Technical Report

Estimating Children’s Physical Activity and Sedentary Behavior during the School Day: A Comparison of the Jawbone UP Move and ActiGraph Accelerometry

Suggested Citation:

**Background:** As consumer based physical activity tracking devices evolve, their ability to provide accurate data must continually be evaluated. Much of the research examining the validity of the data from these devices has focused on adult populations. However, there is expanding interest in the use of wearables to track the physical activity (PA) and sedentary behavior patterns of children, both in the research and school practice contexts. Few studies have focused on the ability of commercially available accelerometers to quantify children’s activity in a free-living context. **Objective:** This study examined the validity of the Jawbone UP Move triaxial accelerometer for measuring school day PA (moderate to vigorous physical activity [MVPA], steps) and sedentary behavior (sedentary minutes, maximum length of sedentary bouts [MLSB]) in children compared to the ActiGraph GT3X accelerometer. **Methods:** Fifty-nine elementary school children ranging in school year from kindergarten to fifth grade wore both devices on the waist during one school week. Concurrent validity of the Jawbone was examined with paired t-tests, intraclass correlations, and Bland-Altman plots. Kappa and phi statistics were used to assess agreement for categorical outcomes. **Results:** A total of 108 wear time days were available for analysis. Paired t-tests revealed significant differences in each of the outcomes measured, with effect sizes ranging from small to large. The ActiGraph measured 5.6 more wear time minutes (Cohen’s $d = 0.07$) and 568.1 more steps (Cohen’s $d = 0.28$) than the Jawbone, on average. The Jawbone measured 12.2 more MVPA minutes (Cohen’s $d = 0.66$) and 110.8 more sedentary minutes (Cohen’s $d = 1.95$) than the ActiGraph. The average MLSB was 34.8 minutes less (Cohen’s $d = 1.62$) for the ActiGraph. The Jawbone outcomes of steps (ICC = 0.93), MVPA (ICC= 0.82) and sedentary time (ICC=0.62) had significant agreement with ActiGraph outcomes, but MLSB did not. Bland-Altman plots showed proportional bias for all outcomes but steps. Inter-device agreement for classification within categories was substantial for steps ($\kappa = 0.73$) and fair for MVPA ($\kappa = 0.30$). **Conclusions:** The Jawbone UP Move yielded step outcomes that were relatively accurate, but showed low absolute bias for undercounting steps when compared to ActiGraph accelerometers. Users should exercise caution when interpreting the device’s MVPA minutes, sedentary minutes, and sedentary bout outcomes.

**Keywords:** physical activity, sedentary behavior, digital health, children’s health, wearables

**Abbreviations**
CI: confidence interval
ICC: intraclass correlation coefficient
MLSB: maximum length of sedentary bouts
MVPA: moderate to vigorous physical activity
PA: physical activity
Introduction

Addressing the low levels of physical activity (PA) and high levels of sedentary behavior in youth is a matter of utmost concern in the United States (US). Children who engage in more minutes of moderate to vigorous physical activity (MVPA) have fewer cardiovascular risks than their peers [1], and sedentary behavior is associated with increased likelihood of weight gain over time for children who are already above the 50th percentile for body mass index [2]. Guidelines state that for optimal health and development, children need to engage in at least 60 minutes of PA that is moderate to vigorous in intensity [3], and minimize lengthy periods of sitting [4,5]. Yet, only less than half of US children ages 6-11 are meeting recommendations for engaging in PA and limiting sedentary screen time [6]. Recently, a subcommittee of the National Physical Activity Plan Alliance outlined several areas where improvements in policy and environment could help facilitate an increase in children’s PA and a decrease in sedentary behavior. Of these, one target area is schools [7]. It is recommended that children achieve at least half of their daily PA during the school day, since most spend at least half of their waking hours at school [8]. However, to continue to build the research base surrounding the implementation and effectiveness of school day PA interventions, more large-scale, high quality studies are needed. And to conduct those studies, researchers and health promotion professionals need appropriate tools for measuring children’s PA in school settings.

Background

There are several commonly used methods for PA measurement in children, but with a range of characteristics in terms of reliability and validity, feasibility of use, and the ability to capture all aspects of children’s PA [9]. Accelerometry and pedometry are currently the most appropriate methods for the objective quantification of PA in large samples of children over long periods of time [9,10], such as the duration of a school week. While several research-grade pedometers have evidence of validity and reliability for collecting step measurements [11], pedometers generally require sealing to prevent tampering and to decrease the likelihood of reactivity from child users [12]. Additionally, many of the traditional spring-loaded models (which are the lowest in cost) capture only total step count, and need to be manually queried, logged, and reset at the end of each day. This presents significant logistical challenges for researchers wishing to capture data over multiple days, from many students.

Accelerometers are among the most frequently used instruments for quantifying children’s PA in free living settings [13], and have shown highly accurate PA measurement [10,14,15]. The depth of data that accelerometers provide on PA and sedentary behavior, along with their long battery life and data storage capabilities, make them preferable to pedometers for lengthy periods of measurement among many participants. However, commonly used accelerometers such as Actical (Philips Respironics, Murrysville, PA, USA) and ActiGraph (ActiGraph, Pensacola, FL, USA) are priced at more than $300 per device, making them cost prohibitive for many users. Thus, there remains a need to examine the validity evidence related to the use of commercially-available instruments that are time and cost-effective for the purpose of PA quantification in large samples of children.
With the emergence of new technology over the past decade, PA tracking applications and devices have become more affordable and accessible to the general public. Companies such as Jawbone, FitBit, Garmin, and Apple sell activity monitors for a range of prices. They are meant to provide consumers with information regarding the quantity and intensity of daily activities, including outcomes such as MVPA minutes, time spent sitting, and steps taken. The relatively fast-paced emergence and refinement of these technologies, contrasted with the longer time needed to conduct and publish quality research, makes it difficult to consistently find up-to-date information regarding the validity and reliability of these devices and their analytic algorithms.

The Jawbone UP Move (Jawbone, San Francisco, CA, USA) is a relatively new device (first sold in 2014) that is commercially available for ~$20 [16]. Though it is not marketed to youth, the UP Move has properties that make it suitable for children’s activity tracking (simple interface, lightweight, clips easily to clothing). To our knowledge, only one published study to date has examined the validity of PA data provided by the Jawbone UP Move device [17], and this study solely analyzed the accuracy of the device’s step count, not the outcomes of inactive time and active time, which are also provided by the Jawbone. Further, the sample in this study included ambulatory adults with multiple sclerosis, which limits the generalizability of the results.

The Goal of This Study
With an affordable price point and the ability to collect multiple days of data at a time, The Jawbone UP Move might provide a feasible, economical option for researchers and practitioners seeking to track PA in children. However, thus far little information is available about the measurement characteristics and accuracy of this device. The primary purpose of this investigation was to examine the validity of the Jawbone UP Move triaxial accelerometer for measuring children’s school day PA (steps and MVPA minutes) and sedentary behavior. The secondary purpose of this investigation was to provide a description of the feasibility of the Jawbone UP Move device for use in research and practical applications.

Methods

This study assessed the concurrent validity of an inexpensive, commercially available accelerometer (Jawbone UP Move) compared to a research grade accelerometer (ActiGraph GT3X, Pensacola, FL) with extensive validity evidence to support its use for the measurement of children’s PA and sedentary behavior. The present study was part of a larger project examining the implementation of classroom activity breaks in elementary schools. The investigation was approved by the Institutional Review Board at Boise State University. Parents were provided notification about the study with the opportunity to opt-out, and children were also verbally informed of the project and given the opportunity to decline assent.
Participants
Participants included 59 elementary school students from 17 classrooms at four public elementary schools in the Intermountain West. Participants wore the devices for up to five consecutive days within one school week, and data were collected over a total of 8 weeks during the 2015-16 and 2016-17 school years. The grade levels of the students ranged from kindergarten to fifth grade, which typically encompasses the ages of 5-11 years. However, data regarding specific birth dates were not collected in this study. After removing unusable data (see details below), validation data were available from 53 unique students in 16 classrooms, on a total of 108 days of wear time. A description of these participants is included in Table 1.

Table 1. Characteristics of sample

<table>
<thead>
<tr>
<th>Grade level</th>
<th>Percentage</th>
<th>Number</th>
<th>Percentage</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kindergarten</td>
<td>20.4</td>
<td>22</td>
<td>17</td>
<td>9</td>
</tr>
<tr>
<td>First</td>
<td>13.9</td>
<td>15</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td>Second</td>
<td>15.7</td>
<td>17</td>
<td>13.2</td>
<td>7</td>
</tr>
<tr>
<td>Third</td>
<td>15.7</td>
<td>17</td>
<td>17</td>
<td>9</td>
</tr>
<tr>
<td>Fourth</td>
<td>16.7</td>
<td>18</td>
<td>18.9</td>
<td>10</td>
</tr>
<tr>
<td>Fifth</td>
<td>17.6</td>
<td>19</td>
<td>18.9</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th>Percentage</th>
<th>Number</th>
<th>Percentage</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>64.8</td>
<td>70</td>
<td>66</td>
<td>35</td>
</tr>
<tr>
<td>Male</td>
<td>35.2</td>
<td>38</td>
<td>34</td>
<td>18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cases by school</th>
<th>Percentage</th>
<th>Number</th>
<th>Percentage</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>School A</td>
<td>13.9</td>
<td>15</td>
<td>13.2</td>
<td>7</td>
</tr>
<tr>
<td>School B</td>
<td>29.6</td>
<td>32</td>
<td>26.4</td>
<td>14</td>
</tr>
<tr>
<td>School C</td>
<td>56.5</td>
<td>61</td>
<td>60.4</td>
<td>32</td>
</tr>
</tbody>
</table>

In Table 1, “n = 108; refers to all observations in the sample” and “n = 53; refers to a reduced sample with one observation per participant.”

ActiGraph GT3X Characteristics
The ActiGraph GT3X triaxial accelerometer is one of the most widely used instruments for capturing PA patterns in youth [18]. As of 2013, ActiGraph accelerometers had been used in more than 100 studies for PA and sedentary behavior measurement in children 6-11 years old [13]. The ActiGraph accelerometers used in this study were initialized to measure vertical, horizontal, and perpendicular axes at an epoch length of 60 seconds and a frequency of 30 Hz. The device measures $4.6 \times 3.3 \times 1.5$ cm and weighs 27 grams, uses a lithium ion rechargeable 4.20-volt battery and functions on frequencies of 30-100 Hz. The GT3X must be initialized before every new use. ActiLife software version 6.11.9 was used for data analysis.
Jawbone UP Move Characteristics
The Jawbone UP Move is a commercially available triaxial accelerometer. To initialize the device, it must be paired via Bluetooth to its own account within the Jawbone UP mobile application. Researchers created a unique account and password for each device for this purpose. The accelerometers must be synced using a mobile device such as a smart phone or a tablet, as they do not have any accompanying cables to hook up externally. After this initial setup, the device syncs when the user has logged into the app with the appropriate password and the device is in range. To set up the account, parameters of height, weight, gender and age must be entered. Because these data were not collected in this study, each Jawbone device’s account was given the same anthropometric data. The calculations which depend on these data (e.g. calories burned) were not outcomes of interest in this study. The user interface on the device features a single circular button with LED lights around the perimeter of the button which, when pressed, will show progress toward a goal specified within the app by illuminating a portion of the lights. In addition, two small lights show when the tracker is in active versus sleep mode. The device uses a non-rechargeable lithium ion battery that will last up to six months, per manufacturer specifications. The sensor and clip are fit into a silicone casing which allows the device to stay attached to clothing such as straps, pockets, and waistbands. The device measures 4.5 x 3 x 1.4 cm, and weighs ~5 grams. Versions 4.8-4.26 of the Jawbone UP application were used for wireless syncing of data.

Data collection
For ease of wear, the ActiGraph accelerometers were attached to elastic belts with plastic buckles to be worn around the waist. The Jawbone accelerometers, which according to manufacturer specifications can be worn attached to clothing anywhere on the body, were sewn onto the same elastic belts, directly adjacent to the ActiGraphs. At the beginning of each school week during which data were collected, the researchers dropped off the devices to two separate classrooms (with 6 paired devices in each classroom). Per classroom, there were three ActiGraph/Jawbone belts to be worn by female students, and three paired belts to be worn by male students. Apart from tracking the grade level and gender of the students who wore each device, the data collection was completely anonymous. Selection of which three male and three female students who wore the paired belts was left up to each teacher. Students received the same accelerometer for each day of measurement. Teachers were asked to prompt the students to put the devices on at the start of each school day, and to remove the devices immediately prior to the dismissal of school in the afternoon (i.e., to not wear them home). Students were to wear the devices at their right hip and teachers were asked to monitor the students for misuse of the devices during the class day (e.g., taking them off, shaking them). Class schedules and school start/end times were obtained from each school’s master calendar. Upon pickup of the accelerometers from each classroom, the research staff verbally solicited the opinions of teachers regarding the use of the accelerometers. Any feedback the teachers provided was written down and are subsequently described in the feasibility section.

Data import
ActiGraph GT3X accelerometer data were uploaded to a computer via USB cable and imported into ActiLife software. The Jawbone UP Move accelerometers were synced wirelessly using a
Bluetooth connection to an iPad with the Jawbone UP mobile application. Once the accelerometers were synced within the app, the data were automatically stored on Jawbone’s remote server. To access the data in .csv format, researchers logged into the Jawbone web site using the unique accounts created for each device and downloaded the data spreadsheets from Jawbone’s dynamic API linked web page within the user account. The spreadsheet provided the sums of each activity outcome for each day the device was worn. Activity outcomes of interest for the current analysis included the following: a) total steps; b) total active time; c) total inactive time; and d) longest continuous inactive time. For comparison purposes, Jawbone outcomes of total active time, total inactive time, and longest consecutive inactive time were interpreted as total MVPA, total sedentary minutes, and MLSB, as done in previous work examining the validity of other Jawbone accelerometers [19,20]. The activity cut points used to calculate these outcomes are not publicly available on the Jawbone web site, nor are detailed definitions of the outcomes provided. Features of the Jawbone accelerometer data that were only available in the mobile interface (not the downloaded spreadsheet) included “start time” and “end time.” The start and end times listed for each day were interpreted as the beginning and end of Jawbone wear time, as the activity graphs within the app showed no movement before the start time or after the end time. The timestamps for start and end time were manually entered into the data spreadsheet for each participant on each day.

Data reduction and scoring
Several steps were taken to prepare the data for analysis. First, wear time outcomes were calculated. Parameters developed by Choi and colleagues were used for the wear time validation of the ActiGraph accelerometers. These parameters designate non-wear time as at least 90 continuous minutes of zero counts, with a 2 minute activity spike tolerance, provided that the activity spike is preceded by and followed by 30 minutes of zero activity [21]. Evenson cut points, scaled to 60 second epochs, were used to score the vertical axis data of the ActiGraphs: any minute with less than 100 counts was considered sedentary, any minute with 101-2295 counts was considered light activity, and any minute with more than 2295 counts was considered moderate to vigorous activity [22]. Because they were not otherwise available, wear time minutes for the Jawbone accelerometers were manually calculated by taking the difference in minutes between the start and end times of activity listed within the mobile application (i.e., a start time of 8:30 and end time of 15:30 would result in a calculation of 7 hours [or 420 minutes] of wear time). Any weekend days or other non-school days that acquired incidental activity from device transit were removed.

It was clear upon return of the devices that some of the Jawbone accelerometers had been removed from the belt and thus may not have been worn paired with the ActiGraph, in accordance to protocol. Prior research examining the validity of commercial accelerometers in a free-living context has controlled for wear time by using a criterion of both devices having “valid days,” i.e., meeting a threshold of minutes of wear time [20,23]. To filter out cases in which the separation of the devices caused large discrepancies in wear time, observations were omitted if the difference in total wear time between devices was greater than 15 minutes for any given day. This resulted in omitting 6 participants, leaving 53 participants and 108 days of data with matching wear time.
Eight cases were omitted from the MVPA analyses due to participant tampering with the “timed workout” Jawbone feature, which resulted in inaccurate calculation of active time. The analyses on MLSB were limited to 103 cases because the ActiGraph bout analysis was programmed to report only sedentary bouts of at least 10 minutes, and in five cases the Jawbone data reported MLSB less than five minutes. Participants provided an average of 2.03 days of data, with 20 participants providing 1 day of data, 14 providing 2 days, 16 providing 3 days, and 3 providing 4 days of data.

**Statistical analysis**
All statistical tests were performed using SPSS Statistics version 23 (IBM Corp, Armonk, NY, USA). First, descriptive analyses were calculated, including means and standard deviations for the ActiGraph GT3X and Jawbone UP Move wear times, and the outcomes of active time, inactive time, MLSB, and total steps. Paired t-tests and Cohen’s $d$ effect sizes [24] were calculated to assess differences in each outcome. Intraclass correlation coefficients (ICCs) and 95% confidence intervals (CIs) were calculated using two way mixed method with absolute agreement. ICCs were used instead of Pearson correlations to estimate the strength of agreement between the GT3X and Jawbone UP Move because ICCs take into account the absolute level of agreement when estimating the size of the correlation, while Pearson correlations look only at the level of relative agreement. Because the purpose of was to examine the level of absolute agreement between the two devices, with the GT3X considered to provide the criterion measures, ICCs were chosen to examine validity evidence rather than Pearson correlations. The strength of ICCs was interpreted as: $<0.4$ poor, $0.4$ to $<0.59$ fair, $0.6$ to $<0.74$ good, and $>0.75$ excellent [25]. Alpha was set a-priori at 0.05.

Bland-Altman plots were used to visually examine the distribution of differences between the outcomes of the two devices and quantify the mean and 95% limits of agreement of those differences (i.e. absolute bias). Systematic bias is demonstrated in a Bland-Altman plot when the difference between methods of measurement varies in relation to the magnitude of the reference measure (i.e. proportional bias), or when the differences consistently differ from zero (i.e. fixed bias). To construct Bland-Altman plots, the differences (Jawbone outcome - GT3X outcome) were plotted on the y-axis against the GT3X outcomes on the x-axis as the criterion or reference measure, rather than plotting the mean of both measures on the x-axis [26].

Criterion-referenced agreement was analyzed using Cohen’s kappa and phi coefficient. The standards used were the minimum recommendations for school day PA for children 6-11 years old; 30 minutes of MVPA [8] and 6,000 steps, which corresponds to half of the 12,000 steps per day recommended for children [27–29]. To compare the categorizations, two new variables were coded for each device based on the activity levels reported in the MVPA and steps outcomes (meets standards = 1; does not meet standards = 0). Kappa statistics were evaluated based on the criteria defined by Landis and Koch: slight agreement ($0.0 – 0.20$) fair agreement ($0.21 – 0.40$) moderate agreement ($0.41 – 0.60$), substantial agreement ($0.61 – 0.80$), and almost perfect agreement ($0.81 – 1.00$) [30].
Because some participants provided more than one day of data for the paired devices, and such clustering has the potential to impact calculations that rely on standard errors (e.g., intraclass correlations), the analyses were also re-computed with a de-duplicated set of cases that included only one observation per student (see Tables 1-3, n=53 observations). Finally, factors related to the feasibility of using the devices were synthesized from teacher feedback and research notes. A description is provided in these feasibility notes regarding any challenges that occurred with the wearing and use of the devices, and the process of accessing and managing the resulting data.

Results

Figure 1 (top) depicts an example of the user interface for the Jawbone application after the device was synced, including the timeline of activity (the data parameters used to generate the graph are unspecified). Figure 1 (bottom) shows Axis 1 activity over time generated by the ActiGraph software from the same participant on that day. Visual comparison of the two activity plots indicates a similar pattern of PA. Next, statistical analyses were conducted to examine corresponding results for the two devices. The tables show analyses with both sets of data; however, for parsimony the interpretation of results focuses on the full set of observations. Additionally, the Bland-Altman plots from the de-duplicated dataset were nearly identical to the main dataset, and thus not presented in the results.

Figure 1. Comparison of visual displays yielded by the Jawbone and ActiGraph, on the same day and time period
Paired t-tests
Descriptive statistics for each of the outcome variables are listed in Table 2. Paired t-tests for the overall wear time minutes as well as for each of the activity outcomes for the Jawbone UP Move and the ActiGraph showed significant differences between the devices. Mean wear time for the ActiGraph was 5.6 minutes longer than the Jawbone mean wear time \((t(107) = 10.27, P < 0.001, \text{Cohen’s } d = 0.07)\). The ActiGraph measured 568.1 more mean steps than the Jawbone \((t(107) = 7.731, P < 0.001, \text{Cohen’s } d = 0.28)\). The average MVPA recorded was 12.2 minutes less for the ActiGraph compared to the Jawbone \((t(99) = -10.73, P < 0.001, \text{Cohen’s } d = 0.66)\). Mean sedentary minutes was 110.8 minutes less for the ActiGraph than the Jawbone \((t(107) = 23.12, P < 0.001, \text{Cohen’s } d = 1.95)\). The ActiGraph measured the MLSB at an average of 34.8 minutes less than the Jawbone devices \((t(102) = -11.34, P < 0.001, \text{Cohen’s } d = 1.62)\).

**Table 2. Descriptive analyses for Jawbone and ActiGraph outcomes**

<table>
<thead>
<tr>
<th></th>
<th>Jawbone(^a) Mean (SD)(^c)</th>
<th>GT3X(^a) Mean (SD)(^c)</th>
<th>Jawbone(^b) Mean (SD)(^c)</th>
<th>GT3X(^b) Mean (SD)(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wear time minutes(^c)</td>
<td>417.7 (75.7)</td>
<td>423.3 (74.9)</td>
<td>411.0 (82.7)</td>
<td>415.5 (81.9)</td>
</tr>
<tr>
<td>Steps</td>
<td>5,011.2 (2,092.8)</td>
<td>5,579.3 (1,962.7)</td>
<td>4,778.1 (2,134.2)</td>
<td>5,358.3 (2,095.1)</td>
</tr>
<tr>
<td>MVPA minutes</td>
<td>39.7 (16.6)</td>
<td>27.5 (19.6)</td>
<td>38.7 (18.3)</td>
<td>25.5 (18.1)</td>
</tr>
<tr>
<td>Sedentary minutes</td>
<td>306.3 (48.9)</td>
<td>195.5 (63.7)</td>
<td>319.1 (51.2)</td>
<td>199.6 (65.9)</td>
</tr>
<tr>
<td>MLSB</td>
<td>61.8 (27.1)</td>
<td>27.0 (13.8)</td>
<td>64.6 (26.4)</td>
<td>28.7 (14.6)</td>
</tr>
</tbody>
</table>

\(^a\)\(n= 108\); refers to all observations in the sample  
\(^b\)\(n=53\); refers to a reduced sample with one observation per participant  
\(^c\)Mean school day minutes from calendar data for all observations was 407 +/- 27

Correlations
Intraclass correlation coefficients (Table 3) showed excellent agreement for steps \((ICC=0.93, P < 0.001)\) and MVPA minutes \((ICC=0.81, P < 0.001)\) and good agreement for total minutes spent sedentary \((ICC=0.62, P < 0.001)\). There was no agreement between the estimates of MLSB \((ICC=-0.05, P = 0.70)\).

**Table 3. Intraclass correlations (ICCs) for Jawbone and ActiGraph outcomes**

<table>
<thead>
<tr>
<th></th>
<th>Jawbone(^a) ICC (95% CI)</th>
<th>GT3X(^a) ICC (95% CI)</th>
<th>Jawbone(^b) ICC (95% CI)</th>
<th>GT3X(^b) ICC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steps</td>
<td>0.93 (0.89, 0.95)</td>
<td>0.96 (0.93, 0.98)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MVPA minutes</td>
<td>0.81 (0.73, 0.87)</td>
<td>0.89 (0.81, 0.94)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sedentary minutes</td>
<td>0.62 (-0.48, 0.72)</td>
<td>0.71 (-0.55, 0.82)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLSB</td>
<td>-0.05 (-0.24, 0.14)</td>
<td>-0.13 (-0.39, 0.15)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)\(n= 108\); refers to all observations in the sample  
\(^b\)\(n=53\); refers to a reduced sample with one observation per participant  
\(^c\)All correlations were significant at \(P < 0.001\) except for MLSB, where \(P = 0.70\) (108 observations), and 0.83 (53 observations)
Bland-Altman Plots
The Bland-Altman plots (Figures 2-5) show the means (solid line) and 95% limits of agreement (dashed lines) for the differences (Jawbone-ActiGraph) of all outcomes. The limits of agreement ranged from -2064.98 to 928.69 \((M = -568.14)\) for steps, from -10.03 to 34.29 \((M = 12.13)\) for MVPA minutes, from 13.18 to 208.52 \((M = 110.85)\) for sedentary minutes, and from -26.04 to 95.44 \((M = 34.7)\) for MLSB. Analysis of the regression lines (red lines) for each scatterplot indicated that there was no pattern of proportional bias for steps (slope = -0.007, \(P = 0.848\)), though the Jawbone showed low absolute bias toward underestimation of steps compared to the ActiGraph. The regression lines demonstrated patterns of proportional bias for MVPA minutes (slope = -0.314, \(P < 0.001\)), sedentary minutes (slope = -0.511, \(P < 0.001\)), and MLSB (slope = -1.123, \(P < 0.001\)). The Jawbone showed moderate absolute bias toward overestimation of these outcomes.

Figure 2. Bland-Altman plot of step count differences (y-axis) versus GT3X steps (x-axis). Note: Solid line is the mean of differences (Jawbone-ActiGraph), dashed lines represent upper and lower 95% limits of agreement, red line indicates line of best fit.
Figure 3. Bland-Altman plot of step MVPA differences (y-axis) versus GT3X MVPA (x-axis).

Figure 4. Bland-Altman plot of differences in sedentary minutes (y-axis) versus GT3X sedentary minutes (x-axis).
Figure 5. Bland-Altman plot of differences in maximum length of sedentary bouts (y-axis) versus GT3X maximum length of sedentary bouts (x-axis).

Classification for Meeting Activity Guidelines
Agreement between the Jawbone and ActiGraph for the classification of meeting school day recommendations for PA is shown in Table 4. Agreement between the devices for the classification of meeting the 6,000 steps per school day goal was substantial ($\kappa = 0.73$, $P < 0.001$; $\phi = 0.75$, $P < 0.001$), with the Jawbone correctly classifying 69% (24/35) of the days that met the 6,000 step-goal and 99% (72/73) of the days that did not meet the goal. The overall proportion of agreement for the steps classification was 89%. Agreement in classification for meeting 30 minutes of MVPA was fair ($\kappa = 0.30$, $P < 0.001$; $\phi = 0.42$, $P < 0.001$). Using the full sample, the Jawbone correctly classified 100% (30/30) of the days that met 30 minutes of MVPA, and 42% (29/70) of the days that did not meet 30 minutes of MVPA. Thus, the overall proportion of agreement for the MVPA classification was 59%.

Table 4. Agreement between devices in criterion-referenced standard categorization

<table>
<thead>
<tr>
<th></th>
<th>$\kappa$</th>
<th>$P$</th>
<th>$\phi$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meets 6,000 steps</td>
<td>0.73</td>
<td>&lt;0.001</td>
<td>0.75</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Meets 30 minutes MVPA</td>
<td>0.30</td>
<td>&lt;0.001</td>
<td>0.42</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

$^a$ $n = 108$; refers to all observations in the sample
$^b$ $n = 53$; refers to a reduced sample with one observation per participant
Feasibility
We noted several economic and logistic advantages to using the Jawbones for data collection. The mobile app used for syncing the Jawbone, as well as the companion website for data download, are both free. When each device is paired with a unique account, the data upload is simple, with Bluetooth syncing of the Jawbones to a mobile device such as a tablet or smartphone, after which the data are stored on the Jawbone site. Then, up to a year’s worth of daily data can be downloaded in one spreadsheet format. This makes the collection of multiple days of data relatively straightforward. However, there are some limitations. Because each device must be synced to its own account on the Jawbone mobile application to upload the activity data, the syncing process cannot be done in bulk (unless the users possess multiple mobile devices with the application installed on each), which is time intensive. Currently, there is no way to filter analyses by epochs of wear time (i.e., downloaded data only provides daily totals). Further, the data spreadsheet does not provide a start hour and end hour; although this information is available from the mobile application, it needs to be recorded and input manually into the data spreadsheet. These factors make large scale data cleaning more complex than using ActiGraph software, which allows for wear-time validation of each device based on downloaded data. Jawbone does offer a contact for research support (research@jawbone.com), as well as an online portal that can be utilized by software developers, neither of which were utilized for the current study.

The device features also presented some drawbacks for data collection. Some students enabled the “timed workout” feature of the device by pressing and holding the central button. While we originally thought this would not cause a discrepancy in the activity tracking of the device, upon analyses we noted that timing a workout artificially inflates the active minutes measured by the device by classifying all minutes during the timed workout as active minutes. It is worth noting that if the device is worn on clothing as intended by the manufacturer, younger children might require the help of a teacher or research assistant to put the devices on. This was not an issue for the validation devices examined here, which were sewn onto an elastic band. As an alternative to wearing the Jawbone UP Move on the hip, the same tracker can be purchased with an adjustable wristband, which might improve ease of use; however, no validity data exists for this wear location.

Discussion

Key Results
Steps. The evidence related to the use of the Jawbone UP Move for estimated step outcomes indicates a high relative agreement with the ActiGraph (ICC = 0.93). The mean difference shows the Jawbone UP Move consistently undercounts the number of steps by approximately -570 steps but the spread of the data seen in the Bland-Altman plot indicates no other systematic bias. Consistent with its bias for undercounting, the step outcomes from the Jawbone caused some misclassification of individuals who met the 6,000-step standard compared to the ActiGraph but overall the agreement was substantial. Taken together, the evidence supports the use of the Jawbone UP Move step count data by researchers and practitioners to examine
total student steps at school, as a suitable low cost option for multi-day step monitoring. However, researchers should recognize the potential impact of the undercount when comparing step counts to criterion-referenced standards, or across studies using different instruments for measuring steps.

**MVPA minutes:** Similar to steps, the relative agreement between devices was high for the MVPA outcome (ICC = 0.81). However, a mean difference of approximately 12 minutes between the two devices shows less acceptable absolute agreement, with the Jawbone overestimating MVPA. This overestimation led to a misclassification of individuals according to the Jawbone that did not meet MVPA recommendations according to the ActiGraph. The Bland-Altman plot revealed a significant negative trend in differences (slope = -0.314) relative to the magnitude of the ActiGraph measurement. With the high volume of individuals misclassified as not meeting standards, in addition to the moderate proportional bias, we conclude that there is not sufficient validity evidence for the use of the Jawbone UP Move’s MVPA estimates for research purposes.

**Sedentary Minutes:** Though the ICCs showed moderate but significant (ICC=0.62) relative accuracy, the Jawbone provided large overestimates of the absolute amount of sedentary time compared to the ActiGraph measure. The mean difference was approximately 110 minutes, with the Jawbone overestimating sedentary minutes. Additionally, the Bland-Altman plot depicted a significant moderate proportional bias (slope = -0.511). Results from these analyses indicate that there is not sufficient validity evidence to support the use of the Jawbone’s sedentary minutes outcome for sedentary behavior measurement in children.

**Sedentary Bouts:** The ICCs indicated there was no agreement between the devices for the measurement of MLSB. The mean difference was approximately 35 minutes between the two devices, and a significant negative trend of proportional bias (slope = -1.123) was shown on the Bland-Altman plot. Thus, similar to the sedentary minutes outcome, there is insufficient validity evidence to support the use of the Jawbone’s MLSB outcome for the measurement of children’s sedentary behavior.

**Comparison with Prior Work**
To date, few studies have examined the accuracy of the data produced by commercially available accelerometers when worn by children under 11 years old [31,32]. There are only a handful of PA interventions using consumer wearables in youth as well, highlighting the overall shortage of research using commercially available accelerometers in this population [33]. In a recent literature review of studies examining the validity of commercially available activity trackers, only 2 of the 22 studies included were conducted with youth participants, and these two studies were focused on sleep tracking outcomes [34]. Research examining the validity evidence to support the use of the Jawbone UP (a wrist-worn triaxial accelerometer available from Jawbone) to measure PA outcomes produced similar patterns of results to those of the current study. The Jawbone UP and UP24 models have been evaluated for accuracy of their step count measurement in several studies conducted in both free-living [19,20] and laboratory [35,36] conditions. Though these studies exclusively tested the wrist-worn Jawbone devices
with adults, the results for the metrics of agreement for step counts were comparable to those found in this study: ICC = 0.99 compared to research grade accelerometers [19]; systematic undercount of steps (~100 steps less) compared to direct observation [35]; mean absolute percentage error 2-10% compared to a validated biomechanics sensor [36]. These same models have shown slightly lower accuracy (ICC = 0.70 [19] ICC = 0.56 [20]) for estimating MVPA minutes, consistent with the present research on the Jawbone UP Move.

**Strengths**

Strengths of this study include that it is the first of its kind to consider the validity of a low-cost commercially available PA tracking device for measuring activity levels in free-living children during the school day. Data were collected from children of several different grades and different gender, as well as across multiple schools, adding to the generalizability of these findings to children in elementary school settings. Participants wore the devices for all activities of the school day, which allowed for the inter-device comparison of a wide range of activities, including more vigorous intensity activities such as recess and physical education, and sedentary activities such as seated classroom lessons. A detailed description of the setup and procedures necessary to use the Jawbone device was provided, as well as notes on feasibility which can not only help researchers, but also consumers and practitioners.

**Limitations**

When quantifying children’s movement in a free-living context, there is a potential for behavior that is not in compliance with research protocol, as opposed to data that are gathered in a controlled laboratory setting. In the broader study, some teachers reported deviations in protocol such as children taking off devices, or not wearing the devices at all. We acknowledge that some of these deviations may have occurred in our study. One strategy we employed to mitigate these potential errors in the data was to remove observations where large discrepancies in wear time existed between the Jawbone and ActiGraph. Generalizability of these results to “real life” wearing of the Jawbone devices is somewhat limited because the Jawbone devices were not worn exactly as specified by the manufacturer; they were sewn onto a belt and worn over clothing rather than attached directly to clothing.

The act of adapting PA behavior in response to wearing a PA tracking device, also known as reactivity, is a concern in PA interventions. While assessing the level of reactivity caused by wearing one device versus the other was not a primary aim of this study, it does warrant consideration. The Jawbone UP Move device can provide some feedback if the button is pushed, as there is a lighted display with a visual representation of steps achieved (similar to a clock face). Devices that do not have a digital readout may be less likely to cause reactivity than are devices with a user interface that details activity outcomes of the user [37], though this has not been consistently shown [38]. Studies examining sealed [39,40] and unsealed [41] pedometers worn by children have not shown reactivity to cause irregularities in physical activity outcomes. However, the type of device worn (i.e., devices with a readable interfaces versus those without) may cause other complications which impact school-based research, such as behavioral issues arising from the distraction of checking a device at inappropriate times.
This is an important limitation to consider for the Jawbone’s feasibility for utilization in schools, whether for practice or research purposes.

**Conclusions and Implications for Research and Practice**

Given the difference in cost and the strong evidence of relative consistency between the two devices, we see the Jawbone UP Move as a viable alternative to the ActiGraph GT3X accelerometer to measure step counts of children 6-11 years old during the school day, for the purposes of comparing within individuals across time, and across groups. However, the outcomes of MVPA, sedentary minutes and MSLB did not yield sufficient evidence of validity for use in research. Further, we suggest that the sedentary minutes and MLSB outcomes be interpreted with caution by consumers. Because the analysis software used by the Jawbone may change over time, future users of the device should examine the web site for information on new software releases that may alter the data interpretation that is performed by Jawbone’s algorithms.

As the consumer market for PA tracking technology expands and advances, it is important to continue to evaluate the potential for commercially available PA trackers to provide valid data for PA research. Many of the research-grade pedometers currently available do not have the capacity to capture and store multiple days of data, nor do they record start and end times, which is important for researchers who need to be able to conduct quality assurance checks that the devices were worn during the correct time periods (i.e., worn for an entire school day, but not worn home after school). At a minimum, confirming the validity of affordable PA tracking devices that can capture multiple days of step count data would provide researchers with an economical and logistically-viable approach to obtaining data on one key PA outcome measure among relatively large samples of children.

**Acknowledgements**

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A150277. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

**Conflicts of Interest**

None declared.
References


[16] https://www.jawbone.com


