

Assessing Health-Related Fitness Tests in the School Setting: Reliability, Feasibility and Safety; The ALPHA Study

Authors

V. España-Romero^{1,2}, E. G. Artero¹, D. Jimenez-Pavón^{1,3,4}, M. Cuenca-García¹, F. B. Ortega^{1,2}, J. Castro-Piñero^{2,5}, M. Sjöström², M. J. Castillo-Garzon¹, J. R. Ruiz^{1,2}

Affiliations

Affiliation addresses are listed at the end of the article

Key words

- reliability
- feasibility
- safety
- health-related fitness test
- school setting

Abstract

The aim of this study was to determine the reliability, feasibility and safety of a health-related fitness test battery administered by Physical Education (PE) teachers in the school setting. Six PE teachers, from three primary schools and three secondary schools, assessed twice (7 days apart) the 20 m shuttle run, handgrip and standing long jump tests, as well as weight, stature, triceps and subscapular thickness and waist circumference in 58 children (age: 6–11 yr) and 80 adolescents (age: 12–18 yr). Feasibility and

safety were assessed by researchers by means of direct observation. Significant inter-trial differences were found for the standing long jump test (3.8 ± 12.7 cm, $P < 0.05$) and for stature (0.73 ± 0.8 cm, $P < 0.001$) in children, and for waist circumference in both children and adolescents (-0.82 ± 1.2 cm and -0.35 ± 0.8 cm respectively, $P = 0.001$). The feasibility and safety items assessed presented a successful answer. Therefore, the results indicate that health-related fitness tests administered by PE teachers are reliable, feasible and safe to be performed in the school setting.

Introduction

Physical fitness is considered an important health-related marker already in youth [23,27]. Physical fitness can be objectively and accurately measured through laboratory methods, yet due to the high cost, necessity of sophisticated instruments, qualified technicians, and the time constraints, laboratory tests are still not feasible to be used at population level. In contrast, field-based fitness tests are easy to administer, involve minimal equipment, low cost and a larger number of participants can be evaluated in a relatively short period of time [24,26,30]. In the school setting, field-based fitness tests are a practical and feasible option to assess physical fitness. Validity and reliability are two characteristics that need to be assured for any measurement tool [8]. Validity refers to the ability of a test to reflect what it is designed to measure, and reliability refers to the reproducibility of a test in repeated trials on the same individual/s. A reliable test is considered when an individual performs a test on two or more occasions under the same conditions and close proximity in time and obtains similar results [13].

Several studies determined the reliability of health-related fitness tests assessed by trained researchers under standardized and controlled circumstances in youth [9,10,15,18–22,25,28]. Whether fitness tests are still reliable enough when performed in the school setting by physical education (PE) teachers remain to be investigated. The school setting is an ideal place to carry out fitness monitoring because virtually all the children and adolescents attend elementary and secondary school and therefore can be systematically monitored [32]. Schools may play an important role in identifying children with low fitness via standardized field tests and promoting positive fitness-enhancing behaviours. If this is the case, feasibility and safety are major concerns in testing [33], however studies examining these issues in children and adolescents are scarce. The aim of the present study was to determine the reliability, feasibility and safety of a health-related fitness test battery administered in the school setting by PE teachers.

accepted after revision
February 26, 2010

Bibliography

DOI <http://dx.doi.org/10.1055/s-0030-1251990>
Published online: 2010
Int J Sports Med
© Georg Thieme Verlag KG
Stuttgart · New York
ISSN 0172-4622

Correspondence

Vanesa España-Romero
University of Granada,
Department of Medical
Physiology, School of Medicine,
Granada, Spain, Avd/Madrid sn
18008 Granada
Spain
Tel.: +34/958/24 3540
Fax: +34/958/24 9015
vanespa@ugr.es

Methods

The present study was performed under the framework of the ALPHA (Assessing Levels of Physical Activity and Fitness) study. The ALPHA study is a European Union-funded project aiming to provide a set of instruments for assessing levels of physical activity as well as health-related physical fitness in a comparable way within the European Union. A detailed description of the ALPHA study has been published elsewhere [17].

PE teachers randomly chosen, with 5-years university degree in Sport Sciences and qualified for PE teaching in the school, were informed about the aims and methodology of the study and volunteered to participate. Likewise, a comprehensive verbal description of the nature and purposes of the study was given to the children, adolescents and their parents. Both students and their parents or guardian signed a written consent form to participate in the study. It was performed in accordance with the ethical standards of the International Journal of Sport Medicine [12].

Six PE teachers from three primary schools and three secondary schools (2 women and 4 men) with a teaching experience ranging from 3 to 12 years, administered twice (7 days apart) a health-related fitness test battery in 58 children (aged 6–11.9 years) and 80 adolescents (aged 12–18 years) in Granada (Spain). The physical fitness measurements were administered during the PE lessons. The fitness test battery was administered in two standard PE lessons, plus other two additional lessons to administer the re-test measurement one week later. The two physical fitness measurements were performed at the same time of day and by the same PE teacher.

Physical fitness assessment

A detailed standard operating procedure was designed and thoroughly read by every PE teacher one week before the data collection started. Questions were solved in a consensus meeting. In addition, an individual (one PE teacher at time) one-day (5 h) training session was carried out before data collection by two experienced researchers. They were trained on (i) anthropometric measurements (how to identify the Frankfort plane, how to identify the bone landmarks and how to measure the skinfolds); (ii) fitness assessment (how to measure the fitness level in children and adolescents, assessment preparation, and how to distinguish correct and incorrect trials); (iii) and about how to handle the questionnaires. This process was performed to ensure standardization of data collection. As the measurements were performed in the real school setting, every PE teacher organized the tests according to their own school circumstances, i.e., test inside or outside, 20m shuttle run test in one or more groups, anthropometry in a specific room or in the classroom, etc.

The health-related physical fitness components assessed were: cardiorespiratory fitness, musculoskeletal fitness and body composition (weight, stature, triceps skinfold, subscapular skinfold and waist circumference). The inclusion of body composition as a component of health-related physical fitness was completed following the recommendations of the Toronto model [6] and the ACSM guidelines [2]. The scientific rationale for the selection of the physical fitness tests was based on their evidence-based predictive validity [27] and criterion-related validity [7].

Cardiorespiratory fitness

It was assessed by means of the 20m shuttle run test. The participants performed the test as previously described by Léger

et al. [14]. Participants were required to run between two lines 20m apart, while keeping pace with audio signals emitted from a pre-recorded CD. The initial speed was 8.5 km/h, and was increased by 0.5 km/h per minute (one minute equals one stage). Participants were instructed to run in a straight line, to pivot on completing a shuttle, and to pace themselves in accordance with the audio signals. The test was finished when the participant failed to reach the end lines concurrent with the audio signals on two consecutive occasions, or when the subject stopped because of fatigue. The participants were encouraged to keep running as long as possible throughout the course of the test. The test was performed once and the last completed stage or half-stage at which the subject dropped out was scored. A gymnasium or space large enough to mark out a 20m track was used to perform the test.

Upper body muscular strength

It was assessed by means of the handgrip strength test using a hand dynamometer with adjustable grip (TKK 5101 Grip D; Takey, Tokio Japan). This dynamometer presents a high validity and reliability when compared with calibrated known weights [10]. The participant squeezed gradually and continuously for at least two seconds, performing the test with the right and left hands in turn, and with the elbow in full extension as described elsewhere [10]. The grip-span of the dynamometer was adjusted according to the hand size for determining the maximum handgrip strength using the equations specifically developed for children [9] and adolescents [29]. In children those equations were $y = x/4 + 0.44$ for boys and $y = 0.3x - 0.52$ for girls while that in adolescents were $y = x/7.2 + 3.1$ for males and $y = x/4 + 1.1$ for females, where x is the hand size (in cm), and y is the grip span (in cm). The test was performed twice and the maximum score for each hand was recorded in kilograms. The sum of the scores achieved by left and right hands was used in the analysis.

Lower body muscular strength

It was assessed by means of the standing long jump test. From a starting position immediately behind a line, standing with feet approximately shoulder's width apart, the participant jumped as far forwards as possible on a non-slip hard surface. The test was performed twice and the best score was recorded in centimeters.

Anthropometry

Participants were barefoot and wore a t-shirt and short trousers. Weight was measured with an electronic scale (Type SECA 861; range, 0.05 to 130 kg; precision, 0.05 kg), and stature was measured in the Frankfort plane with a telescopic stature measuring instrument (Type SECA 225; range, 60 to 200 cm; precision, 1 mm).

Triceps and subscapular skinfold thickness were measured on the right side of the body with a Holtain caliper (range, 0–40 mm; precision, 0.2 mm). Triceps was raised in a vertical fold halfway between the acromion process and the superior head of the radius, in the posterior aspect of the arm, and subscapular about 20 mm below the inferior angle of the scapula and 45° to the lateral side of the body. It was performed according to Lohman's anthropometric standardization reference manual [16]. Waist circumference was measured with a non-elastic tape (Seca 200; range, 0–150 cm; precision, 1 mm), at the level of the natural waist, in a horizontal plane, which was the narrowest part of the torso, as seen from a front view. In some obese participants, it

was difficult to identify the waist circumference, therefore we measured it in the midpoint between the superior iliac spine and the costal edge in the midaxillary line [21]. The measurements were taken at the end of a normal expiration, without the tape compressing the skin. The measurements were carried out twice, but not consecutively; all the anthropometric variables were measured in order, and then the same measurements were repeated once more. The mean of the two measurements was used in the analyses.

Feasibility

The following items requiring yes/no answer were collected by the PE teachers or researchers: 1. Whether the children and adolescents wore appropriate sport clothes to perform the health-related fitness tests. 2. Information about whether the tests' instructions were properly understood by the participants. 3. If any participant rejected to perform the measurements, and the reason/s (1 = shyness, 2 = lack of motivation). 4. We also asked the PE teachers about (i) the school's appropriate facilities to perform the tests, (ii) if they considered the tests easy to administer, and (iii) their previous experience in administering these tests. 5. Furthermore, we also registered the time required by the PE teacher to prepare and administer the tests. An acceptable level of feasibility was considered when the items were 'positively' answered in at least 95% of the cases [35].

Safety

We recorded the following items requiring yes/no answer: 1. Instrument allergy, in the case of anthropometric and hand-grip assessments. 2. Sick feeling, in all the tests. 3. Pain in hand or forearm in handgrip test. 4. We also recorded musculoskeletal injuries during or after the physical fitness assessment. 5. Delayed-onset muscle soreness (DOMS) was assessed with a questionnaire that was completed 2 to 3 days after testing took place. This questionnaire included questions about the experience and severity of DOMS, the location of the pain or soreness, the particular test that could cause the DOMS, and the functional consequences of DOMS in usual activities. The DOMS-related questionnaire was completed by individual interviews. An acceptable level of safety was considered when the items were successfully answered in the 99% of the cases. The level of safety was higher than the level of feasibility because we considered the safety an important aspect of the physical fitness tests. Feasibility and safety of the tests were assessed by two trained researchers with previous experience by means of direct observation, i.e. they were in the classroom when the physical fitness tests were being administered by the PE teachers.

Statistical analysis

Test and retest (hereafter called T1 and T2) were compared between boys and girls and between children and adolescents by means of two-way repeated measures analysis of variance (ANOVA), with sex and age group as fixed factors. Since no sex-specific effect on reliability of the studied physical fitness tests was found (all $P > 0.2$), the analyses were performed for girls and boys together. However, age group-specific effect on reliability was found in three of the study tests (all $P < 0.05$), therefore the analyses were performed for the whole sample as well as for children and adolescents separately.

Potential systematic error between test and retest (also called bias) was analyzed by means of repeated measures ANOVA. The differences between test and retest scores was also examined

through different error measures. Suppose that N cases are available to evaluate the error measurements, where \hat{y} is T2 and y is T1. The sum of squared errors (SSE) was calculated as:

$$SSE = \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

We also calculated the mean sum of squared errors (MSE):

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

The root mean sum of squared errors (RMSE) was calculated by converting MSE into domain units by taking the root square:

$$RMSE = \sqrt{MSE}$$

The percentage error was calculated as:

$$\%Error = \frac{RMSE}{y_{max} - y_{min}} \times 100$$

We also calculated the standard error of estimate [31] as follows:

$$SEE = SD_y \sqrt{(1 - R_{yy}^2)}$$

where SD is the standard deviation of the T1 for every test, and R^2 is the explained variance between the measured T1 and T2. Moreover, the inter-trials agreement was also examined graphically by plotting the difference between each pair of measurements against their mean, according to the Bland and Altman approach [4, 5]. The 95% limits of agreement for all the physical fitness variables were calculated as the inter trials mean difference ± 1.96 SD of the inter trial differences.

The presence of heteroscedasticity was examined by using one-way ANOVA in order to determine whether the absolute inter-trials difference was associated with the magnitude of the measurement (i.e., inter-trials mean). For that purpose, age group-specific quartiles of the inter-trials mean were established. A significant association ($P < 0.05$) would confirm heteroscedasticity, which means that the inter-trials difference of a test would differ with the physical fitness level groups. All analyses were performed using SPSS v.16 software for Windows. For all analyses, the significant level was set at 5%.

Results

Physical fitness assessment

Mean and SD values for the test and retest trials according to age groups are shown in **Table 1**. The inter-trials measurement error for the whole sample is shown in **Table 2**. Significant differences between test and retest were found in stature and waist circumference ($P < 0.001$).

Significant inter-trials differences were found in stature for children ($P < 0.001$), in waist circumference for both children and adolescents ($P = 0.001$), and in the standing long jump test only for children ($P < 0.05$) (**Table 3**).

The Bland-Altman plots (**Fig. 1**) graphically show the reliability patterns, in terms of systematic error (bias or mean inter-trial differences) and random error (95% limits of agreement), in

Table 1 Test and retest measurements (mean \pm SD) in children (age range: 6–11.9 years) and adolescents (age range: 12–18 years).

	All		Children		Adolescents	
	Trial 1	Trial 2	Trial 1	Trial 2	Trial 1	Trial 2
weight (kg)	48.9 \pm 15.4	49.20 \pm 15.0	36.8 \pm 10.2	37.1 \pm 10.1	57.6 \pm 11.8	57.4 \pm 11.5
stature (cm)	153.2 \pm 16.0	153.8 \pm 15.6	138.7 \pm 12.6	139.8 \pm 12.5	163.3 \pm 8.2	163.3 \pm 8.1
triceps skinfold (mm)	18.1 \pm 8.3	18.2 \pm 8.0	17.1 \pm 7.1	17.1 \pm 6.3	18.7 \pm 9.0	18.9 \pm 8.8
subscapular skinfold (mm)	12.2 \pm 6.1	12.6 \pm 6.3	10.9 \pm 5.2	11.7 \pm 5.4	13.1 \pm 6.5	13.3 \pm 6.8
waist circumference (cm)	66.1 \pm 9.0	65.8 \pm 8.7	60.8 \pm 7.4	60.2 \pm 7.2	70.3 \pm 7.6	69.5 \pm 7.4
handgrip (kg)	43.7 \pm 18.9	43.3 \pm 18.8	29.9 \pm 4.9	28.4 \pm 3.8	50.5 \pm 14.6	51.9 \pm 15.3
standing long jump (cm)	143.0 \pm 36.3	142.1 \pm 34.6	135.1 \pm 26.9	131.4 \pm 21.8	180.2 \pm 31.9	186.6 \pm 33.6
20 m shuttle run (stages)	4.7 \pm 2.3	4.6 \pm 2.1	4.3 \pm 1.7	4.0 \pm 1.5	6.2 \pm 2.1	5.9 \pm 2.0

Table 2 Inter-trials measurement error in the whole sample.

	Inter-trial difference (T2-T1)	SSE	MSE	RMSE	% Error	SEE
weight (kg)	0.07 \pm 0.6	44.1	0.34	0.58	0.75	0.58
Stature (cm)	0.38 \pm 0.8*	93.05	0.73	0.85	1.23	0.85
triceps skinfolds (mm)	0.10 \pm 2.6	816.42	6.80	2.61	7.47	2.60
subscapular skinfolds (mm)	0.24 \pm 2.2	568.06	4.77	2.18	6.41	2.18
waist circumference (cm)	-0.56 \pm 1.0*	157.89	1.35	1.16	2.22	1.16
handgrip (kg)	-0.10 \pm 4.0	499.31	3.96	1.99	2.28	1.99
standing long jump (cm)	0.33 \pm 13.4	22273.75	160.5	12.7	6.30	13.32
20 m shuttle run (stages)	0.05 \pm 1.0	125.91	0.97	0.98	9.84	0.98

* Significant differences were found between T1 and T2 (repeated measures ANOVA)

T2-T1 refers to Trial 2 minus Trial 1. Values are means \pm standard deviation

SSE indicates sum of squared errors; MSE, mean of squared errors; RMSE, root mean squared errors; SEE, standard error of estimate

Interaction between sex \times test assessment was not found in any variable

Interaction between children and adolescent \times test assessment was found in height, waist circumference and standing broad jump

those measurements which presented a significant inter-trial difference. Those measurements were stature and standing long jump in children (0.73 \pm 0.8 cm and 3.8 \pm 12.7 cm, respectively), and waist circumference for both children and adolescents (-0.82 \pm 1.2 cm and -0.35 \pm 0.8 cm, respectively).

The heteroscedasticity analysis showed a positive association between inter-trial differences and inter-trial mean in weight for adolescents ($P < 0.01$), in stature for children ($P < 0.05$) and in skinfolds thickness (triceps and subscapular) for both children and adolescents ($P < 0.05$) (data not shown).

Feasibility

Most of the feasibility items presented a positive answer in $\geq 95\%$ of the cases (see **Table 4**). The only exception was the body composition assessment, specially triceps and subscapular skinfold thickness, which was considered by the PE teachers as not easy to measure in 17.8% of the cases, mainly because they were not familiarized with the equipment.

The maximum time required to prepare the tests was approximately 3 min in the anthropometrical assessment and 3 min in the physical fitness assessment (only handgrip strength and standing long jump). Preparing the 20 m shuttle run test took approximately 5 min. The mean time required to perform the anthropometric assessment was around 5 min per subject, approximately 1 min for the took and stature and 4 min for waist circumference, triceps and subscapular. Handgrip and standing long jump tests took 90 and 50 s, respectively. It took around 10 min to assess the 20 m shuttle run test in a group of 20 individuals.

Overall, the average time required to prepare and administer the whole battery to a group of 20 individuals was 2 h and 30 min, which equals to approximately 3 PE sessions of ~55 min. The

time required was shorter if triceps and subscapular skinfold thickness were not included in the battery: ~2 h to assess 20 persons (i.e. 2 PE sessions of ~55 min).

The experience reported by the PE teachers in administering health-related fitness tests was as follows: previous experience measuring weight was reported by 50% (3 teachers), 66.7% (4 teachers) for stature, 83.3% (5 teachers) for standing long jump test and 66.7% (4 teachers) for 20 m shuttle run test. None of the teachers reported to have ever assessed the handgrip test, skinfolds thickness or waist circumference.

Safety

Most of the safety items presented a successful answer in $\geq 99\%$ of the cases (see **Table 4**). No complications occurred during the testing procedure. Only one participant interrupted the 20 m shuttle run test because of lower body muscle cramp.

The responses rate for the DOMS questionnaire was 133 participants (96.4%). Among the responders, 22 (37.9%) children and 25 (33.3%) adolescents experienced some degree of DOMS, from which six children (10.3%) and four adolescents (5.3%) indicated that their DOMS was severe. Three (2.3%) subjects reported having severe pain in the upper body, 29 (21.8%) in lower body and 14 (10.5%) in the whole body. Most of them (39 participants; 29.3%) assumed that the 20 m shuttle run test could be the cause. For 11 (19%) children and 14 (18.7%) adolescents, DOMS caused difficulties in daily activities, especially stair climbing and walking.

Table 3 Inter-trials measurement error by age group.

	Inter-trial difference (T2-T1)		SSE		MSE		RMSE		% Error		SEE	
	Children	Adolescents	Children	Adolescents	Children	Adolescents	Children	Adolescents	Children	Adolescents	Children	Adolescents
weight (kg)	0.02±0.5	0.13±0.6	12.27	31.82	0.23	0.42	0.48	0.65	1.20	1.11	0.48	0.65
stature (cm)	0.73±0.8*	0.11±0.6	68.58	24.47	1.29	0.33	1.14	0.57	2.31	1.56	1.14	0.57
triceps skinfolds (mm)	-0.31±2.5	0.44±2.7	330.60	485.82	6.12	6.48	2.47	2.55	11.40	7.29	2.46	2.54
subscapular skinfolds (mm)	0.36±1.7	0.16±2.5	162.23	405.83	3.06	5.41	1.74	2.33	8.88	6.82	1.74	2.32
waist circumference (cm)	-0.82±1.2*	-0.35±0.8*	110.12	47.78	2.04	0.64	1.43	0.80	4.68	1.95	1.43	0.80
handgrip (kg)	-0.5±2.2	0.2±4.9	68.91	430.40	1.28	5.66	1.13	2.38	2.48	3.43	1.13	2.38
standing long jump (cm)	3.8±12.7*	-2.19±13.4	9467	12806.75	166.09	173.06	12.89	13.16	13.86	7.03	12.76	11.12
20m shuttle run (stages)	-0.11±0.9	-0.01±1.0	51.66	74.25	0.91	1.00	0.95	1.00	11.90	12.52	0.95	0.99

* Significant differences were found between T1 and T2 (repeated measures ANOVA)

T2-T1 refers to Trial 2 minus Trial 1. Values are means ± standard deviation

SSE indicates sum of squared errors; MSE, mean of squared errors; RMSE, root mean squared errors; SEE, standard error of estimate

Discussion

The main findings of this study suggest that a health-related fitness test battery including the 20m shuttle run, handgrip and standing long jump tests and measures on weight, stature, triceps and subscapular skinfolds, and waist circumference administered by PE teachers in the school setting can be considered reliable, feasible and safe.

Our results concur with those observed in previous studies. The studied fitness tests also have an acceptable level of reliability when administered by a trained staff in children [9, 18] and adolescents [10, 21, 22, 28]. No significant differences were reported between test and retest in handgrip strength in both children [9] and adolescents [22, 28]. The standing long jump test also seem to be reliable in adolescents [22], whereas no information exists for children. No differences between test and retest were reported in the 20m shuttle run test in children and adolescents aged 8–15 years [15, 22, 25]. Regarding anthropometric assessment, the error has been traditionally considered unavoidable [19] and has been frequently studied by means of technical error of measurement [19–21]. The inter- and intra-observer measurements errors for skinfolds and circumferences can be minimized by paying close attention to every aspect of the data collection process.

With the aim of making the data as close to the reality as possible, we asked PE teachers to administer the measurements in the school setting, with only one previous training day. We observed that the reliability was high in all the study tests except in stature and standing long jump in children and waist circumference in children and adolescents. Our findings concur with a similar study in boys and girls aged 5–12 years [32]. The authors compared the weight and stature measurements performed by school nurses with those performed by a trained research staff. They reported that weight measurements had a good reliability (<0.5kg between school nurses and trained research staff), while stature measurements required more attention especially among younger children because of the importance of proper positioning in order to obtain an accurate measurement [32]. Several studies have indicated that the reliability of stature measurement may improve in children when two or more measures are obtained and averaged [1, 34].

We observed a relatively low reliability in waist circumference and standing long jump in children. In waist circumference it could be due to the fact that this measure was performed by a person with just one day of previous training. It is well established that there is always certain anthropometric measurements error [19] and it is difficult to take a correct measurement [35]. Therefore more days of training could be needed so that the waist circumference measurement would be easier and more reliable. A low reliability in standing long jump could be due to performance on the day or even learning effects (positive systematic bias) exist among standing long jump tests when repeated measurements were performed for children. Ortega et al. [22] examined the reliability of standing long jump in adolescents. They suggested that neither learning nor fatigue (negative systematic bias) effects occurred in their study, on a test-retest basis. To our knowledge, there are no studies examining the reliability of standing long jump in children.

Although significant differences were found between test and retest in stature for children, waist circumference for both children and adolescents, and standing long jump test for children, the mean inter-trial difference for these measurements was

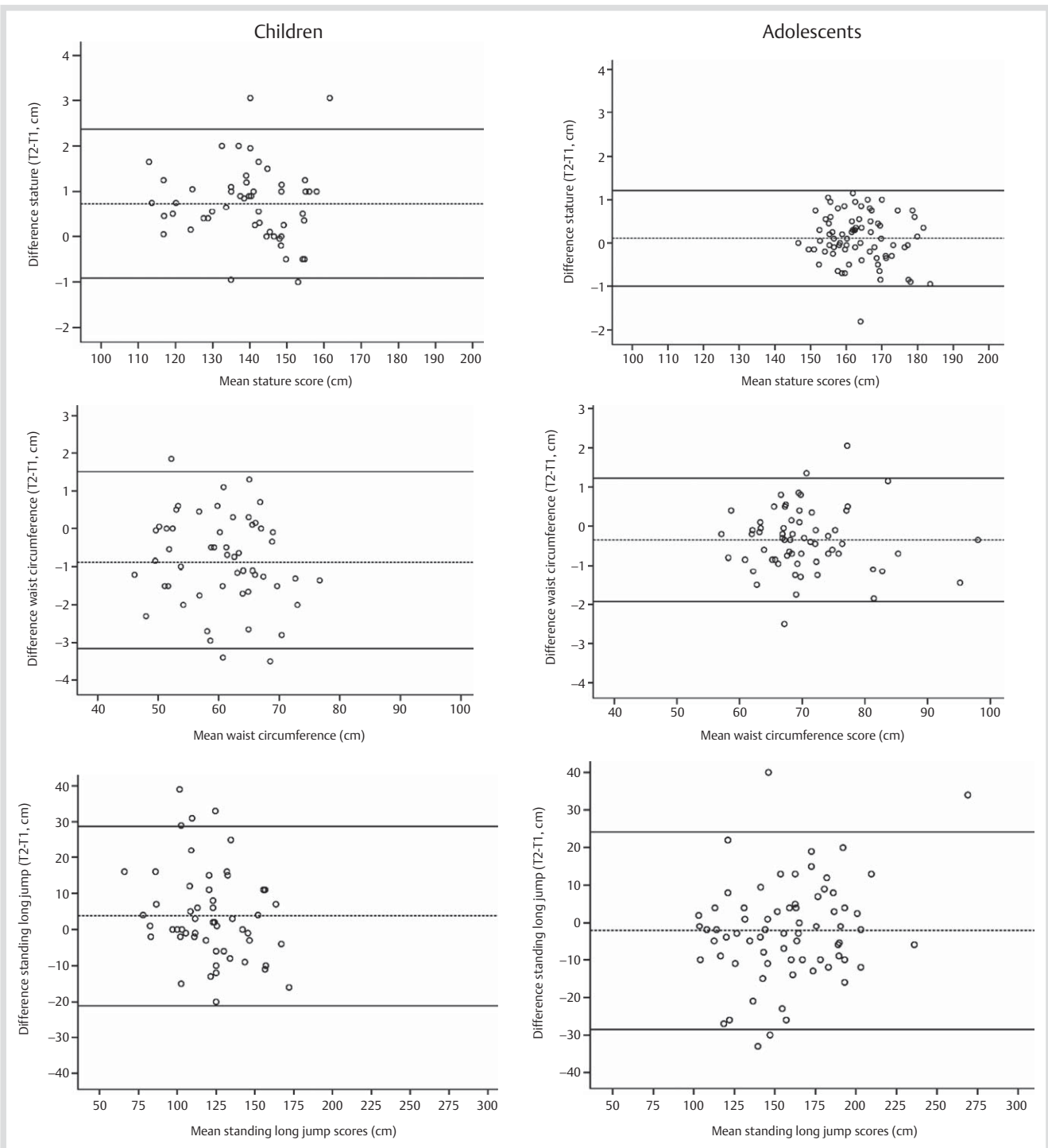


Fig. 1 Bland-Altman plot of the stature, waist circumference and standing long jump test in children (C1) and adolescents (A1). The central line represents the mean differences between the second trial (T2) and the first trial (T1); the upper and lower dotted lines represent the upper and lower 95% limits of agreement (means differences ± 1.96 SD of the differences), respectively.

nearly 0.5 cm for stature and waist circumference and 4 cm for standing long jump (Table 3). This suggests that despite the significant differences between test and retest, from a practical point of view these tests can be used in the school setting to evaluate children and adolescents.

The heteroscedasticity analysis showed significant association between the magnitude of the measurement and the difference between test and retest in weight for adolescents, in stature for children and in skinfolds thickness (triceps and subscapular) for

both children and adolescents. It is well known that accuracy in skinfolds thickness and circumferences is higher in lean subjects compared to obese subjects [7,11,26]. Differences in heteroscedasticity for weight and stature between children and adolescents can be due to their differences in the homogeneity (i.e. SD and range) between age groups. No heteroscedasticity was observed in physical fitness tests.

The information regarding the feasibility of the tests indicate the presence of some rejections because of shyness (especially for

Table 4 Feasibility and safety of the measurement protocol (no retest measurement).

	Body composition		Handgrip		Standing long jump		20m shuttle run	
	Yes	No	Yes	No	Yes	No	Yes	No
<i>Feasibility</i>								
appropriate sport clothes	100 (138)	0 (0)	99.2 (131)	0.8 (1)	99.2 (131)	0.8 (1)	99.2 (131)	0.8 (1)
test's instructions understood	–	–	98.5 (130)	1.5 (2)	99.2 (130)	0.8 (1)	100 (132)	0 (0)
reject	2.9 (4)	97.1 (134)	0 (0)	100 (132)	0.8 (1)	99.2 (131)	1.5 (2)	98.5 (130)
shyness	2.2 (3)	97.8 (135)	0 (0)	100 (132)	0.8 (1)	99.2 (131)	0 (0)	100 (132)
lack of motivation	0 (0)	100 (138)	0 (0)	100 (132)	0 (0)	100 (132)	0.8 (1)	99.2 (131)
appropriate facilities*	100 (6)	0 (0)	100 (6)	0 (0)	100 (6)	0 (0)	100 (6)	0 (0)
easy to measure*	82.2 (111)	17.8 (24)	100 (6)	0 (0)	100 (6)	0 (0)	100 (6)	0 (0)
<i>Safety</i>								
instrument allergy	0 (0)	100 (137)	0 (0)	100 (131)	–	–	–	–
sick feeling	0.7 (1)	99.3 (135)	0 (0)	100 (131)	0 (0)	100 (131)	0.8 (1)	99.2 (131)
pain in hand or forearm	–	–	0.8 (1)	99.2 (131)	–	–	–	–
injury	–	–	0	100	0 (0)	100 (131)	0.8 (1) ⁺	99.2 (131)

Values are % (number of individuals)

* Questions answered by the physical education teacher

+ Muscle cramp

body composition) and lack of motivation, yet it occurred in less than 3% of the cases. In contrast, all the feasibility items presented a positive answer in $\geq 95\%$ of the cases, indicating that the health-related fitness tests administered by PE teachers in the school setting seem to be feasible. All PE teachers reported that the school facilities were appropriate for performing the physical fitness tests. Moreover, all the physical fitness tests were considered easy to measure except for the body composition assessment, specially triceps and subscapular skinfold thickness, which were reported by PE teachers as difficult in 17.8% (n=24) of the participants. This could be due to the fact that the body composition assessment is more difficult to administer than handgrip, standing long jump or 20m shuttle run tests.

Regarding the safety of the measurements, $\geq 99\%$ of the items presented a successful answer, indicating that the fitness tests seem to be safe when administered by PE teachers in the school setting. All the physical fitness tests were well tolerated by the children and adolescents. Occurrence of severe DOMS in 10 participants was our major concern. The 20m shuttle run test was reported to cause severe DOMS and difficulties in daily activities, especially in stair climbing and walking, among a small number of subjects (25). DOMS is considered to be typically experienced by all individuals regardless of fitness level, and is a normal physiological response to increased exertion and the performance of unfamiliar physical activities [3].

From a practical point of view, this study proposes a health-related fitness test battery that is reliable, feasible and safe to be administered by PE teachers in the school setting. Given the well-documented evidence about physical fitness as a marker of health status in children and adolescents [23,27] the provided battery will help to identify children and adolescents with low physical fitness levels, which will be useful for health promotion and disease prevention programs.

In conclusion, the present study indicates that a health-related fitness test battery including the 20m shuttle run, handgrip and standing long jump tests and measures on weight, stature, triceps and subscapular skinfolds, and waist circumference administered by PE teachers is reliable, feasible and safe to be performed in the school setting.

The results presented in this study, together with experiences from the previous projects across Europe (e.g. AVENA, EYHS, HELENA, IDEFICS) indicate that the fitness test battery examined could be a standardized battery, namely the ALPHA Fitness tests battery, to be used in schools as well as in population based studies.

Acknowledgements

This study takes place under the frame of the ALPHA study, which takes place with the financial support of the public Health Executive Agency, DG Sanco, Health Information Strand (Ref: 2006120); The Spanish Ministry of Education (EX-2008-0641, AP-2005-3827, AP-2005-4358, EX-2007-1124, AP-2008-03806), and the Swedish Council for Working Life and Social Research (FAS) and the HELENA study funded by the European Community Sixth RTD Framework Programme (Contract FOOD-CT-2005-007034). We gratefully acknowledged the help of all the children and adolescents that took part in this study, and specially the physical education teachers involved in field work for their efforts and great enthusiasm.

Affiliations

¹ University of Granada, Department of Medical Physiology, School of Medicine, Granada, Spain

² Department of Biosciences and Nutrition, Unit for Preventive Nutrition, Karolinska Institutet, Huddinge, Sweden

³ Department of Health and Human Performance, Faculty of Physical Activity and Sport Sciences-INEF, Technical University of Madrid, Spain

⁴ EUMI Sagrado Corazón, University of Córdoba, Spain

⁵ School of Science, University of Cádiz, Department of Physical Education, Puerto Real, Spain

References

- 1 Ahmed ML, Yudkin PL, Macfarlane JA, McPherson K, Dunger DB. Are measurements of height made by health visitors sufficiently accurate for routine screening of growth? Arch Dis Child 1990; 65: 1345–1348
- 2 Armstrong LE, Whaley MH, Brubaker PH, Otto RM. ACSM' Guidelines for Exercise Testing and Prescription. 7th edn., Lippincott Williams & Wilkins, American College of Sport Medicine, Guidelines Skin Fold Measurement, Philadelphia PA; 2005
- 3 Bakhtiary AH, Safavi-Farokhi Z, Aminian-Far A. Influence of vibration on delayed onset of muscle soreness following eccentric exercise. Br J Sports Med 2007; 41: 145–148

- 4 Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet* 1995; 346: 1085–1087
- 5 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307–310
- 6 Bouchard C, Shepard RJ, Stephens T. *Physical Activity, Fitness and Health*. Human Kinetics Publishers, Champaign, IL; 1994
- 7 Castro-Pinero J, Artero EG, Espana-Romero V, Ortega FB, Sjostrom M, Suni J, Ruiz JR. Criterion-related validity of field-based fitness tests in youth: A systematic review. *Br J Sports Med* 2009
- 8 Currell K, Jeukendrup AE. Validity, reliability and sensitivity of measures of sporting performance. *Sports Med* 2008; 38: 297–316
- 9 España-Romero V, Artero EG, Santaliesra-Pasias AM, Gutierrez A, Castillo MJ, Ruiz JR. Hand span influences optimal grip span in boys and girls aged 6 to 12 years. *J Hand Surg [Am]* 2008; 33: 378–384
- 10 España-Romero V, Ortega Porcel F, Vicente-Rodriguez G, Artero EG, Rey-López JP, Ruiz JR. Elbow position affects handgrip strength in adolescents: Validity and reliability of Jamar, DynEx and TKK dynamometers. *J Strength Cond Res* 2010; 24: 272–277
- 11 España-Romero V, Ruiz JR, Ortega Porcel F, Artero EG, Vicente-Rodriguez G, Moreno LA, Castillo MJ, Gutiérrez Sainz A. Body fat measurement in elite sport climbers: comparison of skinfold thickness equations with dual energy X-ray absorptiometry. *J Sports Sci* 2009; 27: 469–477
- 12 Harris DJ, Atkinson G. *International Journal of Sports Medicine – Ethical standards in sport and exercise science research*. *Int J Sports Med* 2009; 30: 701–702
- 13 Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med* 2000; 30: 1–15
- 14 Leger LA, Mercier D, Gadoury C, Lambert J. The multistage 20 metre shuttle run test for aerobic fitness. *J Sports Sci* 1988; 6: 93–101
- 15 Liu NY, Plowman SA, Looney MA. The reliability and validity of the 20-meter shuttle test in American students 12 to 15 years old. *Res Q Exerc Sport* 1992; 63: 360–365
- 16 Lohman TG, Roche AF, Matorell R. *Anthropometric standardization reference manual*. Human Kinetics, Champaign IL 1991
- 17 Meusel D, Ruiz JR, Ortega F, Hagstromer M, Bergman P, Sjostrom M. Assessing Levels of Physical Activity in the European Population – the ALPHA project. Selección: *Revista Española e Iberoamericana de la Medicina de la Educación Física y el Deporte* 2007; 16: 9–12
- 18 Molenaar HM, Zuidam JM, Selles RW, Stam HJ, Hovius SE. Age-specific reliability of two grip-strength dynamometers when used by children. *J Bone Joint Surg Am* 2008; 90: 1053–1059
- 19 Moreno LA, Joyanes M, Mesana MI, Gonzalez-Gross M, Gil CM, Sarria A, Gutierrez A, Garaulet M, Perez-Prieto R, Bueno M, Marcos A. Harmonization of anthropometric measurements for a multicenter nutrition survey in Spanish adolescents. *Nutrition* 2003; 19: 481–486
- 20 Moreno LA, Rodriguez G, Guillen J, Rabanaque MJ, Leon JF, Arino A. Anthropometric measurements in both sides of the body in the assessment of nutritional status in prepubertal children. *Eur J Clin Nutr* 2002; 56: 1208–1215
- 21 Nagy E, Vicente-Rodriguez G, Manios Y, Beghin L, Iliescu C, Censi L, Dietrich S, Ortega FB, De Vriendt T, Plada M, Moreno LA, Molnar D. Harmonization process and reliability assessment of anthropometric measurements in a multicenter study in adolescents. *Int J Obes* 2008; 32 (Suppl 5): S58–S65
- 22 Ortega FB, Artero EG, Ruiz JR, Vicente-Rodriguez G, Bergman P, Vriendt TD, Eniko N, Konsta O, Rey P, Polito A, Dietrich S, Plada M, Beghin L, Manios Y, Sjostrom M, Castillo MJ. Reliability of health-related physical fitness tests in European adolescents. The HELENA study. *Int J Obes* 2008; 32: S49–S57
- 23 Ortega FB, Ruiz JR, Castillo MJ, Sjostrom M. Physical fitness in childhood and adolescence: a powerful marker of health. *Int J Obes* 2008; 32: 1–11
- 24 Paineau D, Chiheb S, Banu I, Valensi P, Fontan JE, Gaudelus J, Chapalain V, Chumlea C, Bornet F, Boulier A. Comparison of field methods to estimate fat mass in children. *Ann Hum Biol* 2008; 35: 185–197
- 25 Pitetti K, Fernhall B, Figono S. Comparing two regression formulas that predict VO₂ peak using the 20-m shuttle run for children and adolescents. *Pediatr Exerc Sci* 2002; 125–134
- 26 Rodriguez G, Moreno LA, Blay MG, Blay VA, Fleta J, Sarria A, Bueno M. Body fat measurement in adolescents: comparison of skinfold thickness equations with dual-energy X-ray absorptiometry. *Eur J Clin Nutr* 2005; 59: 1158–1166
- 27 Ruiz JR, Castro-Pinero J, Artero EG, Ortega FB, Sjostrom M, Suni J, Castillo MJ. Predictive validity of health-related fitness in youth: A systematic review. *Br J Sports Med* 2009; 43: 909–923
- 28 Ruiz JR, Espana-Romero V, Ortega FB, Sjostrom M, Castillo MJ, Gutierrez A. Hand span influences optimal grip span in male and female teenagers. *J Hand Surg (Am)* 2006; 31: 1367–1372
- 29 Ruiz JR, Espana-Romero V, Ortega FB, Sjöström M, Castillo MJ, Gutiérrez A. Hand span influences optimal grip span in male and female teenagers. *J Hand Surg (Am)* 2006; 31: 1367–1372
- 30 Ruiz JR, Ramirez-Lechuga J, Ortega FB, Castro-Pinero J, Benitez JM, Arauzo-Azofra A, Sanchez C, Sjostrom M, Castillo MJ, Gutierrez A, Zabala M. Artificial neural network-based equation for estimating VO_{2max} from the 20 m shuttle run test in adolescents. *Artif Intell Med* 2008; 44: 233–245
- 31 Saelens BE, Seeley RJ, van Schaick K, Donnelly LF, O'Brien KJ. Visceral abdominal fat is correlated with whole-body fat and physical activity among 8-y-old children at risk of obesity. *Am J Clin Nutr* 2007; 85: 46–53
- 32 Stoddard SA, Kubik MY, Skay C. Is school-based height and weight screening of elementary students private and reliable? *J Sch Nurs* 2008; 24: 43–48
- 33 Suni JH, Miilunpalo SI, Asikainen TM, Laukkanen RT, Oja P, Pasanen ME, Bos K, Vuori IM. Safety and feasibility of a health-related fitness test battery for adults. *Phys Ther* 1998; 78: 134–148
- 34 Vegelin AL, Brukx LJ, Waelkens JJ, Van den Broeck J. Influence of knowledge, training and experience of observers on the reliability of anthropometric measurements in children. *Ann Hum Biol* 2003; 30: 65–79
- 35 Waninge A, van der Weide W, Evenhuis JJ, van Wijck R, van der Schans CP. Feasibility and reliability of body composition measurements in adults with severe intellectual and sensory disabilities. *J Intellect Disabil Res* 2009; 53: 377–388