

Analizando la validación del International Physical Activity Questionnaire en jóvenes adolescentes: Un protocolo modificado para la recogida de los datos

Testing the validity of the International Physical Activity Questionnaire in early Spanish adolescents: A modified protocol for data collection

Alberto Aibar¹, Luis García-González¹, Alberto Abarca-Sos², Berta Murillo³ y Javier Zaragoza¹

¹University of Zaragoza (Huesca, Spain) ²University of Zaragoza (Teruel, Spain) ³University of San Jorge (Zaragoza, Spain)

Resumen: La versión corta del International Physical Activity Questionnaire (IPAQ-SF) ha sido recomendada como un método eficiente para evaluar la actividad física (AF), pero su validez no ha sido analizada en adolescentes españoles. Este artículo, el cual está compuesto de dos estudios, pretende en primer lugar testar la validez del IPAQ-SF en una muestra de adolescentes españoles, y en un segundo lugar testar y comparar la validez del IPAQ-SF utilizando una versión modificada del protocolo adaptada a adolescentes.

Dado que este trabajo se divide en dos estudios, se utilizaron dos muestras independientes. Una muestra de 270 adolescentes en el estudio I (55% chicos), y otra de 151 adolescentes en el estudio II (44.6% chicos), con una edad entre 11 y 14 años de seis institutos de Huesca (España). Los participantes llevaron un acelerómetro durante siete días consecutivos y completaron el IPAQ-SF el último día. Se calculó el coeficiente de correlación de Pearson entre ambos métodos. Se utilizó el método Bland-Altman para proporcionar el error total entre los dos métodos en ambos estudios.

Se encontraron correlaciones significativas entre el IPAQ-SF y el acelerómetro para la AF moderada, vigorosa, y moderada-vigorosa. La AF moderada-vigorosa fue significativamente sobreestimada por el IPAQ-SF. Los gráficos Bland-Altman mostraron que los resultados del IPAQ-SF realizado con el protocolo modificado obtuvieron mejores niveles de acuerdo para la AF vigorosa y moderada-vigorosa.

El IPAQ-SF puede ser una herramienta válida para medir la AF en jóvenes adolescentes, especialmente cuando se aplican modificaciones metodológicas en el protocolo de recogida de datos.

Palabras clave: Acelerometría, Evaluación, Cuestionario, Actividad Física, Validación.

Abstract: The short version of the International Physical Activity Questionnaire (IPAQ-SF) has been recommended as an efficient method to assess physical activity (PA), but its validity has not been investigated in Spanish adolescents. This article, which is comprised of two studies, pursues firstly to test the validity of the IPAQ-SF in a sample of Spanish adolescents, and secondly to test and compare the validity of the IPAQ-SF using a modified protocol version specifically adapted to adolescents.

Given that this work is divided into two studies, two independent samples were used. One sample of 270 adolescents for Study One (55% males), and another sample of 151 adolescents for Study Two (44.6% males), aged 11–14 years old, were recruited from six secondary schools in Huesca (Spain). All subjects wore an accelerometer for seven consecutive days and completed the IPAQ-SF on the last day. Pearson correlation coefficients were calculated between both methods. The Bland and Altman method was also used to provide the total error between the two methods in both studies.

Significant correlations between IPAQ-SF and accelerometer data were found for time spent on moderate, vigorous and moderate-to-vigorous PA. Moderate-to-vigorous PA levels were significantly overestimated by IPAQ-SF. Bland Altman plots showed that results from the IPAQ-SF, conducted with a modified adapted protocol for early adolescents, obtained much better levels of agreement for vigorous and moderate-to-vigorous PA intensities. The IPAQ-SF may be a valid tool to measure PA levels in early adolescents, especially when methodological modifications are applied in the data collection protocol.

Keywords: Accelerometry, Assessment, Questionnaire, Physical Activity, Validity.

Introduction

Regular physical activity (PA) has consistently been associated with physical health benefits (Borrego-Balsalobre, López-Sánchez y Díaz-Suárez, 2015a; Borrego-Balsalobre, López-Sánchez y Díaz-Suárez, 2015b; Borrego-Balsalobre, López-Sánchez y Díaz-Suárez, 2015c; López-Sánchez, Borrego-Balsalobre y Díaz-Suárez, 2013). Comparing global surveillance and international physical activity (PA) levels between different populations is difficult due to the use of different measurement techniques (Montoye, Cunningham,

Hugh y Epstein, 2000). The International Physical Activity Questionnaire (IPAQ) is a self-report measure that has been developed to estimate regular PA levels across different countries in the world and in different socio-cultural environments (Craig et al., 2003).

The short version of IPAQ (IPAQ-SF) was developed to be used by public health institutions to monitor entire worldwide populations. Its reliability and validity has initially been tested on 18 to 65 year-old adults in diverse settings in more than a dozen of countries (Craig et al., 2003). Later on, several studies replicated its validation in other countries (Hallal y Victora, 2004; Qu y Li, 2004) and/or other populations, such as adolescents (Guedes, Lopes y Guedes, 2005; Lachat et al., 2008; Rangul, Holmen, Kurtze, Cuypers y Midthjell, 2008).

Dirección para correspondencia [Correspondence address]: Alberto Aibar Solana. Universidad de Zaragoza. Faculty of Social Sciences and Education. Departamento de Expresión Musical, Plástica y Corporal. Calle Valentín Carderera, nº 4. 22003 Huesca (Spain). E-mail: aibar@unizar.es

However, it should be pointed out that the IPAQ-SF has still not been validated in a Spanish adolescent population.

It should be known that despite the fact that the IPAQ-SF is widely used in current international research (Medina, Janssen, Campos y Barquera, 2013; Rosenberg, Bull, Marshall, Sallis y Bauman, 2008) due, in part, to the general inherent positive features linked to self-report questionnaires (Kemper, Montoye, Saris y Washburn, 1996), such as its accessibility and its data collection simplicity, several studies have documented some limitations in its use (Faulkner, Cohn y Remington, 2006; Lee et al, 2011; Medina, Barquera y Janssen, 2013). It is possible that the IPAQ-SF may overestimate (Ainsworth et al., 2006; Rzewnicki, Van den Auweele y De Bourdeaudhuij, 2003), or even underestimate (Hallal y Victora, 2004; Qu y Li, 2004; Rutten et al., 2003) PA levels. These limitations become especially important in adolescent populations (Guedes et al., 2005; Rangul et al., 2008), who may have conceptual difficulties in differentiating some terms, such as vigorous or moderate PA (Bauman et al., 2009; Hagströmer et al., 2008; Papathanasiou et al., 2009), especially depending on the fitness level of participants (Johnson, Tillgren y Hagstromer, 2009). Likewise, and also considering our previous experience with the application of the IPAQ-SF, people may have some limitations recalling their weekly activities. These limitations may also possibly be greater in early adolescents (Guedes et al., 2005). It should be finally noted that, in order for adolescents to obtain health benefits, the World Health Organization (WHO, 2010) has made official international PA recommendations proposing 60 minutes of moderate-to-vigorous PA per day.

This study, which is comprised of two independent studies, pursues two main goals. These objectives are two-fold: firstly, to test the validity of the IPAQ-SF in a Spanish adolescent sample. Secondly, to test and compare the validity of the IPAQ-SF using a modified protocol version, specifically adapted to an adolescent population.

Materials and methods

Study sample and design

Two independent samples were used for each study. One sample of 270 adolescents for Study One (55% males), and another sample of 151 adolescents for Study Two (44.6% males), aged 11 – 14 years old, were recruited from six urban secondary schools in Huesca (Spain), on the basis of their willingness to cooperate in the study. Before participating in these studies, parents had to provide written informed consent, and adolescents had to give written assent. These studies were approved by the Ethics Committee of the University of Zaragoza.

To make it easier to compare data between the self-report IPAQ-SF and the accelerometer data, different inclusion criteria

were established in both studies. Before further processing, each self-report IPAQ-SF had to satisfy criteria according to Guidelines for Data Processing and Analysis of the IPAQ Committee (2005). It should be pointed out that, in order to be registered, a period of activity had to last for at least 10 minutes. Moreover, each accelerometer file had to satisfy the following criteria before further processing: (1) the subject must have worn the monitor for 7 consecutive days (including weekends), except at night while sleeping or during water-based activities; (2) a “*full day of wearing*” was defined as at least 10 hours’ continuous monitoring per day during weekdays, and at least 8 hours’ continuous monitoring at weekends (Rowlands, 2007). These accelerometer criteria are consistent with wearing compliance recommendations and reliability of PA variables derived from free-living conditions (Ward, Evenson, Vaughn, Rodgers y Troiano, 2005). The 7-consecutive-day criterion was quite restrictive. However, owing to the fact that the IPAQ-SF gave us 7-day output data, we considered it important to preserve the same number of days from the accelerometer data. Time spent on light physical activity (LPA), moderate physical activity (MPA), vigorous physical activity (VPA) and moderate-to-vigorous physical activity (MVPA) was averaged across all subjects’ “*full wearing days*”.

Applying these criteria, a total of 40 adolescents (55% male) and 67 adolescents (37.32% male) with valid IPAQ-SF and accelerometer data were included in Study One and Study Two, respectively.

Measurement

Self-report physical activity questionnaire

IPAQ-SF was used (Craig et al., 2003) to assess PA of the last 7 days. It should be noted that in Study Two the IPAQ-SF was also applied to evaluate PA using a modified version of the protocol, thus becoming the IPAQ-SF-MP. The modified protocol will be explained in the following sections. Nevertheless, it should be noted that the structure of the questionnaire was entirely similar. The IPAQ-SF is made up of eight open-ended questions and its information permits estimating the time spent per week on PA at different intensities. The Spanish version of IPAQ-SF was used by following the procedures recommended by the International Consensus Group (www.ipaq.ki.se).

Accelerometer

The uni-axial GT1M accelerometer (Actigraph, Pensacola, Florida, USA) was used to continuously assess objective PA during waking hours, over a 7-day period. The GT1M accelerometer is widely used in PA research and has been shown to be a valid and reliable tool for quantifying PA (Freedson, Pober y Janz, 2005; Reilly et al., 2006). The accelerometer was attached to an adjustable elastic belt and worn on the right side of the hip in

accordance with the guidelines suggested by Trost, Mciver and Pate (2005). For this study, monitors were initialised as described by the manufacturer, and the epoch was set at 30 seconds.

Procedure

After the 7-day monitoring period, accelerometers were collected and data were immediately downloaded and filed. Data were analysed using specific software and “activity counts” were translated into time (min/day) spent on PA at different intensities (vigorous, moderate and MVPA) according to Treuth cut-points (Treuth et al., 2004) due to their applicability to children of any age (Andersen et al., 2006). It should be pointed out that data with time periods of more than 20 minutes of zero values were excluded from the analysis and that a recording of more than 20,000 counts per minute was seen as a potential malfunction of the accelerometer and was also excluded from the analysis.

For the IPAQ-SF application in Study One, the adolescents were gathered into groups of 25 subjects. Participants received the questionnaire with instructions on how to fill it in as well as recommendations (www.ipaq.ki.se). No time limit was given for the task and possible doubts were promptly explained by the professional in charge of the data collection. Whilst filling in the questionnaire, the adolescents had no communication with other participants, in an attempt to avoid possible undesirable interferences in their responses.

Statistical analysis

Means, standard deviations and differences were calculated for

minutes of light, moderate, vigorous and moderate-to-vigorous physical activity (LPA, MPA, VPA, MVPA) per week, using data obtained from each one of the two instruments in the two studies. Owing to the normal distribution of the PA data, Pearson correlation coefficients were calculated between total minutes of different PA intensities measured by the accelerometer and the IPAQ-SF (both with its original protocol and its modified protocol). We used Cohen’s guidelines (Cohen, 1998) for classifying correlation values: (i.e. $r < 0,30$, small; $0,31 > r < 0,50$, moderate; $r > 0,51$, large). The Bland and Altman method (Altman, 1991; Atkinson y Nevill, 1998) was used to provide the total error between the two methods in both studies. An indication of the systematic random error and the 95% limits of agreement were used in order to describe it. Bland and Altman analyses were conducted on minutes spent on MPA, VPA and MVPA. Statistical significance was set at $p \leq 0,05$. The Statistical Package for Social Sciences (SPSS, version 19.0) for Windows and the Analyse-it Software for Microsoft Office Excel 2007 for Windows were used to perform all statistical analyses.

Study One Results

Table 1 includes a summary of subjective (i.e., IPAQ) and objective (i.e., accelerometer) data of time spent at different PA intensities (study one and study two). Study one showed that, in comparison with the ActiGraph, self-reported levels of vigorous ($54,50 \pm 36,00$ vs. $6,37 \pm 5,33$ min/week) and MVPA ($68,03 \pm 42,66$ vs. $39,39 \pm 17,05$ min/week), were significantly overestimated, while MPA ($13,53 \pm 18,97$ vs. $23,02 \pm 12,66$ min/week) were substantially underestimated by IPAQ-SF.

Table 1. Mean scores and Standard Deviations (SD) of the PA data obtained by the IPAQ and the accelerometer Actigraph in trial one and trial two.

	Study One (n=40)						Study Two (n=67)					
	ActiGraph			IPAQ-SF			ActiGraph			IPAQ-SF-MP		
	All	Boys (n=22)	Girls (n=18)	All	Boys (n=22)	Girls (n=18)	All	Boys (n=25)	Girls (n=42)	All	Boys (n=25)	Girls (n=42)
<i>Light activity</i>												
Weekday	266.83*** (47.30)	279.67*** (41.81)	251.13*** (50.00)	36.85 (39.16)	44.55 (39.97)	27.44 (37.06)	282.36*** (47.16)	295.76*** (49.23)	274.38*** (44.57)	37.22 (25.89)	25.61 (26.57)	44.13 (23.12)
Weekend	288.22*** (78.17)	313.92*** (76.21)	256.81*** (70.33)	45.44 (49.74)	53.52 (53.73)	35.56 (43.82)	247.64*** (75.52)	261.82*** (77.26)	239.20*** (74.09)	50.82 (47.18)	45.00 (45.02)	54.29 (48.63)
Total	272.94*** (49.66)	289.45*** (49.66)	252.75*** (52.45)	39.30 (39.09)	47.11 (38.87)	29.76 (38.25)	272.44*** (48.11)	286.06*** (47.38)	264.33*** (47.24)	41.11 (27.47)	31.15 (27.20)	47.03 (26.17)
<i>Moderate activity</i>												
Weekday	23.02** (12.66)	27.31** (13.78)	17.77 (8.94)	13.53 (18.97)	15.41 (22.81)	11.22 (13.12)	29.87*** (15.72)	43.09** (14.34)	22.01 (10.37)	20.51 (14.16)	26.08 (16.87)	17.19 (11.22)
Weekend	27.53 (18.89)	36.49 (16.63)	16.58 (15.68)	20.44 (37.58)	19.20 (41.80)	21.94 (32.82)	19.73 (13.98)	28.07** (14.37)	14.76 (11.22)	15.63 (21.39)	12.80 (15.21)	17.32 (24.36)
Total	24.31** (13.44)	29.94** (14.06)	17.43 (8.89)	15.50 (19.13)	16.50 (23.16)	14.29 (13.21)	26.98** (13.69)	38.80*** (11.85)	19.94 (9.13)	19.12 (12.38)	22.29 (13.23)	17.23 (11.59)

	Study One (n=40)						Study Two (n=67)					
	ActiGraph			IPAQ-SF			ActiGraph			IPAQ-SF-MP		
	All	Boys (n=22)	Girls (n=18)	All	Boys (n=22)	Girls (n=18)	All	Boys (n=25)	Girls (n=42)	All	Boys (n=25)	Girls (n=42)
<i>Vigorous activity</i>												
Weekday	6.37 (5.33)	8.17 (6.44)	4.16 (2.13)	54.50*** (36.00)	65.64*** (36.02)	40.89*** (31.83)	6.20 (5.73)	9,26 (6,06)	4.39 (4.71)	25.31*** (22.86)	35.32*** (26.03)	19.36*** (18.63)
Weekend	6.66 (8.32)	9.66 (9.48)	3.00 (4.71)	43.63*** (47.51)	52.39*** (46.38)	32.92* (47.96)	4.04 (4.84)	6,32 (5,84)	2.68 (3.57)	18.02*** (25.10)	25.70** (28.61)	13.45** (21.85)
Total	6.45 (5.60)	8.59 (6.63)	3.83 (2.08)	51.39*** (33.80)	61.85*** (32.58)	38.61*** (31.59)	5.58 (4.97)	8,42 (5,13)	3.90 (4.06)	23.23*** (18.99)	32.57*** (18.58)	17.67*** (17.14)
<i>MVPA</i>												
Weekday	29.39 (17.05)	35.48 (19.20)	21.93 (10.22)	68.03*** (42.66)	81.05*** (45.18)	52.11* (34.13)	36.08 (20.03)	52,35 (17,95)	26.39 (14.13)	45.82** (27.21)	61.40 (29.52)	36.55** (21.12)
Weekend	34.19 (25.43)	46.15 (23.33)	19.58 (20.01)	64.25** (67.09)	71.59 (69.71)	55.28* (64.55)	23.76 (16.84)	34,39 (17,57)	17.44 (12.88)	33.66* (31.27)	38.50 (35.00)	30.77** (28.88)
Total	30.76 (18.06)	38.53 (19.41)	21.26 (10.46)	66.95*** (42.94)	78.34*** (46.25)	53.02* (34.85)	32.56 (17.32)	47,22 (14,49)	23.83 (12.29)	42.35** (23.42)	54.86 (23.60)	34.90** (20.11)

*Significantly differences between the IPAQ and ActiGraph within each study. *p<.05; **p<.01; ***p<.001. Data are expressed in minutes.

IPAQ-SF-MP: International Physical Activity Questionnaire short Form Modified Protocol. IPAQ-SF: International Physical Activity Questionnaire Short Form

Note: significant differences are expressed between IPAQ and Actigraph within each study. Study one independent of study two.

Pearson correlation coefficients (see table 2) between MVPA data of IPAQ-SF and ActiGraph accelerometer were $r = 0,55$ ($p < .01$). Meanwhile, there were moderate Pearson's correlation coefficients between instruments for time spent on MPA ($r = 0,37$, $p < .05$) and VPA ($r = 0,34$, $p < .05$).

Table 2. Pearson correlations between the IPAQ and the accelerometer data for Trial one and two.

	Study one (n=40)			Study two (n=67)		
	ActiGraph vs IPAQ-SF			ActiGraph vs IPAQ-SF-MP		
	All	Boys (n=22)	Girls (n=18)	All	Boys (n=25)	Girls (n=42)
<i>Light</i>						
Weekday	.107	.280	-.221	-.164	-.154	-.051
Weekend	.231	.433*	-.237	.037	.069	.042
Total	.156	.369	-.251	-.188	-.208	-.090
<i>Moderate</i>						
Weekday	.413**	.523*	.008	.114	.030	-.301
Weekend	.273	.289	.445	.042	.097	.113
Total	.370*	.410	.318	.036	-.083	-.181
<i>Vigorous</i>						
Weekday	.257	.165	.119	.340**	.072	.403**
Weekend	.237	.140	.282	.398**	.521**	.128
Total	.342*	.277	.154	.332**	.077	.293
<i>MVPA</i>						
Weekday	.535**	.534*	.264	.358**	.055	.169
Weekend	.431**	.380	.520*	.305*	.409*	.163
Total	.547**	.521*	.387	.311**	.016	.140

*Significantly correlates between the IPAQ and ActiGraph within each study.

IPAQ-SF-MP: International Physical Activity Questionnaire Short Form Modified Protocol. IPAQ-SF: International Physical Activity Questionnaire Short Form

Figures 1, 2 and 3 show the Bland-Altman plots for time spent on MPA, VPA, and MVPA. The solid line in the plot reflects the mean of the difference between the questionnaire and the accelerometer: 16.0 min/week for vigorous PA; -28,7 min/week for moderate PA and 12,7 min/week for MVPA. The dashed lines show the 95% limits of agreement: -15,5 to 47,6 min/week for vigorous PA, -65,6 to 8,2 min/week for moderate PA and -33,7 to 59,0 min/week for MVPA. These limits of agreement values are larger for MVPA (from -33,7 to 59,0 min/week) than for vigorous and moderate PA (from -15,5 to 47,6 min/week and from -65,6 to 8,2 min/week respectively).

Figure 1. Bland-Altman plot of VPA minutes estimated by accelerometer and the IPAQ-S7T in the Trial 1.

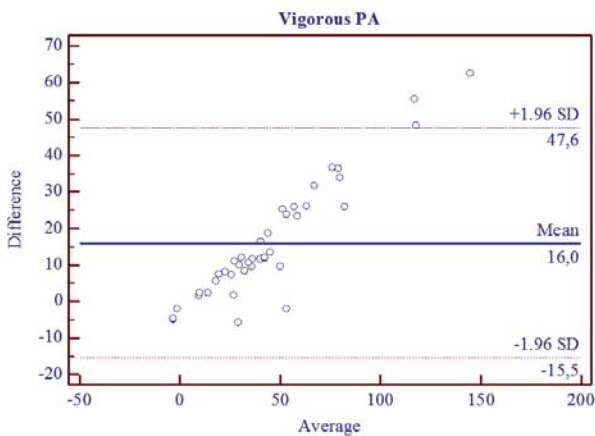


Figure 2. Bland-Altman plot of MPA minutes estimated by accelerometer and the IPAQ-S7T in the Trial 1.

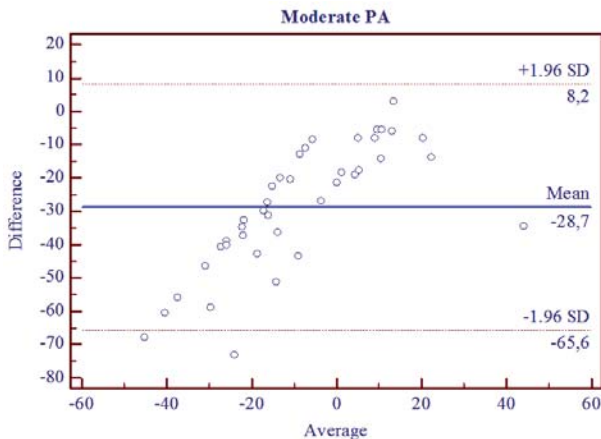
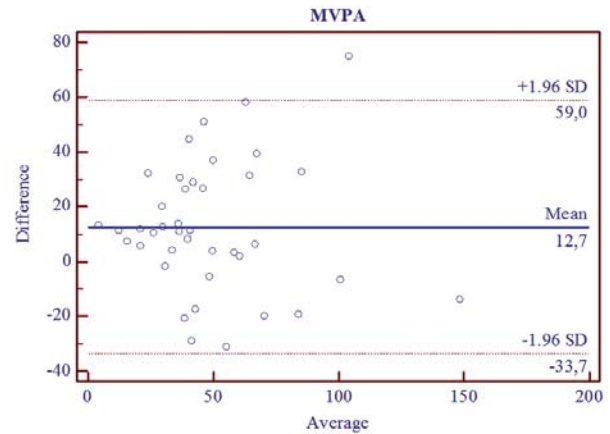


Figure 3. Bland-Altman plot of MVPA minutes estimated by accelerometer and the IPAQ-S7T in the Trial 1.



Possible sources of self-reported estimation error of IPAQ-SF

Once the IPAQ-SF had been completed in Study One, adolescents were asked open-ended questions about problems or difficulties that they had found whilst filling in the questionnaire. Researchers retrieved information from this data collection process (Study One), regarding problems or aspects that may have been misunderstood by students. Afterwards, several research team meetings took place to analyse and study those problems reported by the participants and those problematic aspects perceived by the researchers. Several requirements were identified (see table 3), namely: a) difficulties in recalling the activity of the previous week; b) difficulties in differentiating PA intensities: especially between vigorous and moderate; c) difficulties in estimating the number of minutes spent at different PA intensities (i.e. MPA and VPA). Table 3 also shows some hypothetical sources of error that researchers had detected.

Table 3. IPAQ-SF: requirements and potential sources of error obtained from Trial one.

Requirements and potential sources of error	
IPAQ requirements	Potential sources of error
1. Recall of activity during previous week	1.1. Instructions incorrectly interpreted as meaning the previous full week (Monday to Sunday) rather than the previous seven days. 1.2. Instruction incorrectly interpreted due to the high number of people who fill in the questionnaire at the same time; noisy work environment.
2. Differentiation of PA intensities: vigorous, moderate and walking.	2.1. Lack of understanding of the concept relating to intensity discrimination: Vigorous PA, moderate PA and walking.
3. Estimation of the number of minutes spent in different PA intensities.	3.1. Recall error owing to a lack of encoded information relating to activity frequency and duration, especially in the walking question. 3.2. Lack of perceived relevance may cause reduced motivation to complete the questionnaire.

Protocol modification: "IPAQ-SF-MP"

In order to test and compare the validity and agreement of the IPAQ-SF and to reduce hypothetical self-reported recall errors, a modified protocol was created and used as a questionnaire for Study Two. From then onwards, the name of the questionnaire was modified and called the IPAQ-SF-MP (modified protocol). The modifications were as follows: a) whereas groups of 25 subjects were established in Study One to answer the questionnaires, data was collected in groups of 5 subjects in Study Two. Students' recall was expected to be favoured with this modification as well as their capacity to discriminate different PA intensities. b) Before completing the questionnaire, participants had to recall all the activities that they had done the previous week and fill in a short weekly log diary. They had 15 minutes to correctly fill in the log diary. With this strategy it was hoped that students would find it easier to recall their daily and weekly routines. c) To facilitate their capacity of distinguishing between PA intensities. In order to do that, researchers gave a previous structured image-support explanation of the different PA intensities, followed by question time where participants could ask any questions.

Study Two Results

It can be seen in Table 1 how differences between different PA intensities between IPAQ-SF and accelerometer estimates are less important in Study Two than in Study One. Study Two showed that, in comparison with the accelerometer data, self-reported levels of VPA ($23,23 \pm 18,99$ vs. $5,58 \pm 4,97$ min/week) and MVPA ($42,35 \pm 23,42$ vs. $32,56 \pm 17,32$ min/week), were overestimated to a lesser extent than in Study One. Moreover, the underestimation found in MPA values ($19,12 \pm 12,38$ vs. $26,98 \pm 13,69$ min/week) was also smaller than in Study One. Pearson correlation coefficients in Study Two (see table 2) did not differ substantially from Study One, showing a moderate correlation ($r= 0,31$, $p<0,01$) in the MVPA value. We did not find significant correlations in

MPA values but correlations in VPA values were greater than in Study One.

Figures 4, 5, and 6 show the Bland-Altman plots for time spent on MPA, VPA, and MVPA in Study Two respectively.

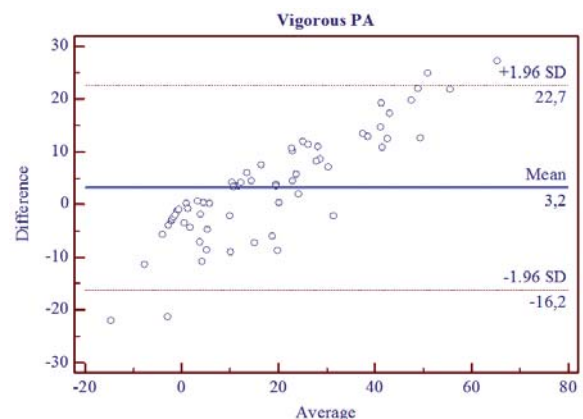
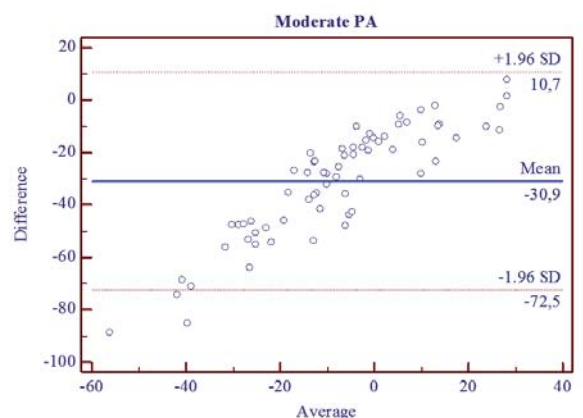
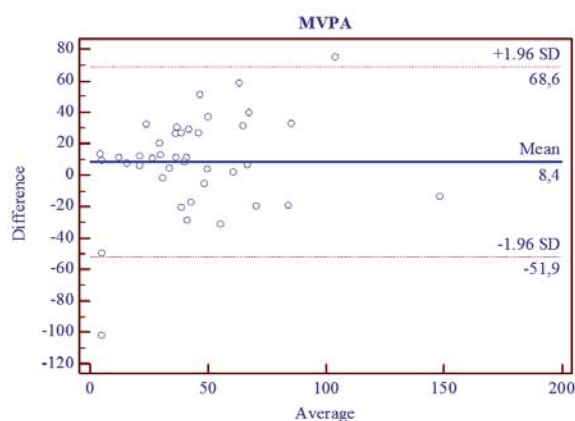
Figure 4. Bland-Altman plot of VPA minutes estimated by accelerometer and the IPAQ-S7T-MP in the Trial 2.**Figure 5.** Bland-Altman plot of MPA minutes estimated by accelerometer and the IPAQ-S7T-MP in the Trial 2.

Figure 6. Bland-Altman plot of MVPA minutes estimated by accelerometer and the IPAQ-S7T-MP in the Trial 2.



The Bland-Altman plots indicated much greater agreement at vigorous and MVPA intensities in Study Two than in Study One. The difference between the self-report questionnaire and the accelerometer was: 3,2 min/week for VPA and 8,4 min/week for MVPA. Meanwhile, we found similar agreement at MPA: -30,90 min/week. As in Study One, limits of agreement values were greater for MVPA (from -51,9 to 68,6 min/week) than for VPA and MPA (from -16,2 to 22,7 min/week and from -72,5 to 10,7 min/week respectively).

Discussion

Two main findings can be rescued from these two studies. Firstly, the findings supported acceptable levels of validity of the IPAQ-SF when compared with data obtained from the accelerometers. Nevertheless, it should be pointed out that MVPA levels were significantly overestimated by IPAQ-SF. Secondly, results from the IPAQ-SF-MP, conducted with a modified adapted protocol for early adolescents, obtained much better levels of agreement for vigorous and moderate-to-vigorous PA intensities.

PA self-reported validity studies in young people typically report correlation coefficients ranging from ,20 to ,60 (Kohl, Fulton y Caspersen, 2000; Martínez-Gómez et al., 2009). Correlation coefficients for moderate and vigorous intensity in the IPAQ-SF were similar to what is typically reported for these kinds of self-report questionnaires (Bauman y Merom, 2002). Our findings are comparable with other validation studies conducted worldwide, which show that, when compared with objective PA data obtained from accelerometer, self-report questionnaires have acceptable validity, but typically overestimate total PA (Brown, Bauman, Chey, Trost y Mummery, 2004; Craig et al., 2003; Hagstromer, Oja y Sjostrom, 2006; Mäder, Martin, Schultz y Marti, 2006; Moy, Scragg, McLean y Carr, 2008). In addition, it has

been shown in literature that the use of IPAQ-SF specifically overestimates PA (Ekelund et al., 2006; McFarlane, Lee, Ho, Chan y Chan, 2006).

Overestimation can be due to a variety of factors. Firstly, self-report questionnaires are essentially dependent on the accuracy of participant recall, and for this reason an error rate may exist. Furthermore, in today's health and exercise conscious society, individuals often over-report socially desirable behaviours such as PA, thus influencing the outcomes of self-report research (Adams et al., 2005). Recent data suggests that social-desirability bias is associated with an overestimation of PA duration by approximately 4–11 minutes per day over a 7-day period (Adams et al., 2005). This problem could be due to features of the short version of IPAQ, such as asking for average times or using a criterion of perceived intensity of breathing to estimate time at several PA intensities. Given that the IPAQ was developed for use in adults from 18 years old, it might not be easy for adolescents to understand and correctly interpret the PA concepts, either (Hagströmer et al., 2008). Good measurements of light and moderate PA using self-report measures have been shown to be the most difficult to obtain for adolescents (Bassett, 2000; Washburn, Heath y Jackson, 2000). Moreover, it has been speculated that individuals with low PA patterns tend to overestimate their responses in self-report PA questionnaires (Ekelund et al., 2006; Rutten et al., 2003). According to comparative international studies, the Spanish adolescent population is one of the most inactive of Europe (Arribas, Izaskun, Gil de Montes y González, 2013; Hallal, Victora, Wells y Lima, 2003; Rzewnicki, Van den Auweele y De Bourdeaudhuij, 2003). This factor may also explain the differences between the two methods. Therefore, if we improve the ability of the IPAQ to capture moderate and vigorous PA intensities, the validity of the instrument would be enhanced, especially in a population with low levels of PA.

Several investigations in young people (Troped et al., 2007) have similarly shown poor agreement between accelerometers and self-report PA questionnaires, in particular with the IPAQ-SF (Rangul et al., 2008). In our study, after applying a modified adapted protocol for early adolescents, the agreement between instruments was enhanced, especially for more intense activities. Whereas the difference between subjective and objective mean estimates remained the same at MPA intensity (Liou, Jwo, Yao, Chiang y Huang, 2008), differences decreased in VPA and MVPA. This fact indicates that applied modifications may have improved the agreement between the two methods, giving more validity to the use of the IPAQ-SF in early adolescents. According to Moy et al. (2008), clarifying the terminology used to define moderate and vigorous PA intensities may possibly have helped to minimise bias resulting from misinterpretation. Nevertheless, MPA intensity is still possibly difficult to identify in

free-living activities due to the difficult identification of its symptoms and the intermittence of these activities in adolescence (Liou et al., 2008). In spite of this fact, we strongly recommend the use of specific methodological adaptations in the protocol, as other studies have already implemented (Rzewnicki et al., 2003; Zanovec, Johnson, Marx, Keenan y Tuuri, 2009), in order to reduce possible sources of error if the IPAQ-SF is to be applied in early adolescents.

The Bland-Altman technique presents higher sensitiveness in the detection of differences between measurement methods when compared to conventional analyses that involve parametrical statistical resources, such as, the Pearson correlation coefficient. In theory, it is admitted that systematic differences detected between measurement methods should not affect the magnitude of any correlation coefficient. However, they could significantly restrict the concordance degree between both methods (Atkinson y Nevill, 1998). In our study, results indicated much better agreement at vigorous and MVPA intensities with the IPAQ-SF_MP than with the IPAQ-SF. However, we also suggest caution in the interpretation of the amplitude between the extreme limits of the confidence intervals due to the reduced number of subjects involved in this study.

This study has several limitations. Firstly, the sample size of both studies was small due to the restrictive criteria inclusion. However, it could be said that both samples were homogeneous in terms of age and gender. Secondly, the use of a specific accelerometer cut-off point to determine moderate and vigorous PA levels may create a measurement error when comparing these results with other studies (Guinhouya et al., 2006). In this study, we used Treuth et al. (2004) cut-off points, as they are applicable to children of any age (Andersen et al.,

2006). Thirdly, although accelerometers are frequently used in studies as criterion measures of PA to validate self-report questionnaires, there are a number of limitations associated with accelerometers. The accelerometer is known to underestimate PA levels for specific activities and this may lead to the underestimation of total energy expenditure (Hendelman, Miller, Baggett, Debold y Freedson, 2000). The reason for this is that uniaxial activity monitors can only provide information about acceleration on the vertical plane. Moreover, in our study, the accelerometer was placed on the waist and this could represent a potential problem for activities requiring mainly waist movement, or static exercise activities (bike exercise, water activities or upper limb movements) (Leenders, Sherman, Nagaraja y Kien, 2001). They may not be sensitive to many of the complex movement patterns exhibited by children during free play. Nevertheless, accelerometers provide an objective and non-reactive tool for assessing PA and are considered an acceptable means to validate self-report questionnaires (Reilly et al., 2006).

In conclusion, our results show that the IPAQ-SF presents acceptable validity properties for assessing PA levels of Spanish adolescents. The IPAQ-SF may be a valid tool to measure PA levels in early adolescents, especially when methodological modifications are applied in the collection data protocol. Therefore, the IPAQ-SF-MP could allow us to do a better worldwide measurement and comparison of PA levels in early adolescents.

Acknowledgements.- We would like to thank all the adolescents who took part in the study, as well as their parents and teacher, for their collaboration.

References

- Adams, S.W., Matthews, C.E., Ebbeling, C.B., Moore, C.G., Cunningham, J.E., Fulton, J., et al. (2005). The effect of social desirability and social approval on self-reports of physical activity. *American Journal of Epidemiology*, 161, 389-98.
- Ainsworth, B.E., Macera, C.A., Jones, D.A., Reis, J.P., Addy, C.L., Bowles, H.R., y Kohl, H.W (2006). Comparison of the 2001 BRFSS and the IPAQ Physical Activity Questionnaires. *Medicine & Sciences in Sports & Exercise*, 38, 1584-1592.
- Altman, D (1991). *Practical Statistics for Medical Research*. London: Chapman & Hall.
- Andersen, L.B., Harro, M., Sardinha, L.B., Froberg, K., Ekelund, U., Brage, S., et al. (2006). Physical activity and clustered cardiovascular risk in children: A cross-sectional study (the european youth heart study). *Lancet*, 368, 299-304.
- Arribas, S., Izaskun, L., Gil de Montes, L., y González, O. (2013). Cuestionario de hábitos de práctica de actividad física y deportiva (C-PAFYD). *SPORT TK: Revista Euroamericana de Ciencias del Deporte* 1 (2), 9-22.
- Atkinson, G., y Nevill, A.M (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports & Medicine*, 26, 217-238.
- Bassett Jr, D,R (2000). Validity and reliability issues in objective monitoring of physical activity. *Research Quarterly for Exercise and Sport*, 71, 30-336.
- Bauman, A., Ainsworth, B.E., Bull, F., Craig, C.L., Hagstromer, M., Sallis, J.F., et al. (2009). Progress and pitfalls in the use of the international physical activity questionnaire (IPAQ) for adult physical activity surveillance. *Journal of Physical Activity & Health*, 6, 5-8.
- Bauman, A., y Merom, D (2002). Measurement and surveillance of physical activity in Australia- an introductory guide. *Australian Epidemiology*, 9, 2-5.
- Borrego-Balsalobre, F.J., López-Sánchez, G.F., Díaz-Suárez, A. (2015a). Effects of a vigorous physical activity program in the strength of primary schoolchildren. *TRANCES: Revista de Transmisión del Conocimiento Educativo y de la Salud* 7(3), 387-406.
- Borrego-Balsalobre, F.J., López-Sánchez, G.F., Díaz-Suárez, A. (2015b). Effects of a vigorous physical activity program in the endurance of primary school children. *ATHLOS: International Journal of Social Sciences of Physical Activity, Game and Sport* 8, 31-46.
- Borrego-Balsalobre, F.J., López-Sánchez, G.F., Díaz-Suárez, A. (2015c). Influence of a vigorous physical activity program on cholesterol level of primary schoolchildren. *AGON: International Journal of Sport Sciences* 5(2), 60-71.

13. Brown, W., Bauman, A., Chey, T., Trost, S., y Mummery, K (2004). Comparison of surveys used to measure physical activity. *Australian and New Zealand Journal Public Health*, 28, 128–34.
14. Cohen, J (1998). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
15. Craig, C.L., Marshall, A.L., Sjoström, M., Bauman, A.E., Booth, M.L., Ainsworth, B.E., et al. (2003). International physical activity questionnaire: 12-country reliability and validity. *Medicine & Science in Sports & Exercise*, 35, 1381-1395.
16. De Courten, M (2002). Developing a simple global physical activity questionnaire for population studies. *Australian Epidemiology*, 9, 6–9.
17. Ekelund, U., Sepp, H., Brage, S., Becker, W., Jakes, R., Hennings, M., et al. (2006). Criterion-related validity of the last 7-day, short form of the International Physical Activity Questionnaire in Swedish adults. *Public Health Nutrition*, 9, 258-265.
18. Faulkner, G., Cohn, T., y Remington, G (2006). Validation of a physical activity assessment tool for individuals with schizophrenia. *Schizophrenia Research*, 82, 225-231.
19. Freedson, P., Pober, D., y Janz, K.F (2005). Calibration of accelerometer output for children *Medicine & Science in Sports & Exercise*, 37, 523-530.
20. Guedes, D.P., Lopes, C.C., y Guedes, J.E (2005). Reprodutibilidade e validade do questionário internacional de atividade física em adolescentes. *Revista Brasileira De Medicina do Esporte*, 11, 151-158.
21. Guinhouya, C.B., Hubert, H., Soubrier, S., Vilhelm, C., Lemdani, M., y Durocher, A (2006). Moderate-to-vigorous physical activity among children: Discrepancies in accelerometry- based cut-off points. *Obesity*, 14, 774-777.
22. Hagströmer, M., Bergman, P., DeBourdeaudhuij, I., Ortega, F.B., Ruiz, F.B., Manios, Y., et al. (2008) Concurrent validity of a modified version of the International Physical Activity Questionnaire (IPAQ-A) in European adolescents: the HELENA Study. *International Journal of Obesity*, 32, 42–48.
23. Hagstromer, M., Oja, P., y Sjoström, M (2006). The International Physical Activity Questionnaire (IPAQ): A study of concurrent and construct validity. *Public Health Nutrition*, 9, 755–762.
24. Hallal, P.C., Victora, C.G., Wells, J.C., y Lima, R.C (2003). Physical inactivity: Prevalence and associated variables in Brazilian adults. *Medicine & Sciences in Sports & Exercise*, 35, 1894-1900.
25. Hallal, P.C., y Victora, C.G (2004). Reliability and validity of the international physical activity questionnaire (IPAQ). *Medicine & Science in Sports & Exercise*, 36, 556-556.
26. Hendelman, D., Miller, K., Baggett, C., Debold, E., y Freedson, P (2000). Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Medicine & Sciences in Sports & Exercise*, 32, 442-449.
27. IPAQ committee (2005). *Guidelines for data processing and analysis of the International Physical Activity Questionnaire (IPAQ)*. November 2005. Available from: www.ipaq.ki.se
28. Johnson, I., Tillgren, P., y Hagstromer, M (2009). Understanding and interpreting the concept of physical activity - a focus group study among Swedish women. *Scand J Public Health*, 37, 20-27.
29. Kemper, H.C.G., Montoye, H.J., Saris, W.H.M., y Washburn, R.A (1996). Measuring physical activity and energy expenditure. *Human Kinetics Publishers Champaign Illinois*, 191, 146-159.
30. Kohl, H.W., Fulton, J.E., y Caspersen, C.J (2000). Assessment of physical activity among children and adolescents: A review and synthesis. *Preventive Medicine*, 31, 54-76.
31. Lachat, C.K., Verstraeten, R., Khanh le, N.B., Hagstromer, M., Khan, N.C., Van Ndo, A., et al. (2008). Validity of two physical activity questionnaires (IPAQ and PAQA) for vietnamese adolescents in rural and urban areas. *International Journal of Behavioral Nutrition and Physical Activity*, 5, 37- 45.
32. Lee, P.H., Yu, Y.Y., McDowell, I., Leung, G.M., Lam, T.H., y Stewart, S.M (2011). Performance of the international physical activity questionnaire (short form) in subgroups of the Hong Kong Chinese population. *International Journal of Behavioral Nutrition and Physical Activity*, 8, 81-90.
33. Leenders, N.Y., Sherman, W.M., Nagaraja, H.N., y Kien, C.L (2001). Evaluation of methods to assess physical activity in free-living conditions. *Medicine & Sciences in Sports & Exercise*, 33, 1233-1240.
34. Liou, Y.M., Jwo, C.J., Yao, K.G., Chiang, L.C., y Huang, L.H (2008). Selection of appropriate Chinese terms to represent intensity and types of physical activity terms for use in the Taiwan version of IPAQ. *JNR*, 16, 252-263.
35. López-Sánchez, G.F., Borrego-Balsalobre, F.J., Díaz-Suárez, A. (2013). Effects of a physical activity program on body composition of school children of 3-5 years. *SPORT TK: Revista Euroamericana de Ciencias del Deporte* 2(2), 41-44.
36. Mäder, U., Martin, B.W., Schutz, Y., y Marti, B (2006). Validity of four short physical activity questionnaires in middle-aged persons. *Medicine & Science in Sports & Exercise*, 38, 1255-1266.
37. Martínez-Gómez, D., Martínez-De-Haro, V., Del-Campo, J., Zapatera, B., Welk, G.J., Villagra, A., Marcosa, A., Veiga, O.L. (2009). Validez de cuatro cuestionarios para valorar la actividad física en adolescentes españoles. *Gaceta Sanitaria*, 23(6), 512-517.
38. McFarlane, D.J., Lee, C.C.Y., Ho, E.Y.K., Chan, K.L., y Chan, D.T.S (2006). Convergent validity of six methods to assess physical activity in daily life. *Journal of Applied Physiology*, 101, 1328–1344.
39. Medina, C., Barquera, S., y Janssen, I (2013). Validity and reliability of the International Physical Activity Questionnaire among adults in Mexico. *Revista Panamericana de Salud Publica*, 34, 21-28.
40. Medina, C., Janssen, I., Campos, I., y Barquera, S (2013). Physical inactivity prevalence and trends among Mexican adults: results from the National Health and Nutrition Survey (ENSANUT) 2006 and 2012. *BMC Public Health*, 13, 1063-1072.
41. Montoye, H.J., Cunningham, D.A., Hugh, G. y Epstein H.G (2000). Laboratory methods of assessing metabolic capacity in a large epidemiologic study. *American Journal of Epidemiology*, 91, 38-47.
42. Moy, K.L., Scragg, R.K., McLean, G., y Carr, H (2008). The New Zealand physical activity questionnaires: Validation by heart-rate monitoring in a multiethnic population. *Journal of Physical Activity & Health*, 5, 45–61.
43. Papatthasiou, G., Georgoudis, G., Georgakopoulos, D., Katsouras, C., Kalfakakou, V., y Evangelou, A (2009). Reliability Measures of the Short International Physical Activity Questionnaire (IPAQ) in Greek Young Adults. *Hellenic Journal Cardiology*, 50, 283-294.
44. Qu, N.N., y Li, K.J (2004). Study on the reliability and validity of international physical activity questionnaire (chinese version, IPAQ). *Zhonghua Liu Xing Bing Xue Za Zhi = Zhonghua Liuxingbingxue Zazhi*, 25, 265-268.
45. Rangul, V., Holmen, T.L., Kurtze, N., Cuypers, K., y Midthjell, K (2008). Reliability and validity of two frequently used self-administered physical activity questionnaires in adolescents. *BMC Medical Research Methodology*, 8, 47- 56.
46. Reilly, J.J., Kelly, L.A., Montgomery, C., Jackson, D.M., Slater, C., Grant, S., et al. (2006). Validation of actigraph accelerometer estimates of total energy expenditure in young children. *IJPO: An Official Journal of the International Association for the Study of Obesity*, 1, 161-167.
47. Rosenberg, D.E., Bull, F.C., Marshall, A.L., Sallis, J.F., y Bauman, A.E (2008). Assessment of sedentary behavior with the international physical activity questionnaire. *Journal of Physical Activity & Health*, 5, 30-44.
48. Rowlands, A.V (2007). Accelerometer assessment of physical activity in children: An update. *Pediatric Exercise Science*, 19, 252-266.
49. Rutten, A., Vuillemin, A., Ooijendijk, W.T., Schena, F., Sjoström, M., Stahl, T., et al. (2003). Physical activity monitoring in Europe. the eu-

- ropean physical activity surveillance system (EUPASS) approach and indicator testing. *Public Health Nutrition*, 6, 377-384.
50. Rzewnicki, R., Vanden Auweele, Y., y De Bourdeaudhuij, I (2003). Addressing overreporting on the International Physical Activity questionnaire (IPAQ) telephone survey with a population sample. *Public Health Nutrition*, 6, 299-305.
 51. Treuth, M.S., Schmitz, K., Catellier, D.J., McMurray, R.G., Murray, D.M., Almeida, M.J., et al. (2004). Defining accelerometer thresholds for activity intensities in adolescent girls. *Medicine & Science in Sports & Exercise*, 36, 1259-1266.
 52. Troped, P.J., Wiecha, J.L., Fragala, M.S., Matthews, C.E., Finkelstein, D.M., Kim, J., et al. (2007). Reliability and validity of YRBS physical activity items among middle school students. *Medicine & Sciences in Sports & Exercise*, 39, 416-425.
 53. Trost, S.G., McIver, K.L., y Pate, R.R (2005). Conducting accelerometer-based activity assessments in field-based research. *Medicine & Science in Sports & Exercise*, 37, 531-543.
 54. Ward, D.S., Evenson, K.R., Vaughn, A., Rodgers, A.B., y Troiano, R.P (2005). Accelerometer use in physical activity: Best practices and research recommendations. *Medicine & Science in Sports & Exercise*, 37, 582-588.
 55. Washburn, R.A., Heath, G.W., y Jackson, A.W (2000). Reliability and validity issues concerning large-scale surveillance of physical activity. *Research Quarterly for Exercise and Sport*, 71, 104-113.
 56. World Health Organization (2010). *Global recommendations on physical activity for health*. Geneva: World Health Organization.
 57. Zanovec, M., Johnson, L., Marx, B.D., Keenan, M.J., y Tuuri, G (2009). Self-reported physical activity improves prediction of body fatness in young adults. *Medicine & Sciences in Sports & Exercise*, 41, 328 - 335.