

Original Research Article

Interrater reliability of fitness gram test to be used as a reliable tool for assessing physical fitness for school children in Guwahati urban society of North-East India

Pranjal Gogoi^{1*}, Nirmal C. Bhattacharyya²

¹Department of Physiotherapy, Downtown Hospital Ltd, Guwahati, Assam, India

²Department of Pediatric Surgery, Gauhati Medical College, Guwahati, Assam, India

Received: 14 November 2019

Revised: 26 November 2019

Accepted: 07 December 2019

*Correspondence:

Dr. Pranjal Gogoi,

E-mail: Pranjal_gogoi25@rediffmail.com

Copyright: © the author(s), publisher and licensee Medip Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Background: The assessment of physical fitness using a specific tool has become an important part to find out the fitness level of children and adolescence. Fitness gram is a health-related fitness test that utilizes criterion-referenced standards on health-related components. Through the years research has shown that Fitness gram has become one of the most widely used programs in the United States, though it's used in India is not popular.

Methods: Seventy school students in the age group of 5 to 14 years were included in this study and randomly allocated for physical fitness test where 62 students have completed the test. Two physiotherapists as a rater were allotted and children were tested by them with Fitnessgram test battery in a gap of 1 week. Outcome measures includes Fitness gram test battery. Statistical analysis was done by Cronbach's Alpha value has been computed for the Interrater reliability.

Results: Sixty-two students with both gender (male 56.3% and female 43.7%, age -10.12±2.72) were assessed for physical fitness using Fitness gram test battery. All the test variables showed an excellent reliability (Cronbach's alpha =0.91-0.95).

Conclusions: The Fitness gram test battery has been found to be reliable tool and can be implemented for physical fitness test for the children.

Keywords: Fitness gram test, Inter-rater reliability, Pediatric population, Physical fitness

INTRODUCTION

Physical fitness is generally considered to be "the ability to perform daily tasks without fatigue." Definition of physical fitness is given differently by different people. It is defined as the ability to meet life's demand and still have enough energy to respond to unplanned events.¹ Physical fitness includes cardiorespiratory fitness, muscular endurance, muscular strength, flexibility, coordination, and speed. It has been seen that a high level of physical fitness in childhood and adolescence is related to better health-related outcomes, which are concerned

with present and future risk for obesity, cardiovascular disease, skeletal health and mental health.² As such the assessment of physical fitness using a specific tool has become an important part to find out the fitness level of children and adolescence.

There were several fitness test batteries for the assessment of physical fitness in children and their reliability and validity has been established by different author in accordance to different geographical area. The appropriate fitness test that promoted exercise and fitness was still under examination.³ In response to this, a study

concluded that any fitness test which promotes enjoyment and motivation to do physical activity should be accepted and encouraged.⁴ Such a fitness test is Fitness gram. Through the years research has shown that Fitness gram has become one of the most widely used programs in the United States, though it's used in India is not popular.⁵

Fitness gram is a health-related fitness test that utilizes criterion-referenced standards on health-related components. Fitness gram includes a variety of health-related physical fitness tests designed to assess cardiovascular fitness, muscle strength, muscular endurance, flexibility, and body composition. These standards are age and gender specific and are established based on how fit children need to be for good health. Fitness gram was developed in 1982 by The Cooper Institute in response to the need in physical education for a detail and comprehensive assessment, and with the vision of helping enhance the effectiveness of school-based physical education.^{6,7} In accordance to the present health and reduced risk of disease, the Fitnessgram standards were developed and the results were the minimal levels of fitness which is consistent.⁸

Therefore, the aim of the current study was to evaluate the Interrater reliability of Fitnessgram test battery to be used in pediatric population of Guwahati urban society.

METHODS

A small exploratory study was conducted to assess the inter rater reliability of Fitness gram test in a school of Guwahati urban society, a capital city of north eastern India where children ages between 5 to 14 years were included, 70 subjects were included in the study by using random sampling method with both the genders randomly selected and allocated for the screening. All grades covering 5 to 14 years participated in the Fitness gram® assessment test for physical fitness (Cooper Institute for Aerobics Research, 1999).

Inclusion criteria

- School children with age between 5 to 14 years with either gender or those having prior consent from their parents were included in the study.

Exclusion criteria

- The criteria for exclusion in the study were that the students had a history of medical, neurological, orthopedic (including wore any type of orthotic device), balance or visual disorders, any acute illness, recent trauma or fall and those children with difficulty in understanding about the test procedure were excluded from the study.

Table 1: Procedure of the Fitness gram test battery.

Aerobic capacity (Pacer)	Student was instructed to run as long as possible with continuous movement back and forth across a 20-meter space at a specified pace that gets faster with each minute.	
Body composition	Skin fold measurement	The triceps skin fold is measured on the back of the right arm over the triceps muscle, midway between the elbow and the acromion process of the scapula. The calf skinfold is measured on the inside of the right leg at the level of maximal calf girth. The right foot is placed flat on an elevated surface with the knee flexed at a 90° angle.
	Body mass index	BMI = weight (kg) / height (m) ²
Abdominal strength and endurance (Curl up)	The student is made to lie in supine position on the mat, knees bent at an angle of approximately 140°, feet flat on the floor, legs slightly apart, arms straight and parallel to the trunk with palms of hands resting on the mat. A measuring strip is placed on the mat so that fingertips are just resting on the nearest edge of the measuring tape. Keeping heels in contact with the mat, the student curls up slowly; sliding fingers across the measuring strip until fingertip reach the other side. Movement should be slow, and the student continues without pausing until he or she no longer continue or has completed 75 curl ups.	
Trunk extensor strength and flexibility (Trunk lift)	The student being tested lies on the mat in a prone position with hands placed under the thighs. Maintaining the head in a neutral alignment with the spine the student lifts the upper body off the floor, in a very slow and controlled manner, to a maximum height of 12-inches. The distance from the floor to the student's chin is determined.	
Upper body strength and endurance (90° Push up)	The student being tested assumes a prone position on the mat with hands placed under or slightly wider than the shoulders, fingers stretched out, legs straight and slightly apart, and toes tucked under. The student pushes up off the mat with the arms until arms are straight, keeping the legs and back straight.	
Flexibility (Shoulder stretch)	The students need to touch the fingertips together behind the back by reaching over the shoulder and under the elbow. To test the right shoulder student reaches with the right hand over the right shoulder and down the back as if to pull up a zipper or scratch between the shoulder blades. At the same time places the left hand behind the back and reaches up, trying to touch the fingers of the right hand.	

Eight participants were excluded from this analysis leaving a final sample of 62 participants. In order to participate in the study written consent were taken from the parents. Appropriate ethical clearance taken from Institutional ethical committee of Gauhati Medical College, Guwahati, Assam. Two physiotherapists as a rater were allotted for Interrater reliability test and the children were screen by the two raters with Fitnessgram test batteries in a gap of one week. The first rater conducts the fitness test among the selected students and the same subjects were again assessed by the second rater after 1 week.⁹ The two raters were blinded from each other findings. The subjects were included randomly irrespective of any musculoskeletal pain, pyrexia, and any inflammatory joint disease, clumsy child as referred by parents or any neurological problems. If any red flags found in the subjects, then immediate referral to a hospital or a concern medical setup was made.

Using Fitnessgram for assessing the physical fitness of school children. The selected students were assessed with Fitnessgram test battery. The physiotherapist followed the procedures outlined in the Fitnessgram® manual updated fourth edition to provide feedback on correct form.⁶ Each test of Fitnessgram was demonstrated to the students prior to their participation and corrective test position has been maintained for each and every selected student. The tests were conducted in a group of 6 children at a time to avoid any chaos and confusion. The detail procedure of the test battery was given in (Table 1).

Statistical analysis

Data were pooled and put under statistical analysis. The two observer's score were lined up in the excel sheet where the qualitative data of shoulder reach test was coded into quantitative data. Each component of the Fitnessgram test battery were analysed for inter rater reliability. Interrater reliability was done with independent ratings of the same event with two raters. No discussion or collaboration occurred when reliability is tested. Reliability is determined by the correlation of the scores from two independent raters (for ratings on a continuum).

Cronbach's Alpha value has been computed for the Interrater reliability. The mean and standard deviation for the age, height and weight were computed. Each component of Fitnessgram test battery were analysed for range, mean and standard deviation and was compared between both the raters. Statistical Package for Social Survey (SPSS) for Windows, version 20 and Microsoft excel was used to find out the reliability coefficient.

RESULTS

Demographic information

There were 70 students involved in the study. However, 62 students had complete data with no missing values,

thus the response rate was 88.57%. The sample population were both males and females constituting male with 56.3% and female with 43.7%. Demographic details including age, height and weight were described in (Table 2). The overall students mean age was 10.12 ± 2.72 years. The mean height of the whole study participants was 140.9 ± 17.1 cm. The overall mean weight of the students was 36.8 ± 12.5 kg. The BMI (Body mass index) mean for the whole study population was 18.46 ± 3.43 (Table 2).

Table 2: Demographic details of the participating school children.

	Mean (62)	SD
Age	10.12	2.72
Height	140.9	17.1
Weight	36.8	12.5
BMI	18.46	3.43

Overall, raters demonstrated acceptable to excellent agreement on each of variable in Fitnessgram for the total sample. The two assessor's mean and SD for different variables of Fitnessgram has minimal difference. An excellent reliability has been found when two raters assessed the individual variables i.e. $\alpha \geq 0.9$. As shown in the (Table 3). The VO_2 max calculated by both the raters on the same students were in the range of 32.2 mL/kg/min to 43.9 mL/kg/min where the difference between minimum values is 0.5 and the maximum values is also 0.5. The overall mean of the VO_2 max as measured by rater 1 and rater 2 were 38.157 ± 2.28 and 38.24 ± 2.25 respectively (Table 3). The curl up repetitions were measured and the minimum and maximum values has some differences where some students showed increase in repetition on second attempt with the second rater though the mean and standard deviation doesn't have significant difference. The mean number of curl ups as measured by rater 1 and rater 2 were 17.1667 ± 7.15 and 17.8636 ± 7.31 (Table 3). The number of repetitions of push up doesn't have differences between the first attempt and the second attempt. The range of 0 to 30 numbers of pushups by the students with mean value 4.86 ± 5.88 and 5.63 ± 6.08 during first and second attempt as measured by rater 1 and rater 2 respectively has shown an excellent inter-rater reliability of the test (Table 3). The trunk lifts by the students in the both instances have similar raise (in inches) of the trunk above the ground with excellent inter-rater reliability (≥ 0.9) of the test. The mean inches raise of the ground was 8.8182 ± 2.74 inch and 9.7576 ± 2.73 inch as measured by rater 1 and rater 2 respectively (Table 3).

The shoulder reach test has result in qualitative form and was quantified into numerical 0 and 1. The contact of the student's fingertip of both the hand at the back has been coded 0 and the one which doesn't have been coded 1. The inter-rater reliability was found to be acceptable (0.73). The body composition measured with skin fold measurement and Body Mass Index were having similar

scores measured by both the raters and Cronbach's alpha value of 0.99 found. The overall mean of skin fold measurement was 12.51 ± 6.50 mm and 12.57 ± 6.46 mm as measured by rater 1 and rater 2 respectively (Table 3).

Table 3: Inter rater reliability of different components of Fitnessgram test.

N=62		Min-max	Mean±SD	Cronbach's alpha
Pacer	Rater 1	32.3-43.4	38.157±2.28	0.959
	Rater 2	31.7-43.9	38.24±2.25	
Skin fold	Rater 1	4-35	12.51±6.50	0.99
	Rater 2	4-35	12.57±6.46	
Curl up	Rater 1	5-32	17.166±7.15	0.949
	Rater 2	2-39	17.863±7.31	
Push up	Rater 1	0-27	4.8636±5.88	0.941
	Rater 2	0-30	5.6364±6.08	
Trunk lift	Rater 1	3-12	8.8182±2.74	0.914
	Rater 2	4-11	9.7576±2.73	
Shoulder reach	Rater 1	0-1	0.8939±0.31	0.73
	Rater 2	0-1	0.9091±0.28	

DISCUSSION

This small exploratory study investigates the inter-rater reliability of Fitnessgram test battery to be used for assessing the physical fitness of children with reference to the pediatric population of Guwahati, a capital city of north east India. This is the first study to report the reliability of a physical fitness test.

The components of Fitnessgram test battery consist of different variables where each of the variables counts its own importance and the general physical fitness of a children depends not only on one variable instead has to consider all the components of Fitnessgram.

The measurement of aerobic capacity through PACER where number of laps measuring 20 meters were covered by the students in two different instances and was found to be similar.

The actual VO_2 max and VO_2 max measured by PACER laps does not have significant difference.¹⁰ Measuring aerobic fitness using the criterion measure VO_2 max requires expensive equipment and is thus not feasible for administration in a school or many clinical settings. Instead, several field tests are commonly used.¹¹ One of the study provide an excellent example of determining the criterion-referenced reliability of the Fitnessgram's PACER and 1-mile walk/run items.⁸ The assessment is evaluated using criterion-referenced standards that reflect the amount of fitness needed for good health. In the study the aerobic capacity VO_2 max was compared with the Criterion-referenced standards. A quadratic equation was used to find the VO_2 max where gender of the child, BMI and number of laps were used.¹⁰ The quadratic equation used for calculating aerobic capacity is,

$$VO_2 \text{ max} = 41.77 + 0.49 (\text{laps}) - 0.0029 (\text{laps})^2 - 0.62(\text{BMI}) + 0.35 (\text{gender} \times \text{age}),$$

Where gender, 0 for girls, 1 for boys was used. Criterion-referenced standards are more useful for fitness evaluation since it makes it possible for individuals to compare their overall fitness to an absolute criterion (The Cooper Institute, 2011). From age 5 years to 9 years the number of laps were not recommended to find out the aerobic capacity though the Fitnessgram recommended for participation in the run. These standards help to place the individual in either the Health Fitness Zone (HFZ) or the Needs Improvement Zone (NI). In this study the results of VO_2 max by both the rater were within the range as provided in the criterion referenced standards for healthy fitness zone. The curl up, trunk lift and push up were done to assess the strength and endurance of muscles. They all showed an excellent reliability. This may be due to the reason that muscle physiology of the individual doesn't get change until any injury or pathological changes takes place. The reliability coefficient of shoulder stretch test was found to be 0.73 which is acceptable but the reason for this variation from other components may be due to some minimal error from the part of the assessor and the students. The body composition measured by means of skin fold measurement and Body mass index were found to be similar as measured by both the raters and this absolute similar result was found as there was no loss of weight in the said duration and nor has chance of reduction of height. Shin fold variation is less likely to change in this short duration of gap. In the fitness testing, most of the tests involve maximum effort that results in lasting fatigue (e.g., PACER Test, mile run/walk, 900 push-up test, curl-up test) and due to this the student's maximum aerobic capacity, strength and endurance can be easily determined.⁹ The scores of Fitnessgram test can be influenced by some of the factors and to make the scores meaningful the tester need to fulfil certain condition like maintaining privacy more in the case of body composition measurement and also the scores can be influenced by making the fitness test as an integral part of teaching and providing health related concepts.⁸ The scores result of this study can be compared with the criterion referenced standard scores to evaluate the relationship between them using an appropriate statistical test.

Additional research on establishment of normative data of physical fitness of pediatric population in India is needed by using Fitnessgram test battery.

CONCLUSION

Study has enlightened the requirement of Fitnessgram test for assessing the physical fitness of the pediatric population in Guwahati urban society which has excellent intra-rater reliability. Based upon this study Fitnessgram test can be used in other parts of the country or as a whole to established a geographical based criterion

reference scores in order to determine the association of fitness and other health outcomes, evaluate the effectiveness of training programs designed to increase fitness, and determine the prevalence of adequate levels of fitness. In school settings, Fitnessgram tests can be implemented for other purposes, to provide individualized feedback to students about their fitness levels and make recommendations for increasing or maintaining current fitness levels.

ACKNOWLEDGEMENTS

Authors would like to thank the principal and the entire faculty member of Shankardev shishu niketan, Maligaon, Guwahati, Assam, India for their support and their arrangement.

Funding: No funding sources

Conflict of interest: None declared

Ethical approval: The study was approved by the Institutional Ethics Committee of Gauhati Medical College, Guwahati, Assam, India

REFERENCES

1. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Rep. 1985 Mar;100(2):126-31.
2. Ortega FB, Ruiz JR, Castillo MJ. Physical fitness in childhood and adolescence: a powerful marker of health. Int J Obes. 2008;32:1-11.
3. Fox KR, Biddle SJ. The use of fitness tests: Educational and psychological considerations. J Physic Edu, Recrea Dance. 1988 Feb 1;59(2):47-53.
4. Ernst MP, Corbin CB, Beighle A, Pangrazi RP. Appropriate and inappropriate uses of FITNESSGRAM®: a commentary. J Phys Act Health. 2006 Jan 1;3(s2):S90-100.
5. Meredith MD, Welk GJ. In: Fitnessgram and activitygram Test Administration Manual. Dallas, TX: Cooper Institute Aerobics Res. 2010:9-10.
6. Mahar MT, Guerieri AM, Hanna MS, Kemble CD. Estimation of aerobic fitness from 20-m multistage shuttle run test performance. Am J Prevent Med. 2011 Oct 1;41(4):S117-23.
7. Morrow JR, Falls JR HB, Kohl HW. In: The Prudential Fitnessgram technical reference manual. 4th edition. III. (Eds.). Dallas; TX: Cooper Institute Aerobics Research; 1994:7-8.
8. Mahar MT, Rowe DA. Practical guidelines for valid and reliable youth fitness testing. Measure Physic Edu Exercise Sci. 2008 Jul 29;12(3):126-45.
9. Boiarskaia EA, Boscolo MS, Zhu W, Mahar MT. Cross-validation of an equating method linking aerobic FITNESSGRAM® field tests. Am J Prevent Med. 2011 Oct 1;41(4):S124-30.
10. Safrit MJ, Baumgartner TA, Jackson AS, Stamm CL. Issues in setting motor performance standards. Quest. 1980 Jul 1;32(2):152-62.
11. Beets MW, Pitetti KH. Criterion-referenced reliability and equivalency between the PACER and 1-mile run/walk for high school students. J Phys Act Health. 2006 Jan 1;3(s2):S21-33.

Cite this article as: Gogoi P, Bhattacharyya NC. Interrater reliability of fitness gram test to be used as a reliable tool for assessing physical fitness for school children in Guwahati urban society of North-East India. Int J Res Med Sci 2020;8:312-6.