





Article

Validation of the Apple Watch for Estimating Moderate-to-Vigorous Physical Activity and Activity Energy Expenditure in School-Aged Children

Sunku Kwon ¹, Youngwon Kim ^{2,3}, Yang Bai ¹, Ryan D. Burns ¹, Timothy A. Brusseau ¹ and Wonwoo Byun ^{1,*}

¹ Department of Health and Kinesiology, University of Utah, Salt Lake City, UT 84112, USA; sunku.kwon@utah.edu (S.K.); yang.bai@utah.edu (Y.B.); ryan.d.burns@utah.edu (R.D.B.); tim.brusseau@utah.edu (T.A.B.)

² School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong 999077, China; youngwon.kim@hku.hk

³ MRC Epidemiology Unit, School of Clinical Medicine, University of Cambridge, Cambridge CB2 0SL, UK

* Correspondence: won.byun@utah.edu; Tel.: +1-801-583-1119

Abstract: The Apple Watch is one of the most popular wearable devices designed to monitor physical activity (PA). However, it is currently unknown whether the Apple Watch accurately estimates children's free-living PA. Therefore, this study assessed the concurrent validity of the Apple Watch 3 in estimating moderate-to-vigorous physical activity (MVPA) time and active energy expenditure (AEE) for school-aged children under a simulated and a free-living condition. Twenty elementary school students (Girls: 45%, age: 9.7 ± 2.0 years) wore an Apple Watch 3 device on their wrist and performed prescribed free-living activities in a lab setting. A subgroup of participants ($N = 5$) wore the Apple Watch for seven consecutive days in order to assess the validity in free-living condition. The K5 indirect calorimetry (K5) and GT3X+ were used as the criterion measure under simulated free-living and free-living conditions, respectively. Mean absolute percent errors (MAPE) and Bland-Altman (BA) plots were conducted to assess the validity of the Apple Watch 3 compared to those from the criterion measures. Equivalence testing determined the statistical equivalence between the Apple Watch and K5 for MVPA time and AEE. The Apple Watch provided comparable estimates for MVPA time (mean bias: 0.3 min, $p = 0.91$, MAPE: 1%) and for AEE (mean bias: 3.8 kcal min, $p = 0.75$, MAPE: 4%) during the simulated free-living condition. The BA plots indicated no systematic bias for the agreement in MVPA and AEE estimates between the K5 and Apple Watch 3. However, the Apple Watch had a relatively large variability in estimating AEE in children. The Apple Watch was statistically equivalent to the K5 within $\pm 17.7\%$ and $\pm 20.8\%$ for MVPA time and AEE estimates, respectively. Our findings suggest that the Apple Watch 3 has the potential to be used as a PA assessment tool to estimate MVPA in school-aged children.

Keywords: moderate-to-vigorous physical activity; active energy expenditure; Apple Watch; children



Citation: Kwon, S.; Kim, Y.; Bai, Y.; Burns, R.D.; Brusseau, T.A.; Byun, W. Validation of the Apple Watch for Estimating Moderate-to-Vigorous Physical Activity and Activity Energy Expenditure in School-Aged Children. *Sensors* **2021**, *21*, 6413. <https://doi.org/10.3390/s21196413>

Academic Editor: Ki H. Chon

Received: 19 July 2021

Accepted: 19 September 2021

Published: 25 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Moderate-to-vigorous physical activity (MVPA) in children offers numerous health benefits, including the prevention of childhood obesity [1], improved bone strength [2], and improved cardio-respiratory endurance [3,4]. Given that physical activity (PA) monitoring plays an important role in promoting children's MVPA [5], PA assessment, which focuses on estimating MVPA time and activity energy expenditure (AEE), can be useful for PA monitoring. Such quantification of activity intensities, time, and AEE in real-time has been made possible through the advancement of micro-technology and the deployment of various consumer-based wearable devices, such as activity trackers and smartwatches. Smartwatches, in particular, have recently been recognized as effective tools in monitoring

PA patterns in both clinical and research settings due to wear convenience and real-time monitoring of steps, energy expenditure (EE), and intensity of activities [6–8].

Apple Watch® (Apple Inc., Cupertino, CA, USA) is currently the most popular smart-watch available, possessing approximately 48% of the global market share in 2019 [9]. This device is equipped with accelerometers, gyroscopes, and altimeters that are used to estimate a user's exercise time, total activity, moving distance, active EE, and steps. The Activity app on the Apple Watch offers a daily portrait of how much a user exercises, moves, and stands, and is paired with the accompanying Activity app on a smartphone to track PA patterns daily and in real-time. The apps are designed to allow users to set daily PA goals, earn activity awards, and compete with others in activity-based competitions. For example, Apple Watch users might earn a "Perfect Week (exercise)" when they meet their personal goals. These award features are designed to promote regular PA in users, and it indicated the Apple Watch is indeed a useful tool for increasing PA levels in adults [10]. These features, as well as the Apple Watch's growing popularity and affordability, suggest that the Apple Watch has tremendous potential for use in PA epidemiologic research. Considering the recent increase in smart device use among children, the Apple Watch can also be an attractive tool for children to increase their PA engagement. Therefore, the data from the Apple Watch would allow researchers to better understand the way to promote PA engagement in children.

Previous studies for adults and youth have examined the validity of the Apple Watch's EE estimation, and reported moderate to strong correlations (range: $r = 0.71$ to 0.88), and an acceptable measurement error (14.1 to 24.3%) against the indirect calorimetry [11–13]. In addition, the Apple Watch has been used as a measurement method for PA in a recent large-scale study, named "Apple Heart and Movement Study", which examines any potential factors associated with heart health and PA in the cohort of Apple Watch users. In this particular study, participants were able to self-enroll using their Apple Watch, and researchers can remotely recruit participants and acquire participants' data through the research app.

The rapid rate of advancements in sensor technology promotes the use of device-based PA measurement, and the Apple Watch is one of the most popular and promising wearable devices for measuring PA in epidemiological research. Accordingly, it is essential to assess the validity of the Apple Watch in estimating children's PA. Children typically have intermittent activity patterns and less accurate recall for their behavior, as a result of their less interest in continuous activity and a relatively short span of attention on a given task [14]. Assessing PA in children is challenging due to their intermittent activity patterns and limited ability to recall their behaviors [14–16]. Thus, accelerometry-based activity monitors have been recognized as a standard measure of habitual PA in children due to its objectivity, unobtrusiveness and accuracy. While the Apple Watch can be a useful PA measurement tool in children, the Apple Watch may not record some MVPA times in children due to their intermittent activity patterns. More specifically, Apple Watch captures MVPA time in minutes, and it may not recognize the span as MVPA time if the total amount of activity within 1 min is less than the set volume as MVPA, due to intermittent activity patterns during continuous activity. However, there has been sparse research on whether the Apple Watch accurately estimates the time engaged in MVPA and activity EE (AEE) in school-aged children.

Since 2021, the Apple Watch 3 has become affordable, and has the same features on the monitoring of fitness and PA compared to the Apple Watch 6, which is the newest model. Moreover, the ability to manage multiple Apple Watches from a single iPhone through a family account can support researchers to use of Apple Watch 3 in the research aimed at measuring and promoting children's PA. As the Apple Watch 3 can be utilized as a measurement tool in a large-scale cohort PA study, it is also essential to determine how the Apple Watch performs in children. Therefore, the purpose of this study was to examine the concurrent validity of the Apple Watch 3 for estimating MVPA time and AEE in elementary school-aged children under a simulated activity setting and free-living condition.

2. Materials and Methods

2.1. Participants

A convenience sample of 20 elementary school-aged children (Girls: 50%, Age: 9.7 ± 1.9 years, BMI percentile: $36.9 \pm 23.1\%$) was recruited via email, flyers, and word-of-mouth. Children who were physically disabled or otherwise unable to participate in PA were excluded from this study. The study protocol was approved by the Institutional Review Board (IRB) of the University of Utah (IRB approval number: 00108150). Participants and their parents provided signed assent and informed consent prior to participation in this study.

2.2. Instruments

2.2.1. Apple Watch

The Apple Watch 3 is a wrist-worn smart device (weight range: 26.7 to 52.8 g) that includes a retina OLED display (size: 38 or 42 mm), tri-axial accelerometer (up to 16 g-forces), tri-axial gyroscope, barometric altimeter, optical heart sensor, and global positioning system. This device is water-resistant (up to 50 m) and has up to 18 h of battery life. The Apple Watch 3 is advertised to estimate exercise minutes, active and resting calories, steps, distance, and standing hours by accelerometer, gyroscope, and barometric altimeter in real-time and per day. Moreover, this device can monitor heart rates in real-time using the optical heart sensor. The activity app built in the Apple Watch 3 offers individuals the ability to track PA and set daily PA goals. Further, the app shows total calories by the sum of active and resting calories. The Apple Watch 3 was placed on the dominant wrist of the participants following the manufacturer's recommendations and connected to an accompanied iPhone throughout the study.

2.2.2. Indirect Calorimetry

Cosmed K5 (K5; COSMED, Rome, Italy) served as a criterion method for measuring MVPA and AEE during the lab session. The K5 is a valid and reliable portable indirect metabolic system that can accurately measure respiratory minute volume (VE), oxygen uptake (VO_2), and carbon dioxide production (VCO_2) using a breath-by-breath method [17–19]. Moreover, previous studies in adults and children used the portable indirect calorimeter as a criterion measurement to examine the validity of various wearable activity monitors in estimating PA intensities [20,21]. The main unit of the K5 ($174 \times 111 \times 64$ mm and 900 g, including battery and Oxygen (O_2) sensor) is placed on the participant's upper back using an adjustable harness. The main unit of the K5 communicates with a computer by Bluetooth to record and store the measured data. Before data collection, the K5 was calibrated following the manufacturer's recommendation. The current study assessed MVPA time and AEE of participants using the measured VO_2 values from the K5 [19].

2.2.3. ActiGraph GT3X+

The ActiGraph GT3X+ (GT3X+; ActiGraph, Pensacola, FL, USA) is a small and light ($4.6 \times 3.3 \times 1.5$ cm; 19 g) research-grade accelerometer that can be worn on the wrist or at the waist using a manufacturer-provided wrist-strap or waist strap. This device records raw accelerations in three axes with a dynamic range ± 6 g at a user-specified sampling frequency (30–100 Hz). Also, the GT3X+ can estimate activity and sedentary bout, PA intensity, and steps taken at a user-selected epoch length (1–60 s) [22]. The device has been validated for its accuracy in estimating PA compared to the measures from indirect calorimetry in adults and children [23–25], and has been widely used as a criterion measure for evaluating the validity of consumer-based activity monitors in estimating PA under a free-living condition [26,27]. In the current study, the GT3X+ was used as a criterion measure of MVPA under the free-living condition.

2.3. Procedures

Participants, accompanied by a parent, visited the Physical Activity Research Laboratory at the University of Utah for their lab session. Prior to the lab session, participants completed the informed consent and a demographic questionnaire. Trained research staff measured participants' height (cm), weight (kg), and waist circumference (cm) using a wall stadiometer (ShorrBoard[®], Olney, MD, USA), an electric body scale (Seca 869, Hamburg, Germany), and a tape measure (Baseline[®] Evaluation Instruments, White Plains, NY, USA), respectively. Body mass index (BMI) was calculated using the measured height and weight. The measured anthropometric characteristics were entered in the Apple Watch 3's Activity app, and the K5 software to initialize the Apple Watch 3 and the indirect calorimetry for each individual's testing.

The Apple Watch 3 was placed on a participant's dominant wrist before the participant was fitted with K5 to measure breath-by-breath oxygen uptake during the lab session. MVPA (i.e., exercise minutes) and AEE (i.e., active calories) values of the Apple Watch 3 were recorded at the beginning of the activity protocol. Following these preparations, each participant performed a 50-min activity protocol, which included resting, simulated free-living activities, and 1-min transition periods. Initially, participants were taken a rest in an inclined position for three minutes. The resting was followed by a total of 14 activities in a gymnasium. The activities were selected to simulate typical activities for children in free-living, according to the youth PA compendium [28]. Participants selected their preferred activities to perform across ranges of activity intensities. Research staff tracked the time of each activity, and provided verbal cues to the participants to transition to the next activity.

To evaluate the accuracy of the Apple Watch 3 in estimating time spent in MVPA under the free-living condition in a field setting, five participants were randomly selected as a subsample group following the lab session. The selected participants were asked to wear the GT3X+ and Apple Watch 3 on their non-dominant and dominant wrists, respectively, and went about their daily life in free-living conditions for at least 7 consecutive days. Prior to deployment, the GT3X+ was initialized with a sampling rate of 30 Hz using the ActiLife software (ActiGraph, Pensacola, FL, USA). Participants were required to take off both devices during any aquatic activities and sleep time. Parents recorded their children's non-wear and sleep time on the sleep-activity log sheet.

2.4. Data Processing

Upon completing the lab session, the estimated MVPA and AEE from the Apple Watch 3 were immediately recorded. The metabolic equivalence of tasks (MET) were calculated using the measured VO_2 (mL/min) from the K5 and each participant's body weight (kg). The average value of the metabolic rates during the resting period was used as 1-MET to classify the children's PA intensity. The calculated MET were classified with different activity levels (≤ 1.5 MET = SED, 1.6–2.9 MET = LPA, 3.0–5.9 MET = Moderate PA, and ≥ 6.0 MET = Vigorous PA). For comparison to the Apple Watch 3's AEE estimates, the measured net AEE of the testing day was calculated. First, resting, sedentary activity, and transition periods were removed to leave only active minutes. The measured VO_2 data (mL/min) were multiplied by 1000 to obtain VO_2 in L/min, and then multiplied by 4.867 kcal/L to obtain kcal/min. These calculated kcal/min values were summed to obtain the total AEE of the testing day. In addition, the basal metabolic rate (kcal/day) was predicted for each participant using the Schofield equations [29]. The predicted basal metabolic rate was divided by 1440 min to calculate kcal/min values, then multiplied by the total minutes of active time. The calculated total basal metabolic rate was subtracted from the calculated total AEE to obtain the net AEE for the testing day. The processed K5 data were aggregated to the daily average, then merged and aligned with the Apple Watch 3 data for statistical analyses.

For the subsample group, daily MVPA estimates of the Apple Watch 3 were obtained from the iPhone's Activity app. Data from the GT3X+, under the free-living condition, were

downloaded in a raw acceleration data format and converted into “.csv” files using ActiLife software. The raw acceleration data were processed in R software (<http://cran.r-project.org>; accessed on 22 November 2020) using the GGIR package (version 1.10–10) [30]. The GGIR package calibrated the raw tri-axial accelerations derived by the GT3X+ and converted it to the Euclidean norm minus one (ENMO; $\sqrt{x^2 + y^2 + z^2} - 1g$), which indicates the value of gravity with negative values rounded to zero [31]. The ENMO values were classified as different activity levels per one-minute using the intensity thresholds for ENMO derived by Hildebrand et al. [25,32]. Moreover, periods of non-wear and sleep were identified and excluded using Choi’s algorithm [33]. The self-reported activity/sleep logs from each participant were also excluded. The processed GT3X+ data were aggregated to the daily average, then merged and aligned with the daily MVPA estimates of the Apple Watch 3 for statistical analyses.

2.5. Statistical Analyses

Descriptive analyses were conducted to summarize the demographic and anthropometric characteristics of the participants.

Mean absolute percent errors (MAPEs) were calculated to evaluate the measurement error of the Apple Watch 3 in the estimation of MVPA time and/or AEE compared to those from the criterion measures (i.e., $|(\text{criterion-estimation})/\text{criterion}| \times 100$).

Bland-Altman (BA) plots were used to evaluate the agreement and systematic biases in estimating time spent in MVPA and AEE between the Apple Watch 3 and the K5. The mean bias (i.e., criterion – estimation) was computed to provide the overall overestimation or underestimation of the Apple Watch 3 for MVPA time, as well as the AEE compared to criterion measure. The significance of systematic bias was determined by whether the 95% confidence interval of the mean bias included the line of equality (i.e., criterion – estimation = 0). Moreover, the limits of agreement were calculated as mean bias \pm 1.96 standard deviation for evaluating the individual-level agreement.

Finally, an equivalence test was conducted to determine the equivalence at the group level between the K5 and Apple Watch 3 in estimating MVPA and AEE. The 90% confidence interval (CI) of the estimates from the Apple Watch 3 was compared with the equivalence zone (EZ) from the K5 measures. Given no evidence presently exists of a universally accepted EZ range, the current study established the minimal EZs of the K5 measures that include the 90% CIs of the Apple Watch 3 estimates. Data were analyzed using Stata 14.2 software (StataCorp LLC, College Station, TX, USA) and SAS 9.4 software (SAS Institute, Cary, NC, USA), and statistical significance was determined at $p < 0.05$.

3. Results

Participant characteristics are summarized in Table 1. There were no significant differences in age, height, weight, BMI percentile, and waist circumference between boys and girls. Table 2 presented the mean differences, and MAPE values of MVPA and AEE estimates, in the K5 and Apple Watch 3. As shown in Table 2, the results of the paired t-test revealed no significant differences in MVPA estimates between the K5 and Apple Watch 3. The MAPE in Apple Watch 3 was 1% for MVPA estimate compared to the K5 measures. For the AEE, the Apple Watch 3 underestimated AEE by 4%, but the mean difference was not statistically significant (mean difference: 3.8 kcal, $p = 0.75$). With respect to AEE, the Apple Watch 3 also had a minimal measurement error (4%) relative to the measured AEE from the K5.

Table 1. Descriptive characteristics of participants, mean \pm standard deviation.

Characteristic	All (N = 20)	Boys (N = 11)	Girls (N = 9)	p-Value *
Age (years)	9.7 \pm 2.0	9.7 \pm 1.8	9.7 \pm 2.2	0.95
Height (cm)	138.3 \pm 13.4	138.9 \pm 12.7	137.6 \pm 14.9	0.84
Weight (kg)	31.8 \pm 11.1	32.0 \pm 9.4	31.5 \pm 13.5	0.92
BMI (kg/m ²)	16.3 \pm 3.2	16.3 \pm 2.4	16.2 \pm 4.1	0.95
BMI percentile (%)	36.9 \pm 29.1	35.0 \pm 28.8	39.1 \pm 31.1	0.76
Waist Circumference (cm)	60.9 \pm 9.3	61.4 \pm 9.5	60.2 \pm 9.61	0.78

* p-value for gender difference.

Table 2. Estimated mean (SD), mean difference (SE), mean absolute percent error between indirect calorimetry and the Apple Watch 3 under the simulated free-living condition.

PA Metrics	Cosmed K5 (SD)	Apple Watch 3 (SD)	Mean diff. (SE)	MAPE (%)
MVPA	20.2 min (6.7)	19.9 min (8.3)	−0.3 min (2.3)	1%
AEE	98.2 kcal (25.6)	94.5 kcal (42.9)	3.8 kcal (11.7)	4%

MVPA: moderate-to-vigorous physical activity; AEE: active energy expenditure; MAPE: mean absolute percent error.

Bland-Altman plots (Figure 1) illustrated the agreement between the K5 and Apple Watch 3 for MVPA and AEE estimates by displaying the mean difference and level of agreement. The BA plots showed that there was no apparent bias for the agreement in MVPA estimates between the K5 and Apple Watch 3. For the AEE estimates, however, the Apple Watch 3 had a wide 95% limit of agreement (−100.7 to 108.3 kcal) compared to the K5 measures.

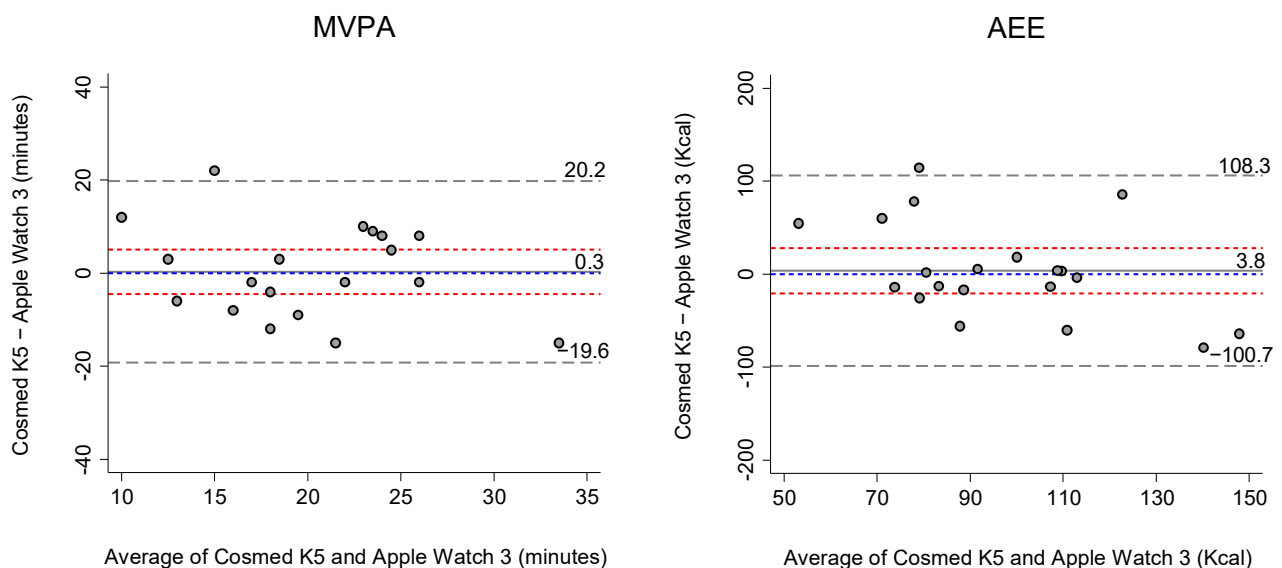


Figure 1. Bland-Altman Plots for comparing Moderate-to-Vigorous Physical Activity (MVPA) and Active Energy Expenditure (AEE) estimates between the K5 and Apple Watch 3. Blue short-dashed lines show the line of equality. Solid lines and red short-dashed lines indicate mean bias and 95% confidence interval of the mean bias, respectively. Dashed lines show 95% limits of agreement (± 1.96 standard deviation).

Figure 2 shows the results of the equivalence tests. The 90% confidence interval (CI) of the MVPA estimates from the Apple Watch 3 (90% CI = 16.63 to 23.36 min) fell within $\pm 17.7\%$ EZ (16.63 to 23.89 min) of the measured MVPA from the K5. Furthermore, the EZ of AEE measured by the K5 was established $\pm 20.8\%$ (77.8 to 118.67 kcal), which included the 90% CI of the AEE estimate (77.87 to 111.03 kcal) from the Apple Watch 3.

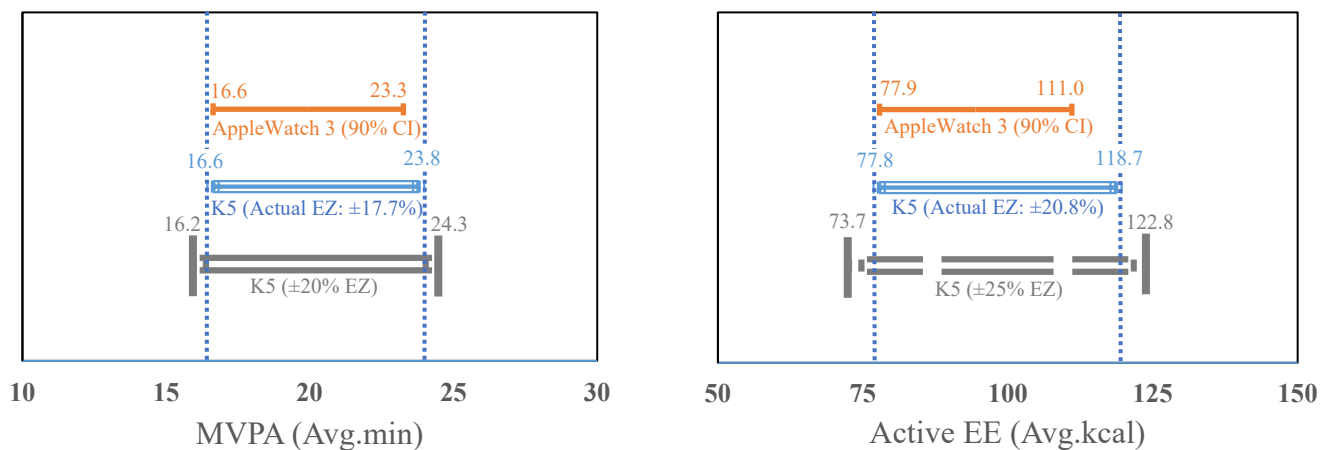


Figure 2. Equivalence Testing for MVPA and AEE estimates between the K5 and Apple Watch 3. CI: Confidence interval; EZ: Equivalence zone; MVPA: Moderate-to-vigorous physical activity; EE: Energy Expenditure; vertical dashed lines show the actual bounds within which the Apple Watch 3 is statistically equivalent to the K5 for MVPA and Active EE estimates.

In a subsample group, we found that the Apple Watch 3 had a relatively large mean bias (Mean difference: -25.3 ± 4.5) and high MAPE (47.5%) in estimating time spent in MVPA compared to the estimates from the GT3X+ in the free-living condition (Table 3).

Table 3. Estimated mean (SD), mean difference (SE), mean absolute percent error between the GT3X+ and Apple Watch 3 under the free-living condition.

PA Metrics	Cosmed K5 (SD)	Apple Watch 3 (SD)	Mean diff. (SE)	MAPE (%)
MVPA	53.3 min (13.0)	78.6 min (21.3)	25.3 min (4.5)	47.5%

MVPA: moderate-to-vigorous physical activity; MAPE: mean absolute percent error.

4. Discussion

To our knowledge, this is the first study to investigate whether the Apple Watch 3 accurately estimates time spent in MVPA and AEE compared to the established criterion measures for elementary school-aged children. The results from this study indicated that the Apple Watch 3 can provide comparable estimates for MVPA time against the indirect calorimetry in elementary school-aged children. We also observed that the Apple Watch 3 had no apparent systematic bias in estimating children's AEE relative to the indirect calorimetry, but the variability of the AEE estimates was relatively large. An advantage of the Apple Watch over the research-grade accelerometers is the ability to provide the incorporated PA data without device retrieving and data processing via remote monitoring. Accordingly, utilizing the data from the Apple Watch facilitates large-scale surveillance, which aims to promote PA in school-aged children. Given the practical applications of the Apple Watch 3 in future PA research, our findings provide important implications for researchers. The Apple Watch 3 can be a considerable device for monitoring children's MVPA.

4.1. Accuracy of the Apple Watch 3 in Estimating MVPA

A notable finding is that the Apple Watch 3 can accurately estimate children's MVPA time. More specifically, the Apple Watch 3 showed a small mean bias (0.3 min) and measurement error (MAPE: 1%) in estimating MVPA time compared to the indirect calorimetry, and the systematic bias was not significant. These findings indicate that Apple Watch 3 has comparable accuracy to the measures from the indirect calorimetry both at the group and individual levels in estimating MVPA in children. Although habitual time spent in MVPA is important to prevent childhood obesity, the MVPA time in children tends to gradually decline from age 8, and the declines are most pronounced at age 9 for both boys and girls [34]. Given that accurate PA monitoring is a key component for the promotion of

children's MVPA level [5], accelerometry-based activity monitoring is a reasonable method to objectively assess MVPA time for school-aged children who have a limited ability to recall their habitual activities [35]. More specifically, wearable activity monitors, embedded in a triaxial accelerometer, may provide valid estimates for children's MVPA, which commonly includes intermittent activity patterns by incorporating accelerations derived from three directions (i.e., vertical, mediolateral, and anteroposterior). Furthermore, an addition to heart rate monitoring may make up for the biomechanical errors associated with accelerometry, thus enhance the accuracy of activity estimates in children during exercise [36,37]. As the Apple Watch 3 includes a heart rate monitor as well as a tri-axial accelerometer, it is speculated that the Apple Watch 3 might determine the user's exercise minutes using a set of algorithms that integrate both the accelerometer and heart rate data to improve the accuracy of MVPA time measurement. Accordingly, the Apple Watch 3 can allow researchers to evaluate the time spent in MVPA during the diverse activities in which children partake. Considering the observed accuracy of the Apple Watch in estimating MVPA time, the Apple Watch 3 would be a viable measurement device in future studies, which aim to increase MVPA levels in school-aged children.

4.2. Validity of the Apple Watch 3 in Estimating AEE

Another important finding is that the present study showed a contrary result to a previous finding that the Apple Watch significantly underestimated AEE in youth [21]. The study by LaMunion et al. reported that the Apple Watch significantly underestimated AEE by 45% (mean difference: -121.8 kcal) and had more than 40% measurement error for AEE estimates compared to indirect calorimetry in youth [21]. However, the present study showed that the Apple Watch 3 similarly estimated AEE (mean difference: -3.8 kcal) with a relatively low measurement error (4%) compared to the indirect calorimetry in children. One possible explanation for the difference in AEE estimation is that the Apple Watch might be sensitive to body movements in estimating AEE [11]. While the current study evaluated the AEE during a simulated free-living activity protocol, LaMunion's study included stationary cycling, which leads to less arm movement during physical activity [21]. Also, LaMunion's study included 39 adolescents between 13 and 18 years old [21]. Adolescents may have relatively less body movement than children when consuming the same calories [38]. It should also be noted that LaMunion's study used the previous generation of the Apple Watch used in the current study. However, given that the proprietary algorithm is confidential, it is unknown if or how the manufacturer updated the energy expenditure prediction algorithm when new models were released.

It is noteworthy that the Apple Watch 3 has a relatively large inter-individual variability for EE estimation in children. The present study defined the expected limits of maximum acceptable bias (i.e., limits of agreement) of the Apple Watch in estimating children's EE through the Bland-Altman plots [39]. The result from the Bland-Altman plots across all activities revealed that there was no significant systematic bias with a relatively small mean bias (3.8 kcal) because the line of equality was within the confidence interval of the mean bias (Figure 1). However, the limit of agreement for the Apple Watch's EE estimation was relatively wider (-100.2 to 108 kcal) than the limits of agreement from previous studies, which include adults (-54.5 to 124.5 kcal) and adolescents (98.1 to 251.2 kcal) [12,21]. For this reason, it is premature to recommend the widespread use of the Apple Watch 3 for the assessment of children's EE in free-living environments. The ability to assess energy expenditure in children is a clinically important component to children's PA research for non-communicable diseases, including malnutrition, obesity, and diabetes. In this regard, the findings from the current study are critical when researchers consider using AEE data from the Apple Watch to facilitate PA in a manner appropriate prevention of childhood obesity. However, future research is still warranted to further explore the practical application of the Apple Watch 3 in estimating children's AEE under free-living conditions.

4.3. Practical Applications and Considerations

A novel aspect of this study was the assessment of the relative equivalency of the Apple Watch for MVPA time and AEE estimations against the criterion measures. Although equivalence testing is a widely accepted analytic method to rigorously examine the agreement to the criterion measures in a dichotomous manner at a group level, the present study has not attempted to assess the agreement in the traditional way, as there is no universally acceptable equivalence zone range for the Apple Watch. Instead, the equivalence test was used to identify the actual equivalence zone where the 90% confidence intervals of the estimates from the Apple Watch 3 completely fell within the mean values from the K5 indirect calorimetry. Accordingly, the relative equivalency approach would make it possible to determine the acceptable measurement error of the Apple Watch in estimating MVPA time and AEE. With this analytic approach, we were able to determine the actual equivalence zone for MVPA time and AEE estimates as $\pm 17.7\%$ (16.63 to 23.89 min) and $\pm 20.8\%$ (77.8 to 118.67 kcal), respectively. We could also identify that the actual equivalence zones of indirect calorimetry to the Apple Watch's AEE estimates from previous studies presented $\pm 18\%$ in adults and $\pm 20\%$ in youth [12,21]. It is noteworthy that the acceptable measurement error of the Apple Watch on AEE estimation would be less than 20.8% for school-aged children, which is higher than that for adults and youth. Therefore, the present study reveals the range of acceptable measurement errors of the Apple Watch in estimating MVPA time and AEE for children at the group level, and these findings are practically significant to researchers given the Apple Watch's real-world application in school-aged children.

In the current study, we examined the validity of the Apple Watch 3 in estimating MVPA under a free-living condition. Compared to the GT3X+, on average, the estimated daily PA time from the Apple Watch 3 (648 min/day) was higher than the estimate of total PA from the GT3X+ (604 min/day). Therefore, researchers need to be aware that the Apple Watch 3 may slightly overestimate the amount of total PA in free-living conditions, compared with the GT3X+.

There are several considerations to be considered when the Apple Watch is used as a measurement tool for PA research in children. First, as observed in previous studies, the Apple Watch may over- or under-estimate AEE and total EE [11,21,40–42]. In addition, current and previous studies indicated that Apple Watch's error in estimating AEE varies widely among children. Moreover, it is likely that the Apple Watch could underestimate the AEE due to the intermittent activities of children. Thus, further investigations are warranted to assess the validity and reliability in estimating AEE under free-living conditions in children. Second, appropriate PA estimates need to be selected for investigating or promoting PA using the Apple Watch in children. The Apple Watch tracks various PA metrics to encourage healthy behaviors. The PA metrics include standing minutes per hour, exercise minutes (i.e., MVPA), and amounts of active calories. Of these, MVPA time tracking is an essential factor in evaluating whether the level of PA is adequate in children. Further, the use of wearable activity monitors is effective at increasing inactive children's PA level by incorporating self-monitoring and goal setting for MVPA [8]. Lastly, using Apple Watch, researchers can collect multiple physiological (i.e., heart rate, blood pressure) and behavioral (i.e., PA) profiles in children with minimal burden. The usability of the device could be greater with the Apple Watch than research-based activity monitors because the Apple Watch provides more age-appropriate and interactive features through various compatible apps. In addition, researchers can encrypt the data collected from the research app, and securely store the collected data in a specific cloud system that technically safeguards the requirement of the Health Insurance Portability and Accountability Act. In support, a recent large-scale clinical trial called "Apple Heart and Movement Study" explores potential factors associated with heart health and PA over time using data from the Apple Watch. Given that Apple Watch users self-enroll in this trial, the project is able to recruit approximately 500,000 participants and remotely acquire participants' data through the research app. In light of the validity and usability, using Apple Watch in research would

help find new interventions that facilitate children to replace their sedentary behavior with an active lifestyle.

4.4. Strengths and Limitations

There are several strengths in this study. The main advantage of this study was the evaluation of accuracy in estimating both MVPA time and AEE in the Apple Watch 3 compared to indirect calorimetry. The indirect calorimetry system is regarded as a gold standard criterion measure to assess time engaged in sedentary behavior and PA, and AEE estimates during activities [11,12,20,21]. Furthermore, the investigation of both MVPA time and AEE estimates could provide better evidence if the Apple Watch 3 comparably estimates net AEE throughout the PA, which results in EE. The current study examined the validity of MVPA time and AEE estimates on the most popular smartwatch. The Apple Watch can be utilized in mobile healthcare systems through continuous updates of operating systems and health-related apps. Moreover, the Apple Watch can be connected to a specific app to deliver health intervention or access PA data monitored in the watch by other smart devices, such as iPhone. Thus, our findings might provide beneficial information to software developers and manufacturers, in order to update the software and hardware of the Apple Watch.

The current study also has several limitations. First, this study included a relatively small number of participants, which may limit the generalizability of the findings from this study; however, we assessed the validity of the Apple Watch 3 using rigorous statistical analytic methods (i.e., MAPEs, BA plots, equivalent tests) [43]. Therefore, the internal validity of this study is not threatened by the sample size. Lastly, given that the criterion measure under the free-living condition (GT3X+) is not completely waterproof, albeit water resistant (1 m, 30 min), this study could not include any aquatic activities that may be popular in children.

5. Conclusions

Compared to the measures from indirect calorimetry, the Apple Watch 3 comparably estimated time the spent in MVPA and AEE under the simulated free-living condition in elementary school students. The findings indicate that the Apple Watch 3 can be used in estimating MVPA time in PA research and health promotion programs. However, it is still unclear whether the Apple Watch 3 can be a valid wearable device for measuring AEE in children. Furthermore, the current study could not reveal whether the Apple Watch 3 accurately estimates MVPA time in children under true free-living conditions. The Apple Watch is a feasible device to measure PA information in research that promotes MVPA and an active lifestyle in children. Further research is warranted to evaluate the validity and reliability of the Apple Watch 3 in estimating children's MVPA time and AEE under free-living conditions.

Author Contributions: Conceptualization, S.K., Y.K. and W.B.; methodology, S.K., Y.K. and W.B.; software, S.K. and W.B.; validation, S.K., Y.K. and W.B.; formal analysis, S.K. and W.B.; investigation, S.K. and W.B.; resources, S.K., Y.K. and W.B.; data curation, S.K. and W.B.; writing—original draft preparation, S.K. and W.B.; writing—review and editing, S.K., Y.K., Y.B., R.D.B., T.A.B. and W.B.; visualization, S.K. and W.B.; supervision, W.B.; project administration, S.K. and W.B. All authors have read and agreed to the published version of the manuscript.

Funding: Not applicable.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board at University of Utah (protocol code: IRB_00108150 and date of approval: 28 February 2018).

Informed Consent Statement: Informed consent was obtained from all participants involved in the study.

Data Availability Statement: The datasets of the current study are available from the authors on reasonable request.

Acknowledgments: The authors thank all the participants of this study and their parents.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Janssen, X.; Basterfield, L.; Parkinson, K.N.; Pearce, M.S.; Reilly, J.K.; Adamson, A.J.; Reilly, J.J. Non-linear longitudinal associations between moderate-to-vigorous physical activity and adiposity across the adiposity distribution during childhood and adolescence: Gateshead Millennium Study. *Int. J. Obes.* **2019**, *43*, 744–750. [CrossRef]
- Janz, K.F.; Letuchy, E.M.; Burns, T.L.; Eichenberger Gilmore, J.M.; Torner, J.C.; Levy, S.M. Objectively measured physical activity trajectories predict adolescent bone strength: Iowa Bone Development Study. *Br. J. Sports Med.* **2014**, *48*, 1032–1036. [CrossRef] [PubMed]
- Carson, V.; Rinaldi, R.L.; Torrance, B.; Maximova, K.; Ball, G.D.C.; Majumdar, S.R.; Plotnikoff, R.C.; Veugelers, P.; Boulé, N.G.; Wozny, P.; et al. Vigorous physical activity and longitudinal associations with cardiometabolic risk factors in youth. *Int. J. Obes.* **2014**, *38*, 16–21. [CrossRef] [PubMed]
- Janz, K.F.; Dawson, J.D.; Mahoney, L.T. Tracking physical fitness and physical activity from childhood to adolescence: The muscatine study. *Med. Sci. Sports Exerc.* **2000**, *32*, 1250–1257. [CrossRef] [PubMed]
- Lee, I.M.; Shiroma, E.J. Using accelerometers to measure physical activity in large-scale epidemiological studies: Issues and challenges. *Br. J. Sports Med.* **2014**, *48*, 197–201. [CrossRef] [PubMed]
- Kheirkhahan, M.; Nair, S.; Davoudi, A.; Rashidi, P.; Wanigatunga, A.A.; Corbett, D.B.; Mendoza, T.; Manini, T.M.; Ranka, S. A smartwatch-based framework for real-time and online assessment and mobility monitoring. *J. Biomed. Inform.* **2019**, *89*, 29–40. [CrossRef] [PubMed]
- Reeder, B.; David, A. Health at hand: A systematic review of smart watch uses for health and wellness. *J. Biomed. Inform.* **2016**, *63*, 269–276. [CrossRef]
- Creaser, A.V.; Clemes, S.A.; Costa, S.; Hall, J.; Ridgers, N.D.; Barber, S.E.; Bingham, D.D. The Acceptability, Feasibility, and Effectiveness of Wearable Activity Trackers for Increasing Physical Activity in Children and Adolescents: A Systematic Review. *Int. J. Environ. Res. Public Health* **2021**, *18*, 6211. [CrossRef] [PubMed]
- Mawston, N. Global Smartwatch Shipments Leap to 14 Million Units in Q3 2019. Available online: <https://www.strategyanalytics.com/strategy-analytics/blogs/wearables/2019/11/06/global-smartwatch-shipments-leap-to-14-million-units-in-q3-2019> (accessed on 6 November 2019).
- Shcherbina, A.; Hershman, S.G.; Lazzaroni, L.; King, A.C.; O’Sullivan, J.W.; Hekler, E.; Moayed, Y.; Pavlovic, A.; Waggott, D.; Sharma, A.; et al. The effect of digital physical activity interventions on daily step count: A randomised controlled crossover substudy of the MyHeart Counts Cardiovascular Health Study. *Lancet Digit. Health* **2019**, *1*, e344–e352. [CrossRef]
- Zhang, P.; Godin, S.D.; Owens, M.V. Measuring the validity and reliability of the Apple Watch as a physical activity monitor. *J. Sports Med. Phys. Fit.* **2019**, *59*, 784–790. [CrossRef]
- Bai, Y.; Hibbing, P.; Mantis, C.; Welk, G.J. Comparative evaluation of heart rate-based monitors: Apple Watch vs Fitbit Charge HR. *J. Sports Sci.* **2018**, *36*, 1734–1741. [CrossRef] [PubMed]
- Dooley, E.E.; Golaszewski, N.M.; Bartholomew, J.B. Estimating Accuracy at Exercise Intensities: A Comparative Study of Self-Monitoring Heart Rate and Physical Activity Wearable Devices. *JMIR mHealth uHealth* **2017**, *5*, e34. [CrossRef] [PubMed]
- Welk, G.J.; Corbin, C.B.; Dale, D. Measurement issues in the assessment of physical activity in children. *Res. Q. Exerc. Sport* **2000**, *71*, S59–S73. [CrossRef] [PubMed]
- Lee, P.H.; Macfarlane, D.J.; Lam, T.H.; Stewart, S.M. Validity of the International Physical Activity Questionnaire Short Form (IPAQ-SF): A systematic review. *Int. J. Behav. Nutr. Phys. Act.* **2011**, *8*, 115. [CrossRef] [PubMed]
- Hussey, J.; Bell, C.; Gormley, J. The measurement of physical activity in children. *Phys. Ther. Rev.* **2007**, *12*, 52–58. [CrossRef]
- Perez-Suarez, I.; Martin-Rincon, M.; Gonzalez-Henriquez, J.J.; Fezzardi, C.; Perez-Regalado, S.; Galvan-Alvarez, V.; Juan-Habib, J.W.; Morales-Alamo, D.; Calbet, J.A.L. Accuracy and Precision of the COSMED K5 Portable Analyser. *Front. Physiol.* **2018**, *9*, 1764. [CrossRef]
- Guidetti, L.; Meucci, M.; Bolletta, F.; Emerenziani, G.P.; Gallotta, M.C.; Baldari, C. Validity, reliability and minimum detectable change of COSMED K5 portable gas exchange system in breath-by-breath mode. *PLoS ONE* **2018**, *13*, e0209925. [CrossRef]
- Crouter, S.E.; LaMunion, S.R.; Hibbing, P.R.; Kaplan, A.S.; Bassett, D.R., Jr. Accuracy of the Cosmed K5 portable calorimeter. *PLoS ONE* **2019**, *14*, e0226290. [CrossRef]
- Reddy, R.K.; Pooni, R.; Zaharieva, D.P.; Senf, B.; El Youssef, J.; Dassau, E.; Doyle Iii, F.J.; Clements, M.A.; Rickels, M.R.; Patton, S.R.; et al. Accuracy of Wrist-Worn Activity Monitors During Common Daily Physical Activities and Types of Structured Exercise: Evaluation Study. *JMIR mHealth uHealth* **2018**, *6*, e10338. [CrossRef]
- LaMunion, S.R.; Blythe, A.L.; Hibbing, P.R.; Kaplan, A.S.; Clendenin, B.J.; Crouter, S.E. Use of consumer monitors for estimating energy expenditure in youth. *Appl. Physiol. Nutr. Metab.* **2020**, *45*, 161–168. [CrossRef]

22. Kim, Y.; Lochbaum, M. Comparison of Polar Active Watch and Waist- and Wrist-Worn ActiGraph Accelerometers for Measuring Children's Physical Activity Levels during Unstructured Afterschool Programs. *Int. J. Environ. Res. Public Health* **2018**, *15*, 2268. [[CrossRef](#)] [[PubMed](#)]
23. Bai, Y.; Welk, G.J.; Nam, Y.H.; Lee, J.A.; Lee, J.M.; Kim, Y.; Meier, N.F.; Dixon, P.M. Comparison of Consumer and Research Monitors under Semistructured Settings. *Med. Sci. Sports Exerc.* **2016**, *48*, 151–158. [[CrossRef](#)]
24. Bai, J.; Di, C.; Xiao, L.; Evenson, K.R.; LaCroix, A.Z.; Crainiceanu, C.M.; Buchner, D.M. An Activity Index for Raw Accelerometry Data and Its Comparison with Other Activity Metrics. *PLoS ONE* **2016**, *11*, e0160644. [[CrossRef](#)] [[PubMed](#)]
25. Hildebrand, M.; VT, V.A.N.H.; Hansen, B.H.; Ekelund, U. Age group comparability of raw accelerometer output from wrist- and hip-worn monitors. *Med. Sci. Sports Exerc.* **2014**, *46*, 1816–1824. [[CrossRef](#)]
26. Redenius, N.; Kim, Y.; Byun, W. Concurrent validity of the Fitbit for assessing sedentary behavior and moderate-to-vigorous physical activity. *BMC Med. Res. Methodol.* **2019**, *19*, 29. [[CrossRef](#)]
27. Byun, W.; Kim, Y.; Brusseau, T.A. The Use of a Fitbit Device for Assessing Physical Activity and Sedentary Behavior in Preschoolers. *J. Pediatr.* **2018**, *199*, 35–40. [[CrossRef](#)] [[PubMed](#)]
28. Butte, N.F.; Watson, K.B.; Ridley, K.; Zakeri, I.F.; McMurray, R.G.; Pfeiffer, K.A.; Crouter, S.E.; Herrmann, S.D.; Bassett, D.R.; Long, A.; et al. A youth compendium of physical activities: Activity codes and metabolic intensities. *Med. Sci. Sports Exerc.* **2018**, *50*, 246–256. [[CrossRef](#)]
29. Schofield, W.N. Predicting basal metabolic rate, new standards and review of previous work. *Hum. Nutr. Clin. Nutr.* **1985**, *39* (Suppl. 1), 5–41.
30. van Hees, V. Accelerometer Data Processing with GGIR. Available online: <https://cran.r-project.org/web/packages/GGIR/vignettes/GGIR.html> (accessed on 22 November 2020).
31. van Hees, V.T.; Gorzelniak, L.; Dean Leon, E.C.; Eder, M.; Pias, M.; Taherian, S.; Ekelund, U.; Renstrom, F.; Franks, P.W.; Horsch, A.; et al. Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. *PLoS ONE* **2013**, *8*, e61691. [[CrossRef](#)]
32. Hildebrand, M.; Hansen, B.H.; van Hees, V.T.; Ekelund, U. Evaluation of raw acceleration sedentary thresholds in children and adults. *Scand. J. Med. Sci. Sports* **2017**, *27*, 1814–1823. [[CrossRef](#)]
33. Choi, L.; Liu, Z.; Matthews, C.E.; Buchowski, M.S. Validation of accelerometer wear and nonwear time classification algorithm. *Med. Sci. Sports Exerc.* **2011**, *43*, 357–364. [[CrossRef](#)]
34. Farooq, A.; Martin, A.; Janssen, X.; Wilson, M.G.; Gibson, A.M.; Hughes, A.; Reilly, J.J. Longitudinal changes in moderate-to-vigorous-intensity physical activity in children and adolescents: A systematic review and meta-analysis. *Obes Rev.* **2020**, *21*, e12953. [[CrossRef](#)] [[PubMed](#)]
35. Evans, E.W.; Abrantes, A.M.; Chen, E.; Jelalian, E. Using Novel Technology within a School-Based Setting to Increase Physical Activity: A Pilot Study in School-Age Children from a Low-Income, Urban Community. *BioMed Res. Int.* **2017**, *2017*, 4271483. [[CrossRef](#)] [[PubMed](#)]
36. Brage, S.; Brage, N.; Franks, P.W.; Ekelund, U.; Wong, M.Y.; Andersen, L.B.; Froberg, K.; Wareham, N.J. Branched equation modeling of simultaneous accelerometry and heart rate monitoring improves estimate of directly measured physical activity energy expenditure. *J. Appl. Physiol. (1985)* **2004**, *96*, 343–351. [[CrossRef](#)] [[PubMed](#)]
37. Corder, K.; Brage, S.; Wareham, N.J.; Ekelund, U. Comparison of PAEE from combined and separate heart rate and movement models in children. *Med. Sci. Sports Exerc.* **2005**, *37*, 1761–1767. [[CrossRef](#)] [[PubMed](#)]
38. Ekelund, U.; Yngve, A.; Brage, S.; Westerterp, K.; Sjöström, M. Body movement and physical activity energy expenditure in children and adolescents: How to adjust for differences in body size and age. *Am. J. Clin. Nutr.* **2004**, *79*, 851–856. [[CrossRef](#)] [[PubMed](#)]
39. Giavarina, D. Understanding Bland Altman analysis. *Biochem. Med.* **2015**, *25*, 141–151. [[CrossRef](#)]
40. Shcherbina, A.; Mattsson, C.M.; Waggott, D.; Salisbury, H.; Christle, J.W.; Hastie, T.; Wheeler, M.T.; Ashley, E.A. Accuracy in Wrist-Worn, Sensor-Based Measurements of Heart Rate and Energy Expenditure in a Diverse Cohort. *J. Pers. Med.* **2017**, *7*, 3. [[CrossRef](#)] [[PubMed](#)]
41. Nuss, K.J.; Thomson, E.A.; Courtney, J.B.; Comstock, A.; Reinwald, S.; Blake, S.; Pimentel, R.E.; Tracy, B.L.; Li, K. Assessment of Accuracy of Overall Energy Expenditure Measurements for the Fitbit Charge HR 2 and Apple Watch. *Am. J. Health Behav.* **2019**, *43*, 498–505. [[CrossRef](#)]
42. Pope, Z.C.; Lee, J.E.; Zeng, N.; Gao, Z. Validation of Four Smartwatches in Energy Expenditure and Heart Rate Assessment During Exergaming. *Games Health J.* **2019**, *8*, 205–212. [[CrossRef](#)]
43. Welk, G.J.; Bai, Y.; Lee, J.M.; Godino, J.; Saint-Maurice, P.F.; Carr, L. Standardizing Analytic Methods and Reporting in Activity Monitor Validation Studies. *Med. Sci. Sports Exerc.* **2019**, *51*, 1767–1780. [[CrossRef](#)] [[PubMed](#)]