Title: Validity and Reliability of the International FItness Scale (IFIS) in preschool children.
Running head: Validity of parent-reported fitness in preschool.

Mairena Sánchez-López[a,b], Antonio García-Hermoso[c,d], Francisco B. Ortega[e,f], Diego Moliner-Urdiales[g], Idoia Labayen[h], Jose Castro-Piñero[i,j], Pedro J. Benito[k], Germán Vicente-Rodríguez[l], Joaquin Sanchis-Moysi[m], Jaume Cantallop[sn], Enrique G. Artero[o,p], Vicente Martínez-Vizcaíno[b,q,*.]

[a]Universidad de Castilla-La Mancha, School of Education, Ciudad Real, Spain.
[b]Universidad de Castilla-La Mancha, Health and Social Research Center, Cuenca, Spain.
[c]Navarrabiomed, IdiSNa, Pamplona, Navarra, Spain.
[d]Laboratorio de Ciencias de la Actividad Física, el Deporte y la Salud, Facultad de Ciencias Médicas, Universidad de Santiago de Chile, USACH, Santiago, Chile.
[e]The PROFITH Research Group, Department of Physical Education and Sports, Faculty of Sports Sciences, University of Granada, Granada, Spain.
[f]Department of Biosciences and Nutrition, Karolinska Institute, Huddinge, Sweden.
[g]LIFE research group. University Jaume I, Castellón, Spain.
hIS (IS-FOOD), Navarra's Health Research Institute (IdiSNA), Department of Health Sciences, Public University of Navarra, 31006 Pamplona, Navarra, Spain.
[i]GALENO research group, Department of Physical Education, School of Education, University of Cádiz, Puerto Real, Spain.
[j]Biomedical Research and Innovation Institute of Cádiz (INiBICA) Research Unit, Cádiz, Spain.
[k]LFE Research Group, Department of Health and Human Performance. Faculty of Physical Activity and Sport Science-INEF. Universidad Politécnica de Madrid.
[l]GENUD (Growth, Exercise, NUtrition and Development) research group, Instituto Agroalimentario de Aragón -IA2-, Centro de Investigación Biomédica en Red de Fisiopatología de la Obesidad y Nutrición (CIBERObn), Faculty of Health and Sport Science (FCSD), Department of Physiatry and Nursing, University of Zaragoza, Spain.
[m]Department of Physical Education and Research Institute of Biomedical and Health Sciences (IUIBS), University of Las Palmas de Gran Canaria, Canary Island, Spain.
[n]Grupo de Investigación en Ciencias de la Actividad Física y el Deporte (GICAFD). Departamento de Pedagogía y Didácticas Específicas (PDE). Facultad de Educación. Universitat de les Illes Balears (UIB).
[o]Department of Education, Faculty of Education Sciences, University of Almería, Almería, Spain.
[p]SPORT Research Group (CTS-1024), CERNEP Research Center, University of Almería, Almería, Spain.
[q]Universidad Autónoma de Chile, Faculty of Medicine, Talca, Chile.

***Corresponding author:**
Vicente Martínez Vizcaíno
C/ Santa Teresa Jornet, s/n, 16071, Cuenca, Spain
Phone number: +34 926 295 300; Fax: +34 969 179 178
E-mail: Vicente.martinez@uclm.es

This is an accepted manuscript of an article published by Taylor & Francis in European Journal of Sport Science on 02 May 2022, available online: https://doi.org/10.1080/17461391.2022.2049884

TITLE: VALIDITY AND RELIABILITY OF THE INTERNATIONAL FITNESS SCALE (IFIS) IN PRESCHOOL CHILDREN.

ABSTRACT

Objectives: Examine the validity and reliability of parent-reported International FItness Scale (IFIS) in preschool-age children.

Method: A cross-sectional study of 3051 Spanish preschoolers (3-5 years). Fitness was measured by PREFIT fitness test battery and reported by parents using an adapted version of the IFIS. Waist circumference was evaluated, and the waist-to-height ratio (WHtR) was calculated. Seventy-six parents of randomly selected schoolchildren completed the IFIS twice (two weeks apart) for a reliability assessment.

Results: ANCOVA, adjusted for sex, age and WHtR, showed that preschoolers who were scored by their parents as having average-to-very good fitness had better levels of measured physical fitness than those preschoolers who were classified as having "very poor/poor" fitness levels (18.1laps to 22.1laps vs 15.6laps for cardiorespiratory fitness; 6.6kg to 7.5kg vs 5.3kg for muscular fitness-handgrip-; 71.7cm to 76.4cm vs 62.0cm for muscular fitness-standing long jump-; 17.2s to 16.2s vs 18.2s for speed/agility; and 11.2s to 15.6s vs 8.7s for balance; $p<0.001$). The weighted kappa for concordance between parent-reported fitness levels and objective assessment was poor ($\kappa \leq 0.18$ for all fitness measures). Overall, the mean values of the abdominal adiposity indicators were significantly lower in high-level fitness categories reported by parents than in low-level fitness categories ($p<0.05$). The test-retest reliability of IFIS items ranged from 0.46 to 0.62.

Conclusions: The reliability of the parent-reported IFIS are acceptable, but the concordance between parents reported and objectively measures fitness levels is poor, suggesting that parents' responses may not be able to correctly classify preschoolers according to their fitness level.

Keywords

Abdominal obesity, preschoolers, physical fitness, parent report.

1. INTRODUCTION

Physical fitness is understood as the functional capability of body systems that allow performance of daily living activities and sports without effort according to age[1]. Good physical fitness level is considered an important marker of current and future health in youth.[1] In this regard, several studies have suggested that low levels of physical fitness in childhood are associated with an increased risk of cardiovascular disease and with musculoskeletal disorders and mental health problems in adulthood.[1–4] Some anthropometric and socio-demographic factors (such as adiposity, physical activity, age or gender) are associated with fitness in childhood[5] and throughout life[6,7], therefore these factors should be taken into account in studies examining children's fitness levels. Although studies focusing on preschool children (aged 3 to 5 years old) are scarce, research suggests that high levels of physical fitness at these early ages are associated with better body composition,[8–10] higher scores for cognitive functions[3,11,12] and, in general, higher health-related quality of life levels.[13]

Given the positive relationship between physical fitness and health at early ages[4,14,15], the assessment of physical fitness in preschoolers has become highly relevant from clinical, educational, and public health perspectives. However, the assessment of physical fitness is not always feasible in large population-based studies in which time, equipment, facilities, and qualified personnel are very often limited.

The International FItness Scale (IFIS), a short and simple scale available in nine different languages, including Spanish, was originally developed for its use in adolescents from nine European countries in the HELENA study. The IFIS provides a measure of fitness based on the answers to five basic questions about the perceived level of general physical fitness and in each fitness component (compared to friends), with answers based on the 5-point Likert-scale (from very poor=1 to very good= 5). This scale showed good validity and reliability in this population[16], as well as in a wide variety of populations, such as young adults,[17] older adults,[18] pregnant women,[19] women with fibromyalgia,[20] and children (aged 9-12 years)[21] from Spain and South America.[22,23] Moreover, fitness levels in children and adolescents using the IFIS have been shown to be strongly associated with adiposity and cardiovascular risk factors.[16,17,22]

However, to accurately complete a questionnaire, the child must have cognitively reached the level of abstract thinking and be able to conceptualise frequency.[24,25] This is not possible in children under 8 years of age[26]; thus, it seems necessary to ask parents. However, parental reports also have limitations, as parents may be more prone to social desirability bias than children, as has been described in studies on health habits[27].

Although researchers quantify validity and reliability in a variety of ways, criterion validity concerns the agreement between the observed value and the true or criterion value of a measure, and re-test reliability concerns the reproducibility of the observed value when the measurement is repeated; both have been considered the two most important aspects of measurement error in sports medicine and science[28]. In addition, convergent validity understood as the extent to which two measures of constructs that theoretically should be related are in fact related, may be another measure of the robustness of the results provided by the IFIS scale and enhance confidence that the construct is being captured.[29]

Therefore, the aim of the present study was to examine the following: 1) the ability of the IFIS, scored by parents, to accurately classify Spanish children aged 3-5 years according to their objectively measured fitness levels (i.e., criterion validity); 2) the associations of the parent-reported IFIS with abdominal adiposity in preschool children (i.e., convergent validity); and 3) the test-retest reliability of the parent-reported IFIS.

2. METHODS

2.1. Study design and participants

This study was conducted under the PREFIT project framework (http://profith.ugr.es/prefit). The main objective of this project was to assess physical fitness and anthropometric characteristics in preschoolers from 10 different cities across Spain. The data collection took place from January 2014 to November 2015. The study protocol was approved by the local Review Committee for Research Involving Human Subjects (n◦845), in accordance with the Declaration of Helsinki 1961 (and the 2013 revision)[19]. Parents or legal guardians of all children included in the study provided written informed consent, and children gave their verbal consent to participate.

A total of 4338 preschoolers and their parents were invited to participate in the PREFIT project. Finally, 3179 parents agreed to participate in the study (73.7% participation rate). No differences were found between the age, sex and anthropometric variables of children who agreed to participate and those who did not. Finally, parent-reported complete data from 3051 children (1,445 girls) were obtained.

For the reliability analysis, a subsample of 76 randomly recruited participants (45 girls and 31 boys) from a school in Granada city, not involved in the PREFIT study, was selected. They did not differ in age, sex, or anthropometric variables from children participating in the study.

The parents of these 76 participants successfully completed the IFIS twice (2 weeks apart). The questionnaires were sent to parents through their children in an open envelope. Once completed at home, parents were asked to put it in the envelope, closed it, and handed it to their child's teacher. After that, the teachers were responsible for sending the questionnaires to the members of the research team. The following instructions were sent to parents to answer the questionnaire: "Please mark with an X the option that best describes your child's fitness level (compared to his/her friends). Please answer all the questions and do not leave any blank. Mark only one answer per question".

2.2. Parent-reported fitness

Parent-reported fitness was assessed by the IFIS, which was originally validated in European adolescents.[16] The original IFIS consists of a five-item Likert-type scale with five response options: very poor (1), poor (2), average (3), good (4) and very good (5). Each item addresses a main self-perceived dimension of fitness (cardiorespiratory fitness, muscular fitness, speed-agility and flexibility), and one item addresses overall fitness (http://profith.ugr.es/IFIS). Taking into account a systematic review[30] showing that in preschoolers, flexibility is not associated with any health indicator and that balance may be a relevant component during earlier childhood, in the version of the IFIS for preschoolers, we decided to replace the item on flexibility with one on balance.

2.3. Objectively measured physical fitness

The physical fitness variables were measured in the schools by experienced researchers under standardized conditions using the PREFIT battery[30,31] as follows:

Cardiorespiratory fitness (CRF) was assessed using the adapted version of the preschoolers' 20 m shuttle run test.[31] Participants were required to run between two lines that were 20 m apart while keeping pace with audio signals emitted from a prerecorded CD. The initial speed was 6.5 kmh^{-1}, which was increased by 0.5 kmh^{-1} (1 min equals one stage). Children were encouraged to keep running as long as possible throughout the course of the test, and the test was finished when the child failed to reach the end lines concurrent with the audio signals on two consecutive occasions. The number of laps completed was recorded as an indicator of his or her CRF.

Muscular fitness (MF) was assessed using two tests: 1) the handgrip test (maximum handgrip strength assessment) using the analog version of a TKK dynamometer (TKK 5001, Grip-A, Takei, Tokyo, Japan) with the grip span fixed at

4.0 cm. The children squeezed gradually and continuously for at least 2-3 s, performing the test with the right and left hands in turn.[32] Children completed two trials (alternately with both hands) with a short rest period between them. The maximum score in kilograms for each hand was recorded, and the average (in kilograms) of both hands was used in the analysis; 2) the standing broad jump test (lower limb explosive strength assessment): from a starting position immediately behind a line, standing with feet approximately shoulder width apart, the schoolchildren jumped horizontally to achieve maximum distance. The best of three attempts was recorded in centimeters.

Speed/agility was measured using the 4x10 shuttle run test in which the child runs as fast as possible from the starting line to the line 10 m away and returns to the starting line, crossing each line with both feet every time. Two evaluators stood at each line, and the preschoolers had to touch the evaluator's hand and return to the starting line as fast as possible. Two attempts were made with an interval of at least five minutes, and only the best mark was used for analysis. The time taken to complete the test was recorded to the nearest tenth of a second. For analyses, this variable was multiplied by -1, as less time represents better results.

Static balance was assessed with the one-leg stance test. The test consisted of standing still on one-leg and bending the other leg at approximately 90°. The beginning of the test starts when one of the legs is no longer in contact with the floor. The children had to maintain the balance position for as long as they could. In accordance with the original protocol, there were no upper-limb movement restrictions. The test finished when the child could not continue in the required position. The children had one attempt with each leg, and the average time was registered in seconds.

2.4. Abdominal adiposity variables

Experienced trained nurses and sports science graduates conducted the waist circumference (WC) and height measurements under standardized conditions.

Waist circumference was calculated as the average of two measurements at the end of expiration at the middle point between the iliac crest and costal margin when the child was upright using a meter tape. Thereafter, the waist-to-height ratio was calculated.

2.5. Statistical analysis

Descriptive statistics included frequencies of each answer for the five questions on the IFIS by sex. The floor and ceiling effects of each item were evaluated by calculating the proportion of cases with minimum and maximum values, respectively.

Because of the small number of participants at the bottom extreme, the categories were merged as "very poor/poor" for the rest of the analyses, except for the reliability analyses, in which the raw data were used.

All objectively measured fitness components were categorized as low, medium, and high according to percentiles (<P25, P25-P75,> P75).[33]

Criterion validity. To examine the ability of the IFIS to categorize children correctly into physical fitness levels, we performed analysis of covariance (ANCOVA), controlling for sex, age, and waist-to-height ratio. Objectively measured fitness variables were entered as dependent variables, and parent-reported fitness variables were entered as fixed factors. In addition, ANCOVA models were also used to test differences in the mean scores for the z-score of each physical fitness component. In addition, to measure agreement between categories of parent-reported fitness levels (i.e., "very poor/poor", "average", "good", and "very good") and objective assessment (according to percentiles, i.e., <P25, P25-P50, P50-P75, >P75), a weighted kappa statistic[34] was used to measure concordance beyond chance.

Convergent validity. Convergent validity was tested using abdominal obesity indicators (WC and waist-to-height ratio) as criteria, since it is one of the main predictors of cardiometabolic risk and has a close relationship with measured physical fitness in children.[8,9] Thus, ANCOVA models controlling for sex and age were used to analyze the mean z-scores for WC and the waist-to-height ratio among categories of parent-reported fitness levels ("very poor/poor", "average", "good" and "very good").

In all ANCOVAs, pairwise posthoc hypotheses were tested using the Bonferroni correction for multiple comparisons.

Analysis of reliability. The test–retest reliability of the IFIS was examined by Cohen´s weighted Kappa (κ) coefficient.[34] Data for imputation into the syntax were generated from cross-tabulation. Weighted Kappa values can vary between -1 and 1. Agreement can be interpreted as follows: κ: < 0.20 = poor, κ: 0.21–0.40 = fair, κ: 0.41–0.60 = moderate, κ: 0.61–0.80 = good/substantial, and κ: 0.81–1.0 = very good/excellent.[35]

Analyses were performed in SPSS v. 25 (IBM Corp, Armonk, NY, USA), and the level of significance was set at $p < 0.05$.

3. RESULTS

Participants were 4.59±0.88 years, they have a mean BMI of 16.49±1.77 and their mean WC was 53.18±5.07 cm. Compared with girls, boys had higher values of/better performance in body weight, height, CRF, handgrip, standing broad jump, and speed-agility. In contrast, girls showed higher values of/better performance in WC, waist-to-height ratio, and balance. There were no differences in age and BMI (Table S1).

We observed a very low percentage (0.1 to 2.3%) of participants reporting having a "very poor/poor" fitness level. Approximately 60.0% of parents answered that their children have "good" fitness (Figure S1).

Criterion validity. Overall, compared with participants reporting "very poor/poor" fitness levels, participants reporting "average", "good", and "very good" CRF, MF, speed-agility and balance had better levels of CRF, MF, speed-agility and balance, respectively ($p<0.001$) (Table 1). Figure S2 shows a dose-response association between parent-reported and measured physical fitness. In addition, the mean z-scores of each measured physical fitness component were significantly higher in preschoolers with a higher parent-reported fitness level. The number of children correctly and incorrectly classified by each method is presented in table 2. The weighted kappa for the concordance between parent-reported and objective assessment was poor k=0.11 (95% confidence interval-CI-: 0.08 to 0.14) for cardiorespiratory fitness, k=0.13 (95% CI: 0.10 to 0.16) for handgrip strength, k=0.08 (95% CI: 0.05 to 0.10) for standing-long jump, k=0.17 (95% CI: 0.14 to 0.20) for speed-agility and k=0.18 (95% CI: 0.15 to 0.21) for balance. And the percentage of agreement ranged from 79.8% to 82.3%.

Convergent validity. Figure 1 shows the association of parent-reported fitness with WC (panel A) and the waist-to-height ratio (panel B), controlling for age and sex. Overall, the mean scores of abdominal adiposity variables were significantly higher ($p<0.05$) in those with lower parent-reported fitness, except for muscular fitness, which had higher mean values in preschoolers classified as "good" or "very good" ($p<0.001$).

Reliability. Table 3 displays the test–retest reliability statistics in children from Granada for the five items that compose the IFIS, i.e., overall fitness and the four main fitness components: CRF, MF, speed-agility, and balance. Weighted Kappa ranged from 0.46 (balance) to 0.62 (CRF), and the average weighted Kappa was 0.56.

4. DISCUSSION

Since fitness at early age predicts fitness levels through adolescence and adulthood[36,37], validating a short and easy-to-apply instrument seems to be a necessary task. To our knowledge, this is the first study to examine the validity and reliability of the parent-reported IFIS in children aged 3 to 5 years. These findings suggest that the reliability (test-retest) scores of the parent-reported IFIS are moderate. However, although the convergent validity values are acceptable, the concordance analysis show that criterion validity is poor, which suggest that parents' responses may not be able to correctly classify preschoolers according to their fitness level.

As in other studies in children and adolescents,[16,21] the distributions of responses to IFIS questions suggest a "ceiling effect" since a high percentage of parents reported that their children had "good" or "very good" fitness levels. This is not surprising considering that at an early age, health problems are unlikely to have appeared, and parents think that their children are healthy. In addition, it is also interesting that in this study, the highest percentage of responses was in the category of "good", while in a previous study in Spanish children aged 9-12 years[21], the highest percentage of responses was in the "very good" category, which suggests that children tend to overestimate their fitness relative to parental perception. However, more studies are necessary to examine this issue in depth.

Given the low number of parents who indicated "very poor" levels of physical fitness (0.1%), the IFIS does not allow the identification of preschoolers with very poor fitness levels. It is likely hard for parents to admit that their children have poor fitness, perhaps due to a social desirability bias[38] since when they rate their children's fitness level as very low, they feel that indirectly, they are recognizing that they are not doing enough to improve it. Although parents answered the questionnaire confidentially, it is likely that they felt the risk of being identified and judged. On the other hand, parents were informed that they were participating in a study on the importance of physical fitness in childhood, so it seems logical that in their response's fitness levels were overestimated and this could be the reason why only a small percentage of parents marked the "very poor" option. Also, parents may not be fully aware of their children's fitness level, probably due to a lack of knowledge about what optimal or poor fitness means.

4.1. Validity and reliability of the International FItness Scale

Consistent with previous studies[17,21,23] and with the original validation study of the IFIS,[11] in the current study, it is observed acceptable agreement between parent-reported and measured fitness in preschoolers in the "average", "good" and "very good" categories using ANCOVA. However, the parent-report IFIS was not a valid tool to detect those preschoolers who had a low or very low level of fitness. Since a low fitness level is not recognized by parents, it seems necessary to calibrate the scale in future research. A potential strategy to do this could be to reword the response options into the following categories: Very poor/poor (1), Average (2), and Good (3). In addition, special attention should be given to ensure confidentiality and that parents have the knowledge to discriminate among fitness levels of their children, and not to give out information about the researchers' stance on fitness status in children.

Three arguments can be put forward to explain the low agreement the observed categories of fitness levels reported by parents and the objective assessment (concordance analysis): first, the categorization of the objective assessment by quartiles, without considering cut-offs according to clinical criteria could misclassified a non-negligible percentage of individuals. Therefore, the concordance would be higher than in other samples where parents would not report poor fitness levels, but more children would be classified as $p<25$ in measured fitness and in the same vein in other categories; second, the high homogeneity of the sample in terms of their fitness levels, as can be seen in table 1, where the ranges of the mean +/- SD intervals of the categories overlap to a large extent, makes it difficult for parents to discriminate among the different categories of fitness; finally, the large number of response options could be another factor that makes it difficult for parents to correctly classify their children, so a smaller number of response options would help parents to identify the physical condition of their children.

In line with previous studies,[17,21,23] which have reported strong associations of the IFIS with adiposity and cardiovascular risk factors. These results show that abdominal adiposity is higher in those preschoolers with "very poor/poor" parent-reported fitness levels (CRF, speed/agility, balance, and overall fitness) than in those participants with "good/very good" fitness. These findings suggest that the IFIS scale has acceptable convergent validity for assessing physical fitness in this age group which makes the scale more robust.

In the present study, abdominal obesity was lower in preschoolers with "very poor/poor" parent-reported MF than in preschoolers with "good/very good" MF. However, when WC is expressed relative to height (i.e., as the waist-to-height ratio), this association disappears. As in previous studies,[16,17,21] these results might suggest that when parents answer this item on the scale, they are thinking of absolute strength. Several studies observed that children and adolescents with overweight/obesity scored higher on tests requiring strength without involvement of body weight.[39,40] Future researchers should consider the direct association between parent-reported MF and abdominal adiposity found in this study to properly interpret their results.

The test-retest reliability of IFIS items ranged from 0.46 to 0.62 (average weighted Kappa= 0.56 for a two-week interval), which can be considered "moderate" to "good" agreement, supporting the reliability of the scale in preschoolers.[35] Therefore, these findings suggest that this tool could provide similar measures in the same individuals at two different points in time, i.e. it has acceptable replicability, showing that it is slightly affected by memory biases, social desirability and learning biases that could have been sources of variation when parents filled the questionnaires. The reliability of the scale was similar to that of the original version of the IFIS (averaged weighted Kappa = 0.58)[16] but lower than that shown in other reliability studies in older children and adolescents.[21–23]

4.2. Limitations and strengths

The present study is of interest for public health since it provides a useful tool to assess physical fitness at a critical stage of life, when it is not possible to objectively evaluate it or when children have difficulties performing the tests correctly due to their level of cognitive and motor development. However, there are some limitations that should be highlighted: 1) the sample included preschool children from a single country, and it is unknown whether this scale would be appropriate for preschoolers from other countries with different characteristics; 2) children´s physical fitness was evaluated by parent reports rather than by self-reports by the preschoolers. This fact may have affected the results since previous studies have shown low agreement between child self-reports and parent proxy reports when measuring health related behaviours[41,42]. Thus, it is debatable whether parents should answer about their children's fitness. Nevertheless, taking into account the cognitive level of children aged 3 to 5 years, it seems necessary to validate a questionnaire answered by parents when it is not possible to assess the level of fitness objectively; 3) convergent validity was tested using indirect measurements (i.e., WC and waist-to-height ratio), and therefore, seem to be necessary more sophisticated modelling to remove the influence of body mass and adiposity. Furthermore, other factors not assessed in this study, such as physical activity or energy intake, may have influenced the results; 4) although some criticisms about the validity and reliability of the 20 m shuttle run test for estimating aerobic capacity because of it is influenced by the leg and stride length, it is also true that it is most suitable field test for estimating CRF in epidemiological population-based studies, as evidenced that this test has been used in more than 177 studies, accumulating more than 1 million children and adolescents[43]. Léger et al. (1988) also developed an equation to indirectly estimate the maximal oxygen consumption (VO2max) from the 20 m shuttle run test-Original[44]. In this study we evaluated CRF using an adapted version of the 20 m shuttle run test, which has been suggested d to be valid and reliable to assess CRF in children under 6 years of age[45,46]; 5) the time interval between the two repeated measures for reliability analysis represents a debatable issue; an interval of two weeks was selected considering the previous literature of similar studies,[47] and also taking into account that it is sufficient for individuals not to remember their first responses and for physical fitness not to have changed, both conditions that must be considered in test-retest reliability studies; and finally, although handgrip strength has known limitations to assess the strength as a single test, is considered as a practical, feasible and scalable functional measure of general strength for clinical and population-based screening and surveillance;[48]

In conclusion, the results of this study suggest that the reliability (test-retest) scores of the parent-reported IFIS are moderately acceptable. However, the agreement between IFIS questionnaire and objectively measured fitness is low, suggesting that parents' perceptions do not seem correctly classify preschoolers on their fitness level.

Practical implications

- The convergent validity and reliability (test-retest) values of the IFIS parent scale are moderately acceptable for assessing physical fitness in children aged 3-5 years.
- However, the results of concordance show that criterion validity is poor suggesting that parents' responses may not be able to correctly classify preschoolers according to their fitness level.
- Considering that the fitness level at these ages is fairly homogeneous, it seems difficult for parents to discriminate between the fitness levels of their children. Therefore, it seems necessary to recalibrate the scale in future work.

Competing interests

The authors declare they have no competing interest.

Funding information

The PREFIT project takes place thanks to the funding linked to the Ramón y Cajal grant held by Ortega FB (RYC-2011-09011).

Acknowledgments

Summit in the manuscript version with author details.

References

1 Ortega F, Ruiz J, Castillo M, et al. Physical fitness in childhood and adolescence: a powerful marker of health. *Int J Obes* 2008; 32(1):1–11. Doi: 10.1038/sj.ijo.0803774.

2 Ruiz J, Castro-Piñero J, Artero E, et al. Predictive validity of health-related fitness in youth: a systematic review. *Br J Sports Med* 2009; 43(12):909–923. Doi: 10.1136/bjsm.2008.056499.

3 Lang JJ, Belanger K, Poitras V, et al. Systematic review of the relationship between 20 m shuttle run performance and health indicators among children and youth. *J Sci Med Sport* 2018; 21(4):383–397. Doi: 10.1016/j.jsams.2017.08.002.

4 García-Hermoso A, Ramírez-Campillo R, Izquierdo M. Is Muscular Fitness Associated with Future Health Benefits in Children and Adolescents? A Systematic Review and Meta-Analysis of Longitudinal Studies. *Sport Med* 2019; 49(7):1079–1094. Doi: 10.1007/s40279-019-01098-6.

5. Magnússon KT, Sveinsson T, Arngrímsson SA, et al. Predictors of fatness and physical fitness in nine-year-old Icelandic school children. *Int J Pediatr Obes* 2008;3(4):217–25. Doi: 10.1080/17477160802169482.

6. Augste C, Lämmle L, Künzell S. Does current behaviour predict the course of children's physical fitness? *Eur J Sport Sci*. 2015; 15(5): 429–435. Doi: 10.1080/17461391.2014.948076

7. Lämmle L, Worth A, Bös K. Socio-demographic correlates of physical activity and physical fitness in German children and adolescents. *Eur J Public Health*. 2012;22(6):880–4. Doi: 10.1093/eurpub/ckr191.

8 Henriksson P, Cadenas-Sanchez C, Leppänen M, et al. Associations of Fat Mass and Fat-Free Mass with Physical Fitness in 4-Year-Old Children: Results from the MINISTOP Trial. *Nutrients* 2016; 8(8):473. Doi: 10.3390/nu8080473.

9 Martinez-Tellez B, Sanchez-Delgado G, Cadenas-Sanchez C, et al. Health-related physical fitness is associated with total and central body fat in preschool children aged 3 to 5 years. *Pediatr Obes* 2016; 11(6):468–474. Doi: 10.1111/ijpo.12088.

10 Niederer I, Kriemler S, Zahner L, et al. BMI Group-Related Differences in Physical Fitness and Physical Activity in Preschool-Age Children. *Res Q Exerc Sport* 2012; 83(1):12–19. Doi: 10.1080/02701367.2012.10599820.

11 Nieto-López M, Sánchez-López M, Visier-Alfonso ME, et al. Relation between physical fitness and executive function variables in a preschool sample. *Pediatr Res* 2020. Doi: 10.1038/s41390-020-0791-z.

12 Latorre-Román PÁ, Mora-López D, García-Pinillos F. Intellectual maturity and physical fitness in preschool children. *Pediatr Int* 2016; 58(6):450–455. Doi: 10.1111/ped.12898.

13 Redondo-Tébar A, Ruíz-Hermosa A, Martínez-Vizcaíno V, et al. Associations between health-related quality of life and physical fitness in 4–7-year-old Spanish children: the MOVIKIDS study. *Qual Life Res* 2019; 28(7):1751–1759. Doi: 10.1007/s11136-019-02136-6.

14 García-Hermoso A, Ramírez-Vélez R, García-Alonso Y, et al. Association of Cardiorespiratory Fitness Levels During Youth With Health Risk Later in Life: A Systematic Review and Meta-analysis. *JAMA Pediatr* 2020; 174(10):952–960. Doi: 10.1001/jamapediatrics.2020.2400.

15 Mintjens S, Menting MD, Daams JG, et al. Cardiorespiratory Fitness in Childhood and Adolescence Affects Future Cardiovascular Risk Factors: A Systematic Review of Longitudinal Studies. *Sports Med* 2018; 48(11):2577–2605. Doi: 10.1007/s40279-018-0974-5.

16 Ortega FB, Ruiz JR, España-Romero V, et al. The International Fitness Scale (IFIS): usefulness of self-reported fitness in youth. *Int J Epidemiol* 2011; 40(3):701–711. Doi: 10.1093/ije/dyr039.

17 Ortega FB, Sánchez-López M, Solera-Martínez M, et al. Self-reported and measured cardiorespiratory fitness similarly predict cardiovascular disease risk in young adults. *Scand J Med Sci Sports* 2013; 23(6):749–757. Doi: 10.1111/j.1600-0838.2012.01454.x.

18 Merellano-Navarro E, Collado-Mateo D, García-Rubio J, et al. Validity of the International Fitness Scale “IFIS” in older adults. *Exp Gerontol* 2017; 95:77–81. Doi: 10.1016/j.exger.2017.05.001.

19 Romero-Gallardo L, Soriano-Maldonado A, Ocón-Hernández O, et al. International Fitness Scale—IFIS: Validity and association with health-related quality of life in pregnant women. *Scand J Med Sci Sports* 2020; 30(3):505–514. Doi: 10.1111/sms.13584.

20 Álvarez-Gallardo IC, Soriano-Maldonado A, Segura-Jiménez V, et al. International FItness Scale (IFIS): Construct Validity and Reliability in Women With Fibromyalgia: The al-Ándalus Project. *Arch Phys Med Rehabil* 2016; 97(3):395–404. Doi: 10.1016/j.apmr.2015.08.416.

21 Sánchez-López M, Martínez-Vizcaíno V, García-Hermoso A, et al. Construct validity and test-retest reliability of the International Fitness Scale (IFIS) in Spanish children aged 9-12 years. *Scand J Med Sci Sports* 2015; 25(4):543–551. Doi: 10.1111/sms.12267.

22 De Moraes ACF, Vilanova-Campelo RC, Torres-Leal FL, et al. Is Self-Reported Physical Fitness Useful for Estimating Fitness Levels in Children and Adolescents? A Reliability and Validity Study. *Medicina (B Aires)* 2019; 55(6):286. Doi: 10.3390/medicina55060286.

23 Ramírez-Vélez R, Cruz-Salazar SM, Martínez M, et al. Construct validity and test–retest reliability of the International Fitness Scale (IFIS) in Colombian children and adolescents aged 9–17.9 years: the FUPRECOL study. *PeerJ* 2017; 5:e3351. Doi: 10.7717/peerj.3351.

24 Burrows TL, Martin RJ, Collins CE. A systematic review of the validity of dietary assessment methods in children when compared with the method of doubly labeled water. *J Am Diet Assoc* 2010; 110(10):1501–10. Doi: 10.1016/j.jada.2010.07.008

25 Mindell J, Coombs N, Stamatakis E. Measuring physical activity in children and adolescents for dietary surveys: Practicalities, problems and pitfalls. *Proceedings of the Nutrition Society. Cambridge University Press*; 73(2), 218–225. Doi:10.1017/S0029665113003820

26 Livingstone MB, Robson PJ. Measurement of dietary intake in children. *Proc Nutr Soc* 2000; 59(2):279–93. Doi: 10.1017/s0029665100000318

27 De Bourdeaudhuij I, Van Oost P. Personal and family determinants of dietary behaviour in adolescents and their parents 2000 *Psychol Health*; 15:6, 751–770. Doi: 10.1080/08870440008405579.

28 Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med* 2000; 30(1):1–15. Doi: 10.2165/00007256-200030010-00001.

29 Kevin DC, Andrew OH. Understanding the Impact of Convergent Validity on

Research Results. *Organizational Research Methods* 2012; 15(1) 17–32. Doi: 10.1177/1094428110392383

30 Ortega FB, Cadenas-Sánchez C, Sánchez-Delgado G, et al. Systematic Review and Proposal of a Field-Based Physical Fitness-Test Battery in Preschool Children: The PREFIT Battery. *Sport Med* 2015; 45(4):533–555. Doi: 10.1007/s40279-014-0281-8.

31 Cadenas-Sanchez C, Martinez-Tellez B, Sanchez-Delgado G, et al. Assessing physical fitness in preschool children: Feasibility, reliability and practical recommendations for the PREFIT battery. *J Sci Med Sport* 2016; 19(11):910–915. Doi: 10.1016/j.jsams.2016.02.003.

32 Sanchez-Delgado G, Cadenas-Sanchez C, Mora-Gonzalez J, et al. Assessment of handgrip strength in preschool children aged 3 to 5 years. *J Hand Surg (European Vol* 2015; 40(9):966–972. Doi: 10.1177/1753193415592328.

33 Cadenas-Sanchez C, Intemann T, Labayen I, et al. Physical fitness reference standards for preschool children: The PREFIT project. *J Sci Med Sport* 2019; 22(4):430–437. Doi: 10.1016/j.jsams.2018.09.227.

34 Cohen J. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychol Bull* 1968; 70(4):213–220. Doi: 10.1037/h0026256.

35 Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 1977; 33(1):159. Doi: 10.2307/2529310.

36 Janz KF, Dawson JD, Mahoney LT. Tracking physical fitness and physical activity from childhood to adolescence: the muscatine study. *Med Sci Sports Exerc* 2000; 32(7):1250–7. Doi: 10.1097/00005768-200007000-00011.

37 Blasquez Shigaki G, Barbosa CC, Batista MB, et al. Tracking of health-related physical fitness between childhood and adulthood. *Am J Hum Biol* 2020; 32(4):e23381. Doi: 10.1002/ajhb.23381

38 Kristiansen CM, Harding CM. The social desirability of preventive health behavior. *Public Health Rep* 1984;99(4):384–8.

39 Artero EG, España-Romero V, Ortega FB, et al. Health-related fitness in adolescents: underweight, and not only overweight, as an influencing factor. The AVENA study. *Scand J Med Sci Sports* 2009; 20(3):418–427. Doi: 10.1111/j.1600-

0838.2009.00959.x.

40 Gulías-González R, Martínez-Vizcaíno V, García-Prieto JC, et al. Excess of weight, but not underweight, is associated with poor physical fitness in children and adolescents from Castilla-La Mancha, Spain. *Eur J Pediatr* 2014; 173(6):727–735. Doi: 10.1007/s00431-013-2233-y.

41 Koning M, de Jong A, de Jong E, et al. Agreement between parent and child report of physical activity, sedentary and dietary behaviours in 9-12-year-old children and associations with children's weight status. *BMC Psychol*. 2018; 6(1):14. Doi: 10.1186/s40359-018-0227-2.

42 Rebholz CE, Chinapaw MJ, van Stralen MM, et al. Agreement between parent and child report on parental practices regarding dietary, physical activity and sedentary behaviours: the ENERGY cross-sectional survey. *BMC Public Health* 2014; 14:918. Doi: 10.1186/1471-2458-14-918.

43 Lang JJ, Tremblay MS, Ortega FB, et al. Review of criterion-referenced standards for cardiorespiratory fitness: what percentage of 1 142 026 international children and youth are apparently healthy? *Br J Sports Med* 2019; 53(15):953–958. Doi: 10.1136/bjsports-2016-096955.

44 Léger LA, Mercier D, Gadoury C, et al. The multistage 20 metre shuttle run test for aerobic fitness. *J Sports Sci* 1988; 6(2):93–101. Doi: 10.1080/02640418808729800.

45 Cadenas-Sánchez C, Alcántara-Moral F, Sánchez-Delgado G, et al. Assessment of cardiorespiratory fitness in preschool children: adaptation of the 20 metres shuttle run test. *Nutr Hosp* 2014 ;30(6):1333–43. Doi: 10.3305/nh.2014.30.6.7859.

46 Mora-Gonzalez J, Cadenas-Sanchez C, Martinez-Tellez B, et al. Estimating VO2max in children aged 5-6 years through the preschool-adapted 20-m shuttle-run test (PREFIT). *Eur J Appl Physiol* 2017; 117(11):2295–2307. Doi: 10.1007/s00421-017-3717-7.

47 Artero EG, España-Romero V, Castro-Piñero J, et al. Reliability of field-based fitness tests in youth. *Int J Sports Med* 2011; 32(3):159–69. Doi: 10.1055/s-0030-1268488.

48 Milliken LA, Faigenbaum AD, Loud RL, et al. Correlates of upper and lower body muscular strength in children. *J Strength Cond Res* 2008; 22(4):1339–46. Doi: 10.1519/JSC.0b013e31817393b1.

Table 1. Means and standard deviation (SD) of measured physical fitness by self-reported physical fitness categories in preschool children.

	Very Poor/Poor (1)		Average (2)		Good (3)		Very good (4)		P*	Pairwise comparisons[†]					
										1-2	1-3	1-4	2-3	2-4	3-4
	Mean	SD	Mean	SD	Mean	SD	Mean	SD							
Cardiorespiratory fitness	*n*=73		*n*=814		*n*=1635		*n*=535								
20-m shuttle run (laps) (*n*=3006; 48% girls)	15.6	9.40	18.1	8.56	20.3	8.09	22.1	9.25	<0.001	ns	<	<	<	<	<
Muscular fitness	*n*=41		*n*=680		*n*=1712		*n*=624								
Handgrip (kg) (*n*=3051; 49% girls)	5.3	1.92	6.6	2.61	7.0	4.14	7.5	2.50	<0.001	<	<	<	<	<	<
Standing long jump (cm) (*n*=3041; 49% girls)	62.0	16.65	71.5	15.65	73.5	16.55	76.4	17.48	<0.001	<	<	<	ns	<	<
Speed-Agility	*n*=54		*n*=746		*n*=1619		*n*=632								
Shuttle run 4 x 10 m (s)[‡] (*n*=3025; 50% girls)	18.2	1.47	17.2	2.73	16.5	4.02	16.2	2.51	<0.001	>	>	>	>	>	>
Balance	*n*=57		*n*=897		*n*=1680		*n*=420								
Standing on one-leg (s) (*n*=3039; 49% girls)	8.7	15.10	11.2	14.97	14.4	16.39	15.6	14.34	<0.001	ns	<	<	<	<	ns

*Analysis of covariance adjusted for sex, age, and waist-to-height ratio. [†]Bonferroni-adjusted pairwise comparisons: the symbol < in the column 1-2, for instance, indicates a significant difference (P<0.05) in the direction 1<2; ns, non-significant.

[‡] The lower the score (time in seconds) the better the performance.

Table 2. Number of agreements between parent-report physical fitness categories and objective physical fitness percentiles in preschool children

	<P25	P25-P50	P50-P75	>P75	Total	Kappa (95% CI)
Cardiorespiratory fitness						0.11 (0.08-0.14)
Very poor/poor	25	21	20	5	71	
Average	241	212	195	150	798	
Good	396	370	435	420	1621	
Very good	97	125	129	165	516	
Total	759	728	779	740	3,006	
Handgrip strength						0.13 (0.10-0.16)
Very poor/poor	27	10	3	1	41	
Average	227	169	162	122	680	
Good	426	422	426	438	1712	
Very good	126	126	151	215	618	
Total	806	727	742	776	3,051	
Standing-long jump						0.08 (0.05-0.10)
Very poor/poor	15	16	8	2	41	
Average	201	180	156	143	680	
Good	428	402	431	435	1696	
Very good	140	148	135	201	624	
Total	784	746	730	781	3,041	
Speed-agility						0.17 (0.14-0.20)
Very poor/poor	29	15	6	4	54	
Average	267	207	173	99	746	
Good	378	414	395	432	1619	
Very good	125	124	132	225	606	
Total	799	760	706	760	3,025	
Balance						0.18 (0.15- 0.21)
Very poor/poor	127	14	10	6	157	
Average	285	241	212	159	897	
Good	291	394	429	463	1577	
Very good	90	108	98	124	420	
Total	793	757	749	752	3,051	

Black indicates perfect agreement; grey indicates a difference in ≤1 category and white a difference > 1 category.

CI= Confidence interval

Table 3. Test-retest (2 weeks apart) reliability of parent-reported fitness measured in a sub-sample of Granada (n = 76; 59.2% girls)

IFIS items	Weighted Kappa coefficients	95% CI
Cardiorespiratory fitness	0.62	0.56 – 0.66
Muscular fitness	0.57	0.54 – 0.62
Speed-agility	0.55	0.52 – 0.60
Balance	0.46	0.43 – 0.51
Overall fitness	0.60	0.55 – 0.63
Average Kappa	0.56	0.52 – 0.60

IFIS, International Fitness Scale; CI, confidence interval

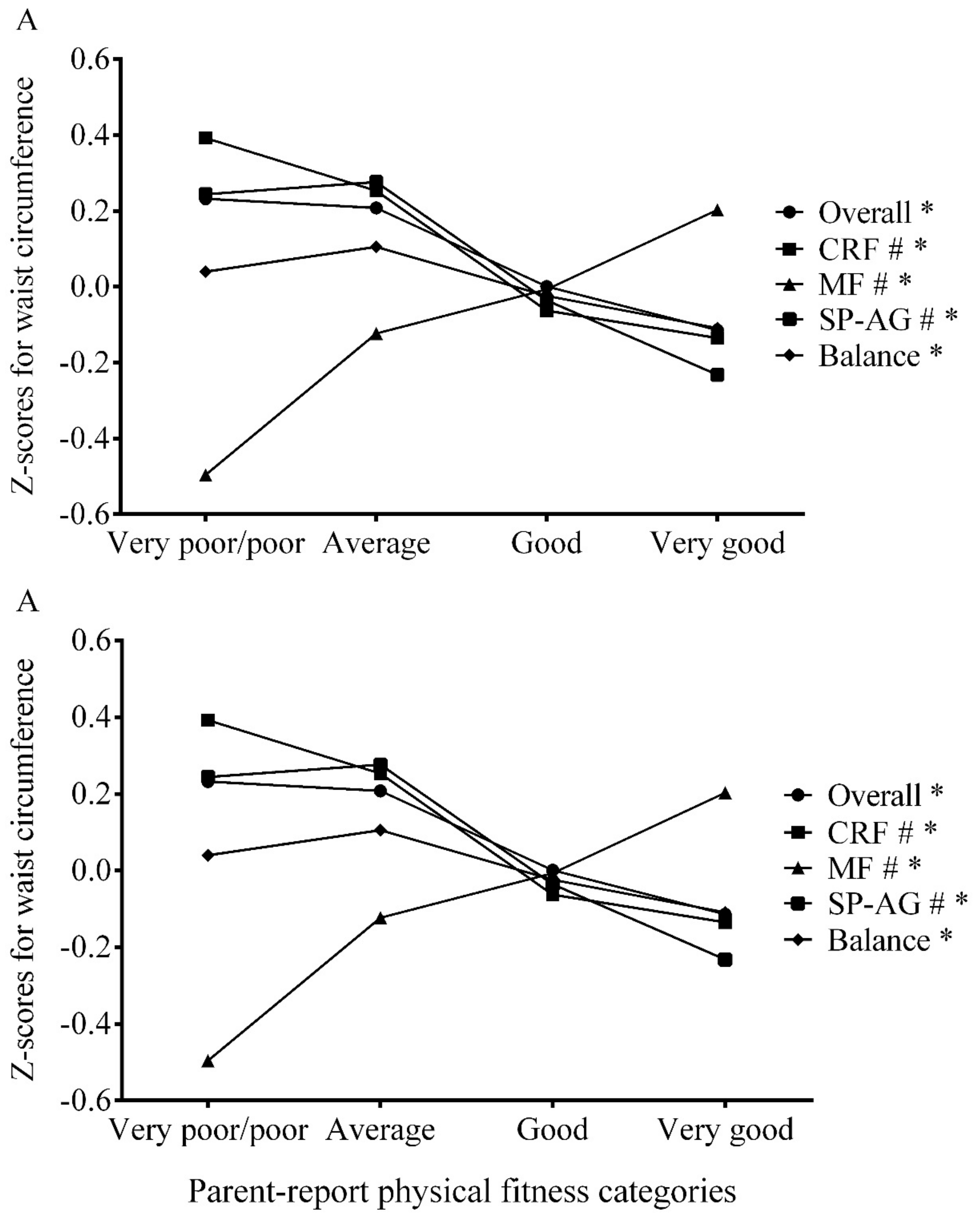

Figure 1. Means of z-score values for waist circumference (A) and waist-to-height-ratio (B) by self-reported physical fitness categories in preschool children. * P<0.05 between "Very poor/poor" *vs* "Good" and "Very good"; # P<0.05 between "Average" *vs* "Good" and "Very good". All z-scores were sex and age specifically computed.

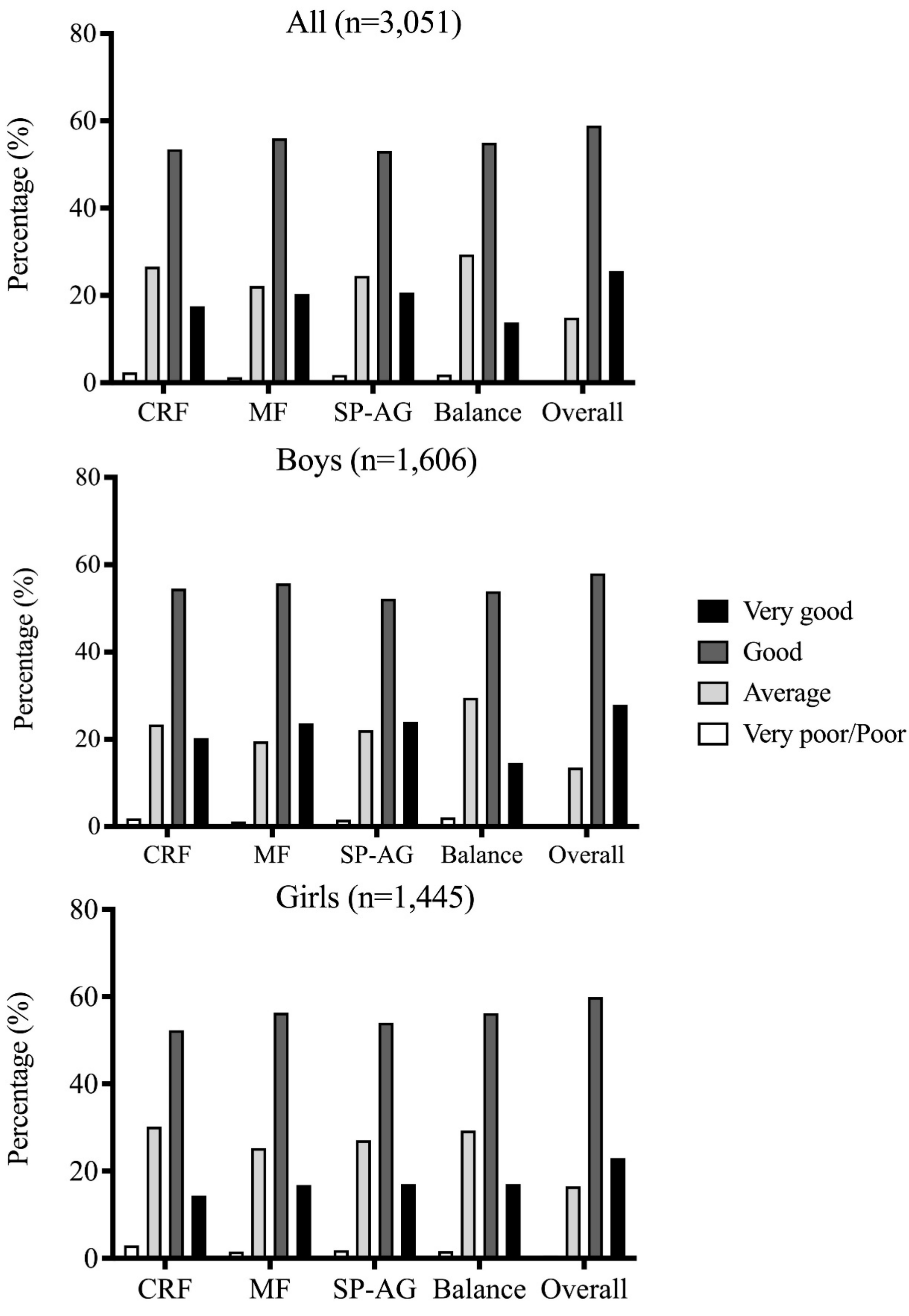

Figure S1. Distribution of the answers for the 5 questions of parent-report IFIS (International Fitness Scale) in boys and girls preschool children. CRF, cardiorespiratory fitness; MF, muscular fitness; SP-AG, speed-agility; Overall, overall physical fitness.

Table S1. Characteristics of the sample

	All (n=3,051)	Boys (n=1,606)	Girls (n=1,445)	p
Age, years	4.59±0.88	4.59±0.88	4.58±0.88	0.753
Body weight, kg	18.99±3.78	19.17±3.84	18.78±3.63	0.003
Height, cm	106.90±7.54	107.44±7.56	106.37±7.42	<0.001
Body mass index, kg/m^2	16.49±1.77	16.49±1.78	16.48±1.77	0.904
Waist circumference, cm	53.18±5.07	52.98±5.00	53.39±5.13	0.025
Waist-to-height ratio	0.50±0.04	0.49±0.04	0.50±0.04	<0.001
Cardiorespiratory fitness, laps	19.92±11.65	21.49±12.38	18.18±10.52	<0.001
Handgrip, kg	7.01±2.49	7.35±2.58	6.63±2.33	<0.001
Standing broad jump, cm	73.62±22.34	77.00±22.15	69.87±21.94	<0.001
Speed-agility, seconds	16.83±2.54	16.52±2.45	17.18±2.59	<0.001
Balance, seconds	13.64±16.96	12.72±16.58	14.66±17.33	0.002

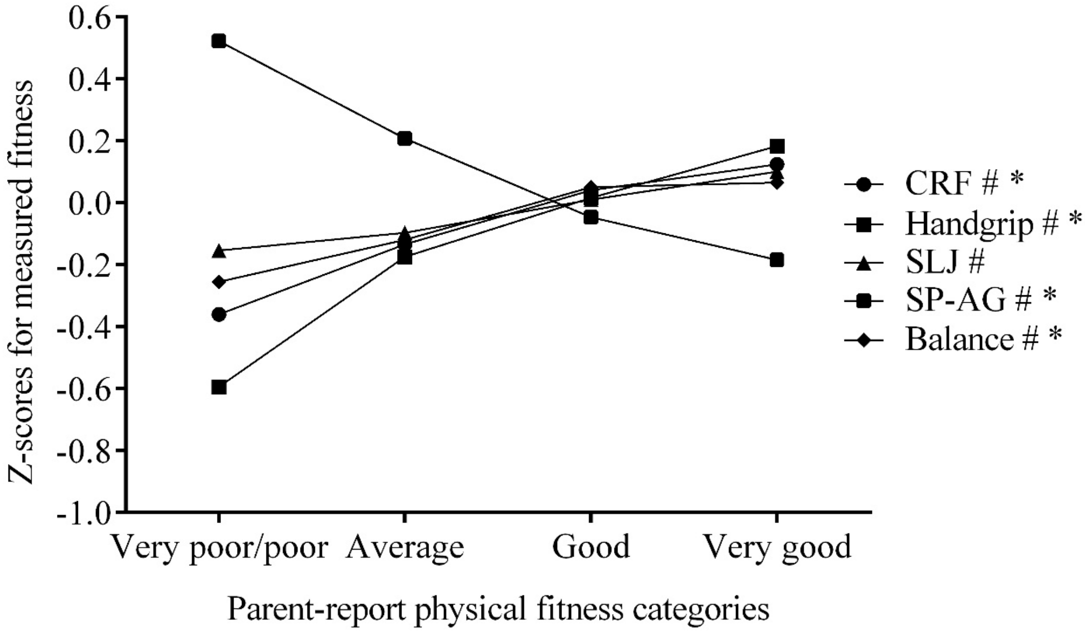

Figure S2. Means of z-score values for measured physical fitness of the children by parent-reported physical fitness categories. CRF, cardiorespiratory fitness; SLJ, standing-long jump; SP-AG, speed-agility. * P<0.05 between "Very poor/poor" *vs* "Good" and "Very good"; # P<0.05 between "Average" *vs* "Good" and "Very good". All z-scores were sex and age specifically computed.