Title	Criterion validity of wearable monitors and smartphone applications to measure physical activity energy expenditure in adolescents
Author(s)	Adamakis, Manolis
Publication date	2020-06-04
Original citation	Adamakis, M. (2020) 'Criterion validity of wearable monitors and smartphone applications to measure physical activity energy expenditure in adolescents', Sport Sciences for Health. doi: 10.1007/s11332-020-00654-2
Type of publication	Article (peer-reviewed)
Link to publisher's version	http://dx.doi.org/10.1007/s11332-020-00654-2 Access to the full text of the published version may require a subscription.
Rights	**© 2020, Springer-Verlag Italia S.r.l., part of Springer Nature. This is a post-peer-review, pre-copyedit version of an article published in Sport Sciences for Health. The final authenticated version is available online at: http://dx.doi.org/10.1007/s11332-020-00654-2**
Embargo information	Access to this article is restricted until 12 months after publication by request of the publisher.
Embargo lift date	2021-06-04
Item downloaded from	http://hdl.handle.net/10468/10366

University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Criterion validity of wearable monitors and smartphone applications to measure physical activity energy expenditure in adolescents

Manolis Adamakis[1,2]

Affiliations:

1. School of Education, University College Cork, Ireland

2. School of Physical Education and Sport Science, National and Kapodistrian University of Athens, Greece

Telephone number: +353 89 965 6773

Email: manosadam@phed.uoa.gr

ORCID: 0000-0003-3881-6892

Abstract

Purpose: This study examined the validity of one researched-based, one GPS and two consumer-based monitors, as well as six freeware Android apps (three pedometer- and three GPS-based apps) in a sample of healthy adolescents, during self-paced outdoor walking and running.

Methods: Twenty-one adolescents (15.9 ± 2.0 years) participated in this cross-sectional study. They walked and ran a distance of 1.2 km for each trial. They were fitted with SenseWear Armband Pro 3, Garmin Forerunner 310XT, Garmin Vivofit, Medisana Vifit, and smartphones running the Runkeeper, Runtastic, Sports Tracker, Pedometer, Accupedo, Pedometer and Pedometer 2.0 apps. Estimation of PAEE was compared to measurement from indirect calorimetry. Repeated measures ANOVA, mean absolute percentage errors and Bland-Altman plots assessed accuracy and proportional bias.

Results: PAEE estimates from all monitors and apps showed large individual errors, ranging from 13.16% for walking (Runtastic) to 37.46% for running (Vifit). For group-level differences, Forerunner, Runkeeper, Runtastic and Accupedo significantly underestimated PAEE for walking, and SenseWear, Forerunner, Runkeeper, Vifit and Pedometer significantly underestimated PAEE for running.

Conclusion: Based on individual errors, none of the monitors and apps tested was accurate for estimating PAEE in adolescents. The only app that had an acceptable error was Runtastic during running. These monitors and apps are not suitable as research measurement tools for recording precise and accurate PAEE estimates.

Keywords

Physical activity monitor, pedometer, GPS, smartphone app, wearable technology, physical activity measurement, energy expenditure.

Abbreviations

PA: Physical activity

EE: Energy expenditure

REE: Resting energy expenditure

PAEE: Physical activity energy expenditure

MAPE: Mean absolute percent error

GPS: Global positioning system

Introduction

Regular physical activity (PA), in addition to disease prevention, provides a variety of extra benefits that help individuals sleep better, feel better, and perform daily tasks more easily [1]. These benefits can be achieved in many ways and walking interventions have shown clinically relevant improvements for many cardiovascular disease risk factors [2]. Furthermore, adequate adherence to exercise is important for the effectiveness of any intervention. Among strategies to improve adherence in PA programmes, the use of electronic devices, health and fitness applications (apps), and wearable monitors have been suggested to help with monitoring of activities and PA promotion [3-5]. The study by Harris et al. [6] further suggests that primary care-based PA interventions may have lasting effects on PA levels, as well as health outcomes.

In order to promote regular PA through walking, the accurate and reliable assessment of PA is necessary [7]. As mentioned by Nelson, Kaminsky, Dickin and Montoye [8], measurement accuracy is important when tracking PA variables to provide meaningful measures of PA. The concern regarding the accuracy of wearable monitors is a continual problem and it is important to balance the cost and feasibility of a monitor against its validity and reliability [9, 10]. There is definitely a need for low-cost, easy-to-use, accurate PA tracking devices to use as both intervention and assessment tools in health promotion research related to PA [11].

A recent meta-analysis examined the accuracy of 40 wrist and arm-worn activity monitors' estimates of total energy expenditure (EE), which revealed that EE estimates differed in accuracy depending on activity type. The inclusion of heart rate and heat sensors in monitors improved EE estimates relative to accelerometry alone, and research-based monitors were more accurate for total EE [12]. In addition, a review of reviews on various techniques of EE measurement in adults concluded that most activity monitors underestimated EE compared to Doubly Labelled Water technique [9].

The above-mentioned reviews included studies with adult populations, and the vast majority examined the accuracy of total EE estimates against criterion measures. Currently limited evidence regarding PA monitors' validity of EE in children and adolescents exists. Furthermore, EE measured during similar study designs included resting energy expenditure (REE), thus providing estimates of total EE, which does not reflect net physical activity EE (PAEE). PAEE may contribute to total EE less, however it is the most variable and unpredictable component of total EE and has the potential to increase it significantly [13].

Because of the relationship between the amount of PA and health outcomes [1], accurate estimation of net PAEE using wearable monitors is required.

Recently, in addition to PA monitors, smartphones have been considered a powerful tool with which to study large-scale population health on a global scale, due to their vast adoption among individuals in developed and developing countries [14]. Many smartphone apps are now available for counting steps, estimating caloric expenditure, and tracking traverced distances [via Global Positioning System (GPS)]. Even though there are some attempts to validate a few smartphone apps, these studies have focused on the accuracy of step detection while walking on a treadmill, rather than validating EE estimates [15-17]. The validation of PA apps is even more urgent nowadays since similar apps with low accuracy and insufficient selection justification [18, 19] are used in large observational studies with big datasets [20].

Precise PA measurement is important for health research and thus, the aim of the present study was to validate the PAEE estimates of one researched-based, one GPS and two consumer-based monitors, as well as six Android apps (three pedometer- and three GPS-based apps) in a sample of healthy adolescents. The main outcome was the criterion validity of PAEE estimates retrieved from the above-mentioned monitors and apps, compared to indirect calorimetry (i.e. Cosmed K4b2). Based on the evaluation framework proposed by Keadle, Lyden, Strath, Staudenmayer and Freedson [21], we used a naturalistic validation study design in real-world conditions, which included self-paced outdoor walking and running. The submaximal outdoor walking and running tests were performed in regular outdoor conditions in Greece with the aim of providing data from uncontrolled and sometimes challenging conditions, where participants would train and perform their regular fitness activities.

Methods

Study design

The design of the study was cross-sectional in nature and 21 healthy adolescents, with no contraindications for exercise and no known orthopaedic limitations that would prevent them from completing the assessments, participated. All adolescents, as well as their parents, read and signed an informed consent document approved by the School of Sport Science and Physical Education of Athens Research Ethics Committee, informing them of the risks and benefits of the study.

Participants reported to the researchers twice. During the first visit, anthropometric and resting EE (REE) measures were obtained in controlled laboratory settings. The second visit (i.e., 2 - 3 days after the first visit) took place in a track and field elliptical stadium. Field tests were performed outdoors between November and January in Greece, in regular winter exercising conditions, that is, during days when it was not raining, and the temperature was above 10 °C. The participants were instructed to wear their own outdoor sports clothing as appropriate for the current weather during the test. These conditions are typical outdoor training conditions in Greece and, hence, provide a good benchmark for challenging real outdoor training conditions that are faced by adolescents while exercising.

The participants were fitted with the portable metabolic analyser (i.e. Cosmed K4b2), four different activity trackers and three smartphones, each one running simultaneously two different apps (one GPS and one pedometer-based app). SenseWear was worn on the nondominant arm. Vivofit was worn on the right wrist and Forerunner on the left wrist. Vifit, as well as the three smartphones, were strapped close to the body on a waist-worn elastic belt over the left hip, near the anterior axillary line, and were counterbalanced for anterior and posterior placement on the hip among participants. All devices were updated with the participants' age, sex, height, dominant hand, weight[1] and length stride. All monitors' firmware and apps' software were updated to the latest available version. In addition to the devices, a heart rate monitor (Garmin HRM-Dual™) was placed around participants' chest to capture exercise heart rate and to incorporate this measurement in the EE algorithm of the two monitors (i.e. Vivofit and Forerunner).

Participants had to perform a total of two field tests in regular outdoor conditions: overground walking and submaximal running, at a self-selected pace. The only limitation that existed was that walking speed should be between 3 and 6 km/h and running speed should be above 8 km/h, following the American College of Sports Medicine [22] recommendations (speed between 6 and 8 km/h is considered a transitional speed between walking and running and should be avoided in experimental procedures). The actual average walking speed, estimated by Forerunnner, was M = 5.27 km/h (SD = 0.62) for walking and M = 11.05 km/h (SD = 1.47) for running, respectively. Between the two trials, participants could rest for 5 minutes and all devices were paused simultaneously. During pause, all apps' specific settings were changed from walking to running option.

[1] In addition to each participant's weight, 1 kg was added in order to take into account the extra weight of the portable analyser (i.e. if a participant weighted 66 kg, a total of 67 (66 + 1) kg was entered in the device).

Distance was recorded with a manual distance measuring wheel [Roadrunner RR182 (Keson)], by measuring the walking route two times and then taking the mean distance for an ending point. The total distance that all participants had to walk, and run was 1.2 km for each trial (2.4 km in total). Smartphones were set to airplane mode to avoid interactions with the mobile phone providers (i.e. no data connection), and all devices were activated simultaneously. In the end of each trial, data initially were stored manually and at a later time were uploaded to the relating devices' software.

Participants

Twenty one healthy adolescents (n = 10 boys, n = 11 girls) with an age range of 12-18 years (15.9 ± 2.0 years), body mass index range of 15.1 - 28.6 kg/m^2 (21.5 ± 4.3 kg/m^2) were screened and participated in the study (with no drop-outs). Power analysis was not conducted because of a lack of previously published (or a priori) identified effect sizes of adolescents' PEAA estimates from smartphone apps.

Anthropometric assessment

Standing height was measured to the nearest 0.1 cm using a wall mounted Harpenden stadiometer (Harpenden, London, UK) using standard procedures. Body mass was measured with participants in light clothes and bare feet on an electronic scale (Omron BF-511) to the nearest 0.1 kg. Body mass index was calculated as weight (kg) / height squared (m^2). Body fat percentage was assessed using a 3-skinfold measurement (i.e. triceps, subscapular and calf skinfolds), taken with a Harpenden caliper (Harpenden, HSK-BI, UK) and the equations for adolescents proposed by Slaughter et al. [23]. All anthropometric measurement results are presented in Table 1.

--- Insert Table 1 approximately here ---

Metabolic parameters assessment

Cosmed K4b2 (Cosmed): The Cosmed (Cosmed S.r.I., Rome, Italy) was used to measure oxygen consumption (VO_2) and carbon dioxide production (VCO_2) during the entire study protocol and was the criterion measure for EE. Breath-by-breath measurements were collected using Hans Rudolph masks (Hans Rudolf, Inc., Kansas City, MO) worn by participants and were used to determine VO_2 and VCO_2 in litres per minute (L/min), which was converted to EE with the Weir equation [24]. Volume and gas calibrations were performed before each trial following manufacturer's instructions.

In order to estimate REE, during the first laboratory visit participants were asked to lay down for 10 min and then were fitted with the portable metabolic analyser for 15 min, as well as a heart rate monitor (Garmin HRM-Dual™) to measure resting heart rate. The estimated REE was expressed as kilocalories per day. The REE measurement was performed in the morning (i.e., 6.00 - 10.00) after a 10-hour fast, following previously published guidelines [25]. Participants were instructed to avoid vigorous exercise the day before the testing and to eat their usual diet.

Wearable monitors

SenseWear Armband Pro 3 (SenseWear): The SenseWear (BodyMedia Inc., Pittsburgh, PA, USA) is an innovative, multisensor research-based high-cost armband monitor that integrates movement data from a three-dimensional accelerometer with various heat-related variables and galvanic skin response to estimate EE. The ProConnect software was used to assess EE data.

Garmin Forerunner 310XT (Forerunner): Forerunner (Garmin International Inc., Olathe, KS, U.S.A.) is a mid-cost GPS-enabled training and heart rate wrist-worn monitor for multisport athletes. It tracks time, distance, average and lap speed and pace, heart rate with a premium heart rate monitor, on land and estimated calorie burn. The main method for EE estimation on Garmin fitness monitors uses the Firstbeat algorithm [26]. The calculation takes into account the user's inputted variables including gender, height, weight and fitness class. It then combines the data with heart rate information from the heart rate strap. More specifically, it evaluates the time between heartbeats (beat to beat) to determine estimated Metabolic Equivalent (MET), which in turn is used to determine actual work expenditure.

Garmin Vivofit (Vivofit): Vivofit (Garmin International Inc., Olathe, KS, U.S.A.) is a mid-cost wrist-based, triaxial accelerometer-based monitor that measures steps taken, distance travelled, calories expended and sleep quality. When paired with a Garmin heart rate chest strap, the device can also measure the user's heart rate and incorporate this measurement in the EE estimation algorithm. The Garmin Connect software was used to assess EE data for both Vivofit and Forerunner.

Medisana Vifit (Vifit): Vifit (Medisana AG, Neuss, Germany) is a low-cost waist-worn accelerometer that counts and keeps track of steps taken and calories burned. By means of a triaxial accelerometer and altimeter technology it records all daily and nightly PA. In comparison to more sophisticated PA monitors, it only has the option to insert walking stride length for PAEE estimation (instead of both walking and running). ViFit also measures the duration and quality of sleep. The VitaDock Online software was used to assess EE data.

Android apps

This study used three Samsung smartphones S8 based on the Android operating system. Inclusion criteria for all applications were: (1) Free of charge indefinitely after download. Applications with a free trial period of finite length were excluded; (2) Full and efficient functionality after downloading, without additional software download being necessary; (3) Functionality only through the built-in accelerometer for the pedometers and GPS for the GPS apps (no 3g/4g signal); (4) Ability to record the number of steps taken, average speed, total distance and energy expenditure; (5) Adjustable sensitivity settings for the pedometers; (6) Manual input of demographic and somatometric data (sex, age, weight, height and step length for walking and running) for accurate EE estimation; (7) Manual choice activity type (i.e. walking or running); (8) Among the most popular and downloadable applications, according to users' ratings and number of downloads from the Google Play Store (as mentioned in the Store on the 23 March 2018). Specifically, for the pedometer apps, they should include an option to capture steps taken during walking and running separately, by inputting different stride length for the two conditions.

Runkeeper: Runkeeper (ASICS Digital, Inc.) is an app designed to log several outdoor sports by GPS, such as running, walking, skiing or skating. This app is able to record all the basic data: duration, distance, speed, pace and elevation. Based on these variables and additional information (i.e., body weight, gender and age), the app also estimates total calories burned [27].

Runtastic running app (Runtastic): Runtastic app (Runtastic GmbH) captures all the basic data: distance, average speed, speed between mile markers, elevation, pace, pace between mile markers, duration, calories burned and route as plotted on a map using GPS. It can be used in many different outdoor or indoor activities such as running, cycling, playing tennis, etc. [28].

Sports Tracker running cycling (Sports Tracker): Sports Tracker app (Sports Tracking Technologies) tracks calories burned, average training, cycling speed and more, with the use of GPS maps, time and distance calculators.

Pedometer: Pedometer app (ITO Technologies, Inc.) records the number of steps walked and displays them along with the number of calories that have been burned, distance, walking time and speed per hour, with the use of the smartphone-based accelerometer.

Accupedo Pedometer (Accupedo): Accupedo (Corusen LLC) is a pedometer app that monitors daily walking and calculates the physical activity level. The accuracy of this app is based on triaxial motion recognition algorithms which track walking patterns by filtering and rejecting non-walking activities. In addition, this app has enough display modes such as steps, distance, minutes and calories [29].

Pedometer 2.0: Pedometer 2.0 (DSD) counts steps, calories, distance, speed, average speed, time in motion, takes all sorts of graphics and split table in different modes, according to BMI. Furthermore, it is the only application with a self-calibration capability, which was used in order to determine the appropriate sensitivity settings for every participant separately.

Statistical analysis

Breath-by-breath data from the indirect calorimetry were transformed with the implementation of longitudinal interpolation [30] into second by second values that were aggregated to provide total estimates of EE for the walking and running trials separately. The net PAEE for each device was then calculated by subtracting REE from total EE.

Descriptive analyses were conducted to examine associations with the criterion measure. Repeated measures analysis of variance (ANOVA) statistical tests were performed to assess differences from all monitors and apps, and criterion measures for PAEE. When the test statistic was significant, post-hoc pairwise comparisons with Bonferroni correction were performed. The significance level was set at $P < 0.05$ and the partial η^2 was presented as a measure of effect size for F-tests. A partial η^2 value between 0.01 and 0.06 was associated with a small effect, between 0.06 and 0.14 with a medium effect, and 0.14 or greater with a large effect [31].

Mean absolute percent errors (MAPE) were also calculated to provide an indicator of overall measurement error {MAPE = [(monitor measurement-criterion measure) / criterion measure] x 100} and was used as an outcome measure. A smaller MAPE represents better accuracy, and less than 10% can be considered acceptable for TEE [32].

To further evaluate individual variations in a more systematic way, Bland-Altman plots with corresponding 95% limits of agreement and fitted lines (from regression analyses between mean and difference) with their corresponding parameters (i.e., intercept and slope) were presented [33]. A fitted line that provides a slope of 0 and an intercept of 0 exemplifies perfect agreement. The statistical analyses were performed with SPSS version 23.0 for Windows (IBM SPSS Corp., Armonk, NY, USA) and MedCalc 12.7 (MedCalc Software bvba).

Results

Participants averaged 64 ± 20 kcal during walking and 75 ± 20 kcal during running, respectively. The repeated measures ANOVA for both walking [$F(5,93) = 3.91$, $P = 0.004$, $\eta^2 = 0.16$] and running [$F(2,48) = 8.82$, $P < 0.001$, $\eta^2 = 0.31$] were statistically significant, with large effect sizes. The post-hoc pairwise comparisons with Bonferroni corrections showed that Forerunner [$F(1,20) = 12.87$, $P = 0.002$], Runkeeper [$F(1,20) = 4.61$, $P = 0.044$], Runtastic [$F(1,20) = 6.42$, $P = 0.020$] and Accupedo [$F(1,20) = 4.75$, $P = 0.041$] significantly underestimated PAEE for walking, while SenseWear [$F(1,20) = 10.17$, $P = 0.005$], Forerunner [$F(1,20) = 18.28$, $P < 0.001$], Runkeeper [$F(1,20) = 7.57$, $P = 0.012$], Vifit [$F(1,20) = 90.79$, $P < 0.001$] and Pedometer [$F(1,20) = 4.54$, $P = 0.046$] significantly underestimated PAEE for running (Table 2).

--- Insert Table 2 approximately here ---

Figure 1 reports MAPE for all monitors and apps tested. During walking the magnitude of errors was least for Vifit (17.39%), while error rates for all other were above 20.0% (20.20% - 34.11%). During running the magnitude of errors was least for Runtastic (13.16%), followed by SenseWear (15.82%) and Runkeeper (18.22%). Error rates for the other monitors ranged from 21.24% to 37.46%.

--- Insert Figure 1 approximately here ---

The Bland-Altman results for PAEE for both walking and running trials are presented in Table 3 (Bland-Altman plots are included in the Supplementary file). For walking, the plots revealed the narrowest 95% limits of agreement for SenseWear (difference = 1.4 kcal) and slightly higher values for Sports Tracker (difference = -2.2 kcal), Vifit (difference = 2.4 kcal) and Pedometer (difference = -2.9 kcal), while values were the highest for Runkeeper (difference = 12.5 kcal) and Forerunner (difference = 15.6 kcal). During running, the narrowest 95% limits of agreement were observed for Runtastic (difference = 1.6 kcal), followed by Vivofit (difference = -5.5 kcal) and Runkeeper (difference = 8.1 kcal). The highest values were observed for Forerunner (difference = 20.7 kcal) and Vifit (difference = 27.3 kcal).

--- Insert Table 3 approximately here ---

Discussion

The aim of the present study was to examine the accuracy of a variety of PA monitors and smartphone apps in estimating PAEE during self-paced outdoor walking and running in a sample of healthy adolescents. To our knowledge, this is the first study to examine these estimates from competing technologies, including both GPS-accelerometer monitors and smartphone apps, in youth. The primary finding from this study was that estimated PAEE from all monitors and apps had large individual errors, with the highest MAPE coming from Vifit during running (37.46%) and the lowest MAPE coming from Runtastic during running (13.16%). During walking, most individual errors ranged between 20% and 30%. Some monitors and apps tended to have lower group-level errors; however, none of these devices performed accurately during both conditions. One important finding was that freeware apps had comparable accuracy levels with the PA monitors.

Due to a lack of studies in adolescents, comparisons for the PA monitors from the current study were limited to studies in adult populations. Only one recent study had tested PA monitors in youth and similarly found that these monitors had large group- and individual-level errors for estimating EE [34]. The results of the current study were comparable to those of studies in adults, in which individual errors ranged between 10% and 30% [e.g. 32, 35, 36].

The inclusion of a research-grade monitor (SenseWear) as a comparison measure provided an advantage of this study, since previous studies had reported low total EE errors [27, 37, 38]. SenseWear was slightly more accurate in total compared to the remaining monitors and apps, however the individual-group errors were significant for a research-grade monitor. These results can be attributed to different, more structured validation protocols and activities in previous studies, as well as the fact that the present study was the first of its kind to validate the most recent algorithm 9.03 developed by BodyMedia company. This algorithm was considered to be more effective for estimating EE in youth, however the current results do not support this hypothesis.

The two Garmin monitors (Forerunner and Vivofit) that were tested included heart rate technology to accelerometry to improve estimates of EE. Forerunner incorporates a sophisticated algorithm developed by Firstbeat Technologies [26], which combines somatometric data with heart rate information for more accurate EE estimates. A recent meta-analysis in adults concluded that the inclusion of heart rate or heat sensors in monitors could improve estimates of EE relative to accelerometry alone [12], however this was not supported by the current results. Vivofit and Forerunner had large individual errors for both walking and running conditions (>20%), while group-level errors were not significant only for Vivofit. A possible explanation for this inconsistency is the variability in estimates of heart rate from heart rate sensors, which is a common finding [12]. It is evident that these heart rate combined monitoring sensors cannot be considered adequate for estimating PAEE in adolescents during walking and running, and the proprietary algorithms used should be further developed and validated.

The low-cost monitor validated in the present study (Vifit) was the only one placed around the waist over the hip, as proposed by the manufacturer. Vifit performed quite accurately during walking, with minimum group- and individual-level errors, however these errors increased significantly during running. A possible explanation for the different results in the two conditions is that this monitor uses one algorithm for estimating PA parameters and this algorithm does not differentiate between various activities (i.e. walking vs running), failing to take into account the elevated energy consumption during running. Vifit can be considered suitable only for light activities in adolescents, such as brisk walking, and this generally supports PA measurement accuracy with hip-worn compared to wrist-worn accelerometers.

A unique aspect of this validation study was the inclusion of freeware GPS- and accelerometer-based smartphone apps. All previous approaches examined the validity of apps regarding step and distance counting (e.g. 15, 16, 39]. In general, apps had larger errors than monitors, based on the individual activity. The most accurate GPS app was Runtastic, which had the lowest error of all tested monitors and apps during running, while all three pedometer-based apps' performance was similar with large individual errors over 20%. Furthermore, the pedometer-based apps showed patterns of proportional systematic bias of underestimating PAEE for running.

Previous studies on apps' step accuracy using Android smartphones showed an unacceptable error percentage of all apps during walking [16, 39]. It is possible that this initial step error existed in the present study and, since step and distance counting are the primary outcome measures of these apps, subsequently the estimation of PAEE was further miscalculated by the respective transformation algorithms. In other words, high step and distance counting errors may lead to unacceptable PAEE errors for Android PA apps, as well as monitors. This can be considered a limitation of the present study, as step and distance estimates were not taken into account during analysis.

Regarding smartphones' position, we adopted waist placement for the smartphones (and subsequently for the apps) strapped close to the body, because the waist is close to the centre of mass of the human body, and the torso occupies the most mass of a human body. According to Yang and Hsu [40], this implies that the accelerations measured by a single sensor at this location can better represent the major human motion and a range of basic daily activities, including walking, can be classified more accurately according to the accelerations measured from a waist-worn sensor. Even though there are contradictory findings regarding the presumed possible impact of the phone's position on the accuracy of step detection during walking at 6.0 km/h [15 vs 17], we are uncertain whether the accuracy would improve if smartphones were placed i.e. around the arm.

When comparing the GPS monitor and the freeware GPS apps, Forerunner outperformed only Sports Tracker in both conditions, and all apps during walking. Runtastic and Runkeeper had substantially lower individual errors than Forerunner during running, while Forerunner also showed the highest group-level errors of all monitors and apps. On the contrary, all pedometer apps performed more accurately during walking than running.

The main strengths of this study included the selection of monitors using various technologies (i.e. accelerometry, GPS and combined sensors) to estimate PAEE, and the comparison to a criterion measure of EE (indirect calorimetry). Other strengths included a sample consisting of adolescents, even distribution of boys and girls, submaximal outdoor walking and running tests in a realistic setting and randomization of the two activities to prevent systematic bias in the measurement. Limitations to this study included the limited sample size consisting of healthy participants performing two PA sessions, while future studies should include more semi- or un-structured activities in free-living environment. In addition, future studies should examine the validity of apps during activities of daily living, preferably over a time frame of 2-4 days to assess the suitability of these devices to be used for long-term accelerometry. Finally, the role of smartphone's optimal position on the human body during exercise should be further investigated.

Conclusion

As growing evidence demonstrates the associations between PA-exercise and morbidity and mortality, more research and refinements in EE estimations and in the ability of PA monitors and apps to record PAEE is clearly needed. Based on individual errors, none of the monitors and apps tested in this study was accurate compared to indirect calorimetry for estimating PAEE in adolescents, during self-paced walking and running. The only app that had an acceptable error was Runtastic during running. Group-level errors were lower for some of the monitors and apps, and freeware smartphone apps had comparable estimates of PAEE with those of consumer-based PA monitors. Group-level and individual errors were greater in adolescents compared to previous studies in adults, further indicating the need for independent validation studies in each population. Even though there is a need for low-cost, easy-to-use tracking devices for assessment and interventions [11], the results of the present study do not support the use of similar monitors and apps for assessment purposes. These devices should not be used as research measurement tools for recording precise and accurate PAEE estimates, and may be only suitable for use in large interventions of behaviour change, due to direct feedback provided to users and the minimum, or no, financial cost of the apps. Since these monitors and apps are widely used in everyday life, practitioners and adolescent trainees and athletes should be aware of these limitations regarding inaccuracy of PAEE estimation. Similar studies should be continuously conducted as technological advancement results in the annual release of new generations and software versions of consumer-based PA monitors and apps.

Disclosure statement

No potential conflict of interest was reported by the author.

References

1. Physical Activity Guidelines Advisory Committee, 2018. Physical Activity Guidelines Advisory Committee Scientific Report. Washington, DC: U.S. Department of Health and Human Services. https://health.gov/paguidelines/second-edition/report/pdf/PAG_Advisory_Committee_Report.pdf
2. Oja P, Kelly P, Murtagh EM et al (2018) Effects of frequency, intensity, duration and volume of walking interventions on CVD risk factors: a systematic review and meta-regression analysis of randomised controlled trials among inactive healthy adults. Br J Sports Med 52:769-775. doi: 10.1136/bjsports-2017-098558
3. Gal R, May AM, van Overmeeren EJ, Simons M, Monninkhof EM (2018). The effect of physical activity interventions comprising wearables and smartphone applications on physical activity: A systematic review and meta-analysis. Sport Med-Open 4(1):42. doi: 10.1186/s40798-018-0157-9

4. Mateo GF, Granado-Font E, Ferré-Grau C, Montaña-Carreras X (2015). Mobile phone apps to promote weight loss and increase physical activity: a systematic review and meta-analysis. J Med Internet Res 17(11):e253. doi: 10.2196/jmir.4836
5. Schoeppe S, Alley S, Van Lippevelde W et al (2016). Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 13:127. doi: 10.1186/s12966-016-0454-y
6. Harris T, Limb ES, Hosking F, et al (2019). Effect of pedometer-based walking interventions on long-term health outcomes: Prospective 4-year follow-up of two randomised controlled trials using routine primary care data. PLoS Med 16(6):e1002836. doi: 10.1371/journal.pmed.1002836
7. Rowlands AV, Eston RG (2007) The measurement and interpretation of children's physical activity. J Sport Sci Med 6:270-276. PMID: 24149412
8. Nelson MB, Kaminsky LA, Dickin DC, Montoye AH (2016) Validity of consumer-based physical activity monitors for specific activity types. Med Sci Sports Exerc 48(8):1619-1628. doi: 10.1249/MSS.0000000000000933
9. Dowd KP, Szeklicki R, Minetto MA et al (2018) A systematic literature review of reviews on techniques for physical activity measurement in adults: A DEDIPAC study. Int J Behav Nutr Phys Act 15:15. doi: 10.1186/s12966-017-0636-2
10. Holden SL, Baghurst TM (2018) Considerations when choosing a fitness tracking device. Strategies, 31,3, 54-56. doi: 10.1080/08924562.2018.1445891
11. Turner-McGrievy G, Jake-Schoffman DE, Singletary C et al (2018). Using commercial physical activity trackers for health promotion research: Four case studies. Health Promot Pract 20(3):381-389. doi: 10.1177/1524839918769559
12. O'Driscoll R, Turicchi J, Beaulieu K et al (2018) How well do activity monitors estimate energy expenditure? A systematic review and meta-analysis of the validity of current technologies. Br J Sports Med 2018 Sep 7. pii: bjsports-2018-099643. doi: 10.1136/bjsports-2018-099643. [Epub ahead of print]
13. Hills AP, Mokhtar N, Byrne NM (2014) Assessment of physical activity and energy expenditure: an overview of objective measures. Front Nutr 1:5. doi: 10.3389/fnut.2014.00005
14. Anthes E (2016) Mental health: There's an app for that. Nature 532:20-23. doi: 10.1038/532020a

15. Höchsmann C, Knaier R, Eymann J et al (2018) Validity of activity trackers, smartphones, and phone applications to measure steps in various walking conditions. Scand J Med Sci Spor 28(7):1818-1827. doi: 10.1111/sms.13074
16. Poojary J, Arora E, Britto A et al (2018) Validity of mobile-based technology vs direct observation in measuring number of steps and distance walked in 6 minutes. Mayo Clin Proc 93(12):1873-1874. doi: 10.1016/j.mayocp.2018.09.003
17. Presset B, Laurenczy B, Malatesta D, Barral J (2018) Accuracy of a smartphone pedometer application according to different speeds and mobile phone locations in a laboratory context. J Exerc Sci Fit 16:43-48. doi: 10.1016/j.jesf.2018.05.001
18. Adamakis M (2019) Physical activity in the era of mHealth big data: Considerations on accuracy and bias. SSP J Sport Sci Med, 2:6-10.
19. Brodie MA, Pliner EM, Ho A et al (2018) Big data vs accurate data in health research: Large-scale physical activity monitoring, smartphones, wearable devices and risk of unconscious bias. Med Hypotheses 119:32-36. doi: 10.1016/j.mehy.2018.07.015
20. Althoff T, Sosič R, Hicks JL et al (2017) Large-scale physical activity data reveal worldwide activity inequality. Nature 547(7663):336-339. doi: 10.1038/nature23018
21. Keadle SK, Lyden KA, Strath SJ, Staudenmayer JW, Freedson PS (2019) A framework to evaluate devices that assess physical behavior. Exerc Sport Sci Rev 47(4):206-214. doi: 10.1249/JES.0000000000000206
22. American College of Sports Medicine (2006) ACSM's guidelines for exercise testing and prescription (7th ed). Lippincott Williams and Wilkins, Baltimore, MD.
23. Slaughter MH, Lohman TG, Boileau RA et al (1988) Skinfold equations for estimating of body fatness in children and youth. Hum Biol 60(5):709-723. PMID:3224965
24. Weir JB (1949) New methods for calculating metabolic rate with special reference to protein metabolism. J Physiol 109:1-9. PubMed: 15394301
25. Compher C, Frankenfield D, Keim N, Roth-Yousey L (2006) Best practice methods to apply to measurement of resting metabolic rate in adults: a systematic review. J Am Diet Assoc 106(6):881-903. doi:10.1016/j.jada.2006.02.009
26. Firstbeat Technologies (2012) An energy expenditure estimation method based on heart rate measurement. Firstbeat Technologies Ltd. https://assets.firstbeat.com/firstbeat/uploads/2015/11/white_paper_energy_expenditure_estimation.pdf

27. Martinez-Nicolas A, Muntaner-Mas A, Ortega FB (2017) Runkeeper: a complete app for monitoring outdoor sports. Br J Sports Med 51:1560-1561. doi:10.1136/bjsports-2016-096678

28. Monroy Antón A, Rodríguez Rodríguez B (2016) Runtastic PRO app: an excellent all-rounder for logging fitness. Br J Sports Med 50:705-706.

29. Milanović Z, Stojiljković N, Pavlović L et al (2016) Accupedo pedometer: daily walking step counter. Br J Sports Med 50:1417-1418.

30. Noor MN, Yahaya AS, Ramli NA, Al Bakri AMM (2013) Filling missing data using interpolation methods: Study on the effect of fitting distribution. Eng Mater, 594-595:889-895. doi: 10.4028/www.scientific.net/KEM.594-595.889

31. Warner RM (2012) Applied statistics: From bivariate through multivariate techniques (2nd ed). Sage, Los Angeles, CA.

32. Lee JM, Kim Y, Welk GJ (2014) Validity of consumer-based physical activity monitors. Med Sci Sports Exerc 46(9):1840-1848. doi: 10.1249/MSS.0000000000000287

33. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 8:307-310. PMID: 2868172

34. LaMunion SR, Blythe AL, Hibbing PR et al (2019) Use of consumer monitors for estimating energy expenditure in youth. Appl Physiol Nutr Metab [Epub ahead of print]. doi: 10.1139/apnm-2019-0129

35. Bai Y, Welk GJ, Nam YH et al (2016) Comparison of consumer and research monitors under semi-structured settings. Med Sci Sports Exerc 48(1):151-158. doi: 10.1249/MSS.0000000000000727

36. Chowdhury EA, Western MJ, Nightingale TE, Peacock OJ, Thompson D (2017) Assessment of laboratory and daily energy expenditure estimates from consumer multi-sensor physical activity monitors. PLoS One 12(2):e0171720. doi: 10.1371/journal.pone.0171720

37. Calabró MA, Stewart JM, Welk GJ (2013) Validation of pattern-recognition monitors in children using doubly labeled water. Med Sci Sports Exerc 45(7):1313-1322. doi: 10.1249/MSS.0b013e31828579c3.

38. Lee J-M, Kim Y, Bai Y, et al (2014) Validation of the SenseWear Mini armband in children during semi-structured activity settings. J Sci Med Sport 19(1):41-45. doi: 10.1016/j.jsams.2014.10.004

39. Leong JY, Wong JE (2017) Accuracy of three Android-based pedometer applications in laboratory and free-living settings. J Sports Sci 35:14-21. doi: 10.1080/02640414.2016.1154592

40. Yang C-C, Hsu Y-L (2010) A review of accelerometry-based wearable motion detectors for physical activity monitoring. Sensors, 10:7772-7788. doi: 10.3390/s100807772

Table 1. Participants' characteristics (Mean ± SD)

	Boys	Girls	Total
	M ± SD	M ± SD	M ± SD
Age (years)	15.2 ± 2.3	16.6 ± 1.7	15.9 ± 2.0
Weight (kg)	70.0 ± 11.7	55.2 ± 7.4	62.3 ± 9.2
Height (m)	1.73 ± 0.11	1.66 ± 0.06	1.69 ± 0.09
Body Mass Index (kg/m^2)	23.0 ± 4.6	20.0 ± 2.0	21.5 ± 4.3
Resting heart rate (bpm)	72.9 ± 9.1	68.4 ± 7.4	70.5 ± 8.4
Body fat (%)	17.7 ± 6.7	20.6 ± 3.0	19.2 ± 5.2
Resting Energy Expenditure (kcals/day)	2354.4 ± 727.7	1964.6 ± 306.4	2150.2 ± 570.1

Table 2. Results of repeated measures ANOVA PAEE (kcal) and comparison with criterion measure.

	Walking					Running				
	M	SD	Pairwise F	Pairwise P	95% CI	M	SD	Pairwise F	Pairwise P	95% CI
Criterion	64	20	-	-	-	75	18	-	-	-
SenseWear	63	17	0.14	0.71	(-13) - 16	65**	22	10.17	0.005	(-2) - 20
Forerunner	49**	15	12.87	0.002	(-1) - 32	54**	16	18.28	< 0.001	(-2) - 40
Runkeeper	52*	16	4.61	0.044	(-10) - 35	66*	14	7.57	0.012	(-3) - 19
Runtastic	54*	16	6.42	0.02	(-5) - 25	73	21	0.43	0.52	(-8) - 11
Sports Tracker	66	14	0.14	0.71	(-25) - 21	86	15	3.53	0.08	(-34) - 12
Vivofit	57	25	1.49	0.24	(-15) - 29	80	25	1.83	0.19	(-21) - 10
Vifit	60	19	1.80	0.19	(-7) - 15	47**	18	90.79	< 0.001	(-16) - 38
Pedometer	67	26	0.43	0.52	(-20) - 14	65*	31	4.54	0.046	(-8) - 28
Accupedo	57*	16	4.75	0.041	(-6) - 20	85	35	3.72	0.07	(-32) - 11
Pedometer 2.0	71	31	2.04	0.17	(-26) - 12	93	60	3.56	0.07	(-57) - 20

*P < 0.05, **P < 0.01

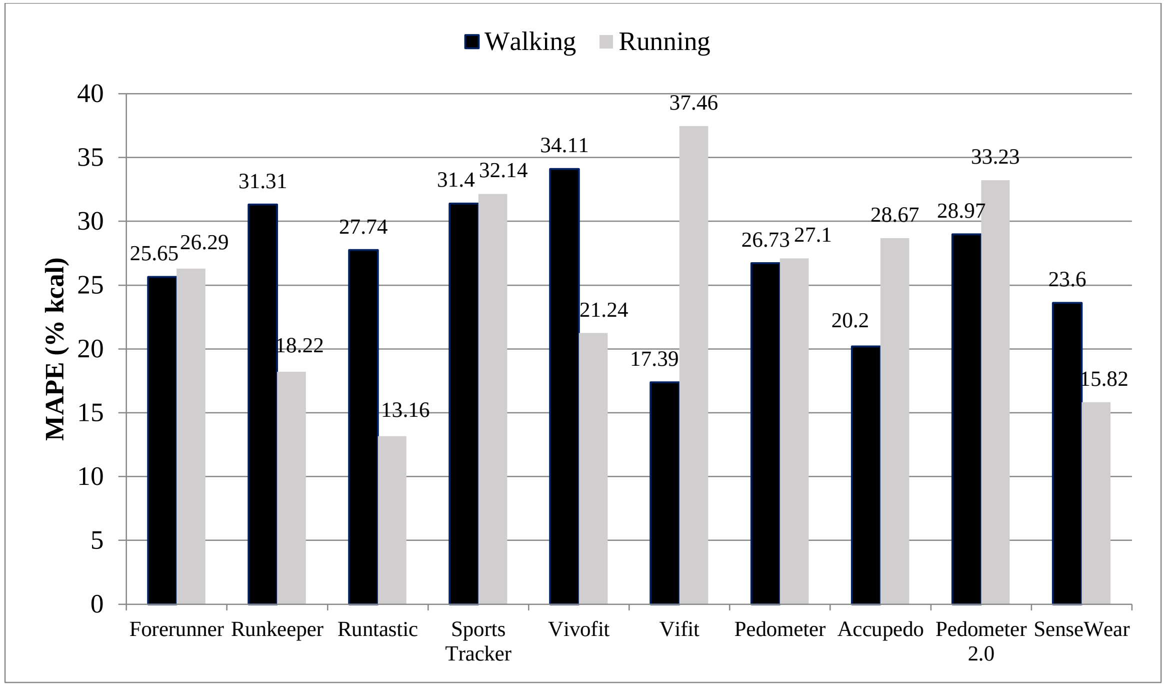

Figure 1. MAPE (% kcal) of PA monitors and apps compared with criterion measure.

Table 3. PAEE Bland-Altman results during walking and running.

	Walking					Running				
	M diff	95% CI	Slope	P	95% CI	M diff	95% CI	Slope	P	95% CI
SenseWear	1.40	(-6.47) - 9.27	0.18	0.47	(-0.32) - .67	9.06	3.14 - 14.99	-0.19	0.22	(-0.51) - 0.13
Forerunner	15.58	6.52 - 24.64	0.36	0.25	(-0.27) - 0.99	20.72	10.61 - 30.83	0.16	0.41	(-0.64) - 0.96
Runkeeper	12.49	0.35 - 24.62	0.48	0.34	(-0.56) - 1.52	8.07	1.95 - 14.19	0.31	0.13	(-0.10) - 0.73
Runtastic	9.82	1.73 - 17.90	0.24	0.34	(-0.280 - 0.77	1.63	(-3.58) - 6.83	-0.18	0.19	(-0.46) - 0.10
Sports Tracker	-2.18	(-14.39) - 10.02	0.96	0.09	(-0.18) - 2.11	-11.04	(-23.30) - 1.22	0.54	0.40	(-0.76) - 1.83
Vivofit	6.93	(-4.90) - 18.77	-0.34	0.30	(-1.00) - 0.32	-5.54	(-17.07) - 2.99	-0.40	0.06	(-0.81) - 0.01
Vifit	2.38	(-2.12) - 9.78	0.05	0.75	(-0.29) - 0.40	27.31	21.33 - 33.29	-0.02	0.92	(-0.39) - 0.35
Pedometer	-2.90	(-12.11) - 6.31	-0.31	0.17	(-0.76) - 0.14	9.77	(0.21) - 19.33	-0.60**	0.001	(-0.92 - (-0.27)
Accupedo	7.33	0.32 - 14.35	0.22	0.30	(-0.21) - 0.66	-10.76	(-22.39) - 0.87	-0.74**	<0.001	(-1.08 - (-0.40)
Pedometer 2.0	-6.90	(-16.96) - 3.17	-0.51**	0.01	(-0.89) - (-0.12)	-18.71	(-39.40) - 1.99	-1.13**	<0.001	(-1.33) - (-0.94)

*P < 0.01, **P < 0.001

Bland-Altman plots

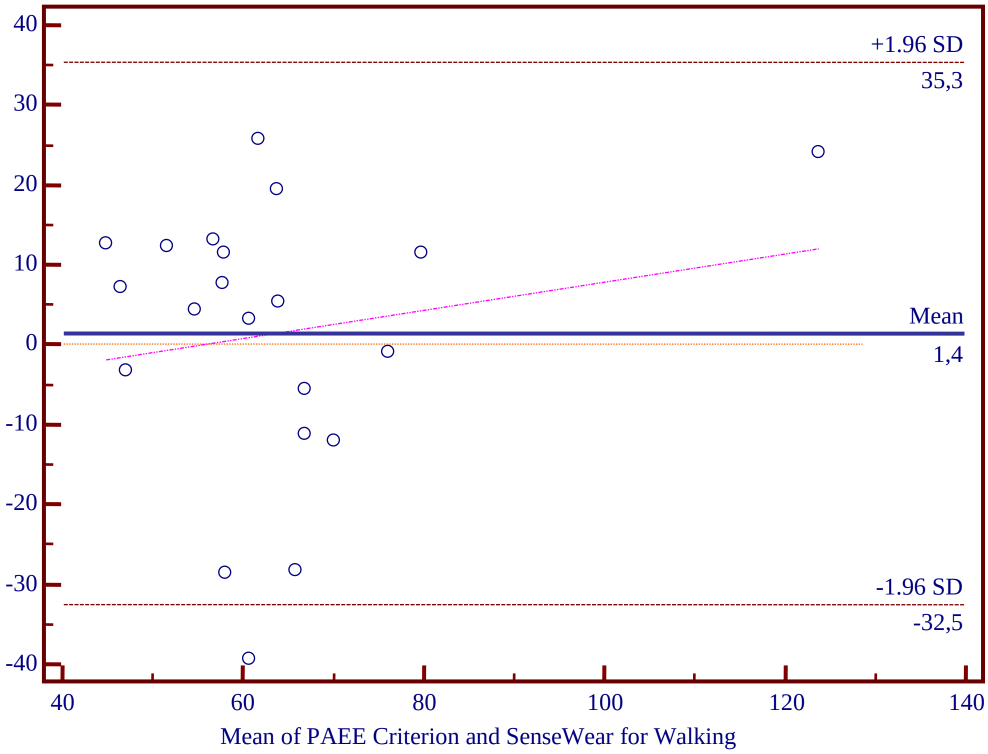

Figure 2. Bland-Altman plot for SenseWear PAEE estimates during walking compared with criterion measure.

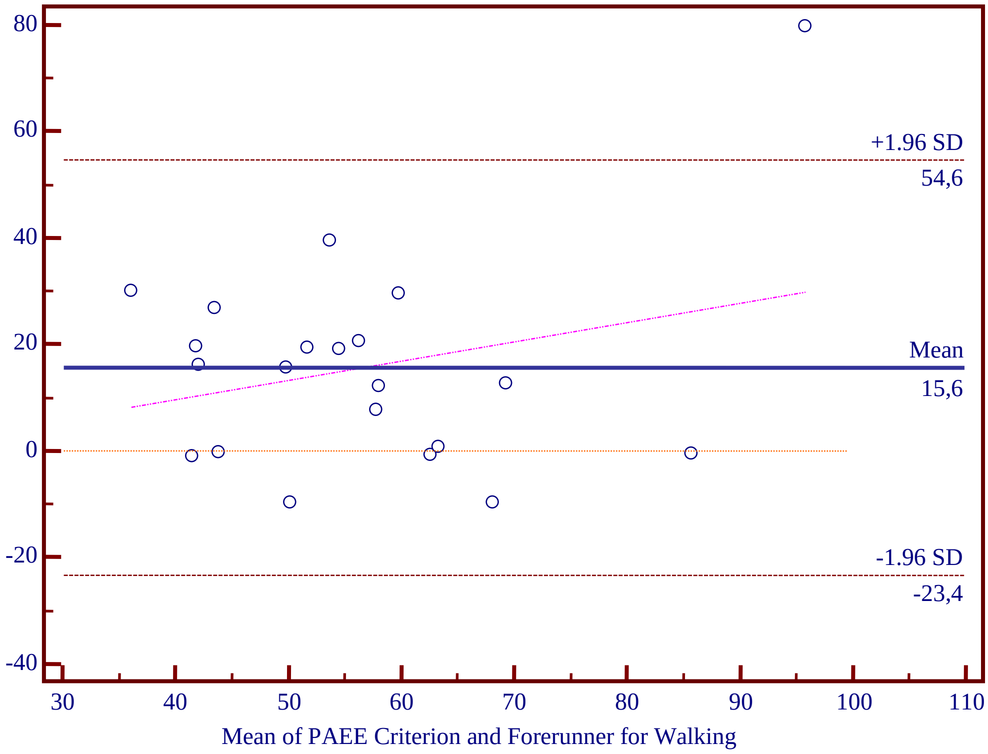

Figure 3. Bland-Altman plot for Forerunner PAEE estimates during walking compared with criterion measure.

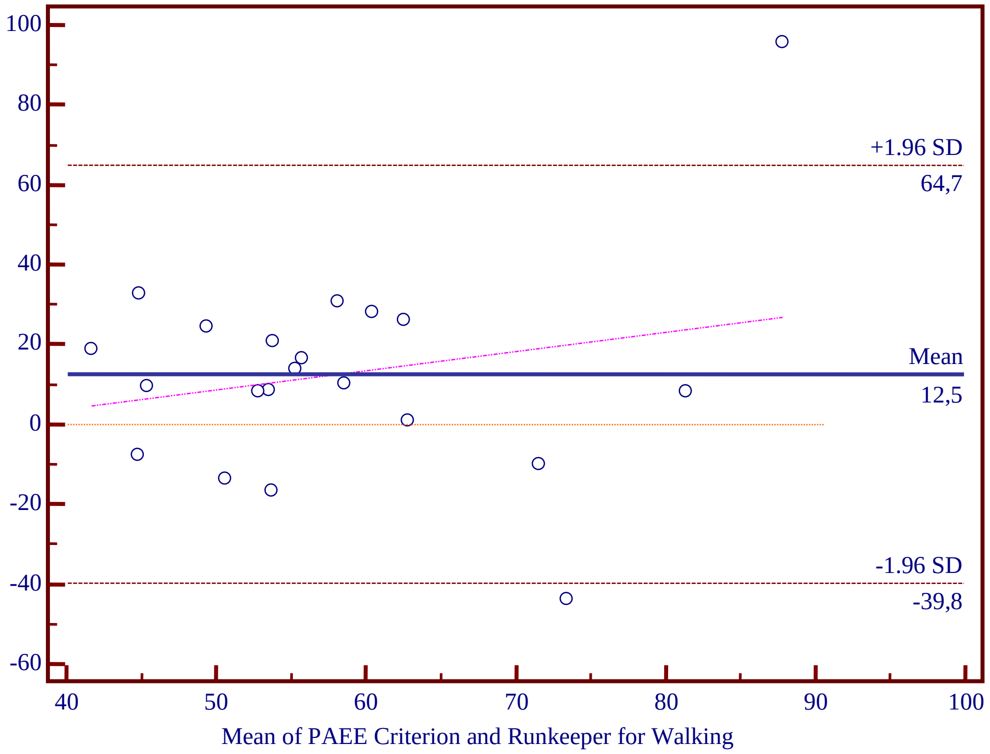

Figure 4. Bland-Altman plot for Runkeeper 2.0 PAEE estimates during walking compared with criterion measure.

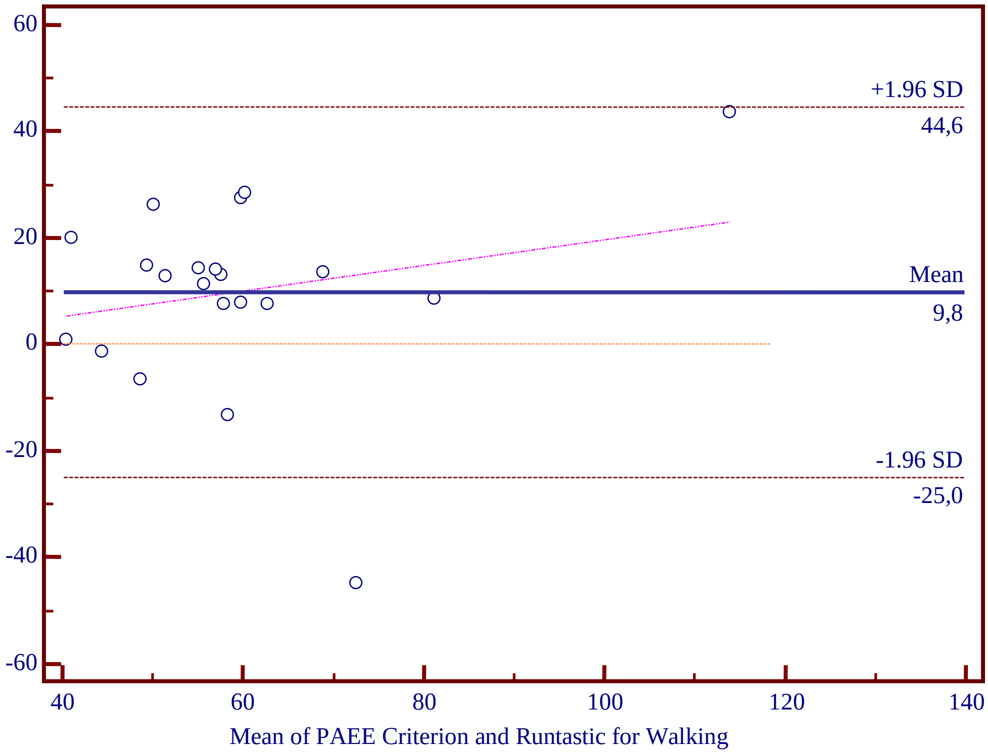

Figure 5. Bland-Altman plot for Runtastic estimates during walking compared with criterion measure.

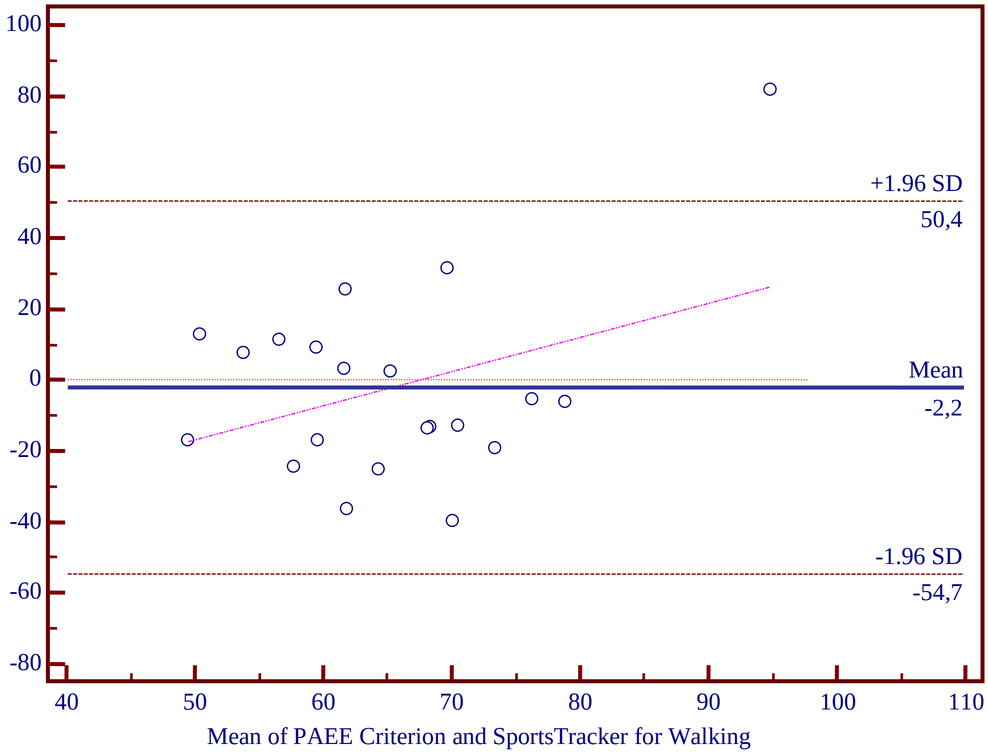

Figure 6. Bland-Altman plot for Sports Tracker PAEE estimates during walking compared with criterion measure.

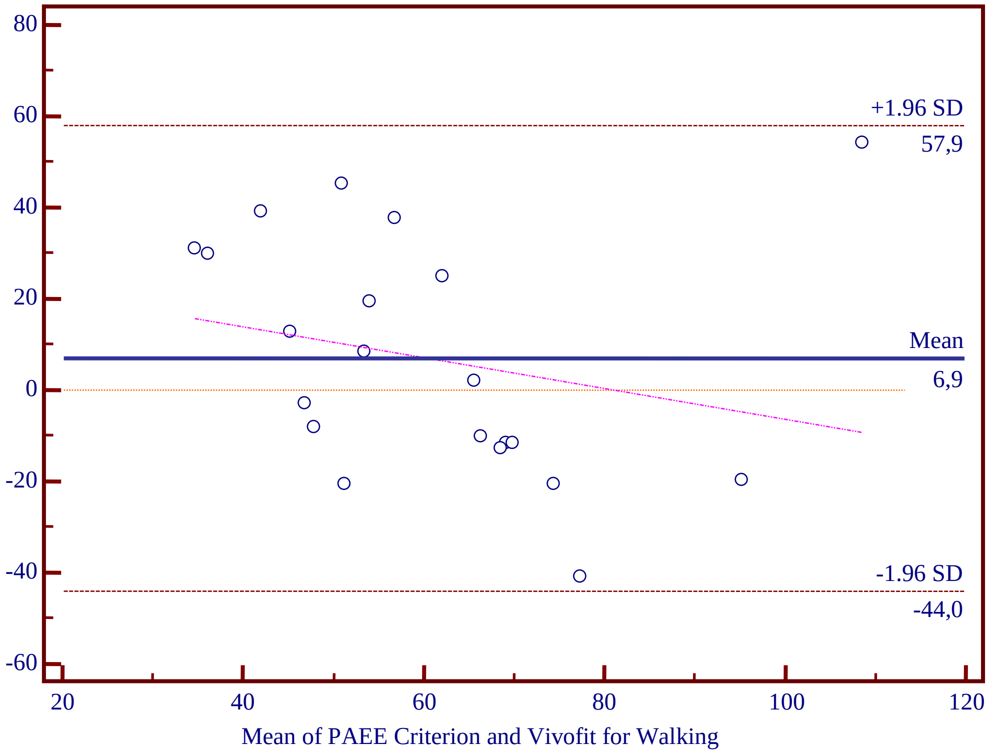

Figure 7. Bland-Altman plot for Vivofit PAEE estimates during walking compared with criterion measure.

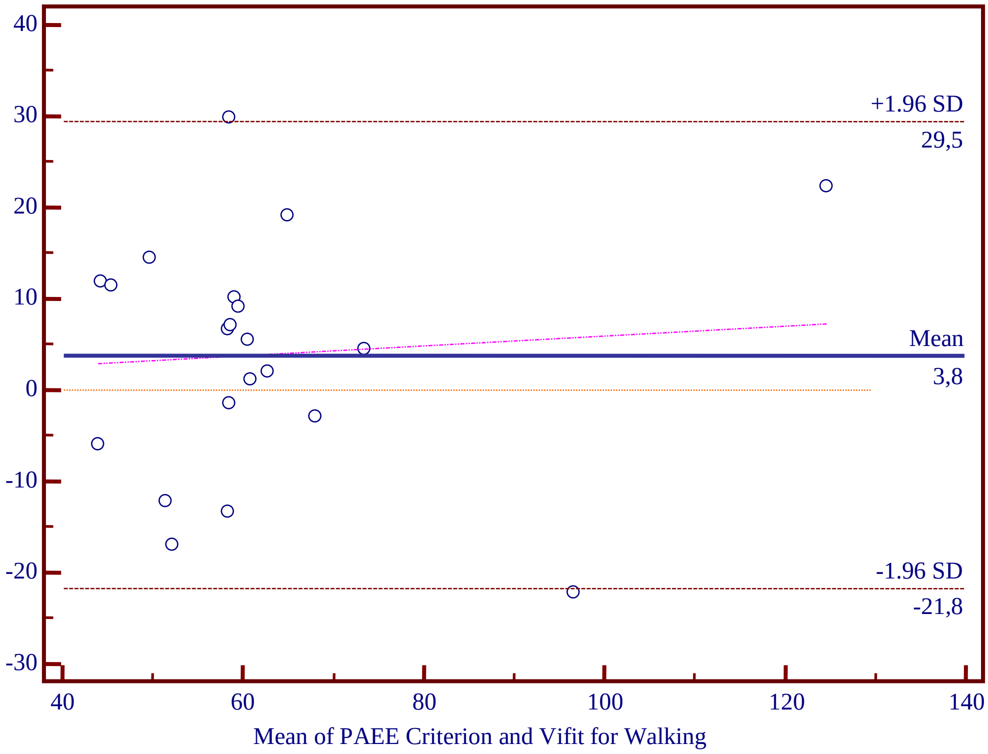

Figure 8. Bland-Altman plot for Vifit PAEE estimates during walking compared with criterion measure.

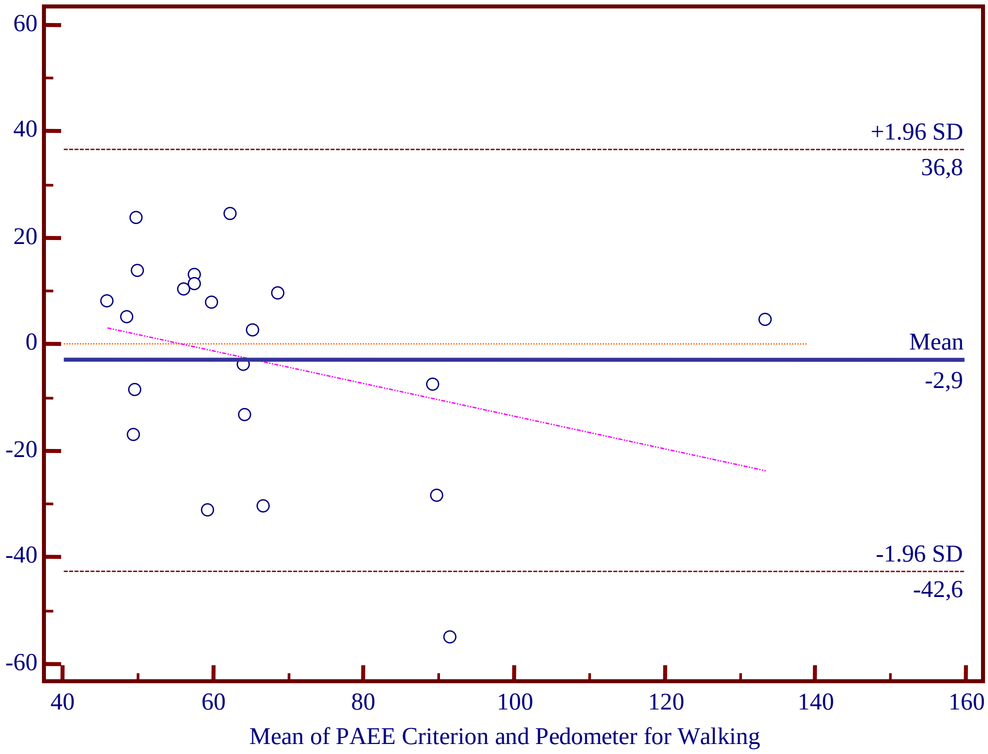

Figure 9. Bland-Altman plot for Pedometer PAEE estimates during walking compared with criterion measure.

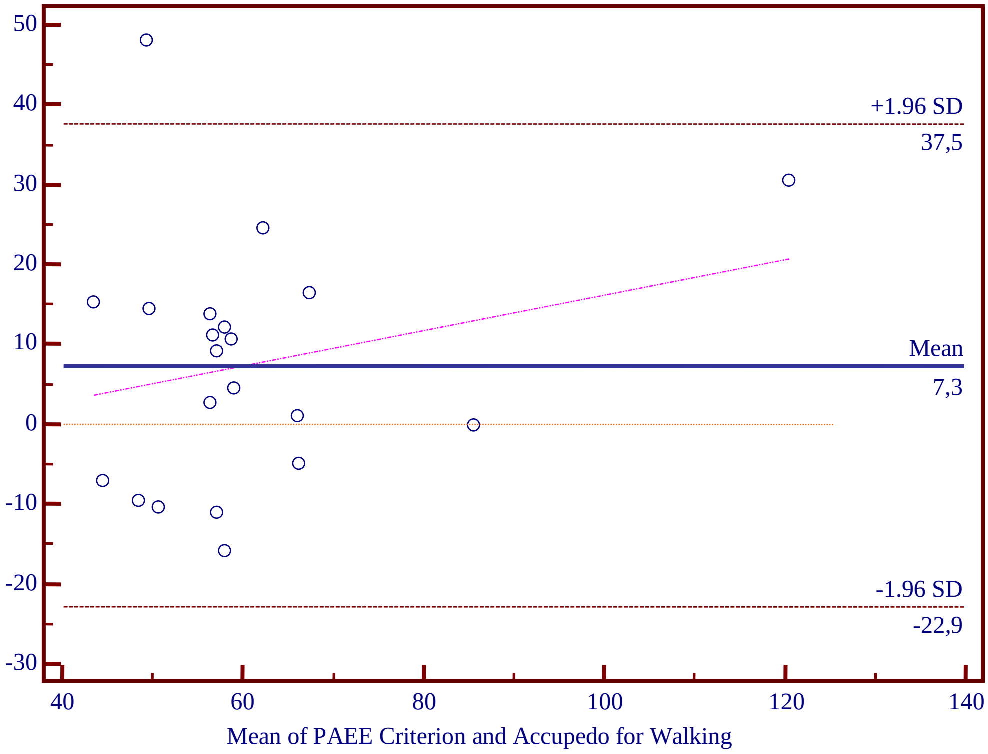

Figure 10. Bland-Altman plot for Accupedo PAEE estimates during walking compared with criterion measure.

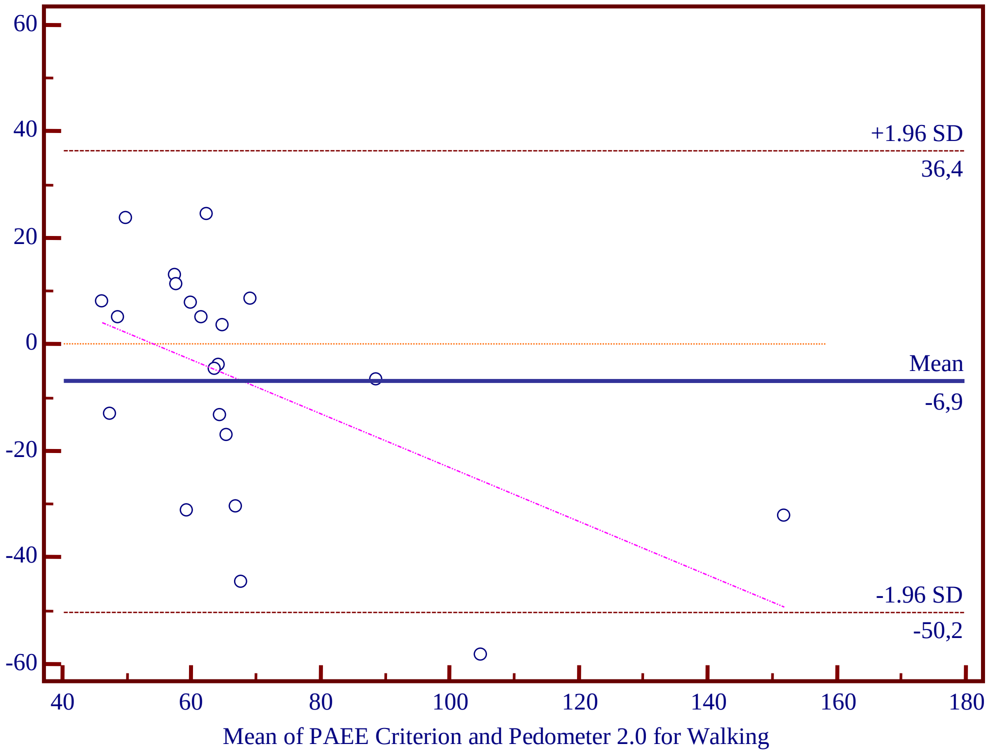

Figure 11. Bland-Altman plot for Pedometer 2.0 PAEE estimates during walking compared with criterion measure.

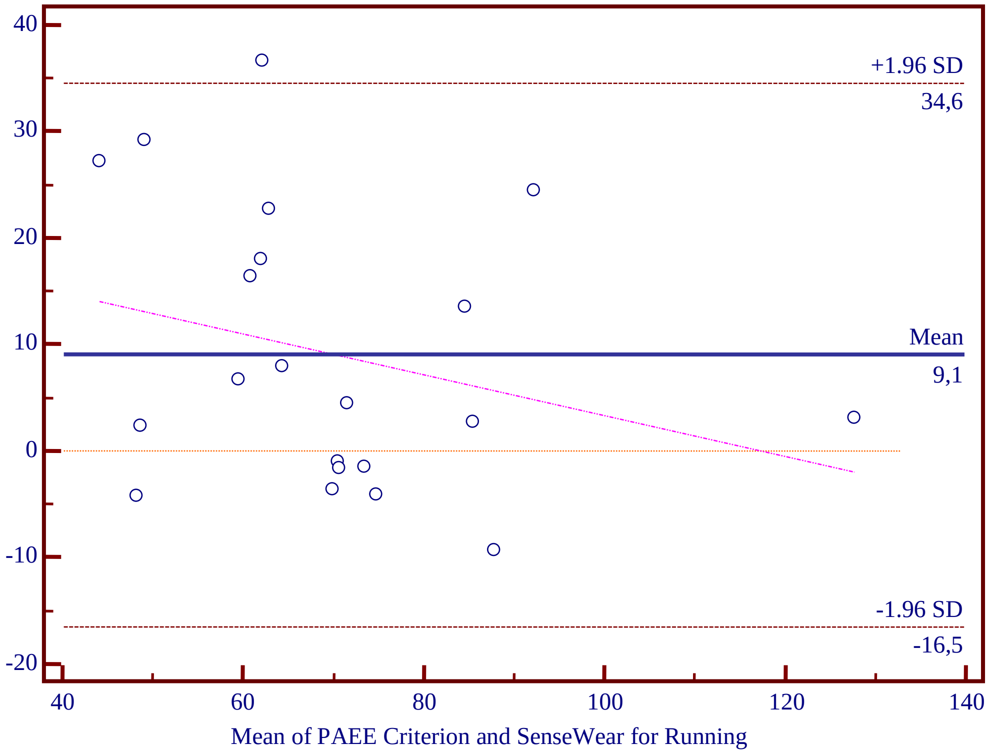

Figure 12. Bland-Altman plot for SenseWear PAEE estimates during running compared with criterion measure.

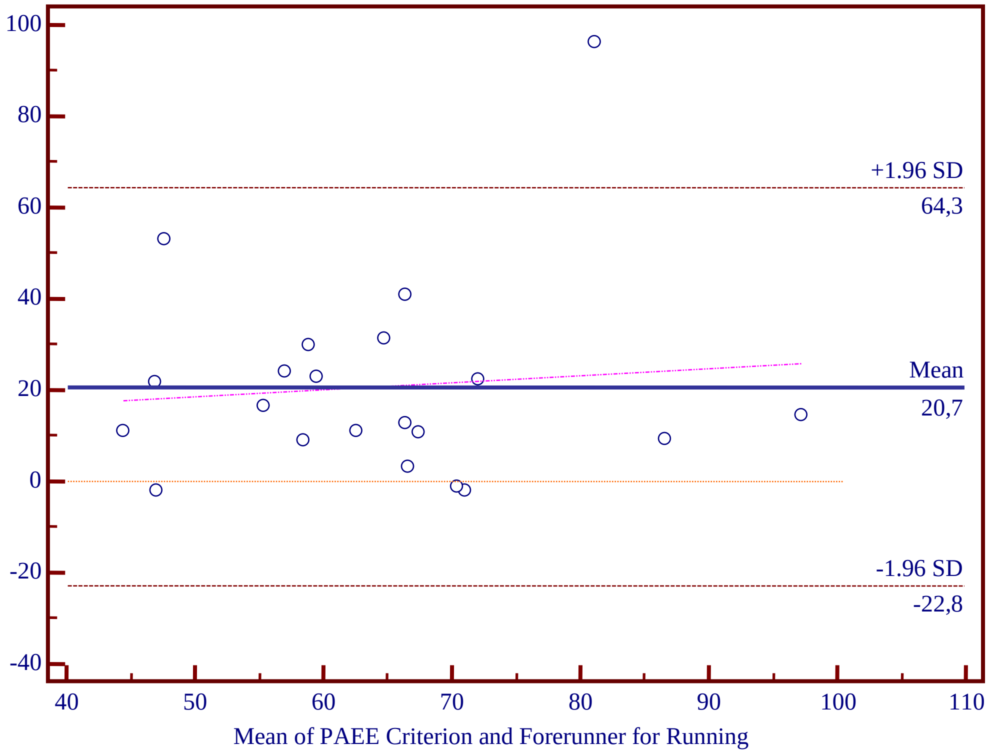

Figure 13. Bland-Altman plot for Forerunner PAEE estimates during running compared with criterion measure.

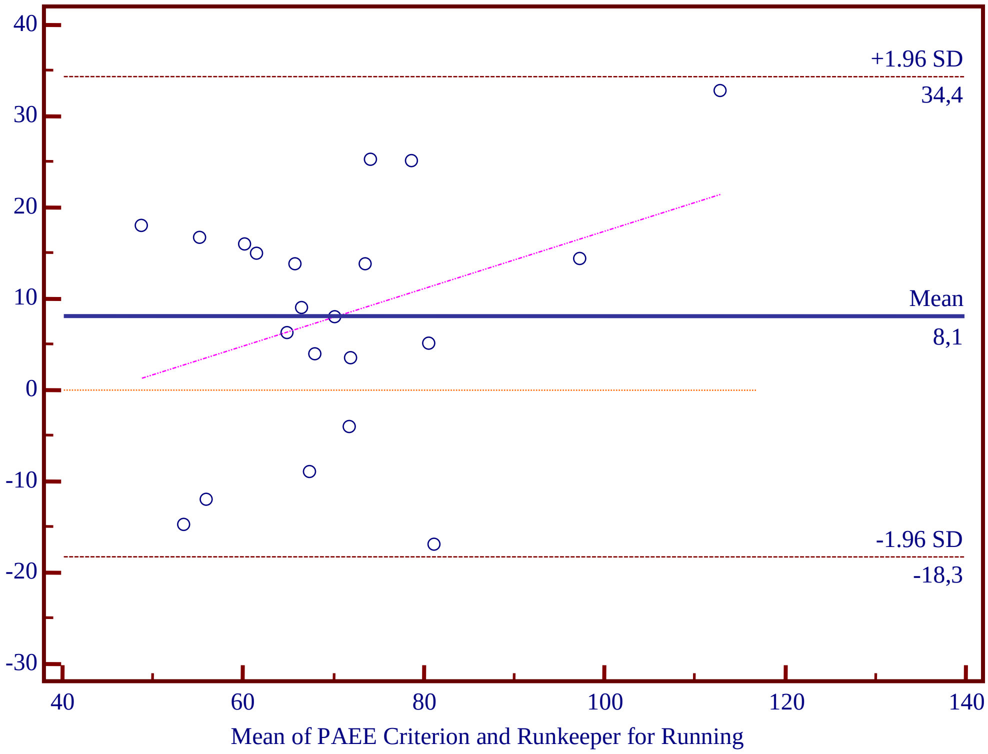

Figure 14. Bland-Altman plot for Runkeeper PAEE estimates during running compared with criterion measure.

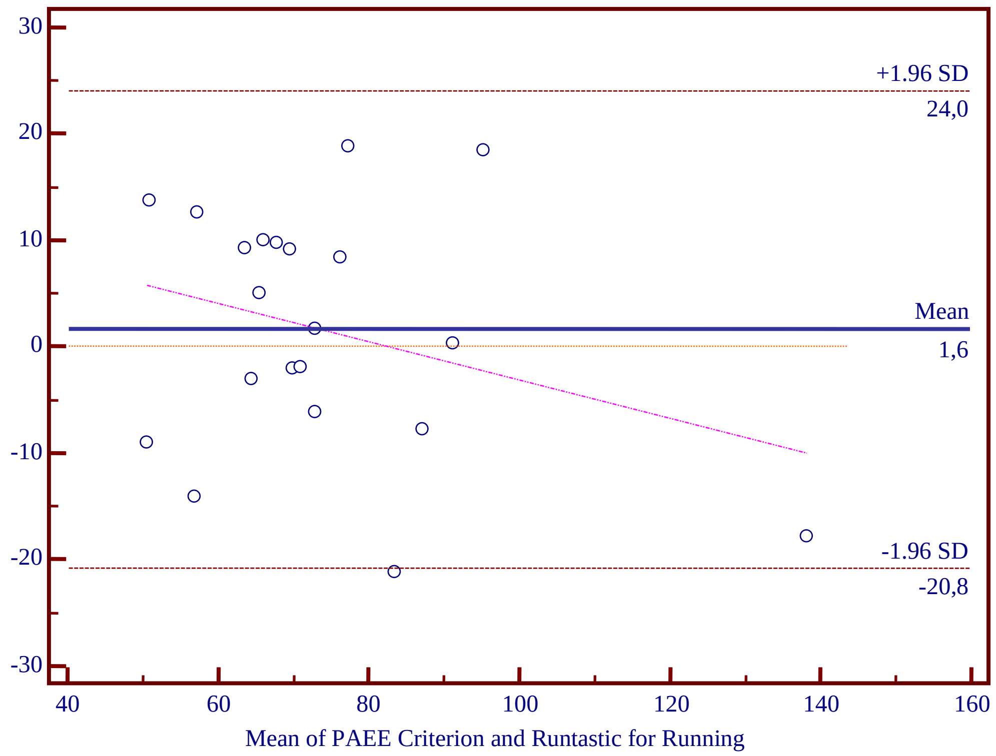

Figure 15. Bland-Altman plot for Runtastic PAEE estimates during running compared with criterion measure.

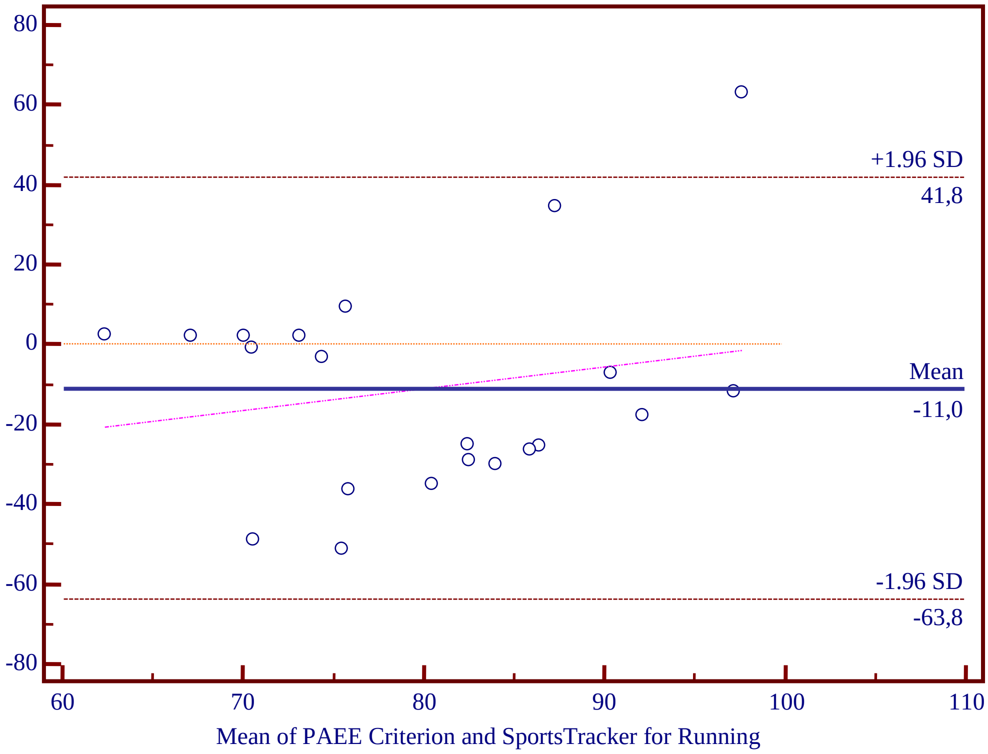

Figure 16. Bland-Altman plot for Sports Tracker PAEE estimates during running compared with criterion measure.

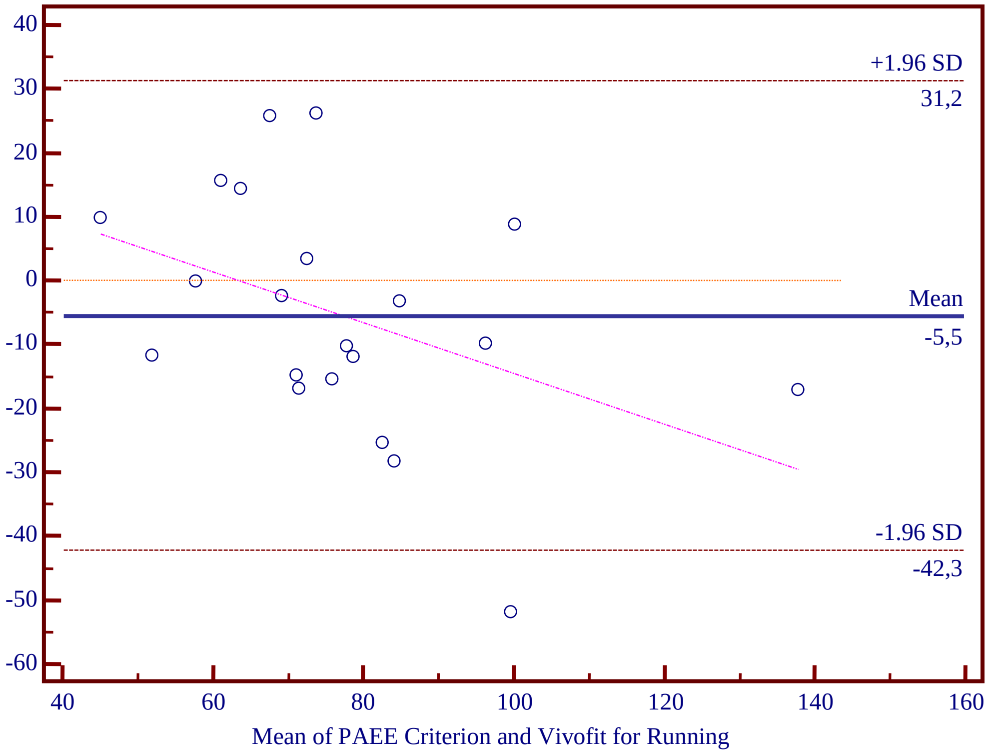

Figure 17. Bland-Altman plot for Vivofit 2.0 PAEE estimates during running compared with criterion measure.

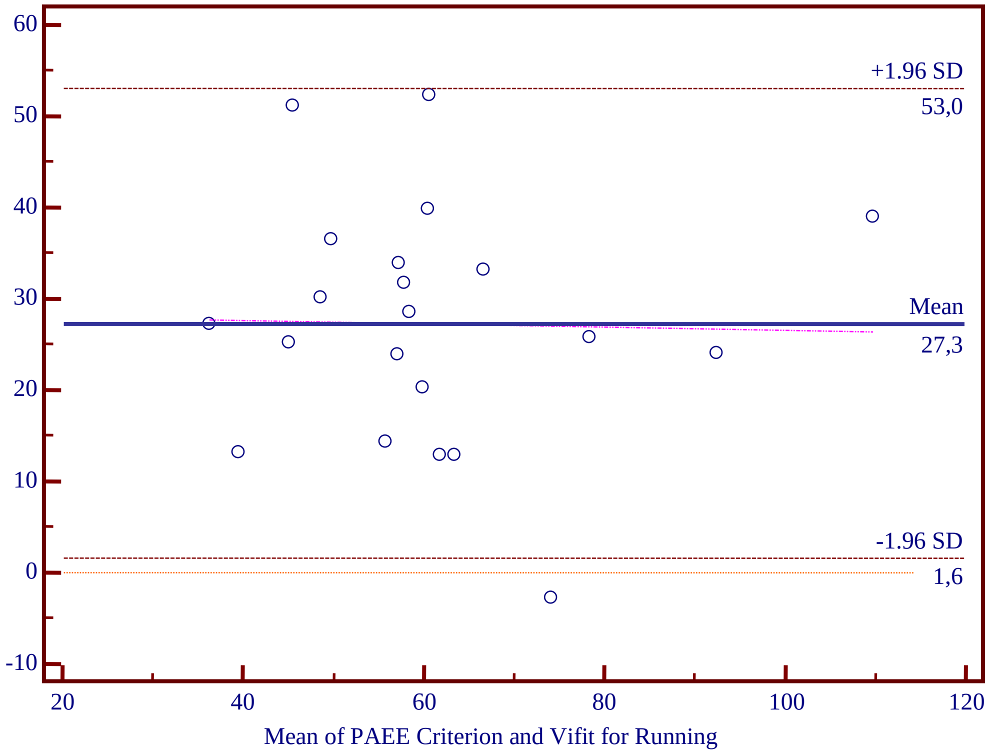

Figure 18. Bland-Altman plot for Vifit PAEE estimates during running compared with criterion measure.

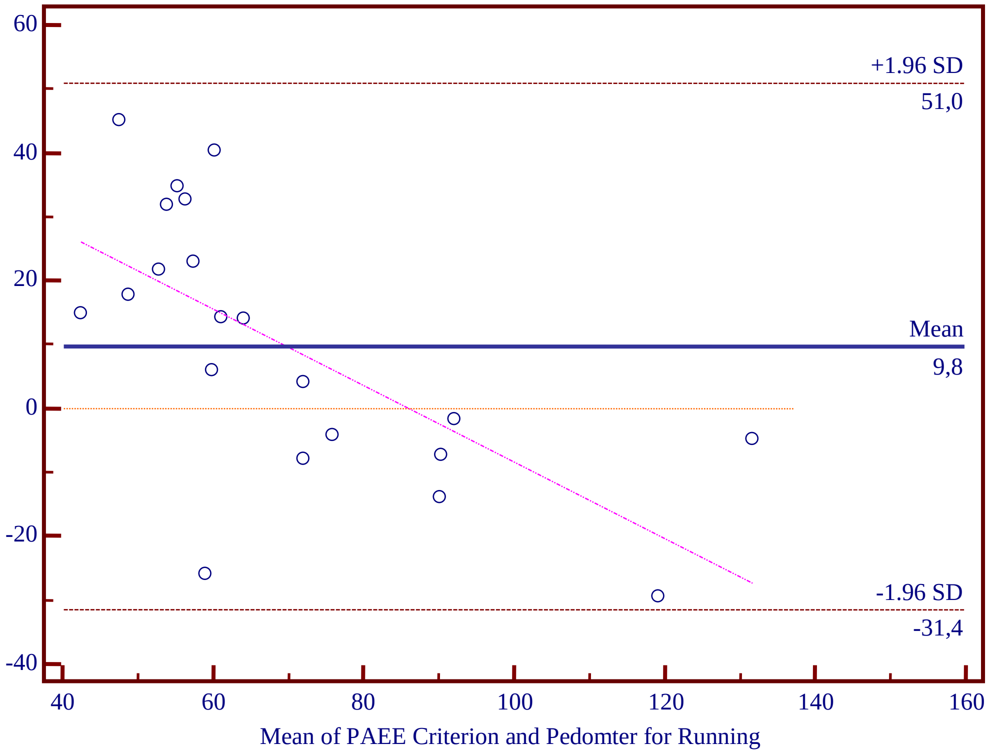

Figure 19. Bland-Altman plot for Pedometer PAEE estimates during running compared with criterion measure.

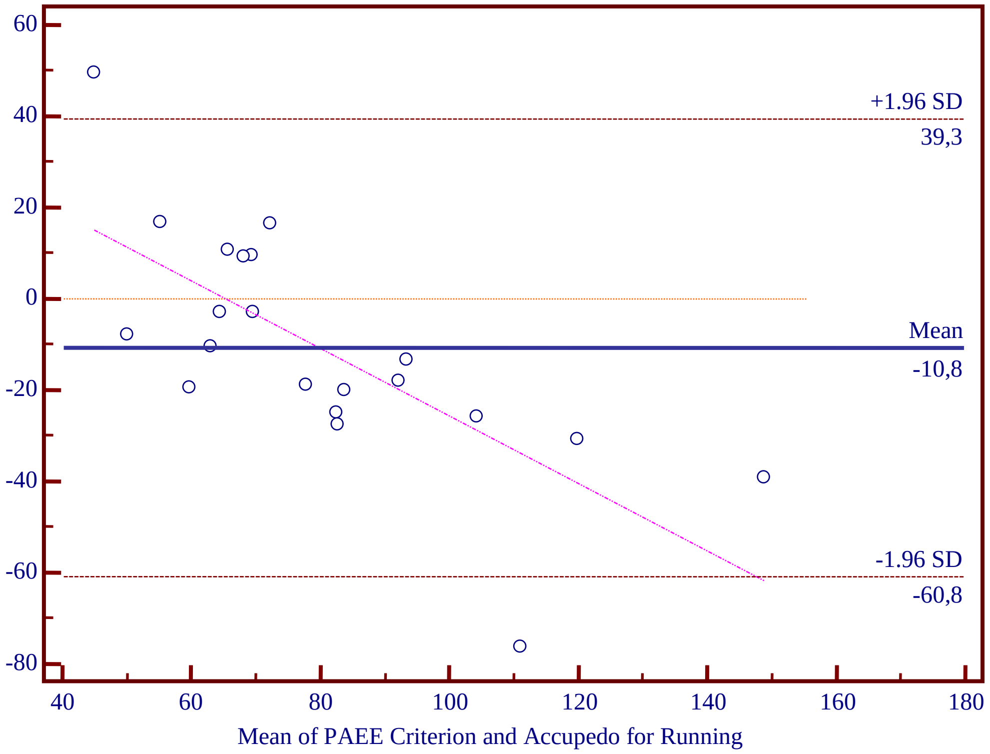

Figure 20. Bland-Altman plot for Accupedo PAEE estimates during running compared with criterion measure.

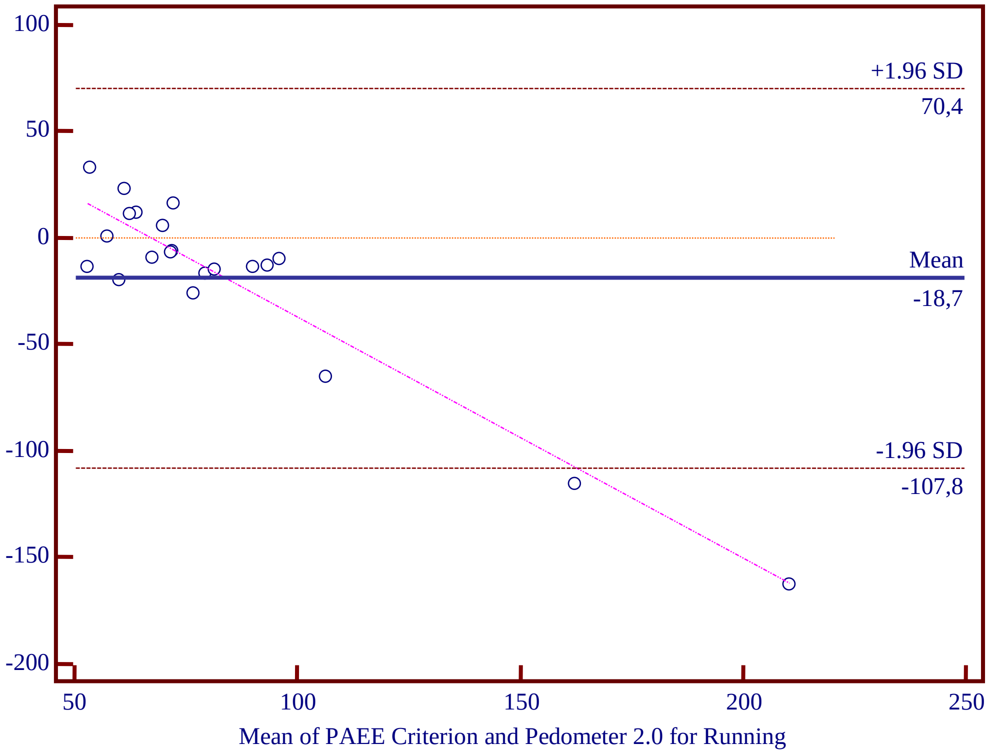

Figure 21. Bland-Altman plot for Pedometer 2.0 PAEE estimates during running compared with criterion measure.

Reply to the 3rd reviewer

Initially we would like to thank the 3rd reviewer for his/her thorough review and for the comments towards the improvement of the manuscript and resubmit this manuscript for further consideration. We attempted to address the reviewer’s comments in the most appropriate way. Below we provide a table that details the specific changes made to the manuscript based on the reviewers’ comments. Each line provides the page number and line number where the change can be located along with a description of what was changed. Changes made to the manuscript based on the reviewer comments are highlight in yellow throughout the manuscript. Any typographical errors noted by the reviewers, as well as some that we identified in our own rereading of the manuscript, have been corrected, but are not addressed in the table below.

No.	Page	Lines	Comments	Changes/Rationale
-	-	-	In addition, they need to update the reference. Some suggestion will be provided in the specific comments. The methods section needs to be expanded with necessary information.	We tried to follow all relevant recommendations, update the references and expand the methods section. We would like to highlight though that the previous 2 reviewers, on the contrary to the 3rd reviewer’s suggestions, had suggested to reduce the word count and not to include more elements. This resulted in almost 1,000 words reduction during the previous reviews.
1	1	1	I suggest to change your title in "Criterion validity of different wearable electronic devices to measure physical activity energy expenditure in adolescents.	We changed the title to “Criterion validity of wearable monitors and smartphone applications to measure physical activity energy expenditure in adolescents”. We believe that this title more adequately captures the purpose of the study, since the general term “devices” usually refers to monitors and not apps.
2	2	36-57	The abstract is not written according to Sport Science for Health Guidelines (purpose, methods, results, conclusions). Please check	We would like to thank the reviewer for this remark. The abstract has been written according to the Journal’s guidelines. We would like to highlight that in the Submission guidelines

			it over.	webpage (https://www.springer.com/journal/11332/submission-guidelines#Instructions%20for%20Authors_Types%20of%20Papers) there are two different approaches for writing the abstract¨ 1. Structured Abstract (except for Editorials, Letters to the Editor, Short Communications): Background; Aims; Methods; Results and Conclusions; 2. Purpose, Methods, Results, Conclusion. This is somehow confusing.
3	2	42	Please add here what is the design of your study.	The design of the study (cross-sectional) has been added.
4	2	42	What does "M" mean? Please define it.	“M” has been deleted (it actually meant mean).
5	2	63	In order to be more friendly reader, I would suggest to add a list of all the abbreviation of the manuscript.	A list of the abbreviations included in the manuscript has been included in lines 61-67.
6	3	75-77 (New 74-80)	Here I would also suggest referring quality and quantity of physical activity also to exercise adherence.	A few systematic reviews regarding the various strategies to promote PA and increase exercise adherence based on apps and wearable devices have been added, based on the reviewer’s suggestion.
7	3	105-114 (New 118-120)	In order to make your introduction more hypothesis driven you have to define better what is the aim of your study précising what is the primary outcome and then the secondary outcomes.	A sentence has been added to highlight the main outcomes of the present study.
8		Methods	The methods must be clear so that the study can be replicate as to equipment, subjects, context of training level and rationale for the design of each independent and dependent variable as we need to know more about	Initially we followed the general approach that similar research papers present the Methods section (including all subsections). We also followed the structure of previously submitted papers in the Sport Science for Health journal, which had a similar structure to ours.

			enrolment, subjects, procedures etc. This needs to be very highly specific as to source of equipment etc. I would suggest reordering your methods.	Since the reviewer believes that the structure should be altered, we followed the specific instruction provided and the Methods section has been restructured according to the guidelines suggested.
9		Discussion	The clarity of your discussion needs to be improved a bit and qualified where appropriate as you need to stick to what your experimental design can tell us with your data and limit speculation or qualify them, and make sure that your statement are referenced.	The Discussion was been slightly altered, including elements of the experimental design. The speculations have been limited, however the qualification of the monitors and apps according to the current findings has not been removed. This is a similar approach followed in many validation studies, because the readers should be aware which devices are more valid and suitable for use in different settings. It is also important for the readers to know whether, or not, low-cost monitors and apps perform similarly to ones that are more expensive. Many references have been included in the manuscript in order to support the arguments presented.
10		Discussion	I would suggest to insert in your conclusions the practical application of your study. What should now physicians, trainers and practitioners now have to do after reading your paper? Does it affect practice is the key factor for this section.	The practical applications of the study have been included in the Conclusion section.