

# Validity of Alternative Fitnessgram® Upper Body Tests Among Adolescent Students

# Kalani Hobayan¹ ◆ Debra Patterson² ◆ Clay Sherman² ◆ Lenny Wiersma²

(1)St Pius X Parish School, Santa Fe Springs, CA • (2)California State University-Fullerton, Fullerton, CA

# **ABSTRACT**

In a society in which the rates of obesity levels have tripled in the past 30 years, the importance of increased fitness levels within the academic setting has become even more critical. The purpose of this study was to investigate the validity of alternative FITNESSGRAM® upper body tests of muscular strength and endurance among seventh and eighth grade males and females. Adolescent males and females (N = 123) in 7th and 8th grades from two urban middle schools were administered all three FITNESSGRAM® muscular strength and endurance assessments on different days. The recommended test item used to assess students is the 90° Push-up (90°PSU). However, the FITNESSGRAM® provides alternative assessments to measure upper body strength: Modified Pull-up (MPU) and Flexed Arm Hang (FAH). The validity was determined by equivalence reliability estimates for the following comparisons: PSU-MPU and PSU-FAH. Both Percentage Agreement (Pa) and Modified Kappa (Kq) were used to determine the relationships between variables. Males exemplified minimal acceptability for both PSU- MPU and PSU-FAH comparisons. Similar to the male's results, results for females indicated unacceptable reliability estimates for both PSU-MPU and PSU-FAH comparisons. As a result of this study, it is imperative that physical educators and administrators are aware that implementing the FITNESSGRAM® alternative assessments of muscular strength and endurance may hinder and/or alter an adolescent's healthy fitness zone classification. Future research regarding the different muscular strength and endurance test items will ultimately promote higher levels of confidence among practitioners when using the different test items interchangeably.

# INTRODUCTION

#### **Physical Education Environment**

#### Lifelong Physical Activity

- Lifestyle of physical fitness/healthy eating habits
- Activities that develop the WHOLE student

#### **FITNESSGRAM®**

- "desirable health standards"
- Assess students' overall fitness and health

#### Muscular Strength and Endurance Test Items:

Three test items:

Recommended Test Item:

90° Push-up (90° PSU)

**Alternative Tests**:

- Modified Pull-up (MPU)
- 2. Flex Arm Hang (FAH)

# Recommended vs. Alternative: Research Findings

#### **Equivalence Reliability Studies:**

- PSU vs. MPU
- Acceptable comparisons for boys; unacceptable comparisons for girls (Romain and Mahar, 2001)
- Acceptable comparisons for boys (PSU-MPU; PSU-FAH)
- Unacceptable comparisons for girls (PSU-MPU; PSU-FAH)
  (Sherman and Barfield, 2006)

#### Validity of Upper Body Tests:

- Invalid measures of absolute strength
- Tests measure "weight relative" muscular strength (Pate, Burgess, Woods, Ross, & Baumgartner, 1993)

# **PURPOSE**

To determine if there a difference between the different muscular strength and endurance test items?

# METHODS

#### Participants:

- Sample of 123 adolescents
- Males & females (between the ages of 12-14)
- Enrolled in grades 7th-8th Physical Education classes
- St. Pius X Parish School, CA (Catholic Archdiocese)
- Kraemer Middle School (Placentia-Yorba Linda School District, CA)
  Instrument/Format:
- FITNESSGRAM® assessment measures of Physical Fitness/Manual
- FITNESSGRAM® Healthy Fitness Zone Classifications

#### Testing Format:

- Conducted during 50 minute class period
- Two days in week 1 (Mon./Thurs.); one day in week 2 (Tues.)
- Counterbalance order of testing
- Review protocols/critical elements for each test item

# ANALYSIS

#### Criterion reference standards:

FITNESSGRAM® Healthy Fitness Zone Classifications:

- 12-14 yr. old male/female
- 3 tests (90°PSU, MPU, FAH)

#### Statistical Output:

- Percentage agreement (Pa) and Modified Kappa (Kq)
- Equivalence Reliability Between Comparisons:
- PSU-MPU
- PSU-FAH

# RESULTS

Percent Agreement and Modified Kappa Values Between the 90° Push-up Test and the Alternative Tests (Modified Pull-up, Flexed Arm Hang) of Upper Body Strength and Endurance

Age	Statistic	PSU- MPU	PSU-FAH
Total Sample (N = 123)	Pa	.68	.62
·	Kq	.37	.24
12 (n = 54)	Pa	.70	.54
	Kq	.40	.07
13 (n = 48)	Pa	.63	.71
	Kq	.25	.43
14 (n = 21)	Pa	.67	.62
	Kq	.33	.25

Note. MPU= modified pull-up; PSU = 90° push-up; FAH=flexed-arm hang. Pa= percent agreement; Kq= Modified Kappa. Age is represented in years. Kq > .75 = Excellent, Good= .60  $\leq$  Kq  $\leq$  .75, Acceptable = .40 $\leq$  Kq  $\leq$  .60.

Males and Females Percent Agreement and Modified Kappa Values Between the 90° Push-up Test and the Alternative Tests (Modified Pull-up, Flexed Arm Hang) of Upper Body Strength and Endurance

Age	Statistic	PSU-	MPU	PSU	-FAH
		M	F	M	F
12 <sup>a</sup>	Pa	.72	.80	.62	.48
	Kq	.44	.60	.24	04
13 <sup>b</sup>	Pa	.69	.68	.62	.63
	Kq	.38	.36	.23	.27
14 <sup>c</sup>	Pa	.81	.30	.81	.50
	Kq	.63	-0.4	.63	0.0
12, 13, and 14 <sup>d</sup>	Pa	.66	.66	.65	.57
	Kq	.35	.31	.29	.14

Note. MPU= modified pull-up; PSU = 90° push-up; FAH=flexed-arm hang. M = males; F = females. Pa= percent agreement; Kq= Modified Kappa. Age is represented in years. Kq > .75 = Excellent, Good= .60  $\leq$  Kq  $\leq$  .75, Acceptable = .40 $\leq$  Kq  $\leq$  .60. an = 29. bn = 25. cn = 11. dn = 65 (males) an = 25. bn = 22. cn = 10. dn = 58 (females)

Summary of Reliability Estimates Between the 90° Push-up Test and the Alternative Tests (Modified Pull-up, Flexed Arm Hang) of Upper Body Strength and Endurance

Age	Gender	PSU- MPU	PSU-FAH	
Total Sample		Unacceptable	Unacceptable	
12	Male	Acceptable	Unacceptable	
	Female	Acceptable	Unacceptable	
13	Male	Unacceptable	Unacceptable	
	Female	Unacceptable	Unacceptable	
14	Male	Acceptable	Acceptable	
	Female	Unacceptable	Unacceptable	
12, 13, 14	Male	Unacceptable	Unacceptable	
	Female	Unacceptable	Unacceptable	

# DISCUSSION AND CONCLUSION

This current study revealed that the validity of the  $90^{\circ}$  PSU and alternative test items were unacceptable in most cases for middle school adolescents. While some reliability estimates within the study revealed signs of statistical acceptability, classification agreement estimates suggest that test items of muscular strength and endurance exemplified unacceptable comparisons. A delimiting factor associated within the study was the size of the total sample. The current study incorporated a total sample of 123 students, a relatively smaller sample size than the one reported in Sherman and Barfield's study (N = 383).

In order for the FITNESSGRAM® to truly measure both muscular fitness and health-related fitness, there needs to be consistent classification across future criteria (Sherman & Barfield, 2006). Due to the limited amount of background information regarding the validity of the FITNESSGRAM® upper body tests, many physical educators are posed with confusion whether to use the muscular strength and endurance test items interchangeably. Romain and Mahar (2001) findings suggested that not only was the reliability of the PSU-MPU unacceptable among subjects in Grades 5 and 6, but criterion-referenced classifications also need to improve in order for test items (90° PSU and MPU) to be deemed valid. Like Romain and Mahar (2001), Sherman and Barfield (2006) implemented all three alternative forms of muscular strength and endurance (FAH, PU, and MPU) and found consistent themes, particularly with the need to modify alternative test items classifications. The current study adds another pillar to the equivalence reliability of the FITNESSGRAM® test. Using a different population as the sample (Grades 7 and 8), concrete comparisons and acceptability was not found between the different test items.

Further studies need to continue the ongoing development of established healthy fitness zone classifications (i.e., criterion-referenced standards). Cuerton and Warren (1990) stated that criterion-referenced assessments have predetermined standards that represent a desired and specified level of performance. As scholars investigate different ways of validating criterion-referenced standards, Sherman and Barfield (2006) note a method of equating all test items of muscular strength and endurance. By equating test items, practitioners will be allowed to determine if a score from one test item relates to that of the "gold standard" (Sherman & Barfield, 2006). The current

HFZ (2011-2012) classifications issued by the FITNESSGRAM® exemplify a transition into the adaptation of merging test item classifications, particularly those of upper body strength and endurance. Modifying standards that are consistent will ultimately promote a higher level of confidence among practitioners when using the different test items interchangeably.

While many suggestions have been implied towards the FITNESSGRAM® tests battery, the Cooper Institute has continued its efforts not only to merge classifications but also to adjust test items to suit present-day society. Although the 90° PSU, MPU, and FAH tests were measured and assessed among subjects, the Pull-up (PU) assessment was eliminated from the current FITNESSGRAM®. One significant problem with the PU assessment arose when looking at the test's inability to differentiate among individuals at the lower end of the scale (Rutherford, 1994). Due to subjects' inability to effectively execute one pull-up repetition, the Cooper Institute has directed physical educators to incorporate current assessments. Similar to the elimination of the PU from the FITNESSGRAM® criterion referenced standards, the FAH test item is another assessment that should be considered for removal based on the results of this study. While passing rates of the FAH were not as significantly low as the PU assessment, the current study revealed that the FAH is not a valid measure of upper body strength and endurance. Again, future editions of the FITNESSGRAM® should consider the effectiveness of the FAH test item.

If the 90° PSU remains as the recommended test item to assess muscular strength and endurance, classification consistency needs to continue to improve within the FITNESSGRAM®. The adaptation of criterion-referenced standards to suit present-day society will increase the level of acceptability of the FITNESSGRAM® among physical educators and other practitioners. As opposed to identifying alternative test items as equivalent measures of muscular strength and endurance, the FITNESSGRAM® should suggest that the alternative assessments support and strengthen the recommended test item. Likewise, practitioners need to be aware of the value of implementing alternative test items in order to obtain an additional measurement of a subject's upper body strength. Further research related to the validity of the different test items is needed in order to develop a greater understanding of the FITNESSGRAM® effectiveness among society in the future.