniors and seniors varying among testing items. However, the cut-off values for juniors and seniors were only slightly higher than those for the first two year students.

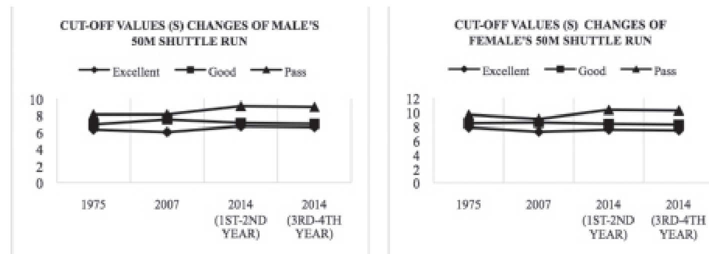

Figure 1. Changes in cut-off values for 50-meter shuttle run in youth fitness test for college students in 1975, 2007 and 2014

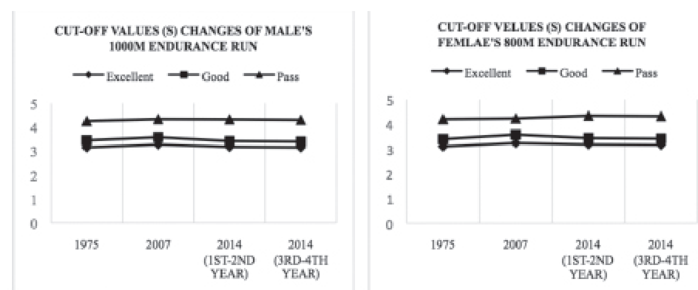

Figure 2. Changes in cut-off values for 1000-meter (males)/800-meter (females) run in youth fitness test for college students in 1975, 2007 and 2014

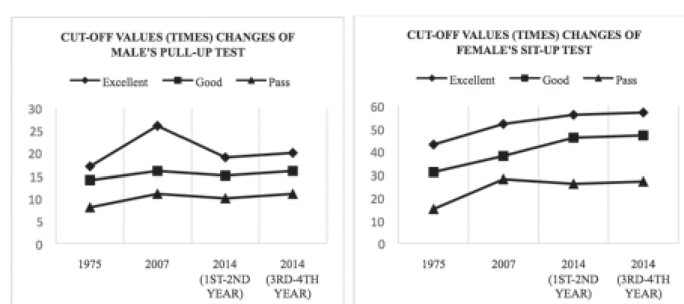

Figure 3. Changes in cut-off values for pull-up (males)/sit-up (females) in youth fitness test for college students in 1975, 2007 and 2014

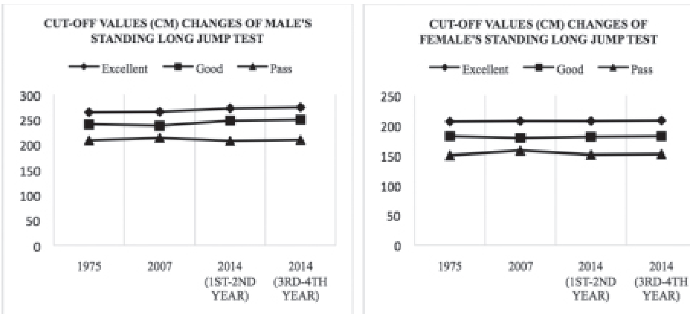

Figure 4. Changes in cut-off values for standing long jump in youth fitness test for college students in 1975, 2007 and 2014

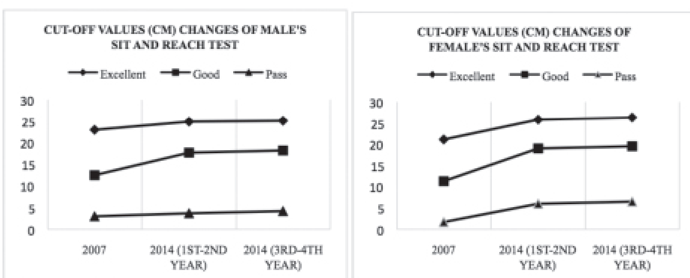

Figure 5. Changes in cut-off values for sit and reach in youth fitness test for college students in 2007 and 2014

Use of Testing Results

Chinese student fitness test scores were used as an important criterion for college graduation (Keating et al., 2003). Apart from being included in a student physical education grade, fitness test scores were included in student grade point averages. Specifically, "good" or "excellent" fitness test scores were required for students to be eligible for scholarships and other academic honors. Physical education scholarships were only awarded to those who had "excellent" fitness test scores. Noticeably, those who did not pass the fitness test had to retake the test within the same semester, and a "F" would be recorded if students still did not pass the test at the second time. According to the Chinese Department of Sports Hygiene and Health Education (2014), college students had to pass the fitness test each year to graduate with a bachelor degree.

There were no requirements or built-in mechanisms to ensure that fitness test results were used for guiding physical education teachers' teaching and/or evaluation, however. Teacher professional development was also not linked to their student fitness test results. Overall, only students were held accountable for passing the fitness test while instructors were only responsible for administering the test. Student parents were also not given the test results each year.

CNPFT Material Distribution

CNPFT testing materials were distributed as official documents from the Chinese Department of Sports Hygiene and Health Education, which belongs to China's Department of Education. The China Student Health Network (2016) was the specific website for the test battery. Each university only had access for data submission once per year. Summarized testing results were not available on the website. Equipment needed for each test and test methods were specified in text format. There were no photos or videos illustrating how each test should be administered. Moreover, no live chat function was available to answer questions providing on-going support.

Discussion

Unlike the United States and many other western countries, college students in China are required to take a fitness test each year. The effects of such unique educational practice on promoting student fitness has not been thoroughly examined as few studies on the topic have been reported. This research strand warrants more attention of professionals in the fields of health, physical education and fitness in order to better understand how youth fitness testing can be used to promote healthy lifestyles among college students.

The data from the current study suggested that test components, test items, and test methods have evolved since 1954. The following results are worth noting: (a) there was an obvious trend to focus on health-related fitness components while fitness knowledge was still absent; (b) technology began to be used in the test battery, resulting in the philosophical change of youth fitness testing, and different test implementation procedures; (c) the cut-off values for each sex and class standing (i.e., first two vs. second two years) were changed with about a half of them lower than those in 2007; (d) the use of test results held students accountable by adding fitness test scores to the grade average point (GPA) for scholarship and graduation, and (e) a website was primarily used for test result reporting only. There was a noted lack of fitness test promotion via the internet.

Test Components

Fitness test components determine what will be tested in schools, which could produce a lifelong impact on next generations' fitness and fitness-related behaviors (Silverman et al., 2008). Researchers believe that skill-related fitness is related to a higher level of performance in sport competition, which is higher than what is needed to be healthy (Morrow et al., 2009). It is clear that the revisions of CNPFT centered more on maintaining the health-related fitness components and eliminating a number of skill-related components from 1954 to 2014. This result is in line with the findings reported by the most recent research on this topic (Vanheist et al., 2014).

The lack of fitness knowledge as part of the test component at the college level, however, is puzzling as it has been suggested that youth fitness test battery should also include a knowledge test (Keating, 2003). Previous research has suggested that students may not be motivated to engage in physical activity on a regular basis when there is a lack of a well-grounded perception about fitness knowledge (Keating, Castro-Piñero, Centeio, Harrison, & Ramirez, 2010). Increased fitness knowledge was found to be associated with an increased understanding of self-regulation, which is a salient factor for behavioral changes (Chen, Sun, Zhu, & Chen, 2014). However, to date, fitness knowledge has not been tested as a fitness component in CNPFT program, undermining the role of knowledge in promoting health-related fitness. Future experimental studies should compare knowledge, and fitness levels among college students to examine if fitness knowledge is strongly correlated to physical fitness.

Test Items

Test items are also critical given that they may result in an ineffective and/or time consuming and/or embarrassing testing experience for students (Cureton & Mahar, 2014). It is necessary to point out that youth fitness testing is different from those for adults in other settings due to its educational purposes (Keating & Silverman, 2009). Costs and implementation procedures are of concern because schools usually have very limited budgets and person power (Craig, Imberman, & Perdue, 2015). To this end, cheap and simple testing items with limited or no equipment have always been selected to ensure that fitness testing is feasible in schools (Castro-Piñero et al., 2010). However, at the same time, it is critical to ensure that testing items have acceptable validity and reliability (Cureton & Mahar, 2014; Silverman et al., 2008). Surprisingly, pull-ups are still used as a compulsory test of arm muscular endurance of college male students when many students could not even perform one pull-up, often resulting in a zero score (Keating et al., 2003). Therefore, this test cannot effectively assess male students' muscular strength and endurance (Castro-Piñero et al., 2010). The difficult requirements of the pull-up test always led to numerous testing errors, and therefore modified pull-ups or other test items have been included to solve the above problem in the US (Cooper Institute for Aerobics Research, 2017). Moreover, pull-up (male) and sit-up (female) tests usually have validity and reliability issues given that many students cannot perform the tests correctly, which may lead to inaccurate test scores. As suggested in the literature (Youm et al., 2015), tracking and recording student scores at the same time causes measurement errors. There is an urgent need to adopt test items with acceptable validity and reliability and limited testing time.

Relatively long distance running (i.e. 800-meter run for females, and 1000-meter run for males) has always been used in the CNPFT since its inception in 1954. The critical elements for effective implementation of such a test are twofold: (a) the appropriate pace of speed. Students must learn how to pace their speed during the continuous running. However, many students are not well trained to pace their speed appropriately, which may result in unpleasant experiences and driving students away from taking the test and ultimately not enjoying running (Wrench & Garrett, 2008); and (b) a relatively large test space. Because all endurance running or skiing, or swimming tests must last for a certain period of time and a large-size class of students (i.e., more than 20 students in each class in general) usually takes part in the test, testing space can be a problem (Cale & Harris, 2005).

It is well documented that the shuttle run (i.e., 20-meter or 50-meter) or PACER in the FitnessGram in the United States could be a better test than continuous running or swimming or skiing because it could be done with a relatively smaller space and a pre-set pace to guide students (Wilkinson, Brown, Graser, & Pennington, 2012). Castro-Piñero and associates (2010) noted that the 20-meter shuttle run generates the best result for testing aerobic fitness and has been widely used in many countries. It is unknown why the 50-meter shuttle run has been chosen for Chinese college student fitness testing. A variety of test items to make the fitness test fun in general and endurance testing in particular is of concern. By repeated use of the same items and practices for many years, many students tend to have low motivation in participating in the fitness test (Hopple & Graham, 1995; Silverman et al., 2008).

Test Methods and New Technology

VO2 max, an internationally accepted variable for assessing aerobic fitness (Hunn, Lapuma, & Holt, 2002), has been added as an item for assessing cardiovascular endurance since 2007. This is certainly a positive change in youth fitness testing if the cost is not an issue. More attention should be given to the use of technology in fitness testing in the future considering that testing time, accuracy, and privacy could be greatly improved (Youm et al., 2015). On the other hand, thousands of college students need to be tested simultaneously, so it might be very costly due to the use of expensive equipment. If only one or a small number of Vo2max machines are used, the large size class of students may take a long time to complete the test, because the Vo2max equipment can only test one student at a time. Much more person power is needed at the beginning to instruct each student to conduct the test correctly. Moreover, a study done by Youm and her colleagues (2015) indicated that new technologies have made it possible to automatically track and record PACER test results. Caution needs to be exercised when using new technology in fitness testing, however, as test feasibility must be of concern.

Cut-off Values for Each Test Item

It has been well documented that cut-off values should be completely determined by what is needed to maintain sound health (Hata et al., 2015; Zhu, Mahar, Welk, Going, & Cureton, 2011). Criterion-referenced evaluation works more efficiently than norm-referenced evaluation (Zhu et al., 2011). By using pre-set cut-off scores, the degree of competition is reduced and students can focus on their fitness rather than being compared with others. As such, it was found that students were able to increase their motivation and self-confidence (Dolosic, Brantmeier, Strube, & Hogrebe, 2016).

The lower cut-off values for the 50-meter shuttle run for the "excellent and pass" scores meant that students could pass the endurance test and receive an excellent score easier than before for the same sex and age. However, this did not mean that college students' aerobic fitness was getting better more recently. The higher cut-off values for the muscular strength and endurance tests [i.e., the pull-up (male) and sit-up (female)], as well as the power test (i.e., standing long jump) indicated that fewer students may pass or get high test scores in those items. Although changes in cut-off values were observed, no data are available to justify such changes. To the best of our knowledge, it seemed to be arbitrary to adjust these cut-off values. Regardless of student passing rates, any testing cut-off value modifications should only be made based on health effects (Zhu et al., 2011). Otherwise, youth fitness testing becomes meaningless.

Test Result Utility

While the fitness testing process is considered as a critical element in physically educating students to be physically fit (Wrench & Garrett, 2008), the use of fitness test results plays a pivotal role in youth fitness testing (Keating & Silverman, 2009; Shape America, 2017). How fitness test results should be used, however, has been debated for more than three decades (Lloyd, Colley, & Tremblay, 2010). To date, many researchers have suggested not to use fitness test results to assess student performance in physical education in the United States (Ernst, Corbin, Beighle, & Pangrazi, 2006). Instead, it has been suggested that teachers should use student fitness test results to guide their fitness related teaching practices (Shape America, 2017). It is important to point out that fitness test results have been used to hold college students accountable in China. However, it is unclear if instructors are also responsible for student health-related fitness because no official documents have indicated that physical education instructors are evaluated based on their students' fitness test results.

Specifically, combined with GPA, scholarship and awards strongly affect college student's choice of coursework, and participation in college activities (Hu, 2010). In other words, scholarship and awards can be the motivation for college students' participation and performance results in physical fitness test. In China, those college students who obtain "excellent" test scores could be eligible for completing physical education scholarships. Moreover, students' test results were used as a critical requirement for scholarships and students' graduation, generating a great impact on the importance of fitness tests (Keating et al., 2003). Requiring a fitness test with an award system has a totally different philosophy compared to those programs without awards, due to its long-term psychological effects on school-aged students (Welk, De Saint-Maurice Maduro, Laurson, & Brown, 2011). It is a cause of concern that instructors are not held accountable for student fitness test results considering that teaching and learning are strongly connected (Ferguson, Keating, Bridges, Guan, & Chen, 2007). More research is urgently needed to examine how test results should be used to motivate students to participate in more PA, and eliminate fitness testing competition at the same time.

Test Program Website

College students have been found to be regular users of websites (Jones, Johnson-Yale, Millermaier, & Seoane Perez, 2009). Previous research has indicated that website richness and interactivity were crucial contributors to users' behavioral intention (Lu et al., 2014). It was suggested that enriching the web content (e.g. 3D images or videos) and adding interactive tools (e.g. internal or external links, or discussion column, etc.) could be helpful in increasing website visit intention (Lu et al., 2014). However, the assessment materials on the Chinese Student Health Network website were not very attractive or interesting, given that only text instructions and a few photos could be found. It was difficult to obtain more specific information and assistance from the website, weakening potential use of this website for the purpose of distributing test program information.

Videos related to the test battery could be the most effective visual tool to help teachers and students grasp the correct ways to implement the tests, and these videos could be utilized to offer information related to test items and equipment needed. Thus, videos should be included in the website. Additionally, it would be more useful if teachers and students could ask questions while watching videos via live chat. Hence, the website quality could be improved by adding interactive functions allowing direct communications between test designers and users.

In summary, as part of the CNPFT, Chinese college student fitness testing practice has been developed over more than 60 years. Although it tends to focus more on the health-related fitness test recently, the majority of testing items have been repeated yearly since 1954. Moreover, fitness knowledge as an important aspect of health-related fitness needs to be added to the fitness test battery in the future. Although test implementation methods, test administration, and material distribution approaches have remained the same for many years, new test methods integrating with technology have been used in the current fitness test. However, more effective ways regarding the development of a web-based test information system is urgently needed. In higher education, the use of fitness test results should not only be used for student's academic assessment and scholarships, but also for teacher evaluation, which could potentially have an impact on the changes of fitness test practices among college students. More attention should be given to new technology as testing methods to improve the accuracy of fitness testing and save time on test implementation.

Limitations

The data of CNPFT in 1954 were incomplete due to missing values. Therefore, it was impossible to fully address the changes of the cut-off values in different test items since its inception. Fitness testing website information was very limited as it was primarily used for data reporting by colleges and universities. It is unknown what additional web information is available for users because the authors did not have a valid user account. Caution needs to be exercised when addressing this issue using the data reported in our study.

References

Cale, L., & Harris, J. (2005). *Exercise and young people: Issues, Implications and Initiatives*. Houndmills: Palgrave Macmillan.

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

Castro-Piñero, J., Attero, E. G., Espan˜a-Romero, V., Ortega, F. B., Sjostrom, M., Suni, J., & Ruiz, J. R. (2010). Criterion-related validity of field-based fitness tests in youth: A systematic review. *British Journal of Sports Medicine, 44*, 934-943.

Castro-Piñero, J., Ortega, F. B., Keating, X. D., Gonzalez-Montesinos, J. L., Sjostrom, M., & Ruiz, J. R. (2011). Percentile values for aerobic performance running/walking field tests in children aged 6 to 17 years. *Nutrición Hospitalaria, 26*, 572-578.

Charmaz, K. (2000). Constructivist and objectivist grounded theory (pp. 509-535). In Denzin NK and Lincoln Y (eds.), *Handbook of qualitative research* (2nd ed.). Thousand Oaks, CA: Sage.

Chen, H. (2015). Fitter students or flabbier standards? Thoughts and reflection on the findings of a recent Chinese fitness study. *Research Quarterly for Exercise and Sport, 86*, 213-216.

Chen, S., Sun, H., Zhu, X., & Chen, A. (2014). Relationship between motivation and learning in physical education and after-school physical activity. *Research Quarterly for Exercise and Sport, 85*, 468-477.

Chinese Student Health Network. (2008). Evolution and development of students' physical health assessment system. Retrieved from: http://www.csh.edu.cn/wtzx/bz/20141226/2c909e854a80abab014a8428886c0002.htm

Chinese Department of Sports Hygiene and Health Education. (2014). *Physical fitness monitoring and research*. Retrieved from: http://www.moe.gov.cn/s78/A17/twys_left/moe_943/

Collins, A., & Halverson, R. (2010). The second educational revolution: Rethinking education in the age of technology. *Journal of Computer Assisted Learning, 26*(1), 18-27.

Craig, S. G., Imberman, S. A., & Perdue, A. (2015). Do administrators respond to their accountability ratings? The response of school budgets to accountability grades. *Economics of Education Review, 49*, 55-68.

Csányi, T., Finn, K. J., Gregory, J., Zhu, W., Karsai, I., Ihasz, F, ... Molnar, L. (2015). Overview of the Hungarian National Youth Fitness Study. *Research Quarterly for Exercise and Sport, 86*(Suppl. 1), S3-S12.

Cureton, K. J., & Mahar, M. T. (2014). Critical measurement issues/challenges in assessing aerobic capacity in youth. *Research Quarterly for Exercise and Sport, 85*, 136-143.

Currell, K., & Jeukendrup, A. E. (2008). Validity, reliability and sensitivity of measures of sporting performance. *Sports Medicine, 38*(4), 297-316.

Cvejic, D., Pejovic, T., & Ostijic, S. (2013). Assessment of physical fitness in children and adolescents. *Physical Education and Sport, 11*(2), 135-145.

De Villiers, C., & van Staden, C. (2011). Shareholder requirements for compulsory environmental information in annual reports and on websites. *Australian Accounting Review, 21*(4), 317-326.

Dolosic, H. N., Brantmeier, C., Strube, M., & Hogrebe, M. C. (2016). Living language, oral production, and domestic immersion. *Foreign Language Annals, 49*(2), 302-316.

Ernst, M. P., Corbin, C. B., Beighle, A., & Pangrazi, R. P. (2006). Appropriate and inappropriate use of FITNESSGRAM: A commentary. *Journal of Physical Activity & Health, 3*(Suppl. 2), S90-S100.

Ferguson, R. H., Keating, X. D., Bridges, D. M., Guan, J., & Chen, L. (2007). California secondary school physical education teachers' attitudes toward the mandated use of the Fitnessgram. *Journal of Teaching in Physical Education, 26*(2), 161-176.

Garrett, R., & Wrench, A. (2008). Fitness testing: The pleasure and pain of it. *ACHPER Australia Healthy Lifestyles Journal, 55*(4), 325-346.

Harris, J., & Cale, L. (2006). A review of children's fitness testing. *European Physical Education Review, 12*, 201-225.

Hata, A., Yonemoto, K., Shikama, Y., Aki, N., Kosugi, C., Tamura, A., ... Funaki, M. (2015). Cut-off values of total Adiponectin for managing risk of developing metabolic syndrome in male Japanese workers. *PLoS One, 10*(2), 1-10.

Hong, Y. (2014). Evolution course and stage feature of student physical fitness test in China. *Journal of Beijing Sport University, 37*(10), 113-118.

Hopple, C., & Graham, G. (1995). What children think, feel, and know about physical fitness testing. *Journal of Teaching in Physical Education, 14*, 408-417.

Hu, S. (2010). Scholarship awards, college choice, and student engagement in college activities: A study of high-achieving low-income students of color. *Journal of College Student Development, 51*(2), 150-161.

Hunn, H. M., Lapuma, P. T., & Holt, D. T. (2002). The influence of pre-test anxiety, personality and exercise on VO2 max estimation. *Journal of Exercise Physiology Online, 5*(1), 5-14.

Jones, S., Johnson-Yale, C., Millermaier, S., & Seoane Perez, F. (2009). U.S. college students' internet use: Race, gender and digital divides. *Journal of Computer-Mediated Communication, 14*, 244-264.

Keating, X. D., Castro-Piñero, J., Centeio, E., Harrison, L. J., & Ramirez, T. (2010). Health-related fitness knowledge and its relation to student physical activity patterns at a large U.S. southern state university. *ICHPER-SD Journal of Research, 5*(2), 3-9.

Keating, X. D. (2003). The current often implemented fitness tests in physical education programs: Problems and future directions. *Quest, 55*, 141-160.

Keating, X. D., Haung, Y., Deng, M., & Qu, S. (2003). A comparative analysis of youth fitness test batteries between the US and the People's Republic of China. *International Sports Studies, 25*(1), 15-22.

Kulavic, K., Hultquist, C. N., & Mclester, J. R. (2013). A comparison of motivational factors and barriers to physical activity among traditional versus nontraditional college students. *Journal of American College Health, 61*(2), 60-66.

Lloyd, M., Colley, R. C., & Tremblay, M. S. (2010). Advancing the debate on fitness testing for children: Perhaps we're riding the wrong animal. *Pediatric Exercise Science, 22*, 176-182.

Lu, Y., Kim, Y., Dou, X., & Kumar, S. (2014). Promote physical activity among college students: using media richness and interactivity in web design. *Computers in Human Behavior, 41*, 40-50.

Mercier, K., & Silverman, S. (2014). High school students' attitudes toward fitness testing. *Journal of Teaching in Physical Education, 33*(2), 269-281.

Morrow, J. J., 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.

Pasek, M., Michalowska-Sawczyn, M., & Nowak-Zaleska, A. (2014). Changes in maximal aerobic fitness and students' attitude towards physical effort during outdoor and indoor school lessons of physical education. *Baltic Journal of Health and Physical Activity, 6*(1), 24.

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.

Ruiz, J. R., Castro-Piñero, J., España-Romero, V., Artero, E. G., Ortega, F. B., Cuenca, M. M., ... Castillo, M. J. (2011). Field-based fitness assessment in young people: the ALPHA health-related fitness test battery for children and adolescents. *British Journal of Sports Medicine, 45*(6), 518-524.

Scott, S. P., De Souza, M. J., Koehler, K., & Murray-Kolb, L. E. (2017). Combined iron deficiency and low aerobic fitness doubly burden academic performance among women attending university. *The Journal of Nutrition, 147*(1), 104-109.

Shape America. (2016). *Instructional framework for fitness education in physical education*. Retrieved from: http://www.shapeamerica.org/standards/guidelines/loader.cfm?csModule=security/getfile&pageid=9264.

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

Vanheist, J., Beghin, L., Czaplicki, G., & Ulmer, Z. (2014). BOUGE-fitness test battery: Health-related field-based fitness tests assessment in children and adolescents. *Revue médicale de Bruxelles, 35*(6), 483-490.

Welk, G. J., De Saint-Maurice Maduro, P. F., Laurson, K. R., & Brown, D. D. (2011). Field evaluation of the new FITNESSGRAM(R) criterion-referenced standards. *American Journal of Preventive Medicine, 41*(Suppl. 4), S131-S142.

Wilkinson, C., Brown, L., Graser, S. V., & Pennington, T. R. (2012). Adolescent girls' preferences pertaining to cardiovascular fitness testing: A comparison between the one-mile run and PACER tests. *Physical Educator, 69*(1), 52-70.

Wrench, A., & Garrett, R. (2008). Pleasure and pain: Experiences of fitness testing. *European Physical Education Review, 14*(3), 325-346.

Youm, S., Jeon, Y., Park, S. H., & Zhu, W. (2015). RFID-based automatic scoring system for physical fitness testing. *IEEE Systems Journal, 9*(2), 326-334.

Zhu, W. M., Mahar, M. T., Welk, G. J., Going, S. B., & Cureton, K. J. (2011). Approaches for development of criterion-referenced standards in health-related youth fitness tests. *American Journal of Preventive Medicine, 41*(Suppl. 4), S68-S76. ■