©Journal of Sports Science and Medicine (2022) **21,** 356-375
http://www.jssm.org DOI: https://doi.org/10.52082/jssm.2022.356

Research article

Are Wrist-Worn Activity Trackers and Mobile Applications Valid for Assessing Physical Activity in High School Students? Wearfit Study

Jesús Viciana [1], Carolina Casado-Robles [1], Santiago Guijarro-Romero [2]✉ and Daniel Mayorga-Vega [3]
[1] Department of Physical Education and Sport, University of Granada, Granada, Spain; [2] Department of Didactic of Musical, Plastic and Corporal Expression, University of Valladolid, Valladolid, Spain; [3] Departamento de Didáctica de las Lenguas, las Artes y el Deporte, Facultad de Ciencias de la Educación, Universidad de Málaga, Málaga, España

Abstract
The purpose was to examine the validity of three wrist-worn commercial activity trackers (Samsung Galaxy Watch Active 2, Apple Watch Series 5, and Xiaomi Mi Band 5) and six mobile apps (Pedometer and Pacer for android and iPhone mobiles, Google Fit for android, and Apple Health for iPhone mobiles) for estimating high school students' steps and physical activity (PA) under free-living conditions. A sample of 56 (27 females; mean age = 14.7 years) and 51 (25 females; mean age = 14.0 years) high school students participated in Study 1 and 2, respectively. Study 1: Students performed a 200-meter course in four different conditions while wearing the wearables. Step counting through a video record was used as the golden standard. Study 2: Students wore the three wrist-worn commercial activity trackers during the waking time of one day, considering ActiGraph model wGT3X-BT accelerometers as a standard of reference. Afterward, the agreement between the PA scores measured by the commercial activity trackers and the video (study 1) or accelerometers (study 2) were calculated as follows: Equivalence test, Limits of Agreement (LOA); Mean Absolute Error (MAE); Mean Absolute Percentage Error (MAPE); and Intraclass Correlation Coefficient (ICC). Results showed that all the wearables presented excellent validity for assessing steps in structured free-living conditions (study 1; MAPE < 5%), although their validity was between poor-excellent based on ICC (95% confidence interval) values (ICC = 0.56-1.00). Regarding Study 2, the Xiaomi wristband and the Samsung Watch presented acceptable-excellent (MAPE = 9.4-11.4%; ICC = 0.91-0.97) validity for assessing steps under unstructured free-living conditions (study 2). However, the Apple Watch presented questionable-excellent validity (MAPE = 18.0%; ICC = 0.69-0.95). Regarding moderate-to-vigorous PA (MVPA) and total PA, only the Apple Watch showed low-acceptable validity for MAPE value and questionable-excellent validity for the ICC values for MVPA assessment (MAPE = 22.6; ICC = 0.67-0.93). All wearables checked in this study have shown adequate validity results in order to assess steps in both structured and unstructured free-living conditions for both continuous and dichotomous variables. Moreover, for assessing MVPA, only the Apple Watch reported valid results for compliance or non-compliance with the daily PA recommendations. However, the results showed low validity for total PA and MVPA as continuous variables. In conclusion, depending on the user's/researcher's aim and context, one or another wearable activity tracker could be more adequate, mainly because of its valid measurements and its costs.

Key words: Consumer-wearables, daily steps, physical activity recommendations, accuracy, validity, schoolchildren.

Introduction

The high prevalence of physical inactivity in all ages (i.e., in adolescence, 77.5% of males and 84.7% of females; Word Health Organization, WHO, 2020; Guthold et al., 2020) and its associated consequences for health is well known (Poitras et al., 2016; WHO, 2020). For instance, physically inactive high school students are more likely to become physically inactive adults who are more susceptible to all-cause mortality (Paluch et al., 2022). Consequently, a current public health priority is to improve the ratio of high school students that meet the daily physical activity (PA) recommendations of 60 minutes of moderate-to-vigorous PA (MVPA) in aerobic activities (WHO, 2020) or its equivalent in steps, that is to say 10,000 per day (Mayorga-Vega et al., 2021).

Additionally, it is not an easy ability for high school students to interpret the intensity level of daily activities and their relationship with MVPA (Crossley et al., 2019). In order to help the population in this task of regulating daily PA and its quantification, the consumer-wearable activity trackers have become a valuable asset as they provide a display for self-monitoring of the users' PA (Strath and Rowley, 2018). Particularly, due to several advantages, two types of wearable activity trackers stand out as being the most popular: (a) wrist-worn activity trackers (including smartwatches and activity wristbands) have shown to be one of the most valued and used type of wearable because of their characteristics, including an attractive display, real-time feedback, low weight and price, audible sedentary break warnings, and goal alerts (Maher et al., 2017); and (b) mobile apps, which are widely used among high school students, are freely available, and let them register many PA parameters. Therefore, wrist-worn activity trackers and mobile apps could be useful, accessible, and feasible devices to objectively assess and promote high school students' daily PA (Casado-Robles et al., 2022; Da Silva et al., 2015). Moreover, these two wearable activity trackers are widely extended all over the world, and the number of users is continuous and exponentially increasing every year (Vailshery, 2021).

Concerning the measurement of these wearables, it is necessary to know to what extent the information that consumers receive from their device regarding PA is accurate and valid. For instance, it is also important for educational researchers and Physical Education teachers in the school setting who are interested in applying wearables in Physical Education-based physical activity promotion

Received: 29 March 2022 / Accepted: 26 May 2022 / Published (online): 21 July 2022

programs, as it affects students' daily physical activity, which is a priority in the educational system of most countries around the world (OECD, 2019). Consequently, valid wearable activity trackers could help researchers and Physical Education teachers to achieve this educational standard with reduced costs. That is, they could better control the PA carried out by high school students, even differentiating between different parts of the day. Additionally, students could have immediate feedback about the PA that they are performing during the day, which may facilitate them to achieve the daily PA recommendations. However, researchers need to know if they can use wearable activity trackers only as a strategy to increase habitual PA or also as a dependent variable.

In regards to PA intensity, the doubly labeled water assessment of energy expenditure is the gold standard (Westerterp, 2017), but the oxygen uptake measurement substitutes this expensive method, and is considered as the "reference standard" for assessing MVPA. Nevertheless, because of viability motives (e.g., small, light, comfortable, and reliable and valid; Migueles et al., 2017), accelerometers have been used as the most appropriate alternative in free-living conditions (Cosoli et al., 2020; Mayorga-Vega et al., 2019; Westerterp, 2013). In fact, ActiGraph accelerometers have been the most widely used method in population and calibration studies with high school students (Romanzini et al., 2014). Regarding the number of steps per day in free-living conditions, the accelerometer is also considered as the reference standard in most previous studies (Westerterp, 2013; 2017). However, it should be considered that accelerometer data for validation of other devices has traditionally been based on proprietary algorithms and therefore still subject to various forms of measurement error, particularly when assessing an outcome like MVPA (Migueles et al., 2017). The gold standard in structured or controlled settings is the video-taped and manual counter of the number of steps in a particular circuit or activity (El-Amrawy and Nounou, 2015; Stamm and Hartanto, 2018). In this sense, the convergence of measurements provided by other wrist-worn activity trackers, mobile apps, and the accelerometer will provide new possibilities and devices to use in those contexts beyond the laboratory, for instance, in Physical Education settings where validity and feasibility must coexist with each other (O'Neill et al., 2017).

Different cross-sectional studies with adults (e.g., Redenius et al., 2019; Toth et al., 2018) and synthesis studies (Brickwood et al., 2019; Gal et al., 2018) have shown the validity of wrist-worn activity trackers and mobile apps. Concerning high school students, the number of publications is scarce, although there are still some publications regarding their validity (Cosoli et al., 2020; Gorzelitz, et al., 2020). Overall results of these above-mentioned publications have shown a great variety of protocols for validation, devices, and methodological characteristics, which provide irregular results barely comparable to each other. Nevertheless, depending on the context to be applied and the required accuracy of the PA parameters, different devices could be recommended. Moreover, particular measurements, such as energy expenditure or PA levels, are more complex constructs, and their measurements present significant errors when assessed by wrist-worn activity trackers and mobile apps. For instance, Adamakis (2020) obtained large errors in all monitors and apps used underestimating the PA energy expenditure from 13.16% to 37.46%. On the contrary, the number of steps is a simpler measurement that can be assessed in a valid manner by mobile apps and wrist-worn activity trackers with a moderate correlation to the reference standard. Although these outputs measured by wearable activity trackers tend to have a lower validity for assessing intensity-related PA than steps among adults because of their high absolute percentage error with respect to reference instrumentation (Cosoli et al., 2020), among high school students, the evidence is currently insufficient (Voss et al., 2017). Moreover, because the components of wrist-worn activity trackers and mobile phones for the estimation of PA levels and steps are in continuous evolution by companies of wearable activity trackers, to analyze new models of them is desirable and recommended. Finally, Apple device validity studies were not found, or were scarce and centered in other parameters such as heart rate or energy expenditure, or measuring steps in a laboratory setting, as some previous synthesis studies have concluded (Fuller et al., 2020; Gorzelitz et al., 2020; LaMunion et al., 2020). However, they did not study unstructured free-living conditions, the measurement of PA intensities, nor the adequacy of its measurement to PA daily recommendations.

Consequently, the overall purpose of the present study was to examine the validity of the mobile apps Pedometer, Pacer, Google Fit for Android mobiles (Samsung); Pedometer, Pacer, and Apple Health for Apple mobiles (iPhone); and the wrist-worn activity trackers Samsung Galaxy Watch Active 2, Apple Watch Series 5, and Xiaomi Mi Band 5 for estimating PA in high school students during free-living conditions. Specifically, the purposes of this study were twofold: *Study 1:* to compare the criterion-related validity of the steps assessed by the nine wearable activity trackers under structured free-living conditions. *Study 2:* to compare the convergent validity of the daily steps, total PA, and MVPA scores assessed by the three wrist-worn activity trackers under unstructured free-living conditions (the comparability between them and their convergent validity with the accelerometer, considered as reference standard).

Methods

The Guidelines for Reporting Reliability and Agreement Studies (GRRAS) have been taken into account in the present study (Kottner et al., 2011), and the protocol followed conforms to the Declaration of Helsinki statements (64th WMA, Brazil, October 2013) and was approved by the Ethical Committee for Human Studies at the University of Granada. Then, the principals of two state high schools (chosen by convenience) from an urban area situated in Granada (Spain) were contacted for Study 1 and 2. They were informed about the project and permission to conduct the study was obtained. After the school's approval was acquired, all the students and their legal guardians were fully informed about the characteristics of the project. Signed written informed assents from the participants and signed written informed consents from their legal guardians' were obtained before starting the study. School A was used for Study 1 (i.e., criterion-related validity under structured free-living conditions) and school B for Study 2 (i.e., convergent validity under unstructured free-living conditions). According to the center's reports, all the students' families had a middle socioeconomic level.

The present validity study followed a cross-sectional design. All the high school students (i.e., 70 and 75 for the school A and B, respectively) enrolled in the selected schools (i.e., 12–18 years old) were invited to participate in the present study. The high schools were located in an urban area of the city center of (deleted for anonymity reasons). For each study, the following inclusion criteria were considered: (a) being enrolled in the selected high schools (i.e., in the 7th to 12h grade); (b) being free of any health disorder that would make them unable to engage in PA normally; (c) presenting the corresponding signed written informed assents of the students, and (d) presenting the corresponding signed written informed consents of their legal guardians. The following exclusion criteria were considered: (a) not having completed and valid data from the three wrist-worn activity trackers (i.e., Samsung Galaxy Watch Active 2, Apple Watch Series 5, or Xiaomi Mi Band 5) or from the six mobile apps (Study 1) or from the three wrist-worn activity trackers (Study 2), and (b) not having completed and valid data from the video-based step counts (Study 1) or from the accelerometer (Study 2).

A priori sample size calculation was estimated with the Arifin´s web-based sample size calculator (Arifin, 2018). Based on steps values, parameters were set as follows: Study 1 and 2: Intraclass Correlation Coefficient (ICC), $\rho_0 = 0.70$ (Nunnally, 1978); $\rho_1 = 0.86$ (Voss et al., 2017), $\alpha = 0.05$, $1 - \beta = 0.80$, $k = 2$, dropout = 10% (Rowlands et al., 2018). Study 2: Kappa, $k_0 = 0.40$ (Cicchetti, 2001); $k_1 = 0.85$ (Mayorga-Vega et al., 2021), $p = 0.25$ (Guthold et al., 2020), $\alpha = 0.05$, $1 - \beta = 0.80$, $k = 2$, dropout = 10% (Rowlands et al., 2018). A final sample size of at least 49 students (minimum initial sample size equal to 55) was used in the final analysis. In addition to exceeding the minimum required sample size, a balance by grade and sex was taken into account for each study´s sample.

Measures

Demographic characteristics. Participants' sex (males/females), age (in years), and non-dominant hand (left/right) information was self-reported.

Anthropometric. Participants' body mass (kg) and height (cm) were first measured following the International Standards for Anthropometric Assessment (Stewart et al., 2011). Participants' body mass and height were measured in shorts, T-shirts, and barefoot. For the body mass measure, participants stood in the center of the scale (Seca, Ltd., Hamburg, Germany; accuracy = 0.1 kg) without support and with their weight distributed evenly on both feet. For the body height assessment, participants stood with their feet together with the heels, buttocks and upper part of the back touching the stadiometer (Holtain Ltd., Crymmych, Pembs, United Kingdom; accuracy = 0.1 cm), and with the head placed in the Frankfort plane. Each measurement was performed twice and the mean was recorded (Stewart et al., 2011). Then, the body mass index was calculated as body mass divided by body height squared (kg/m^2). Finally, students' body weight status was categorized by sex- and age-adjusted body mass index thresholds as overweight/obese or non-overweight/obese (Cole et al., 2000). Body mass index and body weight status scores have shown high evidence supporting validity among high school students (Cole et al., 2000).

Activity wristbands and mobile apps. Participants' PA levels and steps were estimated by the three wrist-worn activity trackers [i.e., Xiaomi Mi Band 5 (Xiaomi, Pekin, China); Samsung Galaxy Watch Active 2 (Samsung, Seul, South Korea); and Apple Watch Series 5 (Apple Park, California, USA)]; and from six mobile apps [i.e., Pedometer (ITO Technologies) and Pedometer Pacer Health for Android and iOS; and Google Fit app for Android (Samsung Galaxy S20+), and the Apple Health app for iOS (iPhone 11 Pro Max)]. Regarding the number of wrist-worn activity trackers, it was considered that three devices were a feasible number that did not interfere with the participants' natural arm swing while walking, brisk walking, or running (Study 1), or daily PA (Study 2). Mobile phones (Study 1) were allocated in two belt bags that changed sides for each participant, which also did not interfere with the participants' movements during the test. The total mass of the three wearable bands was not high (88 grams). According to the user manual of each device brand, the wrist-worn activity trackers were fit snugly on the top of participants' wrist, close and above the wrist bone (7.6 cm width). Concerning the particular chosen wrist-worn activity trackers, the criteria were to study the most used worldwide display-based trackers, choosing the most advanced model (in that moment). Three wrist-worn activity trackers were chosen, one being a more low cost option (i.e., Xiaomi Mi Band 5 ≈ 25€) and the other two were smartwatches from the two main brands [Samsung belonging to Android (Samsung Galaxy Watch Active 2 that uses the Android Wear System) and Apple belonging to iOS (Apple Watch Series 5 that uses the WatchOS Apple System)] (based on the International Data Corporation's Worldwide Quarterly Wearable Device Tracker reports from 2021, and Henriksen et al. 2018). Concerning the mobile apps, the following ones were selected: Pedometer (ITO Technologies) and Pedometer Pacer Health for Android and iOS, the Google Fit app for Android and the Apple Health app for iOS. After revising the previous literature, the criteria followed for the selection were (both for Android and iOS): (a) the most popular and used apps (due to the number of downloads and their user ratings); (b) free download apps; and (c) the included apps of the corresponding mobile phones (i.e., Samsung Google Fit for Android and Apple Health for iOS).

According to the user manual of each device brand, wrist-worn activity trackers were adjusted, in random order, on the participant's wrist of their non-dominant hand. The three chosen devices are characterized to be small and lightweight based in tri-axial built-in accelerometers (Xiaomi Mi Band 5: 22 g, 1.5 x 1.6 x 4.0 cm; Samsung Galaxy Watch Active 2 2: 30 g, 1.09 x 4.4 x 4.4 cm; Apple Watch Series 5: 30.8 g, 1.07 x 4.0 x 4.4 cm). Each wrist-worn activity tracker and mobile phone has its own algorithmic equation to estimate the daily steps taken and the minutes engaged in each specific intensity-related PA. The wrist-worn activity trackers data and mobile phone apps were recorded immediately from the screen (for Study 1), and were also synchronized via Bluetooth for the three wrist-worn activity trackers to their specific applications in order to download and store data (for Study 2) (Xiaomi Mi Band 5: Mi Fit version 5.3.2 for Android; Samsung Galaxy Watch Active 2: Samsung Health version 6.19 for Android; Apple Watch Series 5: Apple Health version 7.6.2 for iOS).

Regarding the data scoring, steps (number) were registered as directly reported in the devices. For the intensity-related PA (minutes), scores were calculated as follows: for the Xiaomi Mi Band 5, the variables “slow walking” and “brisk walking” were calculated by adding the total time spent on all the stretches of “slow walking” and “brisk walking,” respectively. The variable MVPA was calculated by adding the total time spent on all the stretches of “moderate activity” and “vigorous activity”. Aside from the default variable “total activity”, the variable “walking” was also calculated by adding the total time spent on all the stretches of “slow walking” and “brisk walking”; for the Samsung Galaxy Watch Active 2, the variable total PA was calculated considering the default option “total active minutes” and; for the Apple Watch Series 5, the default option “exercise minutes” was considered as the variable MVPA.

Video-based steps count. Participants’ reference standard of steps under structured free-living conditions was determined by step counting the video recording in slow-motion, which is considered to be the golden standard of this measurement. Participants were asked to perform a 200-meter course in four different conditions. The 200-meter course was marked with cones and lines and performed inside the school on a non-slippery sport court with an oval shape and no tight turns. A digital video camera (Go Pro Hero 7, California, USA) with a tripod was situated in the middle of the sports court in order to easily record the participants´ lower limbs during the entire course from the sagittal plane. When the participant was at the starting line, the steps count from the wrist-worn activity trackers were recorded. Then, they were instructed to not move until they started walking/running. They also were asked to always start the course with the contralateral leg to the arm where the wrist-worn activity trackers were attached. For calculating the speed and step cadence of each condition, when the participant started walking/running the manual chronometer was activated and was stopped after he/she crossed the finish line. Participants were requested to stop immediately after the finish line, and a cone was situated five meters beforehand to remind them. Then, the steps counted by the wrist-worn activity trackers were registered. Participants performed the following four conditions: 1) slow pace walking; 2) normal pace walking (also known as self-pace walking); 3) brisk pace walking; and 4) jogging (slow running). Finally, the reference standard step count for each participant in each condition were performed independently by two researchers through the slow-motion video recording projected on a 15.6” screen. When disagreement occurred (8.3%), a third researcher evaluated it.

Accelerometer. Participants’ reference standard of daily steps and intensity-related PA under unstructured free-living conditions were determined by wGT3X-BT accelerometers (ActiGraph, LLC, Pensacola, FL, USA). The ActiGraph model wGT3X-BT is a small (4.6 x 3.3 x 1.5 cm), lightweight (19 g), tri-axial accelerometer. Accelerometers were adjusted on the participants’ right hips. Initializing, downloading, wear time validation and scoring (i.e., PA data filter) were performed using the ActiLife Lifestyle Monitoring System Software version 6.13.3 (ActiGraph, LLC, Pensacola, FL, USA), and were initialized with a sampling frequency of 30 Hz (Evenson et al., 2008; Trost et al., 2011). Data download was carried out with 15-second *epoch* (Evenson et al., 2008). According to previous literature when selecting a set of cut-points to estimate a variable from activity counts, it is recommendable to select the same filter that was used in the validation study for the cut-points employed (Migueles et al., 2017). Valid wear time was set as equal to or higher than 600 minutes per day (Migueles et al., 2017), with non-wear periods set as 60 minutes or more of consecutive zero-count *epochs* with up to a two minute spike tolerance (Oliver et al., 2011).

Regarding the data scoring, steps (number) were assessed by within-instrument processing of the number of cycles in the accelerometer signal or *cycle counts*. The time (minutes) engaged in MVPA ($\geq$ 2,296 *counts*/ min), and total PA ($\geq$ 101 *counts*/min) was calculated by the Evenson’s thresholds (Evenson et al., 2008). According to the cross-validation study performed by Trost et al. (2011), these cut-off points have demonstrated the best evidence supporting score validity for assessing intensity-related PA among high school students. Finally, participants’ steps and MVPA were dichotomized as achieving or not achieving the daily recommendation of at least 10,000 steps and 60 minutes of MVPA, respectively (Mayorga-Vega et al., 2021; WHO, 2020). ActiGraph accelerometer scores have shown high evidence supporting validity for assessing steps and intensity-related PA among high school students (Hickey et al., 2016; Lee et al., 2015; Romanzini et al., 2014).

Procedure

Study 1 and 2: Data collection was carried out by the same researchers, instruments, and protocols. Firstly, participants’ demographic characteristics, anthropometric, and self-reported habitual PA levels were registered. The three wrist-worn activity trackers were adjusted so they would not move (i.e., avoiding over-tightening and clearance), in a random order, on the participants’ wrists of their non-dominant hand. All devices were blocked in order to prevent participants from manipulating any functions and options they had, and to avoid influence in their habitual behavior.

Study 1: Evaluations were carried out during the afternoon in participants’ leisure time from Monday to Friday, and then data were downloaded and batteries charged during the morning. Due to the limitations of instruments, facilities, and time, an average of two or three participants per hour were evaluated one by one during each evaluation session. Apart from the three wrist-worn activity trackers, the two mobile phones were inside two belt bags allocated on each participants' hip, preventing movement, approximately like carrying it in the pants pocket, and alternating sides for each participant. Participants were instructed to walk/run the 200-meter course in the four conditions specified above, at a continuous speed, and with a natural arm swing. Before starting, a demonstration in order to guide each participant was performed. During the five-minute rest between conditions, step count data from the three wrist-worn activity trackers and the mobile phone apps were recorded.

Study 2: The three wrist-worn activity trackers and the accelerometer were adjusted to one participant each day, Monday through Thursday. Then, on Friday, data were downloaded. Due to the limitations of material resources, waves of three participants per day were carried out. For each wave, participants were met at 8:00 a.m. in a room allocated in the school gym, so they could then go and start their school day at the regular time (i.e., 8:30 a.m.). Additionally to the wrist-worn activity trackers, an accelerometer was adjusted on the participants' right hip using an elastic waistband. Participants were instructed to wear the wrist-worn activity trackers and the accelerometer for the whole day until bedtime maintaining their habitual PA levels, and they were asked to take them off only when they took a bath/shower, or to leave them in a plastic box inside their schoolbags just before going to bed. In the case of a participant reporting that he/she was going to engage in aquatic activities that day, he/she was requested to come the next day (and another participant was requested instead). In the morning of the following day, the wrist-worn activity trackers were collected and adjusted to the next three participants following the same protocol.

Statistical analysis

Descriptive statistics for all the variables of the included participants were calculated. Firstly, all the statistical tests assumptions were checked and met (e.g., histograms and Q-Q plots for normality). Furthermore, univariate (i.e., $z \pm 3.0$) and multivariate outliers (i.e., Mahalanobis distance) were removed. Afterward, the agreement between the PA scores (i.e., steps, total PA, and MVPA considered as continuous variables) measured by the wearable activity trackers and the video (study 1)/accelerometers-wearables (study 2) were calculated as follows: Equivalence test with the Confident Interval method (90% CI) (Dixon et al., 2018); (b) Limits of Agreement (LOA) with its confident intervals (95% CI) (Bland and Altman, 1986); (c) Mean Absolute Error (MAE) (Willmott and Matsuura, 2005); (d) Mean Absolute Percentage Error (MAPE) (Johnston et al., 2021); and (e) Intraclass Correlation Coefficient (ICC), and its 95% CI, by a two-way random-effects model with absolute agreement and single measurement [also known as ICC(2,1)] (Koo and Li, 2016). Additionally, LOA plots, which are the individual participant differences between the two scores plotted against the respective individual means, were performed (Bland and Altman, 1986). Heteroscedasticity was also examined objectively by calculating the Pearson's correlation coefficient (r) between the absolute differences and the individual means (Atkinson and Nevill, 1998). Based on Cohen's (1992) benchmarks, a correlation coefficient > 0.50 was considered as indicative of heteroscedasticity. Finally, the agreement between the PA scores (i.e., steps and MVPA) dichotomized as achieving or not achieving the daily recommendations of 10,000 steps (Mayorga-Vega et al., 2021) and 60 minutes of MVPA (i.e., categorical variables) (WHO, 2020) measured by the wearable activity trackers and the accelerometers (Study 2) were calculated as the proportion of agreement [P = number of agreements/ (number of agreements + disagreements)] and kappa coefficient (k) (Hernaez, 2015). Agreement values were interpreted as follows: Equivalence test, the mean reference standard is within ± 15% of the mean activity trackers is considered acceptable (Dixon et al., 2018); MAPE, > 20.0% poor; 15.1-20.0% questionable; 10.1-15.0% acceptable; 5.1-10.0% good; and 0.0-5.0% excellent (Bai et al., 2021; Johnston et al., 2021). ICC, 0.00-0.49 unacceptable; 0.50-0.59 poor; 0.60-0.69 questionable; 0.70-0.79 acceptable; 0.80-0.89 good; and 0.90-1.00 excellent (Nunnally, 1978); k, 0.00-0.39 poor; 0.40-0.59 acceptable; 0.60-0.74 good; and 0.75-1.00 excellent (Cicchetti, 2001). Based on statistical inference, each ICC value was interpreted according to its 95% CI, that means, there was 95% chance that the true ICC value landed on any point between the 95% CI range (Koo and Li, 2016). Regarding kappa values, < 0.00 poor; 0.00-0.20 slight; 0.21-0.40 fair; 0.41-0.60 moderate; 0.61-0.80 substantial; 0.81-1.00 almost perfect (Landis and Koch, 1977). All statistical analyses were performed using the SPSS version 25.0 for Windows (IBM® SPSS® Statistics), except for the equivalence testing where the Jamovi version 2.3 (The Jamovi project, https://www.jamovi.org) was used. The statistical significance level was set at $p < 0.05$.

Results

General characteristics

Figure 1 shows the flow diagram of the participants through the studies. An initial sample of 62 and 61 students agreed to participate and met the inclusion criteria in the Studies 1 and 2, respectively. Since some participants met at least one exclusion criterion, the final sample consisted of 56 (29 males and 27 females) and 51 (26 males and 25 females) participants for the Studies 1 and 2, respectively (i.e., non-compliance rate of 9.7% and 16.4%, respectively). Table 1 shows the general characteristics of the included participants.

Study 1: Validity of the wearable activity trackers during structured free-living conditions

The average speed (SD) in each condition was as follows: Slow pace walking = 1.1 (0.2) m/s [3.9 (0.5) km/h]; normal pace walking = 1.5 (0.1) m/s [5.2 (0.5) km/h]; brisk pace walking = 1.9 (0.2) m/s [6.8 (0.6) km/h]; and running = 3.1 (0.4) m/s [11.0 (1.3) km/h]. The average step cadence (SD) in each condition was as follows: Slow pace walking = 99.7 (7.3) steps/min; normal pace walking = 113.3 (5.0) steps/min; brisk pace walking = 128.7 (7.0) steps/min; and running = 167.2 (19.1) steps/min.

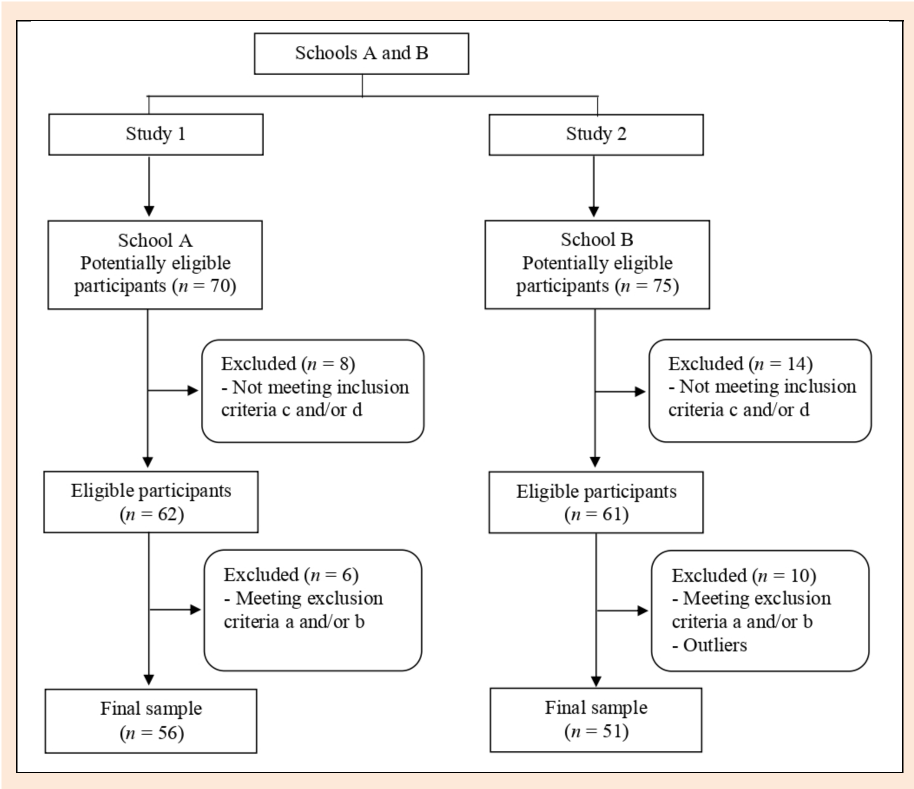


Figure 1. **Flow diagram of the participants through the studies.**

Table 1. **General characteristics of the analyzed participants in each study. Data are reported as mean (standard deviation) or percentage.**

	Study 1 (*n* = 56)	**Study 2 (*n* = 51)**
Age (years)	14.7 (1.7)	14.0 (1.5)
Grade (7^{th}/8^{th}/9^{th} 10^{th}/11^{th}/12^{th})	21.4/17.9/12.5/19.6/14.3/14.3	19.6/21.6/21.6/19.6/17.6/0.0
Gender (males/females)	51.8/48.2	51.0/49.0
Body mass (kg)	58.1 (12.9)	58.6 (13.2)
Body height (cm)	165.0 (11.3)	161.8 (8.3)
Body mass index (kg/m^2)	21.2 (3.3)	22.4 (4.8)
Overweight/obesity (no/yes)	85.7/14.3	31.4/68.6
Non-dominant hand (left/right)	87.5/12.5	90.2/9.8
Self-reported habitual physical activity (days/week)	3.7 (1.3)	2.4 (1.6)

Study 1: Validity in structured free-living conditions; Study 2: Validity in unstructured free-living conditions.

Table 2 shows the validity of the wearable activity trackers for estimating steps during structured free-living conditions. The agreement between the steps assessed by the wearable activity trackers and video-based count tended to be higher for slow pace walking, followed by running, normal pace walking, and brisk pace walking. Regarding the validity results of the steps based on the values of 90% CI of the equivalence test, the 90% confidence interval of all activity trackers scores were inside the equivalence region of reference standard. Similarly, based on the MAPE values, the validity results of the steps assessed by all five devices (three wrist-worn activity trackers and the two mobile phones with their respective apps) in the four conditions were excellent (i.e., < 5%). Note that due to the three Apple iPhone apps registering exactly the same parameters without distinction between them, only "Apple iPhone apps" has been reported in this Results section. Regarding the validity results of the steps based on the values of 95% CI of the ICC, similarly, the Samsung apps Pacer and Google Fit in all conditions ranged between good to excellent. Furthermore, the validity results of the steps assessed by the Apple iPhone and Apple Watch Series 5 were excellent, and by the Samsung Pedometer and Samsung Galaxy Watch Active 2 were good-excellent in the three conditions. However, their validity was worse in one condition: Apple iPhone in the slow pace walking condition was acceptable-excellent; Apple Watch Series 5 in the brisk pace walking was poor-excellent; Samsung Pedometer in the normal pace walking condition was questionable-good; and Samsung Galaxy Watch Active 2 in the brisk pace walking was questionable-excellent. Similarly, while the validity results of the steps assessed by the Xiaomi Mi Band 5 were excellent in the slow pace walking and running conditions, they were acceptable-excellent and poor-good in the normal and brisk pace walking conditions, respectively. Figures 2, 3, 4 and 5 show the LOA plots. Pearson's correlation coefficients showed that with all the instruments there was no heteroscedasticity on any walking/running condition (Table 3).

Table 2. Validity of the wearable activity trackers for estimating steps during structured free-living conditions (n = 56).

Instrument	Mean (SD)	Equivalence test (90% CI)	LOA (95% CI)	MAE	MAPE	ICC (95% CI)
			Slow pace walking			
Video-based count	306.6 (30.9)	-45.99, 45.99	-	-	-	-
Samsung Pedometer	307.6 (32.4)	-2.91, 1.01	-0.9 (-18.1, 16.3)	6.2	2.0	0.96 (0.94, 0.98)
Samsung Pacer	308.0 (29.8)	-3.53, 0.78	-1.4 (-20.2, 17.4)	5.8	1.9	0.95 (0.92, 0.97)
Samsung Google Fit	307.1 (30.2)	-2.48, 1.51	-0.5 (-17.9, 16.9)	5.6	1.8	0.96 (0.93, 0.98)
Apple iPhone apps[a]	307.1 (28.6)	-4.13, 3.17	-0.5 (-32.4, 31.4)	8.2	2.7	0.85 (0.76, 0.91)
Xiaomi Mi Band 5	303.7 (31.1)	1.63, 4.15	2.9 (-7.9, 13.7)	4.1	1.3	0.98 (0.95, 0.99)
Samsung Galaxy Watch Active 2	305.7 (31.0)	-0.18, 1.96	0.9 (-8.5, 10.3)	3.9	1.3	0.99 (0.98, 0.99)
Apple Watch Series 5	304.7 (30.9)	1.15, 2.68	1.9 (-4.8, 8.6)	3.1	1.0	0.99 (0.98, 1.00)
			Normal pace walking			
Video-based count	261.4 (22.6)	-39.21, 39.21	-	-	-	-
Samsung Pedometer	259.9 (23.9)	-1.89, 4.96	1.5 (-28.5, 31.5)	5.6	2.1	0.79 (0.66, 0.87)
Samsung Pacer	262.0 (23.8)	-2.33, 1.11	-0.6 (-15.7, 14.5)	4.2	1.6	0.95 (0.91, 0.97)
Samsung Google Fit	262.8 (22.8)	-2.61, -2.14	-1.4 (-12.2, 9.4)	3.5	1.4	0.97 (0.95, 0.98)
Apple iPhone apps[a]	262.4 (23.9)	-2.71, 0.71	-1.0 (-15.9, 13.9)	4.5	1.7	0.95 (0.91, 0.97)
Xiaomi Mi Band 5	256.3 (25.2)	3.36, 6.93	5.1 (-10.6, 20.8)	6.0	2.4	0.92 (0.79, 0.97)
Samsung Galaxy Watch Active 2	258.1 (22.3)	2.20, 4.37	3.3 (-6.1, 12.7)	4.0	1.5	0.97 (0.89, 0.99)
Apple Watch Series 5	258.3 (22.7)	1.96, 4.22	3.1 (-6.9, 13.1)	3.9	1.5	0.97 (0.91, 0.99)
			Brisk pace walking			
Video-based count	227.6 (18.2)	-34.14, 34.14	-	-	-	-
Samsung Pedometer	230.5 (20.5)	-4.95, -0.87	-2.9 (-20.7, 14.9)	5.8	2.6	0.88 (0.80, 0.93)
Samsung Pacer	229.3 (18.8)	-3.27, -0.19	-1.7 (-15.2, 11.8)	3.7	1.7	0.93 (0.88, 0.96)
Samsung Google Fit	229.3 (18.9)	-3.30, -0.20	-1.8 (-15.5, 11.9)	3.9	1.7	0.93 (0.88, 0.96)
Apple iPhone apps[a]	226.4 (18.1)	0.41, 1.95	1.2 (-5.5, 7.9)	2.5	1.1	0.98 (0.96, 0.99)
Xiaomi Mi Band 5	221.6 (23.2)	2.95, 9.05	6.0 (-20.7, 32.7)	7.3	3.2	0.76 (0.58, 0.86)
Samsung Galaxy Watch Active 2	222.1 (18.1)	3.83, 7.21	5.5 (-9.4, 20.4)	6.8	3.0	0.87 (0.62, 0.95)
Apple Watch Series 5	220.9 (21.1)	4.27, 9.08	6.7 (-14.5, 27.9)	8.1	3.6	0.81 (0.56, 0.90)
			Running			
Video-based count	185.0 (24.0)	-27.75, 27.75	-	-	-	-
Samsung Pedometer	187.7 (24.4)	-4.86, -0.57	-2.7 (-21.5, 16.1)	5.7	3.2	0.92 (0.86, 0.95)
Samsung Pacer	188.3 (25.6)	-4.96, -1.61	-3.3 (-18.0, 11.4)	4.7	2.5	0.95 (0.90, 0.97)
Samsung Google Fit	188.4 (25.7)	-5.01, -1.72	-3.4 (-18.1, 11.3)	4.9	2.7	0.95 (0.89, 0.97)
Apple iPhone apps[a]	185.6 (23.5)	-1.92, 0.77	-0.6 (-12.4, 11.2)	3.3	1.8	0.97 (0.95, 0.98)
Xiaomi Mi Band 5	183.6 (24.6)	0.23, 2.52	1.4 (-8.6, 11.4)	3.6	2.0	0.98 (0.96, 0.99)
Samsung Galaxy Watch Active 2	183.6 (23.5)	0.64, 2.29	1.5 (-5.8, 8.8)	2.8	1.5	0.99 (0.97, 0.99)
Apple Watch Series 5	184.8 (23.8)	-0.28, 0.71	0.2 (-4.1, 4.5)	1.5	0.8	1.00 (0.99, 1.00)

SD = Standard deviation; LOA = Limits of agreement; 90%/95% CI = 90%/95% confident interval; MAE = Mean absolute error; MAPE = Mean absolute percentage error; ICC = Intraclass correlation coefficient; [a] Apple iPhone apps is referred to the three apps activated in the iPhone mobile (i.e., Pedometer, Pacer, and Apple Health).

Table 3. Pearson's correlation coefficient (r) between the absolute differences and the individual means.

Instrument	Slow pace walking (steps)	Normal pace walking (steps)	Brisk pace walking (steps)	Running (steps)
		Study 1 (n = 56)		
Samsung Pedometer	0.37‡	0.00	0.23	0.04
Samsung Pacer	0.30*	-0.06	0.09	0.23
Samsung Google Fit	0.31*	0.01	0.09	0.22
Apple iPhone apps[a]	0.11	0.06	0.20	0.06
Samsung Galaxy Watch Active 2	0.16	0.03	-0.04	0.07
Apple Watch Series 5	0.11	-0.08	-0.14	-0.07
Xiaomi Mi Band 5	0.02	-0.37‡	-0.38‡	-0.14
		Study 2 (n = 51)		
Samsung Galaxy Watch Active 2	0.21	-	0.54†	-
Apple Watch Series 5	-0.01	-0.16	-	-
Xiaomi Mi Band 5	0.39‡	0.55†/0.22[a]	0.62†/0.51†[b]	-

MVPA = Moderate-to-vigorous physical activity; PA = Physical activity; Study 1: Validity in structured free-living conditions; Study 2: Validity in unstructured free-living conditions; [a] Brisk walking time (min); [b] Slow-Brisk walking time (min); [a] Apple iPhone apps is referred to the three apps activated in the iPhone mobile (i.e., Pedometer, Pacer, and Apple Health). * $p < 0.05$, ‡ $p < 0.01$, and † $p < 0.001$

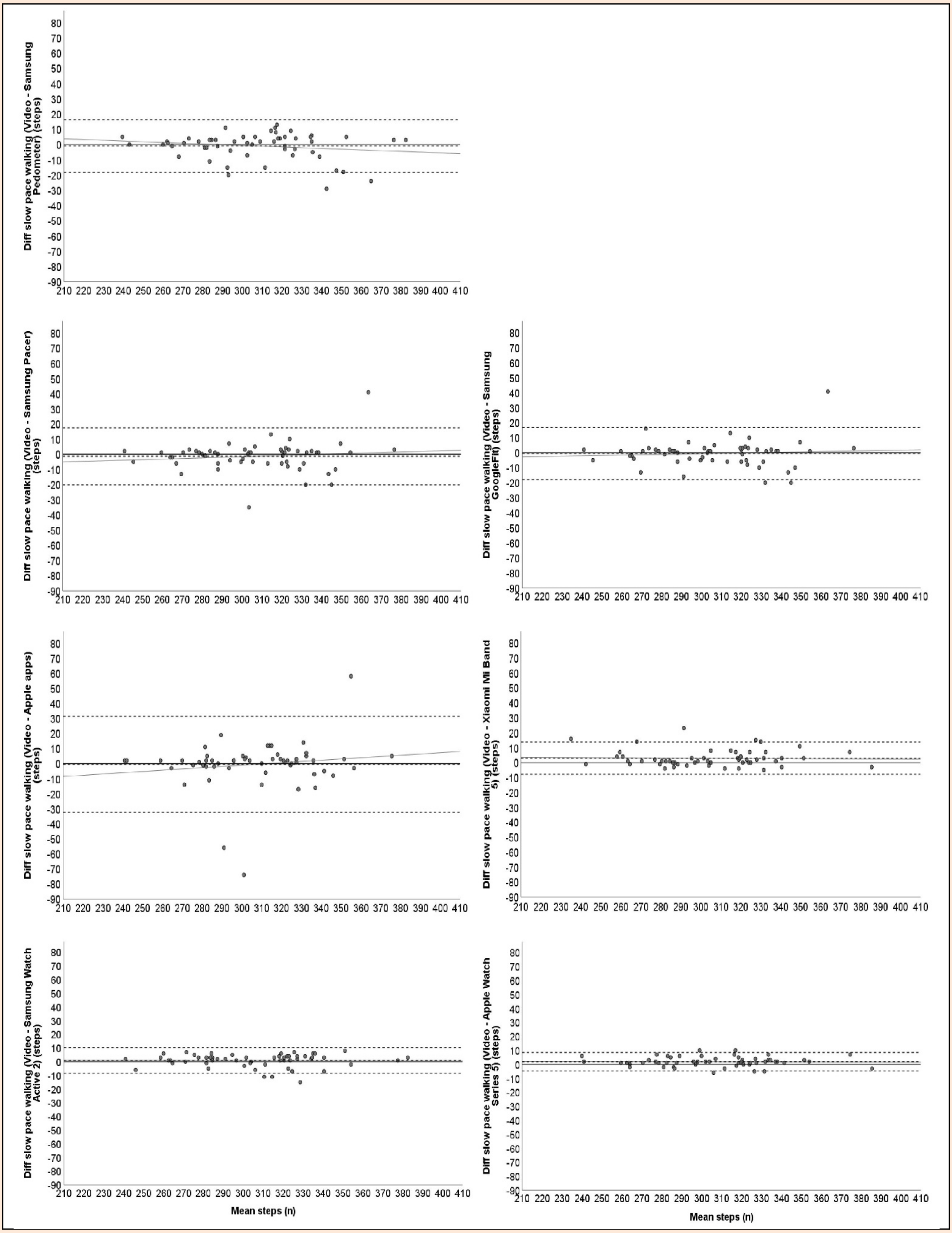


Figure 2. **Bland-Altman plots of the seven devices for measuring steps under a structured free-living setting in slow pace walking condition. The middle line shows the mean difference between the measurements of steps of the activity trackers and video-based step counts and the dashed lines indicate the limits of agreement.**

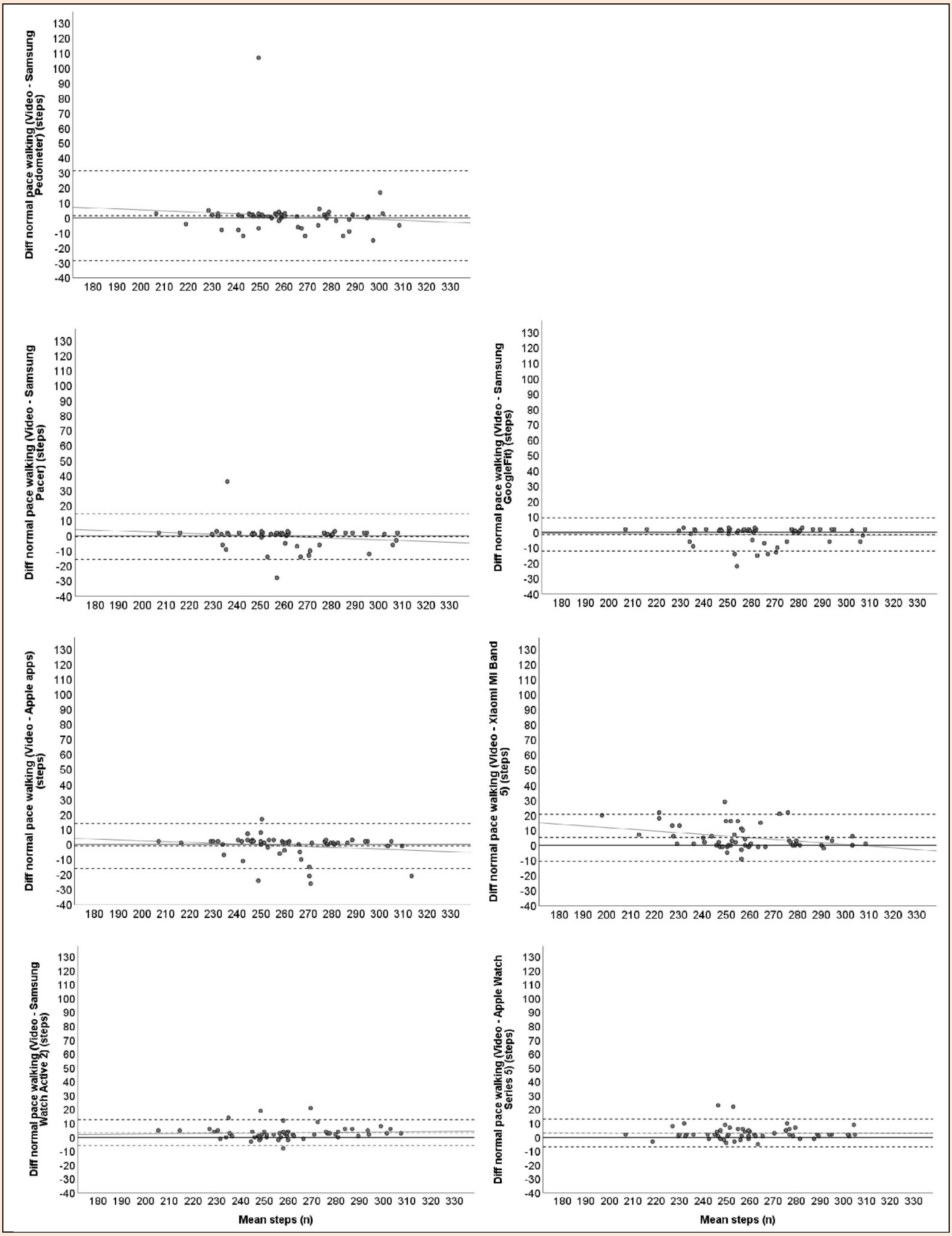


Figure 3. Bland-Altman plots of the seven devices for measuring steps under a structured free-living setting in normal pace walking condition. The middle line shows the mean difference between the measurements of steps of the activity trackers and video-based step counts and the dashed lines indicate the limits of agreement.

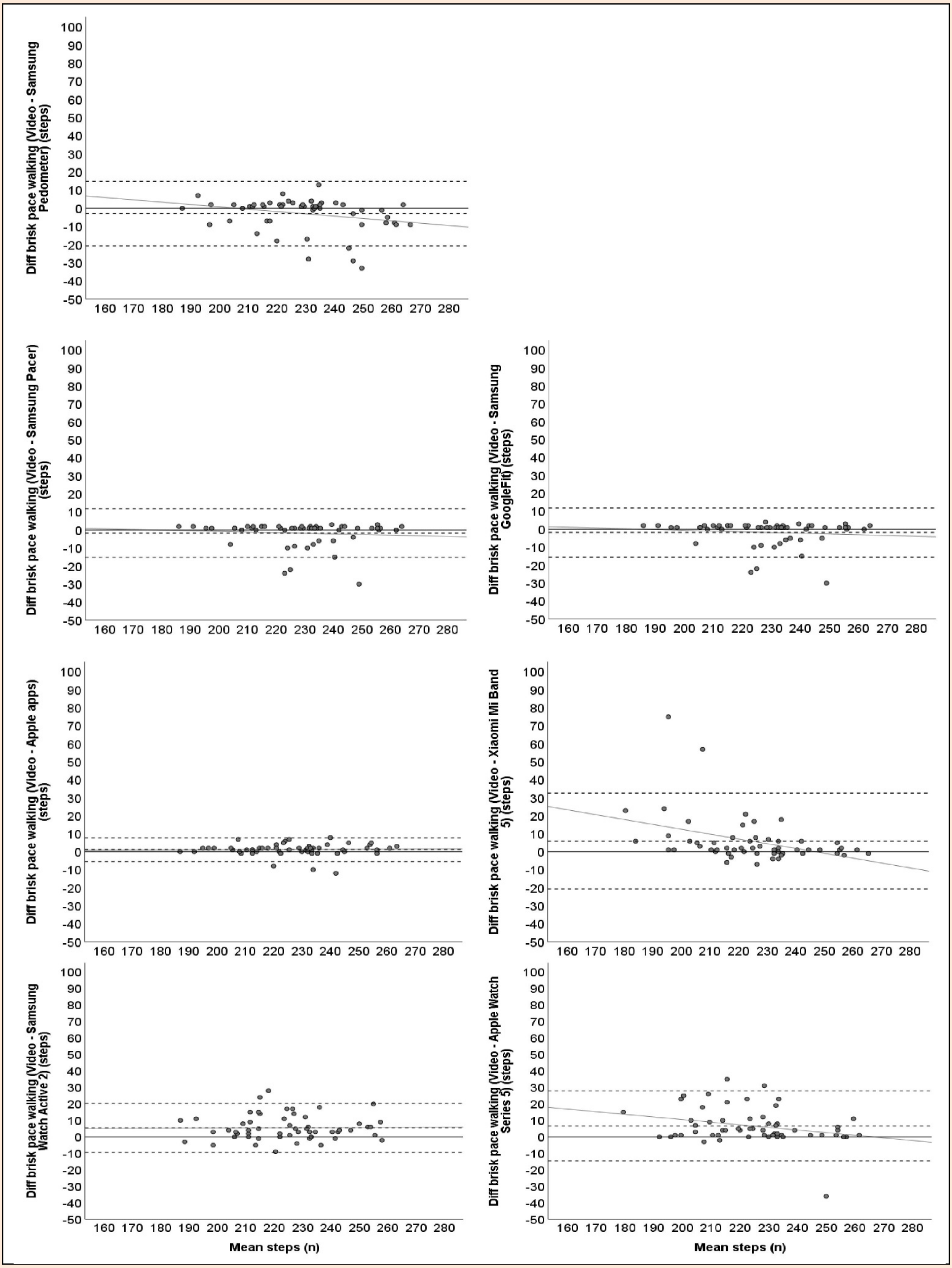


Figure 4. Bland-Altman plots of the seven devices for measuring steps under a structured free-living setting in brisk pace walking condition. The middle line shows the mean difference between the measurements of steps of the activity trackers and video-based step counts and the dashed lines indicate the limits of agreement.

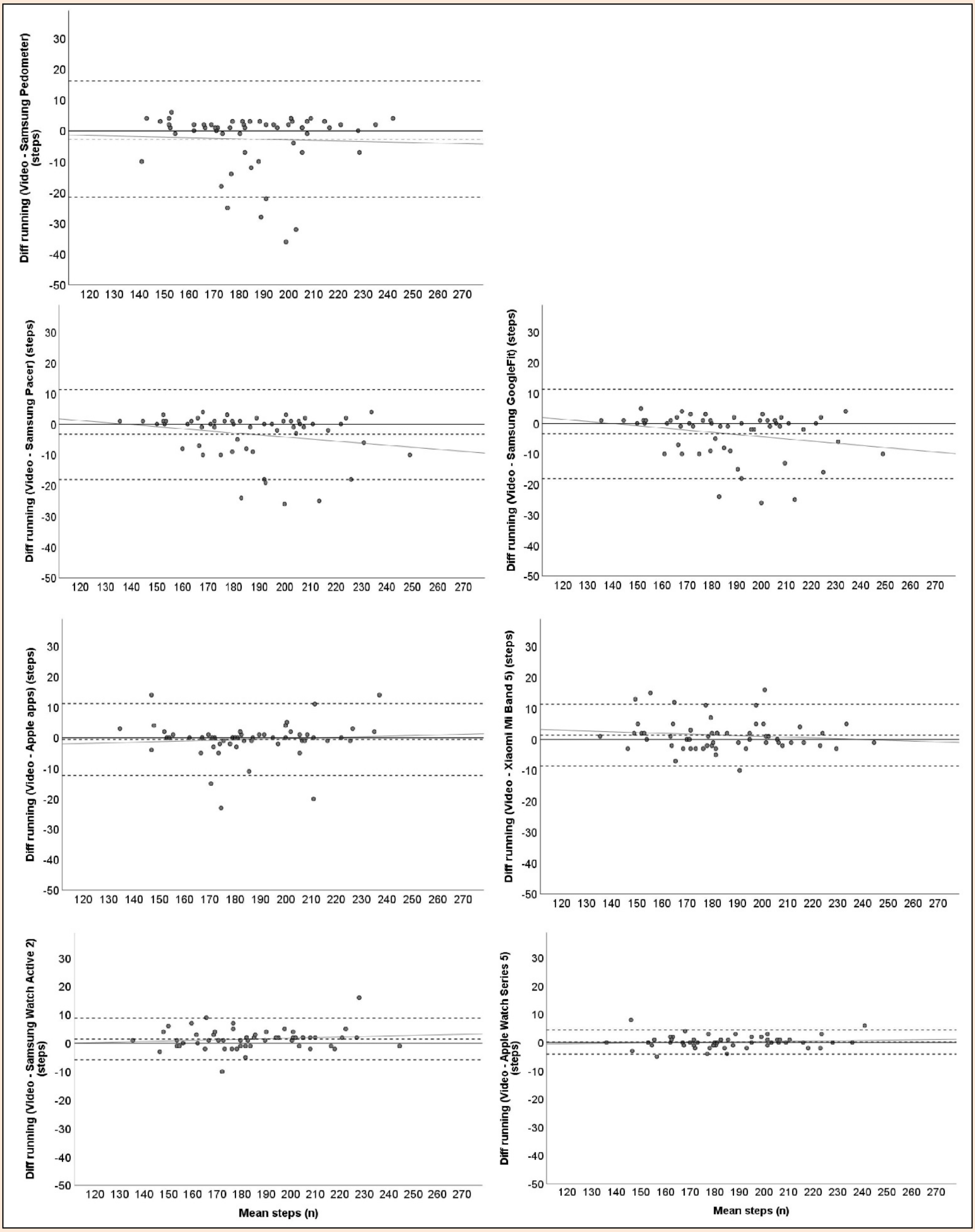


Figure 5. Bland-Altman plots of the seven devices for measuring steps under a structured free-living setting in running condition. The middle line shows the mean difference between the measurements of steps of the activity trackers and video-based step counts and the dashed lines indicate the limits of agreement.

Study 2: Validity of the wrist-worn activity trackers during daily unstructured free-living conditions

Table 4 shows the validity of the wrist-worn activity trackers for estimating daily PA (i.e., steps, moderate-to-vigorous, and total physical activity) in high school students during unstructured free-living conditions.

Regarding the validity results of the steps based on the values of 90% CI of the equivalence test, the 90% confidence interval of the steps assessed by the Xiaomi Mi Band 5 and Samsung Galaxy Watch Active 2 were inside the equivalence region of reference standard. Similarly, based on the values of both the MAPE and 95% CI of the ICC, the validity results of the steps assessed by the Xiaomi Mi Band 5 were good and excellent, respectively; and the validity results of the steps assessed by the Samsung Galaxy Watch Active 2 were acceptable and excellent based on the MAPE and 95% CI of the ICC, respectively. However, as regards the Apple Watch Series 5, the validity results showed that it was not inside the equivalence region of the reference standard (i.e., equivalence test), was poor for the values of the MAPE and questionable-excellent for those with the 95% CI of the ICC. The validity results for the MVPA and total PA assessment for the three wrist-worn activity trackers were not inside the equivalence region of the reference standard (i.e., equivalence test), and were poor for the values of both the MAPE and 95% CI of the ICC (exceptionally was questionable-excellent for the values of 95% CI of the ICC assessed by the Apple Watch Series 5). Figures 6, 7, and 8 show the LOA plots. Pearson's correlation coefficients did not show heteroscedasticity, except with the MVPA with the Xiaomi Mi Band 5 (Table 3).

Table 4. Validity of the wrist-worn activity trackers for estimating daily physical activity (i.e., steps, moderate-to-vigorous, and total physical activity) during unstructured free-living conditions (n = 51).

Instrument	Mean (SD)	Equivalence test (90% CI)	LOA (95% CI)	MAE	MAPE	ICC (95% CI)
Steps (n)						
ActiGraph wGT3X-BT	6562.1 (2662.6)	-984.32, 984.32	-	-	-	-
Samsung Galaxy Watch Active	6290.2 (3002.1)	77.8, 466.0	271.9 (-1349.2, 1893.0)	679.3	11.4	0.95 (0.92, 0.97)
Apple Watch Series 5	7316.0 (2886.6)	-1012.4, -495.4	-753.9 (-2912.8, 1405.0)	1001.1	18.0	0.89 (0.69, 0.95)
Xiaomi Mi Band 5	6688.0 (3155.9)	-347.5, 95.7	-125.9 (-1976.7, 1724.9)	591.7	9.4	0.95 (0.91, 0.97)
Moderate-to-vigorous physical activity (min)						
ActiGraph wGT3X-BT	41.8 (17.8)	-6.27, 6.27	-	-	-	-
Apple Watch Series 5	36.3 (20.8)	3.30, 7.59	5.4 (-12.4, 23.2)	8.2	22.6	0.86 (0.67, 0.93)
Xiaomi Mi Band 5	65.8 (41.8)	-34.05, -13.93	-24.0 (-108.1, 60.1)	37.8	120.6	0.09 (0.00, 0.31)
Xiaomi Mi Band 5[a]	21.3 (15.3)	17.85, 23.07	20.5 (-1.3, 42.3)	20.7	53.5	0.44 (0.00, 0.77)
Total physical activity (min)						
ActiGraph wGT3X-BT	216.1 (61.5)	-32.42, 32.42	-	-	-	-
Samsung Galaxy Watch Active	70.3 (36.6)	134.0, 158.0	145.8 (44.1, 247.5)	145.8	67.6	0.09 (0.00, 0.32)
Xiaomi Mi Band 5	83.0 (34.4)	122.0, 144.0	133.0 (37.7, 228.3)	133.0	61.3	0.12 (0.00, 0.38)
Xiaomi Mi Band 5[b]	64.6 (38.4)	140.0, 163.0	151.5 (52.3, 250.7)	151.5	70.6	0.10 (0.00, 0.34)

SD = Standard deviation; LOA = Limits of agreement; 90%/95% CI = 90%/95% confident interval; MAE = Mean absolute error; MAPE = Mean absolute percentage error; ICC = Intraclass correlation coefficient; a Brisk walking time (min); b Slow-brisk walking time (min).

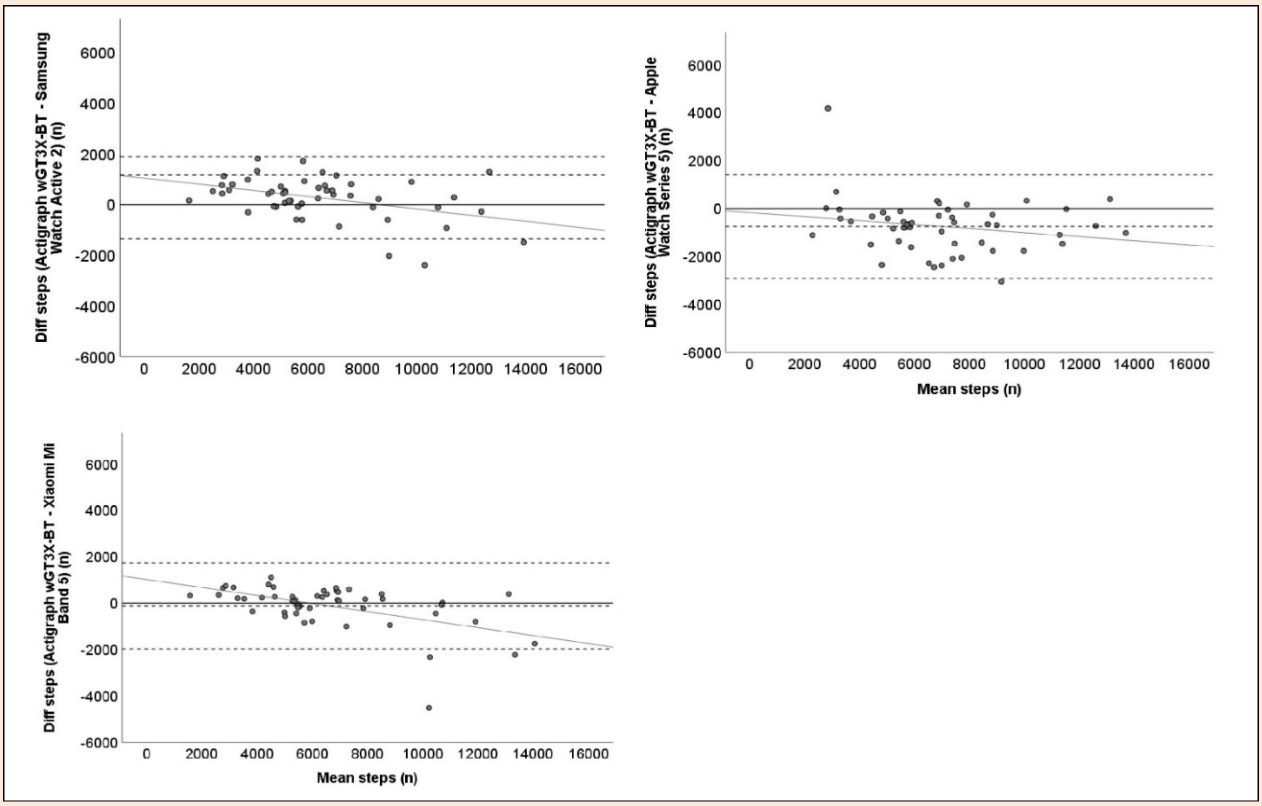


Figure 6. Bland-Altman plots of the three devices for measuring steps under unstructured free-living conditions. The middle line shows the mean difference between the measurements of steps of the three activity trackers and the ActiGraph and the dashed lines indicate the limits of agreement.

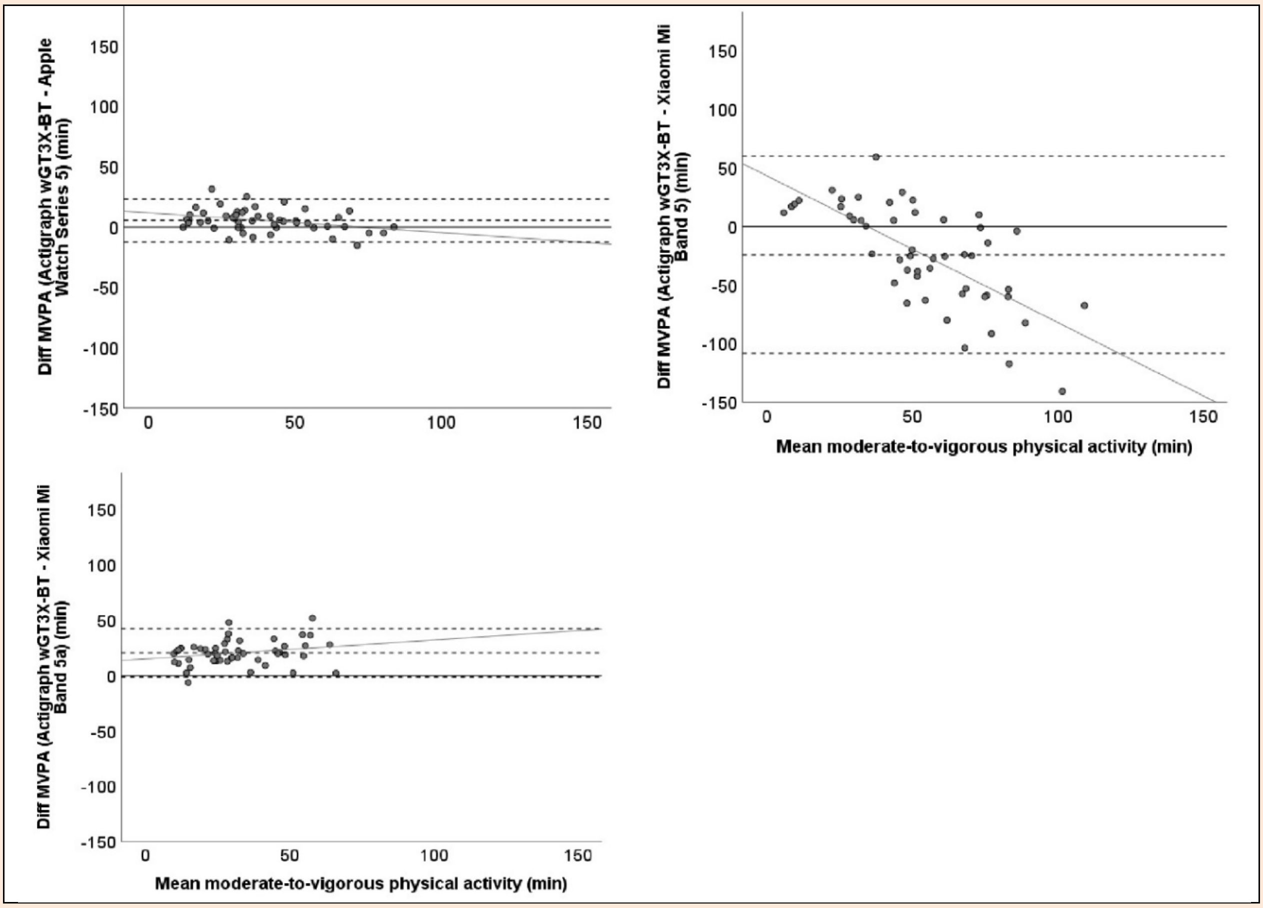


Figure 7. Bland-Altman plots of the Apple Watch Series 5 and Xiaomi Mi Band 5 trackers for measuring moderate-to-vigorous physical activity under unstructured free-living conditions. The middle line shows the mean difference between the measurements of moderate-to-vigorous physical activity of the activity trackers and the ActiGraph and the dashed lines indicate the limits of agreement.

Table 5 shows the comparability of the wearable activity trackers for estimating daily PA during unstructured free-living conditions. Based on the values of both the MAPE and 95% CI of the ICC, the comparability results showed that the steps scores of the Samsung Galaxy Watch Active 2 and Xiaomi Mi Band 5 were good-excellent. Although the Apple Watch Series 5 and Xiaomi Mi Band 5 also had an ICC acceptable-excellent value for steps, with the MAPE the value was questionable. Similarly, while the comparability results showed that the total PA scores of the Samsung Galaxy Watch Active 2 and Xiaomi Mi Band 5 (with the slow-brisk walking time) were good-excellent with the ICC, it was poor with the MAPE. For the rest of results (i.e., Samsung Galaxy Watch Active 2-Apple Watch Series 5 for steps; all the comparison with MVPA; and for total PA Samsung Galaxy Watch Active 2-Xiaomi Mi Band 5 and Xiaomi Mi Band 5 total PA time and slow-brisk walking time) the values were unacceptable-poor.

Table 5. Comparability of the wrist-worn activity trackers for estimating daily physical activity during unstructured free-living conditions (*n* = 51).

Instrument	LOA (95% CI)	MAE	MAPE	ICC (95% CI)
	Steps (n)			
Samsung Galaxy Watch Active 2-Apple Watch Series 5	-1025.8 (-3445.2, 1393.6)	1242.8	23.3	0.86 (0.51, 0.94)
Samsung Galaxy Watch Active 2-Xiaomi Mi Band 5	-397.8 (-1525.2, 729.6)	518.2	8.4	0.98 (0.91, 0.99)
Apple Watch Series 5-Xiaomi Mi Band 5	628.0 (-1895.7, 3151.7)	1103.0	19.4	0.89 (0.78, 0.94)
	Moderate-to-vigorous physical activity (min)			
Apple Watch Series 5-Xiaomi Mi Band 5	-29.4 (-114.7, 55.9)	40.3	86.0	0.10 (0.00, 0.31)
Apple Watch Series 5- Xiaomi Mi Band 5[a]	15.0 (-12.8, 42.8)	16.6	72.1	0.52 (0.00, 0.78)
Xiaomi Mi Band 5-Xiaomi Mi Band 5[a]	44.5 (-48.2, 137.2)	50.7	114.1	0.00 (0.00, 0.12)
	Total physical activity (min)			
Samsung Galaxy Watch Active 2-Xiaomi Mi Band 5	-12.8 (-32.8, 7.2)	14.4	22.9	0.90 (0.26, 0.97)
Samsung Galaxy Watch Active 2- Xiaomi Mi Band 5[b]	5.6 (-21.8, 33.0)	11.5	20.9	0.92 (0.85, 0.96)
Xiaomi Mi Band 5-Xiaomi Mi Band 5[b]	18.4 (-9.4, 46.2)	20.5	33.9	0.82 (0.07, 0.94)

SD = Standard deviation; LOA = Limits of agreement; 95% CI = 95% confident interval; MAE = Mean absolute error; MAPE = Mean absolute percentage error; ICC = Intraclass correlation coefficient; [a] Brisk walking time (min); [b] Slow-brisk walking time (min).

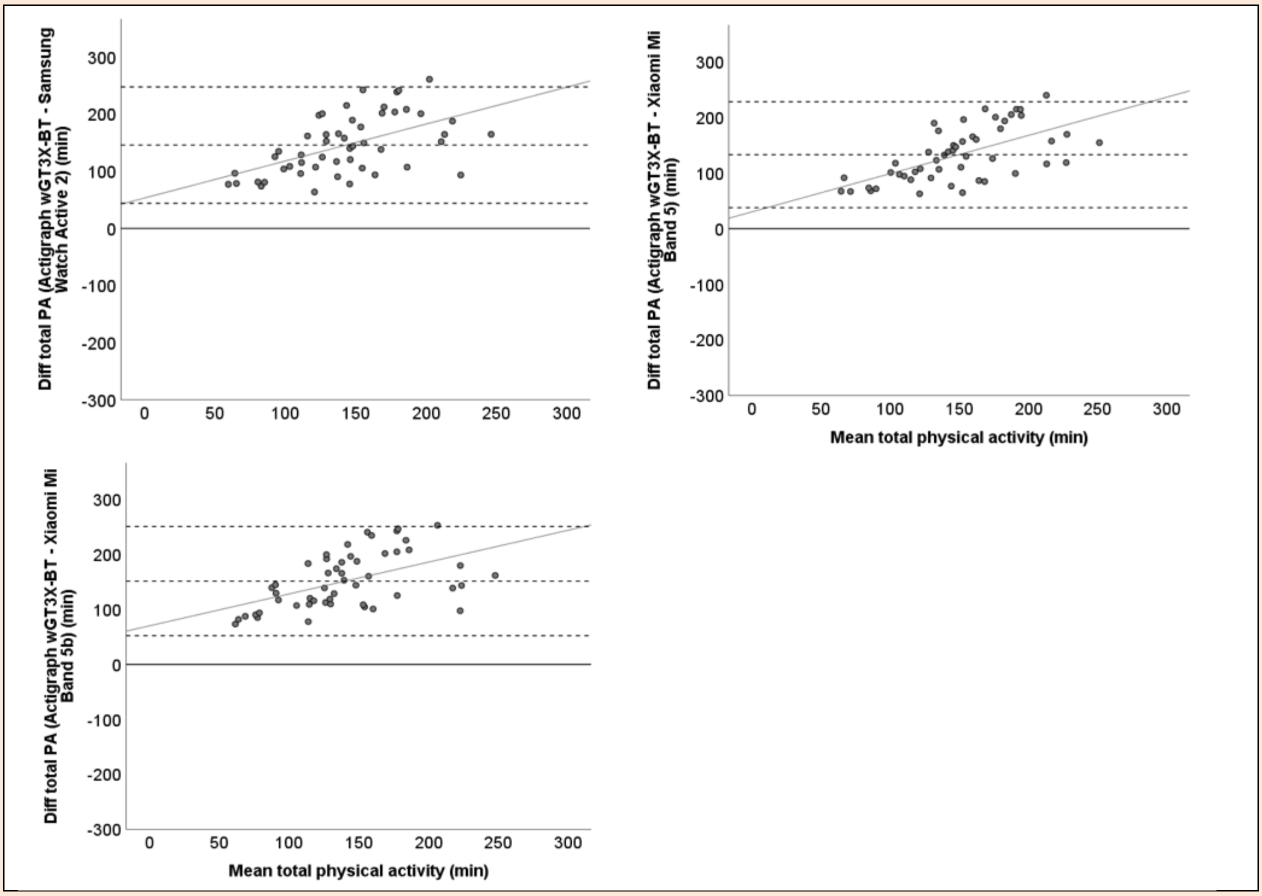


Figure 8. Bland-Altman plots of the Samsung Watch Active 2 and Xiaomi Mi Band 5 trackers for measuring total physical activity under unstructured free-living conditions. The middle line shows the mean difference between the measurements of total physical activity of the activity trackers and the ActiGraph and the dashed lines indicate the limits of agreement.

Table 6 shows the validity of the wrist-worn activity trackers for estimating the daily PA recommendations (i.e., 10,000 steps/day or 60 min. of MVPA) in high school students during unstructured free-living conditions. A total of 13.7% and 17.6% of high school students met the accelerometer-measured step- and MVPA-based recommendations, respectively. The validity results of the daily step-based recommendations assessed by the three wrist-worn activity trackers were excellent. Regarding the daily MVPA-based recommendation, while the validity results with the Apple Watch Series 5 were excellent, for the Xiaomi Mi Band 5 they were poor. Table 7 shows the comparability of the wrist-worn activity trackers for estimating the daily PA recommendations during unstructured free-living conditions. The comparability results of the daily step-based recommendations assessed by the three wrist-worn activity trackers were excellent. However, as regards the daily MVPA-based recommendation, the results with the three wrist-worn activity trackers were poor.

Discussion

Regarding the general objective of the study referring to the validity of the nine wearable activity trackers for estimating PA in high school students during free-living conditions, it is necessary to differentiate the measurement of steps and other PA parameters in structured or unstructured free-living conditions. In regard to the measurement of steps, all wearable activity trackers were valid under the two conditions. However, to measure PA parameters in unstructured free-living conditions only the Apple Watch Series 5 was valued as excellent for estimating the compliance or non-compliance of the daily recommendations based on MVPA.

Table 6. Validity of the wrist-worn activity trackers for estimating the daily physical activity recommendations during unstructured free-living conditions (n = 51).

		ActiGraph wGT3X-BT						
		10,000 steps				60 min of MVPA		
Instrument		%TP	*P*	*k*		%TP	*P*	*k*
Samsung Galaxy Watch Active 2	10,000 steps	13.7	0.96	0.83†	60 min of MVPA	-	-	-
Apple Watch Series 5		15.7	0.94	0.77†		17.6	0.96	0.87†
Xiaomi Mi Band 5		17.6	0.96	0.85†		60.8	0.41	-0.03
Xiaomi Mi Band 5[a]		-	-	-		2.0	0.84	0.17*

MVPA = Moderate-to-vigorous physical activity; %TP= Percentage of total positive cases according to the recommendation; P = Proportion of agreement; k = Kappa coefficient. [a] Brisk walking time (min). * $p < 0.05$; ‡ < 0.01; and † $p < 0.001$.

Table 7. Comparability of the wrist-worn activity trackers for estimating the daily physical activity recommendations during unstructured free-living conditions ($n = 51$).

Instrument		10,000 steps	
		P	***k***
Samsung Watch Active 2-Apple Watch Series 5	10,000 steps	0.98	0.92†
Samsung Watch Active 2-Xiaomi Mi Band 5		0.96	0.85†
Apple Watch Series 5-Xiaomi Mi Band 5		0.94	0.79†
		60 min of MVPA	
		P	***k***
Apple Watch Series 5-Xiaomi Mi Band 5	60 min of MVPA	0.45	0.04
Apple Watch Series 5-Xiaomi Mi Band 5[a]		0.84	0.17*
Xiaomi Mi Band 5-Xiaomi Mi Band 5[a]		0.37	0.00

MVPA = Moderate-to-vigorous physical activity; *P* = Proportion of agreement; *k* = Kappa coefficient. [a] Brisk walking time (min). * $p < 0.05$ and † $p < 0.001$.

Previous research coincided with the present study in regards to steps counting, showing high validity outcomes in youths, both with other wrist-worn activity tracker brands such as Jawbone or Fitbit (Evenson et al., 2015; Kang et al., 2019) and with the ones used in this tudy, although with different models of wrist-worn activity trackers (e.g., Fuller et al., 2020; Hao et al., 2021). However, similar to this study, the measurement of PA intensities has shown inadequate validity results (e.g., Degroote et al., 2020; Evenson et al., 2015; Feehan et al., 2018; Fuller et al., 2020; Voss et al., 2017).

Specifically, regarding the aim of Study 1, all cases of the wearable activity trackers were inside the equivalence region of reference standard (i.e., equivalence test), were excellent for the four conditions evaluated, taking into account the criterion of MAPE values, and good to excellent taking into account the ICC criterion. Perhaps this is the only condition and setting where there is a higher consensus in previous literature regarding wrist-worn activity trackers and their validity for measuring steps accurately (e.g., Fuller et al., 2020; Fokkema et al. 2017; or Hao et al., 2021). There were no clear differences between data obtained among devices. Being the wrist-worn activity trackers more comfortable than mobile phones to wear, and particularly being the Xiaomi Mi Band 5 the cheapest one (around 25€), it seems logical to conclude that this could be the most recommended option for measuring steps under structured free-living conditions. Therefore, in the school setting the Xiaomi Mi Band 5 could provide Physical Education teachers with the best opportunity to measure the high school students' steps in Physical Education-based health promotion programs accurately. In this sense, several aspects have to be highlighted: a) the cost of the wearables analyzed is similar or even less than other materials such as a gymnastics bench or mats which are very used in PE; b) many times, adolescents have their own mobile phones that can be used in the centers with the corresponding authorization. Therefore, it would not be necessary to buy more wearables, but rather to know which applications to use and to know if they are valid, thus being a very feasible way to promote physical activity among this population and in Physical Education classes without spending money; c) in Spain, high school centers have an annual budget distributed among all departments, and d) also, in Spain, external financing called "teaching innovation programs" exist which allow to obtain the necessary budget for buying, for example, enough wearables.

In relation to the mobile apps specifically, it is important to denote that the Apple iPhone 11 used the same algorithm for assessing PA parameters in the three evaluated apps, providing the same results for all of them. Consequently, the possibilities offered by the Apple Company related to PA parameters assessed by mobile apps are reduced in comparison with Android mobile phones.

In regard to the specific aim of Study 2, different results were obtained, and different conclusions could be deduced. First, for the measurement of daily steps, the Xiaomi Mi Band 5 obtained the best validity result (good-excellent; inside the equivalence region of reference standard); followed by the Samsung Galaxy Watch Active 2 (acceptable-excellent; inside the equivalence region of reference standard); and finally, the Apple Watch Series 5 (questionable-excellent; not inside the equivalence region of reference standard). Previous research also obtained good validity results of daily steps in unstructured free-living conditions with high school students (Schneider and Chau, 2016; Šimůnek et al., 2019; Yang et al., 2019), although the first one studied the Fitbit activity wristband detecting an overestimated registration in comparison to the ActiGraph accelerometer; the second studied the Garmin 1 and 3 activity wristband models, obtaining good levels of accuracy in comparison to the Yamax pedometer as the reference standard; and the third one studied a previous version of Xiaomi than the one used in the present study in comparison to the ActiGraph accelerometer. Secondly, for the MVPA measurement, the best validity result was obtained by the Apple Watch Series 5 with questionable-excellent values. The Xiaomi mi Band 5 presented unsatisfactory results and similar to the Apple Watch Series 5 was not inside the equivalence region of reference standard (i.e., equivalence test), and the Samsung Galaxy Watch active 2 did not provide this parameter. Previous literature with high school students' samples is scarce. On the one hand, recent and previous review studies agree with the fact that some specific models of activity wristbands have obtained adequate validity results for MVPA (Gorzelitz et al., 2020), although they only analyzed one study with high school students which examined the validity of the Fitbit activity wristband. On the other hand, a second study with high school students carried out by Yang et al. (2019) obtained inadequate validity results for MVPA. Therefore, further studies are needed in order to achieve a more consensus in this variable, which probably with the new models of activity wristbands that incorporate new algorithms they may provide a more accurate measurement of MVPA in the future. Third, for total PA, none of the wrist-worn activity trackers analyzed provided adequate validity results, and the Apple Watch Series 5 did not provide this parameter.

Finally, for the PA daily recommendations, and taking into consideration the criterion of 10,000 steps, all devices obtained excellent values of validity. In relation to the 60 min of MVPA criterion, while the Apple Watch Series 5 obtained excellent results, the Xiaomi Mi Band 5 obtained poor results (note that the Samsung Galaxy Watch Active 2 did not provide this parameter). Therefore, valuable outputs were obtained for future intervention programs because the three wrist-worn activity trackers were valid for accurately classifying high school students according to whether or not they met the recommended 10,000 steps per day or the 60 min of MVPA only in the case of Apple Watch Series 5.

Consequently, these wearables are also useful for reporting valid feedbacks for being used in high school students' progressive challenges to achieve daily recommendations in health-promotion interventions, and can be used as a motivational strategy to promote PA due to their accurate measurement of PA. Wearables' characteristics of being lightweight, easy to use and understand, and economically affordable (Parra Saldías et al., 2018), together with the results obtained in the present study, makes these devices feasible and applicable to consumers (researchers or not) for promoting better levels of PA in adolescents at any context. In this sense, recent meta-analysis of Casado-Robles et al. (2022) concluded that the use of these devices in PA promotion programs among adolescents are an effective strategy for improving their PA levels. Therefore, and specifically in the school setting, Physical Education teachers may use these wearables in order to: (a) have a better control of their students' extracurricular and out-of-school time PA (i.e., letting students self-monitor their PA and compare it with the daily PA recommendations, and also allow teachers to control students' PA via the wearables' application); (b) individualize health intervention programs for improving all students' PA levels, programming personalized alerts and daily aims throughout the personal wearable; (c) maintain the students' physical fitness levels achieved in previous Physical Education programs; and (d) plan particular students' PA programs for holiday periods, such as Christmas or summertime.

A strength of this study was that, to our knowledge, it was the first study to examine the validity of more advanced and used models of wearable activity trackers (at the moment of the data collection, International Data Corporation's Worldwide Quarterly Wearable Device Tracker Report, 2021) in structured and unstructured free-living conditions and with high school students, analyzing both PA parameters as continuous variables and consensus daily PA recommendations criteria. Furthermore, since the main goal of wearable activity trackers is to assess high school students' daily PA levels or to use them as a motivational tool during daily life, the evaluation in free-living conditions as in the present study are closer to reality and, therefore, they are meaningful and useful (Duncan et al., 2018). Taking into account that the majority of the previous studies have been performed in adults or special samples such as people with different health problems (e.g., Feehan et al., 2018 or Fokkema et al., 2017), this study expands and complements this area of research, especially among the high school student population in which the number of studies are very reduced, which is also applicable to practical interventions in order to improve health promotion and behavior change programs in regards to PA, both in Physical Education and in other health or sport contexts (e.g., Duncan et al., 2012; Strath and Rowley, 2018).

Regarding the limitations, this study used a non-probability and relatively small sample, which limits the generalizability of the obtained results to this particular context. Secondly, ActiGraph accelerometers have been shown as the most common and valid method for objectively assessing high school students' PA levels during free-living conditions (e.g., Shephard and Tudor-Locke, 2016; Van Hecke et al., 2016; Hickey et al., 2016; Trost et al., 2011; Migueles et al., 2017), considering that there is not any golden standard for those PA levels in unstructured free-living conditions. However, some methodological issues have not achieved an evidence-based consensus, such as the use of specific cut-off points which might misclassify students PA levels because they are not relative to the individual students' fitness status. Nevertheless, taking into account this issue, the best current evidence-based decisions were adopted in the present study (Migueles et al., 2017). Therefore, it may contribute to the variability of wearable activity trackers convergent validity obtained results. Finally, the third limitation is related to the models of wearable activity trackers evaluated and its associated consequences. Two repercussions could be mentioned: (a) more advanced models of wearable activity trackers are continuously appearing in the market, and scientific knowledge always lags behind this market progress; and (b) some Companies did not provide their algorithms to classify PA at different intensities (e.g., Xiaomi or Apple companies). In this sense, to overcome these difficulties, companies must be more open about the algorithms they are using to estimate PA intensities. Moreover, it would be useful to standardize the way of taking data and analyzing it, that is, the protocol to carry out the study. To do that, it might be useful to develop interdisciplinary collaborations and open source tools to allow these data to be collected. Likewise, it would be practical to standardize the way of the reports, in order to facilitate the understanding regarding the protocol followed to analyze the validity of specific wearables and how the results obtained have been interpreted. Because of all of these mentioned limitations, further studies should be performed in the future in order to improve the knowledge regarding the accuracy and valid registration of PA parameters by wearable activity trackers, both for consumers' and researchers' decisions of their personal use. Moreover, it would be interesting if future studies show a comparison between males and females in terms of the validity assessment and physical activity levels.

Conclusion

The wearable activity trackers tested in this study have shown adequate validity results in order to assess steps in both structured (i.e., wrist-worn activity trackers and mobile phone apps) and unstructured (i.e., wrist-worn activity trackers) free-living conditions for both continuous and dichotomous variables. However, in unstructured free-living conditions and for assessing MVPA, only Apple Watch Series 5 reported valid results regarding the compliance or non-compliance with the daily PA recommendations, but not for measuring total PA or MVPA as continuous variables. Therefore, depending on the user's/researcher's aim and context, one or another wearable activity tracker could be more adequate, mainly because of its valid measurements and its costs.

Acknowledgements

Authors thank Aliisa Hatten for the English revision. This study is part of the doctoral thesis of Carolina Casado-Robles, as well as of the WearFit project (Reference: B-SEJ-029-UGR18), funded by the FEDER/ Regional Government of Andalusia-Ministry of Economy and Knowledge. The experiments complied with the current laws of the country in which they were performed. The authors have no conflicts of interest to declare. The datasets generated and analyzed during the current study are not publicly available, but are available from the corresponding author who was an organizer of the study.

References

Adamakis, M. (2020) Criterion validity of wearable monitors and smarphone applications to measure physical activity energy expenditure in adolescents. *Sport Sciences for Health* **16**, 755-763. https://doi.org/10.1007/s11332-020-00654-2

Arifin, W. (2018) A web-based sample size calculator for reliability studies. *Education in Medicine Journal* **10**, 67-76. https://doi.org/10.21315/eimj2018.10.3.8

Atkinson, G. and Nevill, A. M. (1998) Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine* **26**, 217-238. https://doi.org/10.2165/00007256-199826040-00002

Bai, Y., Tompkins, C., Gell, N., Dione, D., Zhang, T. and Byun, W. (2021) Comprehensive comparison of Apple Watch and Fitbit monitors in a free-living setting. *Plos One* **16**, e0251975. https://doi.org/10.1371/journal.pone.0251975

Bland, J. and Altman, D. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1**, 307-310. https://doi.org/10.1016/S0140-6736(86)90837-8

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L., Lijmer, J. G., Moher, D., Rennie, D., de Vet, H. C., Kressel, H. Y., Rifai, N., Golub, R. M., Altman, D. G., Hooft, L., Korevaar, D. A. and Cohen, J. F. (2015) STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. *British Journal of Sports Medicine* **351**, 5527. https://doi.org/10.1136/bmj.h5527

Brickwood, K. J., Watson, G., O'Brien, J. and Williams, A. D. (2019) Consumer-based wearable activity trackers increase physical activity participation: Systematic review and meta-analysis. *JMIR mHealth and uHealth* **7**, e11819. https://doi.org/10.2196/11819

Casado-Robles, C., Viciana, J., Guijarro-Romero, S. and Mayorga-Vega, D. (2022) Effects of consumer-wearable activity tracker-based programs on objectively measured daily physical activity and sedentary behavior among school-aged children: A systematic review and meta-analysis. *Sports Medicine-Open* **8**, 18. https://doi.org/10.1186/s40798-021-00407-6

Cicchetti, D. (2001) The precision of reliability and validity estimates re-visited: Distinguishing between clinical and statistical significance of sample size requirements. *Journal of Clinical and Experimental Neuropsychology* **23**, 695-700. https://doi.org/10.1076/jcen.23.5.695.1249

Cohen, J. (1992) A power primer. *Psychological Bulletin* **112**, 155-159. https://doi.org/10.1037/0033-2909.112.1.155

Cole, T. J., Bellizzi, M. C., Flegal, K. M. and Dietz, W. H. (2000) Establishing a standard definition for child overweight and obesity worldwide: International survey. *British Journal of Sports Medicine* **320**, 1240-1243. https://doi.org/10.1136/bmj.320.7244.1240

Cosoli, G., Spinsante, S. and Scalise, L. (2020) Wrist-worn and chest-strap wearable devices: Systematic review on accuracy and metrological characteristics. *Measurement* ***159***, 107789. https://doi.org/10.1016/j.measurement.2020.107789

Crossley, S., McNarry, M. A., Rosenberg, M., Knowles, Z. R., Eslambolchilar, P. and Mackintosh, K. A. (2019) Understanding youths' ability to interpret 3D-printed physical activity data and identify associated intensity levels: Mixed-methods study. *Journal of Medical Internet Research* **21**, e11253.
, https://doi.org/10.2196/11253

Da Silva, M. P., Fontana, F. E., Callahan, E., Mazzardo, O. and De Campos, W. (2015) Step-count guidelines for children and adolescents: A systematic review. *Journal of Physical Activity and Health* **12**, 1184-1191. https://doi.org/10.1123/jpah.2014-0202

Degroote, L., Hamerlinck, G., Poels, K., Maher, C., Crombez, G., De Bourdeaudhuij, I., Vandendriessche, A., Curtis, R. G. and DeSmet, A. (2020) Low-cost consumer-based trackers to measure physical activity and sleep duration among adults in free-living conditions: Validation study. *JMIR mHealth and uHealth* **8**, e16674. https://doi.org/10.2196/16674

Dixon, P. M., Saint-Maurice, P. F., Kim, Y., Hibbing, P., Bai, Y. and Welk, G. J. (2018) A primer on the use of equivalence testing for evaluating measurement agreement. *Medicine and Science in Sports and Exercise*, **50**, 837-845. https://doi.org/10.1249/MSS.0000000000001481

Duncan, M. J., Wunderlich, K., Zhao, Y. and Faulkner, G. (2018) Walk this way: Validity evidence of iPhone health application step count in laboratory and free-living conditions. *Journal of Sports Sciences* **36**, 1695-1704. https://doi.org/10.1080/02640414.2017.1409855

Duncan, M., Birch, S. and Woodfield, L. (2012) Efficacy of an integrated school curriculum pedometer intervention to enhance physical activity and to reduce weight status in children. *European Physical Education Review* **18**, 396-407. https://doi.org/10.1177/1356336X12450799

El-Amrawy, F. and Nounou, M. I. (2015) Are Currently Available Wearable Devices for activity tracking and heart rate monitoring accurate, precise, and medically beneficial?. *Healthcare Informatics Research* **21**, 315-320. https://doi.org/10.4258/hir.2015.21.4.315

Evenson, K. R., Catellier, D. J., Gill, K., Ondrak, K. S. and McMurray, R. G. (2008) Calibration of two objective measures of physical activity for children. *Journal of Sports Sciences* **26**, 1557-1565. https://doi.org/10.1080/02640410802334196

Evenson, K. R., Goto, M. M. and Furberg, R. D. (2015) Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral Nutrition and Physical Activity* **12**, 159. https://doi.org/10.1186/s12966-015-0314-1

Feehan, L. M., Geldman, J., Sayre, E. C., Park, C., Ezzat, A. M., Yoo, J. Y., Hamilton, C. B. and Li, L. C. (2018) Accuracy of Fitbit devices: Systematic review and narrative syntheses of quantitative data. *JMIR mHealth and uHealth* **6,** e10527. https://doi.org/10.2196/10527

Fokkema, T., Kooiman, T. J., Krijnen, W. P., Van Der Schans, C. P. and De Groot, M. (2017) Reliability and validity of ten consumer activity trackers depend on walking speed. *Medicine and Science in Sports and Exercise* **49**, 793-800. https://doi.org/10.1249/MSS.0000000000001146

Fuller, D., Colwell, E., Low, J., Orychock, K., Tobin, M. A., Simango, B., Buote, R., Van Heerden, D., Luan, H., Cullen, K., Slade, L. and Taylor, N. (2020) Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: Systematic review. *JMIR mHealth and uHealth* **8**, e18694. https://doi.org/10.2196/18694

Gal, R., May, A. M., Van Overmeeren, E. J., Simons, M. and Monninkhof, E. M. (2018) The effect of physical activity interventions comprising wearables and smartphone applications on physical activity: A systematic review and meta-analysis. *Sports Medicine-Open* **4**, 1. https://doi.org/10.1186/s40798-018-0157-9

Gorzelitz, J., Farber, C., Gangnon, R. and Cadmus-Bertram, L. (2020) Accuracy of wearable trackers for measuring moderate- to vigorous-intensity physical activity: A systematic review and meta-analysis. *Journal for the Measurement of Physical Behaviour* **3**, 346-357. https://doi.org/10.1123/jmpb.2019-0072

Guthold, R., Stevens, G. A., Riley, L. M. and Bull, F. C. (2020) Global trends in insufficient physical activity among adolescents: a pooled analysis of 298 population-based surveys with 1.6 million participants. *The Lancet Child and Adolescent Health* **4**, 23-35. https://doi.org/10.1016/S2352-4642(19)30323-2

Hao, Y., Ma, X. K., Zhu, Z. and Cao, Z. B. (2021) Validity of wrist-wearable activity devices for estimating physical activity in adolescents: Comparative study. *JMIR mHealth and uHealth* **9**, e18320. https://doi.org/10.2196/18320

Henriksen, A., Haugen Mikalsen, M., Woldaregay, A. Z., Muzny, M., Hartvigsen, G., Hopstock, L. A. and Grimsgaard, S. (2018) Using fitness trackers and smartwatches to measure physical activity in research: Analysis of consumer wrist-worn wearables. *Journal of Medical Internet Research* **20**, e110. https://doi.org/10.2196/jmir.9157

Hernaez, R. (2015) Reliability and agreement studies: A guide for clinical investigators. British Journal of Sports Medicine *Gut* **64**, 1018-1027. https://doi.org/10.1136/gutjnl-2014-308619

Hickey, A., John, D., Sasaki, J. E., Mavilia, M. and Freedson, P. (2016) Validity of activity monitor step detection is related to movement patterns. *Journal of Physical Activity and Health* **13**, 145-153. https://doi.org/10.1123/jpah.2015-0203

International Data Corporation. (2021) *European Wearable Market Showed Positive Growth in 1Q21, Says IDC*. Available at: https://www.idc.com/getdoc.jsp?containerId=prEUR147995121 (Accessed: 15 January 2022).

Johnston, W., Judice, P. B., Molina García, P., Mühlen, J. M., Lykke Skovgaard, E., Stang, J., Schumann, M., Cheng, S., Bloch, W., Brønd, J. C., Ekelund, U., Grøntved, A., Caulfield, B., Ortega, F. B. and Sardinha, L. B. (2021) Recommendations for determining the validity of consumer wearable and smartphone step count: expert statement and checklist of the INTERLIVE network. *British Journal of Sports Medicine* **55**, 780-793. https://doi.org/10.1136/bjsports-2020-103147

Kang, S., Kim, Y., Byun, W., Suk, J. and Lee, J. M. (2019) Comparison of a wearable tracker with actigraph for classifying physical activity intensity and heart rate in children. *International Journal of Environmental Research and Public Health* **16**, 2663. https://doi.org/10.3390/ijerph16152663

Koo, T. and Li, M. (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine* **15**, 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Kottner, J., Audigé, L., Brorson, S., Donner, A., Gajewski, B. J., Hróbjartsson, A., Roberts, C., Shoukri, M. and Streiner, D. L. (2011) Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *Journal of Clinical Epidemiology* **64**, 96-106. https://doi.org/10.1016/j.jclinepi.2010.03.002

LaMunion, S. R., Blythe, A. L., Hibbing, P. R, Kaplan, A. S., Clendenin B. J. and Crouter, S. E. (2020) Use of consumer monitors for estimating energy expenditure in youth. *Applied Physiology Nutrition and Metabolism* **45**, 161-168. https://doi.org/10.1139/apnm-2019-0129

Landis, J. R. and Koch, G. G. (1977) The measurement of observer agreement for categorical data. *Biometrics* **33**, 159-174. https://doi.org/10.2307/2529310

Lee, J. A., Williams, S. M., Brown, D. D. and Laurson, K. R. (2015) Concurrent validation of the Actigraph gt3x+, Polar Active accelerometer, Omron HJ-720 and Yamax Digiwalker SW-701 pedometer step counts in lab-based and free-living settings. *Journal of Sports Sciences* **33**, 991-1000. https://doi.org/10.1080/02640414.2014.981848

Maher, C., Ryan, J., Ambrosi, C. and Edney, S. (2017) Users' experiences of wearable activity trackers: A cross-sectional study. *BioMed Central Public Health* **17**, 880. https://doi.org/10.1186/s12889-017-4888-1

Mayorga-Vega, D., Casado-Robles, C., López-Fernández, I. and Viciana, J. (2021) A comparison of the utility of different step-indices to translate the physical activity recommendation in adolescents. *Journal of Sports Sciences* **39**, 469-479. https://doi.org/10.1080/02640414.2020.1826667

Mayorga-Vega, D., Casado-Robles, C., Viciana, J. and López-Fernández, I. (2019) Daily step-based recommendations related to moderate-to-vigorous physical activity and sedentary behavior in adolescents. *Journal of Sports Science and Medicine* **18**, 586-595.

Migueles, J. H., Cadenas-Sanchez, C., Ekelund, U., Delisle Nyström, C., Mora-Gonzalez, J., Löf, M., Labayen, I., Ruiz, J. R. and Ortega, F. B. (2017) Accelerometer data collection and processing criteria to assess physical activity and other outcomes: A systematic review and practical considerations. *Sports Medicine* **47**, 1821-1845. https://doi.org/10.1007/s40279-017-0716-0

Nunnally, J. (1978) *Psychometric theory*. 2nd edition. New york: McGraw-Hill.

O'Neill, B., McDonough, S. M., Wilson, J. J., Bradbury, I., Hayes, K., Kirk, A., Kent, L., Cosgrove, D., Bradley, J. M. and Tully, M. A. (2017) Comparing accelerometer, pedometer and a questionnaire for measuring physical activity in bronchiectasis: a validity and feasibility study? *Respiratory Research* **18**, 16. https://doi.org/10.1186/s12931-016-0497-2

OECD. (2019) *OECD Future of Education 2030. Making Physical Education dynamic and inclusive for 2030. International curriculum analysis*. OECD Publishing.

Oliver, M., Badland, H. M., Schofield, G. M. and Shepherd, J. (2011) Identification of accelerometer nonwear time and sedentary behavior. *Research Quarterly for Exercise and Sport* **82**, 779-783. https://doi.org/10.1080/02701367.2011.10599814

Paluch, A. E., Bajpai, S., Bassett, D. R., Carnethon, M. R., Ekelund, U., Evenson, K. R., ... and Fulton, J. E. (2022) Daily steps and all-cause mortality: a meta-analysis of 15 international cohorts. *The Lancet Public Health* **7**, e219-e228. https://doi.org/10.1016/S2468-2667(21)00302-9

Parra Saldías, M., Mayorga-Vega, D., López-Fernández, I. and Viciana, J. (2018) How many daily steps are really enough for adolescents? A cross-validation study. *Retos* **33**, 241-246. https://doi.org/10.47197/retos.v0i33.55504

Poitras, V., Gray, C., Borghese, M., Carson, V., Chaput, J., Janssen, I., Katzmarzyk, P. T., Pate, R. R., Connor Gorber, S., Kho, M. E., Sampson, M. and Tremblay, M. S. (2016) Systematic review of the relationships between objectively measured physical activity and health indicators in school-aged children and youth. *Applied Physiology, Nutrition, and Metabolism* **41**, 197-239. https://doi.org/10.1139/apnm-2015-0663

Redenius, N., Kim, Y. and Byun, W. (2019) Concurrent validity of the Fitbit for assessing sedentary behavior and moderate-to-vigorous physical activity. *BioMed Central Medical Research Methodology* **19**, 29. https://doi.org/10.1186/s12874-019-0668-1

Romanzini, M., Petroski, E. L., Ohara, D., Dourado, A. C. and Reichert, F. F. (2014) Calibration of ActiGraph GT3X, Actical and RT3 accelerometers in adolescents. *European Journal of Sport Science* **14**, 91-99. https://doi.org/10.1080/17461391.2012.732614

Rowlands, A. V., Harrington, D. M., Bodicoat, D. H., Davies, M. J., Sherar, L. B., Gorely, T., Khunti, K. and Edwardson, C. L. (2018) Compliance of adolescent girls to repeated deployments of wrist-worn accelerometers. *Medicine and Science in Sports and Exercise* **50**, 1508-1517. https://doi.org/10.1249/MSS.0000000000001588

Schneider, M. and Chau, L. (2016) Validation of the Fitbit Zip for monitoring physical activity among free-living adolescents. *BioMed Central Research Notes* **9**, 448. https://doi.org/10.1186/s13104-016-2253-6

Shephard, R. and Tudor-Locke, C. (2016) The objective monitoring of physical activity: Contributions of accelerometry to epidemiology, exercise science and rehabilitation. 1st edition. Switzerland: Springer. https://doi.org/10.1007/978-3-319-29577-0

Šimůnek, A., Dygrýn, J., Jakubec, L., Neuls, F., Frömel, K. and Welk, G. J. (2019) Validity of Garmin Vívofit 1 and Garmin Vívofit 3 for school-based physical activity monitoring. *Pediatric Exercise Science* **31**, 130-136. https://doi.org/10.1123/pes.2018-0019

Stamm, A. and Hartanto, R. (2018) Feature extraction from mems accelerometer and motion tracking measurements in comparison with smart bands during running. *Proceedings* **2**, 197-202. https://doi.org/10.3390/proceedings2060197

Stewart, A., Marfell-Jones, M., Olds, T. and De Ridder, J. (2011) International standards for anthropometric assessment. New Zealand: International Society for the Advancement of Kinanthropometry.

Strath, S. J. and Rowley, T. W. (2018) Wearables for Promoting Physical Activity. *Clinical Chemistry* **64**, 53-63. https://doi.org/10.1373/clinchem.2017.272369

The International Data Corporation's Worldwide Quarterly Wearable Device Tracker reports (2021) Available from URL: https://www.idc.com/getdoc.jsp?containerId=IDC_P31315

Toth, L. P., Park, S., Springer, C. M., Feyerabend, M. D., Steeves, J. A. and Bassett, D. R. (2018) Video-recorded validation of wearable step counters under free-living conditions. *Medicine and Science in Sports and Exercise* **50**, 1315-1322. https://doi.org/10.1249/MSS.0000000000001569

Trost, S. G., Loprinzi, P. D., Moore, R. and Pfeiffer, K. A. (2011) Comparison of accelerometer cut points for predicting activity intensity in youth. *Medicine and Science in Sports and Exercise* **43**, 1360-1368. https://doi.org/10.1249/MSS.0b013e318206476e

Tudor-Locke, C., Craig, C. L., Beets, M. W., Belton, S., Cardon, G. M., Duncan, S., Hatano, Y., Lubans, D. R., Olds, T. S., Raustorp, A., Rowe, D. A., Spence, J. C., Tanaka, S. and Blair, S. N. (2011) How many steps/day are enough? for children and adolescents. *International Journal of Behavioral Nutrition and Physical Activity* **8**, 78. https://doi.org/10.1186/1479-5868-8-78

Vailshery, L. S. (2021) Wearables Statistics & facts. Availabe at: https://www.statista.com/topics/1556/wearable-technology (Accessed: 15 January 2022).

Van Hecke, L., Loyen, A., Verloigne, M., van der Ploeg, H. P., Lakerveld, J., Brug, J., De Bourdeaudhuij, I., Ekelund, U., Donnelly, A., Hendriksen, I. and Deforche, B. (2016) Variation in population levels of physical activity in European children and adolescents according to cross-European studies: a systematic literature review within DEDIPAC. *International Journal of Behavioral Nutrition and Physical Activity* **13**, 70. https://doi.org/10.1186/s12966-016-0396-4

Voss, C., Gardner, R. F., Dean, P. H. and Harris, K. C. (2017) Validity of commercial activity trackers in children with congenital heart disease. *Canadian Journal of Cardiology* **33**, 799-805. https://doi.org/10.1016/j.cjca.2016.11.024

Westerterp, K. R. (2013) Physical activity and physical activity induced energy expenditure in humans: measurement, determinants, and effects. *Frontiers in Physiology* **4**, 90. https://doi.org/10.3389/fphys.2013.00090

Westerterp, K. R. (2017) Doubly labelled water assessment of energy expenditure: Principle, practice, and promise. *European Journal of Applied Physiology* **117**, 1277-1285. https://doi.org/10.1007/s00421-017-3641-x

Willmott, C. and Matsuura, K. (2005) Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research* **30**, 79-82. https://doi.org/10.3354/cr030079

World Health Organization. (2020) WHO guidelines on physical activity and sedentary behaviour. Geneva: World Health Organization.

Yang, X., Jago, R., Zhang, Q., Wang, Y. Y., Zhang, J. and Zhao, W. H. (2019) Validity and reliability of the wristband activity monitor in free-living children aged 10-17 years. *Biomedical and Environmental Sciences* **32**, 812-822.

Key points

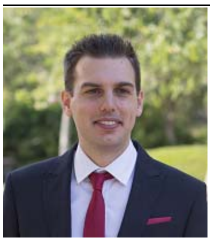
- The Samsung Galaxy Watch Active 2, the Apple Watch Series 5, the Xiaomi Mi Band 5, and the mobile apps (i.e., Pedometer and Pacer for android and iPhone mobiles, Google Fit for android, and Apple Health for iPhone mobiles) were valid to measure high school students' steps under structured and unstructured free-living conditions as a continuous variable.
- The Samsung Galaxy Watch Active 2, the Apple Watch Series 5, and the Xiaomi Mi Band 5 wrist-worn activity trackers were valid to measure the compliance or non-compliance of high school students' daily step-based physical activity recommendation (i.e., 10,000 steps/day) in unstructured free-living conditions.
- The Apple Watch Series 5 wrist-worn activity tracker was valid to measure the compliance or non-compliance of high school students' moderate-to-vigorous physical activity-based recommendation (i.e., 60 minutes of moderate-to-vigorous physical activity/day) in unstructured free-living conditions.
- Considering the user's/researcher's objective and context, one or another wearable activity tracker could be more adequate, mainly because of its valid measurements and its costs.

AUTHOR BIOGRAPHY

Jesús VICIANA
Employment
Department of Physical Education and Sport, University of Granada, Spain.
Degree
PhD
Research interests
Measurement and evaluation, health-enhancing physical activity, sedentary behavior, Physical Education-based interventions, motivation toward physical activity.
E-mail: jviciana@ugr.es

Carolina CASADO-ROBLES
Employment
Department of Physical Education and Sport, University of Granada, Spain.
Degree
PhD
Research interests
Measurement and evaluation, health-enhancing physical activity, sedentary behavior, Physical Education-based interventions, motivation toward physical activity.
E-mail: ccasado@ugr.es

Santiago GUIJARRO-ROMERO
Employment
Department of Didactic of Musical, Plastic and Corporal Expression, University of Valladolid, Valladolid, Spain
Degree
PhD
Research interests
Measurement and evaluation, health-enhancing physical activity, sedentary behavior, Physical Education-based interventions, motivation toward physical activity.
E-mail: santigr93@gmail.com



Daniel MAYORGA-VEGA
Employment
Departamento de Didáctica de las Lenguas, las Artes y el Deporte, Facultad de Ciencias de la Educación, Universidad de Málaga, Málaga, España.
Degree
PhD
Research interests
Measurement and evaluation, health-enhancing physical activity, sedentary behavior, health-related physical fitness, Physical Education-based interventions, motivation toward physical activity.
E-mail: dmayorgavega@gmail.com

✉ **Santiago Guijarro-Romero**
Department of Didactic of Musical, Plastic and Corporal Expression, Faculty of Education and Social Work, University of Valladolid, Paseo de Belén, 1, 47011, Valladolid, Spain