Received: 3 January 2017 | Revised: 14 December 2017 | Accepted: 19 December 2017

DOI: 10.1002/bin.1518

RESEARCH ARTICLE

WILEY

Further evaluation of observational and mechanical measures of physical activity

Carole M. Van Camp | **Diane Berth**

Psychology Department, University of North Carolina, Wilmington, NC, USA

Correspondence
Carole Van Camp, Psychology Department, University of North Carolina Wilmington, 601S. College Road, Wilmington, NC 28403.
Email: vancampc@uncw.edu

Recent behavioral research has used a variety of methods to quantify physical activity, including direct observation using the Observational System for Recording Physical Activity (OSRAC), and automated devices such as heart rate (HR) monitors, pedometers, and Fitbit Accelerometers. The current study evaluated the concurrent validity of these measures, as well as the reliability of the pedometers and Fitbits. Four children engaged in 15 activities listed in the OSRAC coding system while their HRs and steps taken were measured. The results indicated that for most activities, HR covaried with the OSRAC activity levels, although some exceptions were noted. Steps per minute as measured by the Fitbit also covaried with the OSRAC activity levels and Fitbit data were reliable when activities did not involve subjects losing contact with the ground. Pedometers produced similar but less reliable steps per minute data.

1 | INTRODUCTION

The importance of incorporating physical activity in the daily lives of children has been well documented. Children who are overweight or obese are at an increased risk for becoming obese as adults (Boreham & Riddoch, 2001), which in turn is associated with a myriad of health problems, including cardiovascular disease, diabetes, stroke, and even some cancers (see Lauer, Burns, & Daniels, 2006 for a review). Even in children of normal weight, physical activity is important to maintain cardiovascular health and insulin sensitivity (Barbeau, Gutin, & Sothern, 2005). The World Health Organization (WHO, 2011) and the Centers for Disease Control and Prevention (CDC, 2017) recommend that children engage in 60 min a day of moderate-to-vigorous physical activity (MVPA), such as biking, brisk walking, and running. Unfortunately, the majority of children are not meeting these guidelines (Troiano et al., 2008), necessitating the development of effective interventions. Many have suggested behavior modification as a particularly promising approach (Lauer, et al., 2006); indeed, there has been an increase in behavioral assessments and interventions aimed at increasing physical activity in children in the behavior analysis literature (e.g., Hayes & Van Camp, 2015; Kuhl, Rudrud, Witts, & Schulze, 2015; Zerger, Normand, Boga, & Patel, 2016).

An important consideration in such evaluations is how physical activity is measured (Van Camp & Hayes, 2012). Among the most commonly used measures are direct observation and objective measures taken via mechanical

Special thanks to Bruce Pate, Heather Gibson, Alyssa Gandhi, Kaitlin Holloway, Lynda Hayes, and Jenna Price for their assistance in data collection and analysis.

devices such as pedometers, accelerometers, and heart rate (HR) monitors (see Sirard & Pate, 2001 for a review). Increasingly, affordable accelerometers have become available to the public and researchers alike. For example, recent behavioral studies have utilized the Fitbit to measure steps taken throughout the day (Kurti & Dallery, 2013; Washington, Banna, & Gibson, 2014). Reliability and validity of the Fitbit has been established in adults under laboratory conditions (e.g., Fortune, Lugade, Morrow, & Kaufman, 2014; Noah, Spierer, Gu, & Bronner, 2013). To date, published studies on the Fitbit's reliability and validity with school-aged children have been limited to studies of polysomnography (Meltzer, Hiruma, Avis, Montgomery-Downs, & Valentin, 2015; Roane, Van Reen, Hart, Wing, & Carskadon 2015); thus, evaluations of Fitbits when worn by children engaged in physical activity in their natural environment are needed.

Another important consideration in the field of physical activity assessment is how MVPA is defined. Although there is some disagreement with which absolute HR correspond with MVPA, many studies use a cut off of over 139 beats per minute (BPM) for moderate physical activity and over 159 BPM for vigorous physical activity (see Epstein et al., 2001 for a review). With regard to steps per minute (SPM), studies have evaluated steps taken and HR simultaneously where subjects engaged in structured treadmill walking or jogging. A review of such studies has suggested that 100 SPM is approximately the same as 10 min of brisk walking, a moderately intense activity (Tudor-Locke et al., 2011). However, these evaluations were conducted with adults and children over the age of 11 years who were walking on a treadmill. Future research is needed to determine if these criteria are applicable to younger children and if HR and steps covary during activities in the natural environment.

Direct observation is considered the gold standard for activity measurement (Sirard & Pate, 2001). The Children's Activity Rating Scale (CARS; Puhl, Greaves, Hoyt, & Baranowski, 1990) includes five activity levels and expected HR. Calibration of the activity levels was conducted by having 25 young children engage in representative activities for 1 to 5 min while oxygen uptake (VO_2) and HR (BPM) were measured. Representative activities included lying or sitting (Level 1, expected HR below 100 BPM), standing (Level 2, 100–119 BPM), and walking on a treadmill at various speeds and inclines ranging from 2.5 mph, 0% upgrade (Level 3, 120–139 BPM), 2.5 mph, 5% to 10% upgrade (Level 4, 140–160 BPM), to walking at 2.5 mph, upgrade 15% (Level 5, >160 BPM). The Observational System for Recording Physical Activity in Children (OSRAC) is based partly on the CARS (Brown et al., 2006; McIver, Brown, Pfeiffer, Dowda, & Pate, 2009) for use with children in a variety of settings (e.g., home and school). The OSRAC uses similar codes to measure the intensity of physical activity, although the specific behaviors listed for each activity level in the OSCRAC are more varied and include walking, running, swinging from monkey bars, and riding a bicycle. However, although the inclusion of these less structured behaviors allows for the classification of activities one might typically see with children playing at home or at a playground, many have not been evaluated to determine if they are indeed associated with expected HR levels. Nevertheless, the OSRAC activity level codes have been used as a measure of MVPA in several behavioral studies, typically with Levels 4 and 5 combined as a measure of MVPA (e.g., Larson, Normand, Morley, & Miller, 2013; Zerger, Normand, Boga, & Patel, 2016).

Larson, Normand, and Hustyi (2011) conducted a preliminary evaluation of the OSRAC by having four preschool children engage in one activity (twice, for 1 min each) from each of the five activity levels (see Table 2). The children wore New Lifestyles NL-2000 pedometers to measure SPM and Polar F-4 HR monitors to measure BPM. The results indicated that overall SPM and BPM covaried and increased with each higher level activity. Few or no steps were recorded during Levels 1 and 2 activities, and HR remained low. Increases in HR were observed during Levels 3 and 4 activities, and the highest HR were observed during the Level 5 activity. Average SPM and BPM for each subject were not reported, but visual inspection of the graphed data suggests that Levels 4 and 5 activities were consistently associated with HR above 139 BPM for two subjects (Colt and Willie), and activities at Levels 3, 4, and 5 were consistently associated with step counts above 99 SPM for three subjects. There were some individual differences between subjects with regard to the absolute SPM and BPM (e.g., Dakota's HR appeared lower than the other subjects' across each activity). Overall, these data provide some preliminary validation of the OSRAC activity levels. In some cases, however, (see Willie and Dakota), HR at Level 3 were within the range of Level 4 HR, suggesting that for some individuals, there may not be as clear a distinction between what is considered light activity (Level 3) and moderate activity (Level 4). The results also suggest that SPM may be a valid measure of MVPA; however, SPM were somewhat lower during Level 4 compared to Level 3, suggesting that at least this particular Level 3 activity (walking) may be more moderate than light. One particular limitation of the Larson et al. study is that only one activity per level was evaluated and that each of these activities involved being directly on the ground. Activities such as cycling, translocation across monkey bars, and swinging, which are included in the OSRAC coding manual, were not evaluated. In addition, the reliability of the pedometers was not evaluated.

The purpose of the current study was to replicate and extend the Larson, Normand, and Hustyi (2011) study evaluating the construct validity of the OSRAC by comparing children's HR when they were engaged in increasingly more intensive physical activities. We evaluated the same behaviors evaluated by Larson et al., while measuring SPM and HR with the same model of pedometers and HR monitors. We extended the Larson et al. study by evaluating two additional behaviors for each OSRAC level (for a total of 15 behaviors), including behaviors that do not involve translocation. We also evaluated the Fitbit to measure SPM. Our primary purpose was to determine if the devices produced valid measures of physical activity, when compared to the HR measures. A secondary purpose, and further extension, was to evaluate the inter-device reliability of both the pedometer and Fitbit by having subjects wear two of each device at the same time. Finally, we examined the extent to which HR changed during each activity compared to the expected HR zones (Puhl et al., 1990), and the extent to which measured steps were indicative of MVPA (Tudor-Locke et al., 2011).

2 | METHOD

2.1 | Subjects, setting, and materials

Subjects were four typically developing children (three males and one female), ranging in age from 6 to 13 years. Parents provided permission for their children to participate in the study, and each subject provided assent. Parents also provided information regarding age, weight, and health status of their children, and the researchers measured their heights. All of the children were reportedly healthy and their body mass indexes were in the normal range (see Table 1 for detailed demographics).

Sessions took place at local city parks with playground equipment and flat, open spaces. Playground equipment included swings, monkey bars, and slides. Additional activity related equipment included a Radio Flyer wagon (36″ L × 17.5″ W × 6″ H), the children's own two-wheeled bicycles and hand weights (6, 8, and 14 lbs). Materials used to measure physical activity were Polar FT7 HR monitors (which included chest straps and watches), Fitbit Classic accelerometers, and Yamax Digiwalker Fit Solutions SW-200 pedometers. The Polar and Yamax Digiwalker were selected because they were the devices used in the Larson et al. (2011) study.

2.2 | Measures and reliability

Each subject wore a Polar HR monitor on an elastic strap placed against the skin and directly below the pectoral muscles. HR, in BPM, was displayed in real time on the watch display worn by the subject on the left wrist. Each subject also wore a nylon belt around the waist, onto which two Fitbits and two pedometers were clipped (one of each on the

TABLE 1 Subject characteristics

Subject	Gender	Age	Height	Weight (lbs)	BMI
Sam	Male	6	50.5″	58	16.3
Bob	Male	8	55.0″	68	15.8
Aubrey	Female	9	52.5″	72	18.8
Fred	Male	13	68.0″	107	16.3

left front and right front). Devices worn on the right side served as the primary data recorders, and devices on the left side served as the secondary data recorders used in the assessment of inter-device reliability. Fitbit displays showed steps taken, which by default were displayed as cumulative steps taken throughout the day. Observers recorded the number of steps displayed at the beginning and end of each 1-min activity trial, and SPM were calculated by subtracting the smaller beginning reading from the larger ending reading and dividing that value by the number of minutes for that activity (i.e., 1 min). Pedometer displays were reset to zero prior to each activity trial, and observers recorded the number of steps taken at the end of each trial.

Reliability of two Fitbits and two pedometers was assessed during 100% of trials for all subjects. Inter-device agreement (IDA) was calculated for each trial by dividing the smaller number of steps by the larger number of steps and multiplying by 100. Average IDA across all trials for the Fitbits was 81% (range across activities, 39–100%), 88% (range, 22–100%), 85% (range, 66–100%), and 89% (range, 67–100%) for Aubrey, Bob, Fred, and Sam, respectively. Average IDA for the pedometers was 76% (range, 38–97%), 70% (range, 49–100%), 79% (range, 27–98%), and 73% (range, 50–98%) for Aubrey, Bob, Fred, and Sam, respectively. IDA for specific activities are listed in Table 2 and will be further described in the results.

2.3 | Procedure

Fifteen distinct activities derived from the OSRAC (McIver, Brown, Pfeiffer, Dowda, & Pate, 2009) were evaluated, three for each of five activity levels (see Table 2) evaluated in Larson et al. (2011). Level 1 activities were sedentary (standing and lying still and riding in a wagon). Level 2 activities involved minimal effort or movement (standing while moving arms slowly and while holding a moderately heavy object and hanging off monkey bars). Level 3 activities involved light effort (walking and cycling slowly and swinging). Level 4 activities involved moderate effort (walking then jumping, climbing on monkey bars, and hanging off monkey bars while kicking legs). Level 5 activities involved vigorous effort (running, skipping or jumping, and translocation across monkey bars). Prior to activity trials, all activities were described to the subjects, and activities were demonstrated at the subjects' request. Subjects engaged in each activity in the same ways, with few exceptions. For Activity 2B (standing while holding moderately heavy object),

TABLE 2 Activities evaluated per activity level

Level	Activity code	Activity description
1	1A	Standing still
	1B	Lying still*
	1C	Riding in wagon
2	2A	Standing while moving arms slowly*
	2B	Holding a moderately heavy object
	2C	Hanging off bars
3	3A	Walking slowly*
	3B	Cycling slowly
	3C	Swinging w/o assistance or kicking legs
4	4A	Walk then jump alternating*
	4B	Climbing on monkey bars
	4C	Hanging off bars while kicking legs
5	5A	Running at a very fast pace*
	5B	Skipping or jumping
	5C	Translocation across bars while hanging

Note. Activities based on McIver, Brown, Pfeiffer, Dowda, and Pate (2009).

*Activities also evaluated in Larson et al. (2011).

subjects were given a choice of several hand weights, and they were asked to use the weight they felt was heavy, but not so heavy that they could not hold it for several minutes. Aubrey choose 6 lbs, Bob and Sam choose 8 lbs, and Fred 14 lbs. For Activity 5B, subjects were given a choice between skipping and jumping (while still moving forward); Aubrey chose skipping, and the other subjects chose jumping.

Subjects engaged in each activity 3 times, for a total of 45 trials. After each trial, subjects rested as described below. The order of activities was randomized across multiple sessions not exceeding 1 hr each day. Fred, Aubrey, Bob, and Sam completed the study in 3–6 days, respectively.

Prior to each daily session, HR monitors were placed on the subjects—the chest strap with the monitor below the sternum and the watch display on the left wrist. The belt outfitted with two Fitbits and two pedometers was placed around the subjects' waists. Instructions involved asking the subjects to engage in each activity for 1 min. During each 1-min trial, observers told the subjects when there were 30, 20, and 10 s remaining and counted down the last 5 s before saying, “Stop.” Subjects were asked to stop moving immediately when the researcher told them to stop (when the trial was over), and they were asked to remain still until it was time for the next activity to start. They also were instructed not to touch the various monitoring devices. Finally, both subjects and parents were told they could request to end the session at any time.

At the start of each daily session, subjects were asked to relax silently in a seated position for 3 to 5 min to establish baseline HR. The observers sat or stood next to the subject and watched the HR display until the subject's HR no longer increased or decreased for at least 20 s. The observers then made a note of that HR (in BPM) and used that value to determine when to end each subsequent rest period in-between activity trials for that day. New baseline HR measures were taken for each subject at the beginning of each day's sessions. Baseline HR averaged 91 BPM (range, 85–103) for Aubrey, 77 BPM (range, 65–90) for Bob, 86 BPM (range, 82–87) for Fred, and 96 BPM (range, 90–105) for Sam.

Following the establishment of baseline HR, observers recorded the starting step count on the Fitbits, reset the pedometers to zero, and then instructed the subject to engage in the first activity on their randomized list. After 1 min, the subject was instructed to stop moving immediately. To make it more likely that subjects and observers were in the same location during activities that involved moving long distances (e.g., walking, jogging, and cycling), subjects were instructed (when they had only 10 s remaining) to travel back toward the observers while continuing the designated activity, and the observers also walked toward the subjects as needed. At the end of each trial, observers immediately recorded the HR displayed on the subject's watch. Observers then recorded the step count on the Fitbit and pedometer displays and instructed the subject to sit near where the next activity was to take place. During this rest time, subjects could drink water and ask the researchers questions. Once subjects were sitting still and silently, observers reset the pedometers to zero and wrote down the step numbers from the Fitbit display. Observers watched the HR display until the subject's HR returned to baseline levels. At that point, subjects were asked if they were ready, and if so, the next trial began as described above. This sequence repeated until the subjects asked to stop or when 1 hr had elapsed since the beginning of that day's session.

Two researchers were present during sessions and were responsible for assuring subjects complied with all instructions and were indeed engaging in all planned physical activities. First, during the recruitment process, after parents agreed to allow their children to participate, we discussed with the children the specific activities (behaviors) we would be asking them to perform. Part of the assent procedure included a statement that the child agreed to engage in each activity as requested, otherwise we would need to stop. We demonstrated each of the activities and prompted the children to practice the less common activities (holding the weights, dangling off the monkey bars for a minute, etc.) to make sure they were able and willing to engage in those activities once the study began. In addition, we told the subjects which activity was next, we asked them if they were ready to continue, and we reminded them they could ask to stop at any time. Only one subject ever asked to stop (upon being told the next activity was dangling from monkey bars, he stated that his hands were sweaty and he was not sure if he could hold on for a minute, and he requested stop and continue the next day). If at any time the child stopped engaging in the activity (or if they were not still during the breaks in-between activities), we would have prompted them to comply, and if they did not do so immediately, sessions would have been terminated immediately. The four subjects in this study did not need any such prompting as they completed each activity as directed throughout the study.

3 | RESULTS

Average SPM as recorded by pedometers and Fitbits and HR in BPM for three trials at each activity level for each subject are displayed in Figure 1. Horizontal grey bars indicate the HR criteria associated with each activity level (Puhl et al., 1990).

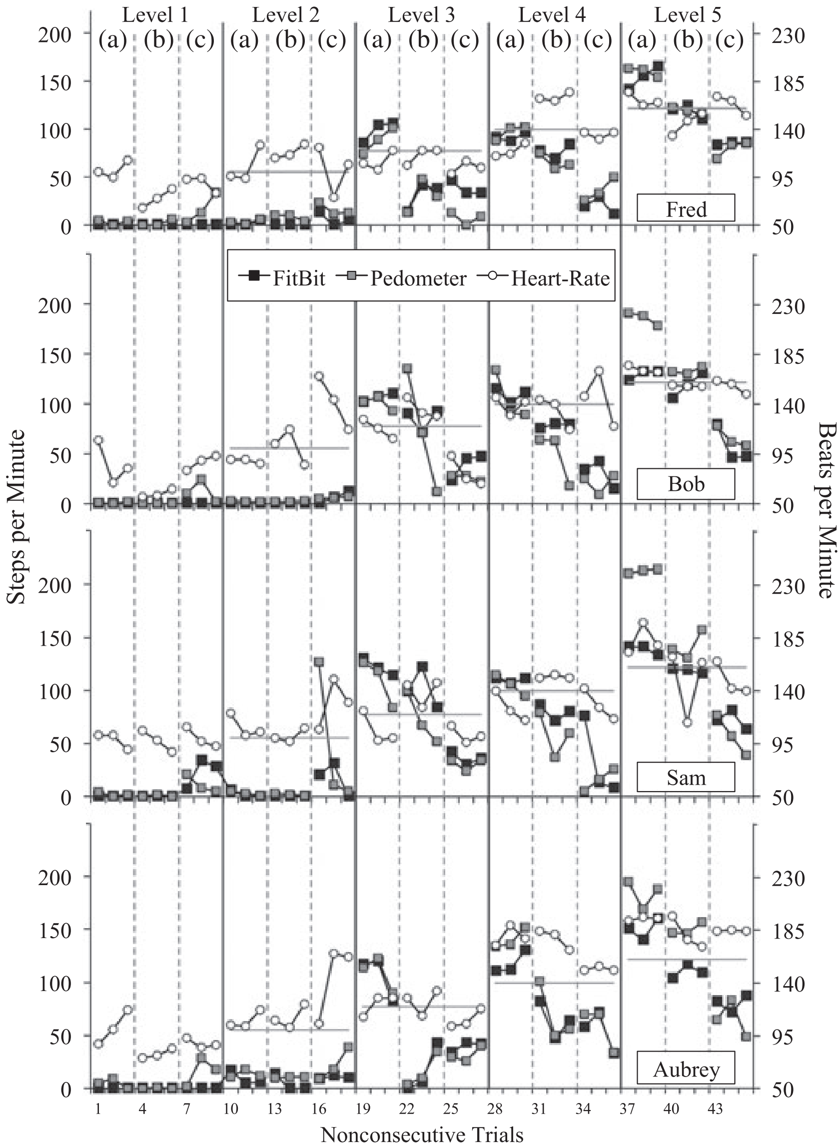

FIGURE 1 Steps per minute (left Y axis) as recorded by Fitbits and pedometers and heart rate in beats per minute (right Y axis), across all trials and subjects. The horizontal grey bars indicate the minimum heart rate criteria associated with each activity level: Level 2 = 100 BPM, Level 3 = 120 BPM, Level 4 and MVPA^{c} = 140 BPM, and Level 5 = 160 BPM

At Level 1, during standing still (1A) and lying still (1B), no steps were recorded by Fitbits, minimal steps were recorded by pedometers (no more than 10 SPM), and average HR (across the three trials for each activity) were 102, 86, 102, and 98 for Aubrey, Bob, Fred, and Sam, respectively. BPM were within or just above minimum L2 criteria (100 BPM) in most cases. Pedometers recorded some steps when all subjects rode in the wagon (1C, average 14 SPM), whereas Fitbits recorded steps only for Sam, and HR across all subjects remained no higher than 100 BPM. Overall, the lowest HR were associated with 1B (lying still), though both 1A and 1C were also associated with low HR. In none of these activities were steps actually taken by subjects; thus, that some steps were recorded by pedometers for all subjects during 1A and 1B is somewhat problematic, although the raw number of steps was minimal. Unexpected steps were also recorded by the Fitbit but for only one subject during one activity (Sam, 1C). IDA was over 80% for 9 of 12 activities (three per subject) for the Fitbits and 2 of 12 activities for the pedometers. IDA between Fitbits was consistently high for 1A and 1B.

At Level 2, during 2A (standing while moving arms slowly) and 2B (standing while holding a moderately heavy object), HR increased above minimum L2 criteria (100 BPM) in most cases. During 2C (hanging off monkey bars), higher HR were recorded for Bob, Sam, and Aubrey, with some trials producing HR over L3 and L4 criteria and MVPA[c] (i.e., over 139 BPM). This activity produced more variability across subjects, as Fred's HR was within L2 criteria. Steps were recorded although no steps were actually taken, but this may be an indication that pedometers and Fitbits were recording some movement when subjects moved their arms or adjusted the weights they were holding. On average, 6 and 3 SPM were recorded during 2A by pedometers and Fitbits, respectively. During 2B, an average of 6 and 1 SPM was recorded by pedometers and Fitbits, respectively. Again, pedometers inaccurately recorded more steps than did Fitbits. Likewise, more variability within subjects in steps per minute was noted during 2C for pedometers (see Sam and Aubrey's data in particular). Average SPM recorded was 23 and 10 for pedometers and Fitbits, respectively. In this case, some more jarring movement was expected given that hanging from monkey bars for 1 min at times necessitated repositioning of hands, though the few steps recorded did not adequately capture the higher degree of effort most subjects exerted (as indicated by their increased HR). IDA was over 80% for 5 of 12 activities for the Fitbits, and 4 of 12 activities for the pedometers. IDA between Fitbits and pedometers was lower than 80% for at least half of the subjects for activity.

At Level 3, for Activity 3A (walking slowly) average HR were 122, 118, 110, and 107 BPM for Aubrey, Bob, Fred, and Sam, respectively. HR were more often within L2 criteria, with some trials for Aubrey and Bob reaching L3 minimum criteria (120 BPM). There was more variability across subjects for Activity 3B (cycling slowly). Aubrey and Fred's HR remained somewhat elevated at L2 or L3 criteria (average 124 and 115 BPM, respectively), whereas Bob and Sam's HR were consistently above L3 criteria (average 136 and 139 BPM, respectively), and on some occasions met MVPA[c]. During Activity 3C (swinging without assistance or kicking legs), HR were at L1 or L2 criteria (average 102 to 109 BPM) and were notably lower than other Level 3 activities for Bob (average 78 BPM). Steps were relatively high as recorded by both Fitbits during 3A (average 108 SPM) and pedometers (average 102 SPM), with steps reaching MVPA[c] (i.e., over 100 SPM) in 67% (Fitbits) and 42% (pedometers) of trials across all subjects. During 3B, pedometers and Fitbits recorded similarly low numbers of steps for both Aubrey and Fred (average 16 and 31 SPM, respectively). More steps were consistently recorded for Bob during 3B by the Fitbits (average 85 SPM, range, 72–92), whereas more variable numbers of steps were recorded via pedometers (average 73 SPM; range, 12–135). During 3C, steps were consistently low and similar across both pedometers and Fitbits (average below 40 SPM) for all subjects, with the exception of Fred, for whom more steps were consistently recorded by Fitbits (compared to those recorded by pedometers). IDA between Fitbits was over 90% 3A (walking slowly) for all subjects, whereas IDA between pedometers was notably lower but still above 80% for three of four subjects. IDA was less consistent for the other activities, with three of eight activities above 80% for the Fitbits and one of eight above 80% for the pedometers.

Level 4 activities were associated with higher HR for most subjects, though not all consistently produced HR meet minimum L4 criteria and MVPA[c] (140 BPM). During Activity 4A (walk then jump), Aubrey's HR was consistently high (average 180 BPM, over L5 minimum expectations), and Bob's reached MVPA[c] in two of three trials (average

139 BPM). Fred and Sam's HRs were somewhat elevated but below $MVPA^c$ and at the low end of L3 criteria (average 120 and 126 BPM for Fred and Sam, respectively). During Activity 4B (climbing on monkey bars), HRs consistently met $MVPA^c$ for three subjects (Aubrey, Fred, and Sam). Bob's average HR was at the high end of L3, Sam's HR was within L4, and Aubrey and Fred's HRs were within L5 criteria. More variability in HRs within and across subjects was observed for Activity 4C (hanging off monkey bars while kicking legs). Aubrey's HR was consistently high and met $MVPA^c$ (average 152 BPM). Bob's HR was elevated and met $MVPA^c$ for two trials (average 146 BPM), Sam's HR met $MVPA^c$ for one trial (average 128 BPM), and Fred's HR did not reach $MVPA^c$ in any trial (average 135 BPM). Aubrey and Bob's HRs were within L4 criteria, and Fred and Sam's were within L3 criteria. Of note is that these general differences across subjects were also observed during 2C (hanging off monkey bars without kicking legs).

During 4A, SPM reached $MVPA^c$ (100 SPM) for two subjects as indicated by both pedometers (average 141 and 105 SPM for Aubrey and Bob, respectively) and Fitbits (average 118 and 109 SPM for Aubrey and Bob, respectively). Fred's steps were lower (average 91 and 97 SPM for Fitbits and pedometers, respectively), and Sam's SPM barely met $MVPA^c$ for all but one trial (average 110 and 105 SPM for Fitbits and pedometers, respectively). During 4B, SPM were consistently elevated but below $MVPA^c$ for all subjects. During 4C, in most cases, lower step counts were noted for both Fitbits and pedometers and at no time did SPM reach $MVPA^c$. IDA for 4A was over 90% for all subjects for the Fitbits and for one of four subjects for the pedometers (the other three were above 80%). IDA for 4B was above 90% for all subjects for the Fitbits and was above 80% for two subjects for the pedometers (the other two were below 80%). IDA was lower for 4C activities for both devices (with Fitbits reaching 80% for only two of four subjects and pedometers reaching 80% for only one subject).

The most consistently elevated HRs were observed during Level 5 activities. During 5A (running), HRs for all subjects (across all trials) reached L5 criteria (160 BPM) and were well above $MVPA^c$ (average 195, 171, 168, and 183 for Aubrey, Bob, Fred, and Sam, respectively). During 5B (skipping or jumping), HRs were elevated but most often within L4 criteria (except for Aubrey). Fred and Sam's HRs reached $MVPA^c$ in two trials (average 146 and 149 BPM, respectively), whereas HRs for Bob and Aubrey were higher and more consistently met $MVPA^c$ (average 156 and 182 BPM, respectively). Finally, during Activity 5C (translocation across monkey bars), HRs were elevated above $MVPA^c$ for all subjects and at L5 criteria for two subjects (average 184, 156, 164, and 149 for Aubrey, Bob, Fred, and Sam, respectively).

During 5A, steps were also high and met $MVPA^c$ as recorded by both Fitbits and pedometers. However, for Bob, Sam, and Aubrey, more steps were consistently recorded by pedometers (on average, 34 to 55 SPM more). During 5B, steps consistently met $MVPA^c$ for all subjects, and again some variability between Fitbits and pedometers was noted (see Aubrey's data in particular). In 5C, as with the other monkey bar related activities (2C and 4C), SPM did not reach $MVPA^c$ for pedometers or Fitbits for any subject (although there was less of a discrepancy between the devices compared to the other Level 5 activities, and generally SPM were higher when translocating across monkey bars compared to hanging from monkey bars). IDA between Fitbits and between pedometers was above 90% for 5A and at least 80% for 5B, for all subjects. More disagreement was noted between pedometers during 5C (IDA was below 80% two of four subjects) compared to the Fitbit (IDA was above 80% for all subjects).

To summarize, HRs and steps were consistently low during all Level 1 activities and Level 2 activities A and B. On average, higher HRs were observed during Activity 2C (hanging off bars), while steps remained low. HRs increased but remained below $MVPA^c$ for Level 3 activities, and SPM also increased. However, only in the case of 3A (walking slowly) did SPM reach $MVPA^c$. More variability in steps was observed for 3B activities (cycling slowly), but it should be noted that Fitbits did detect increased movement (though no literal steps were actually taken). Finally, activity 3C (swinging without assistance or kicking legs) was associated with lower HRs that were more within the range of Levels 1 and 2 sedentary activities. Levels 4 and 5 activities that involved translocation on the ground (4A walking slowly, 5A running, and 5B skipping/jumping) were associated with higher HRs that at times (4A) or always (5A and 5B) met $MVPA^c$, with covarying increases in SPM that also met $MVPA^c$. Most activities that involved interacting with monkey bars (4B climbing, 4C hanging while kicking legs, and 5C translocation) were associated with HRs that met $MVPA^c$, but SPM did not covary consistently with HRs.

Overall, IDA was higher for Fitbits than for pedometers (see Table 3). Fitbit IDA was over 80% for all subjects for seven activities: 1B (laying still), 3A (walking slowly), 4A (walk/jump), 4B (climbing monkey bars), and all Level 5 activities (5A running, 5B skipping, and 5C translocation across bars). IDA for Fitbits that did not meet the 80% criterion was above 60% with three exceptions (2A for Aubrey and 2C for Aubrey and Bob). In contrast, IDA for pedometers was over 80% for all subjects for only three activities (4A, 5A, and 5B), and IDA was below 60% for 13 activities across subjects.

4 | DISCUSSION

The current study measured the physical activity of four children via pedometers, Fitbits, and HR monitors while they engaged in various activities corresponding to OSRAC activity Levels 1 through 5. Generally, results replicated those of Larson et al. (2011) as SPM and BPM covaried with the activity levels, increasing with more intense activities. In particular, all Level 5 activities exceeded MVPA[c], as did many of the Level 4 activities. With the exception of activities that did not involve subjects' feet being on the ground, SPM as measured by the Fitbit appears to be a valid measure of physical activity, as IDA values exceeded 80% and steps covaried with HR. SPM as measured by pedometers were less reliable, often not meeting the 80% minimum criterion, and thus their validity may be questionable. An exception to this general finding is that reliability for pedometers was acceptable (but still lower compared to reliability for the Fitbits) for most subjects during walking, skipping, and running activities.

Some findings involving the new activities not evaluated by Larson et al. (2011) should be noted. HRs produced during certain activities were not consistent with the HR criteria ranges of the CARS coding system (Puhl et al., 1990), upon which the OSRAC is based. For Level 1 activities, all HRs meet criteria (near baseline levels and below 100 BPM), as suggested by the CARS coding system. Level 2 activities A and B meet the CARS criteria (100–119 BPM); however, activity 2C (hanging off the monkey bars) produced HRs more in line with CARS Level 3 (120–139), and in some cases BPM exceeded MVPA levels (i.e., >139). Walking slowly (3A) met CARS Level 3 criteria, but cycling (3B) did not meet Level 3 criteria for half of the subjects. Swinging without assistance (3C) produced HRs in the range of CARS Level 2 (often barely above 100 BPM). Some activities categorized under CARS Level 4 produced individual differences in BPM. For Activity 4A (walk then jump), Sam and Fred's HRs would be classified as Level 3 under the CARS criteria,

TABLE 3 Inter-device agreement percentages

Activity	Aubrey (Fitbit/Ped)	Bob (Fitbit/Ped)	Fred (Fitbit/Ped)	Sam (Fitbit/Ped)
1A	67/50	**100**/50	**100/93**	**100**/50
1B	**100**/38	**100/100**	**100**/67	**100**/67
1C	**100**/42	**100**/56	67/72	78/60
2A	47/76	**100**/57	75/**92**	**92**/54
2B	**90**/72	**100**/59	67/70	67/**81**
2C	39/**87**	22/65	79/**84**	**91**/72
3A	**94/90**	**98**/57	**95/88**	**95/80**
3B	62/67	**95**/62	73/76	76/79
3C	**91/80**	76/72	66/27	**95**/66
4A	**99/89**	**98/87**	**97/88**	**93/95**
4B	**91/81**	**94**/74	**93**/74	**92/87**
4C	**83/83**	60/49	66/69	**88**/55
5A	**94/96**	**93/96**	**98/96**	**91/98**
5B	**96/97**	**93/98**	**95/98**	**84/83**
5C	**83/91**	**89**/75	**98/88**	**89**/71

Inter-device agreement percentages in bold meet minimum acceptable criteria of 80%.

Bob's was nearly at Level 4, and Aubrey's HR would fit best under Level 5 (>160 BPM). Activity 4B (climbing monkey bars) produced HRs consistent with CARS Level 4, but Activity 4C (dangling from monkey bars while kicking legs) was more consistent with CARS Level 3 for half of the subjects. Finally, running (5A) produced HRs that meet CARS Level 5 criteria for all subjects. Skipping or jumping (5B) and translocating across monkey bars (5C) produced HRs that meet MVPA[c] for all subjects, but HRs were more within Level 4 CARS criteria for all but Aubrey (whose HR exceeded 160 BPM, placing her at CARS Level 5). Overall, some of the discrepant results may be due to Aubrey having less experience with (or less upper body strength to navigate) monkey bars, as it was during those activities that her HR was generally above that of the other subjects. In addition, because these activities were less structured than those evaluated by Puhl et al. (1990), who had children walk on treadmills, there may have been some variability in the amount of speed and effort subjects used while jumping, skipping, cycling, etc. Although this lack of control is a limitation of the current study, it is also a reflection of how topographically similar activities may vary from child to child. Indeed, individual differences in HRs were also evident in Larson et al. (2011) results.

A limitation of this study is that reliability between observers transcribing numbers from the devices was not assessed. That is, we did not have two observers simultaneously note the numbers from the Fitbit and Polar displays. However, unlike observing complex behavior in real time, transcribing numbers from devices is relatively simple and less likely to be influenced by observer bias, lack of training, or inadequate operational definitions. In more recent research, we have had multiple observers transcribing numbers from the Fitbit displays, and we have yet to have any disagreements. Still, without such reliability checks in the present study, there remains the possibility that errors were made in transcribing the numbers.

Reliability between two similar devices was assessed, which proved helpful in identifying limitations of the pedometer and Fitbit devices when used to assess physical activity of children in the natural environment. In particular, we noted the relatively low reliability of the pedometers, which produced IDA lower than 80% for all but four activities (4A, 5A, 5B, and 5C). These data limit any conclusions drawn from SPM data produced by the pedometers for many activities that are not considered vigorous physical activity. In contrast, the Fitbits produced IDA above 80% for 11 activities. Activities with lower than 80% IDA were 2A (78%), 2C (58%), 3B (76%), and 4C (74%). Generally, reliability was higher for activities involving translocation on the ground for both devices. This suggests that pedometers and Fitbits may not be the best measure of physical activity when children spend a significant amount of time on monkey bars (or engaging in other activities in which their feet are not on the ground).

The results of this study have implications for the assessment of physical activity in children. First, the study provides additional validation of the OSRAC activity levels, although it may be that some activities (2C, 3A, 3C, 4C, and 5C) should be reclassified on the basis of the HRs they produced in comparison to the CARS criteria. However, 3A (walking) may have produced lower HRs because subjects were asked to walk slowly. Walking at their typical pace or walking quickly may have produced HRs at MVPA levels. Second, the results of this study suggest that some individual differences may be observed with regard to HR and steps associated with particular activities. Although reviews of research in this area suggest a common standard for MVPA as indicated by more than 99 SPM or more than 139 BPM, single-subject research may include, when possible, individualized assessments to set personal HR or step criteria. For example, subjects could be asked to walk slowly, walk briskly and jog for a few minutes while their HR or steps were measured, thereby producing individualized baseline criteria for light activity (walking slowly), moderate activity (walking briskly), and vigorous activity (jogging). Such individualized criteria could then be used to determine the extent to which individual children meet MVPA criteria while engaged in any number of other activities (e.g., soccer, tag, and basketball).

The current results, and recent studies utilizing the OSRAC, have focused on MVPA (Levels 4 and 5; Larson, Normand, Morley, & Miller, 2013; Larson, Normand, Morley, & Miller, 2014); however, this emphasis ignores light activities. Breaks in sedentary behavior are important to consider in a comprehensive approach to healthy living. This is especially important for children who do not engage in the recommended levels of MVPA (Carson, Stone, & Faulkner, 2014); however, studies also have shown that children may meet MVPA guidelines but still be considered overly sedentary by engaging in more than 3 hr a day of physical inactivity (Leatherdale & Wong, 2009). Sedentary behavior has been associated with increased risk of cardiometabolic disease in children (Saunders et al., 2013). Even when controlling for MVPA, diet, and obesity, sedentary behavior has been associated with decreased cardiorespiratory fitness and insulin sensitivity (Mitchell & Byun, 2014). Thus, even engaging in light movements is preferred over sedentary behavior such as watching television, playing stationary video games, etc. Increasingly, researchers have advocated that interventions also should focus on decreasing sedentary time (Healy et al., 2008; Roemmich, Beeler, & Johnson, 2014). In the current study, the data suggest that pedometers were particularly less reliable during light activities, which is consistent with previous research comparing pedometers to direct observation and accelerometers (e.g., Kilanowski, Consalvi, & Epstein, 1999). As such, refinements in both behavioral measures and mechanical measures of physical activity should include light activities as well as MVPA.

In conclusion, results of this study replicate those of Larson et al. (2011) by demonstrating that SPM and HR covaried, and increased, as children engaged in five activities that required increasing physical exertion: laying still, standing while moving arms slowly, walking slowly, walking then jumping, and running. This study extended the Larson study by evaluating two additional activities from each OSRAC level, by comparing HR and SPM to established criteria, by using Fitbits to gather an additional measure of SPM, and by evaluating the IDA of the Fitbits and pedometers. Although many of the additional activities evaluated showed similar increases in SPM and HR across increasing activity levels, the results indicated that some OSRAC behavioral codes may be misaligned with regard to their associated activity levels and expected HR ranges. Specifically, 2C (hanging off the monkey bars) was more consistent with Level 3 HR criteria, 3B (cycling), and 3C (swinging without assistance) were more consistent with Level 2 HR criteria, and 5C (translocation across monkey bars) was more consistent with Level 4 criteria. Other activities (particularly those in Level 4) were associated with such individual differences in HR that topographically similar activities produced HRs that varied below and above $MVPA^{C}$ across subjects. These results suggest that preliminary heart-rate assessments many be needed to determine which behaviors produce HR within beneficial MVPA levels for individual subjects. Also, Fitbits were found to be more reliable (when compared to each other) and more valid (when compared to heart-rate monitors) than previously utilized pedometers. However, both Fitbits and pedometers were less reliable when activities involved subjects' feet not being on the ground (e.g., dangling from monkey bars). Thus, it cannot be assumed that instruments previously validated in structured environments (i.e., on a treadmill) are also valid in less structured environments where they are typically used (e.g., playgrounds). These results suggest that when used as the primary measure of physical activity, inter-device reliability of devices such as pedometers and Fitbits should be assessed during the course of the study.

ORCID

Carole M. Van Camp http://orcid.org/0000-0001-9709-459X

REFERENCES

Barbeau, P., Gutin, B., & Sothern, M. S. (2005). Physical Activity and Adiposity in Children and Adolescents. In *Handbook of pediatric obesity: Etiology, pathophysiology, and prevention*. Boca Raton, FL: CRC Press.

Boreham, C., & Riddoch, C. (2001). The physical activity, fitness and health of children. *Journal of Sports Sciences*, *19*(12), 915–929. https://doi.org/10.1080/026404101317108426.

Brown, W. H., Pfeiffer, K. A., McIver, K. L., Dowda, M., Almeida, M., & Pate, R. R. (2006). Assessing preschool children's physical activity: The observational system for recording physical activity in children-preschool version. *Research Quarterly for Exercise and Sport*, *77*(2), 167–176. https://doi.org/10.5641/027013606X13080769704361.

Carson, V., Stone, M., & Faulkner, G. (2014). Patterns of sedentary behavior and weight status among children. *Pediatric Exercise Science*, *26*(1), 95–102. https://doi.org/10.1123/pes.2013-0061.

Centers for Disease Control and Prevention (2017). Physical activity for everyone: Guidelines. Retrieved from http://www.cdc.gov/physicalactivity/everyone/guidelines/index.html

Epstein, L. H., Paluch, R. A., Kalakanis, L. E., Goldfield, G. S., Cerny, F. J., & Roemmich, J. N. (2001). How much activity do youth get? A quantitative review of heart-rate measured activity. *Pediatrics*, *108*(3), e44–e44. https://doi.org/10.1542/peds.108.3.e44.

Fortune, E., Lugade, V., Morrow, M., & Kaufman, K. (2014). Validity of using tri-axial accelerometers to measure human movement–Part II: Step counts at a wide range of gait velocities. *Medical Engineering & Physics*, *36*, 659–669.

Hayes, L. B., & Van Camp, C. M. (2015). Increasing physical activity of children during school recess. *Journal of Applied Behavior Analysis*, *48*, 690–695. https://doi.org/10.1002/jaba.222.

Healy, G. N., Dunstan, D. W., Salmon, J., Cerin, E., Shaw, J. E., Zimmet, P. Z., & Owen, N. (2008). Breaks in sedentary time: Beneficial associations with metabolic risk. *Diabetes Care*, *31*(4), 661–666.

Kilanowski, C. K., Consalvi, A. R., & Epstein, L. H. (1999). Validation of an electronic pedometer for measurement of physical activity in children. *Pediatric Exercise Science*, *11*(1), 63–68.

Kuhl, S., Rudrud, E. H., Witts, B. N., & Schulze, K. A. (2015). Classroom-based interdependent group contingencies increase children's physical activity. *Journal of Applied Behavior Analysis*, *48*, 602–612. https://doi.org/10.1002/jaba.219.

Kurti, A. N., & Dallery, J. (2013). Internet-based contingency management increases walking in sedentary adults. *Journal of Applied Behavior Analysis*, *46*, 568–581. https://doi.org/10.1002/jaba.58.

Larson, T. A., Normand, M. P., & Hustyi, K. M. (2011). Preliminary evaluation of an observation system for recording physical activity in children. *Behavioral Interventions*, *26*(3), 193–203. https://doi.org/10.1002/bin.332.

Larson, T. A., Normand, M. P., Morley, A. J., & Miller, B. G. (2013). A functional analysis of moderate-to-vigorous physical activity in children. *Journal of Applied Behavior Analysis*, *46*(1), 199–207. https://doi.org/10.1002/jaba.8.

Larson, T. A., Normand, M. P., Morley, J. J., & Miller, B. G. (2014). Further evaluation of a functional analysis of moderate-to-vigorous physical activity in young children. *Journal of Applied Behavior Analysis*, *47*, 1–12. https://doi.org/10.1002/jaba.127.

Lauer, R. M., Burns, T. L., & Daniels, S. R. (2006). *Pediatric prevention of atherosclerotic cardiovascular disease*. New York, NY: Oxford University Press.

Leatherdale, S. T., & Wong, S. (2009). Association between sedentary behavior, physical activity, and obesity: Inactivity among active kids. *Preventing Chronic Disease*, *6*(1), A26.

McIver, K. L., Brown, W. H., Pfeiffer, K. A., Dowda, M., & Pate, R. R. (2009). Assessing children's physical activity in their homes: The observational system for recording physical activity in children-home. *Journal of Applied Behavior Analysis*, *42*(1), 1–16. https://doi.org/10.1901/jaba.2009.42-1.

Meltzer, L. J., Hiruma, L. S., Avis, K., Montgomery-Downs, H., & Valentin, J. (2015). Comparison of a commercial accelerometer with polysomnography and actigraphy in children and adolescents. *Sleep*, *38*(8), 1323–1330. https://doi.org/10.5665/sleep.4918.

Mitchell, J. A., & Byun, W. (2014). Sedentary behavior and health outcomes in children and adolescents. *American Journal of Lifestyle Medicine*, *8*(3), 173–199.

Noah, J., Spierer, D. K., Gu, J., & Bronner, S. (2013). Comparison of steps and energy expenditure assessment in adults of Fitbit Tracker and Ultra to the Actical and indirect calorimetry. *Journal of Medical Engineering & Technology*, *37*, 456–462.

Puhl, J., Greaves, K., Hoyt, M., & Baranowski, T. (1990). Children's Activity Rating Scale (CARS): Description and calibration. *Research Quarterly for Exercise and Sport*, *61*(1), 26–36.

Roane, B. M., Van Reen, E., Hart, C. N., Wing, R., & Carskadon, M. A. (2015). Estimating sleep from multisensory armband measurements: validity and reliability in teens. *Journal of Sleep Research*, *24*, 714–721. https://doi.org/10.1111/jsr.12317.

Roemmich, J. N., Beeler, J. E., & Johnson, L. (2014). A microenvironment approach to reducing sedentary time and increasing physical activity of children and adults at a playground. *Preventive Medicine*, *62*, 108–112.

Saunders, T. J., Tremblay, M. S., Mathieu, M.-È., Henderson, M., O'Loughlin, J., Tremblay, A., … Quality Cohort Research Group. (2013). Associations of sedentary behavior, sedentary bouts and breaks in sedentary time with cardiometabolic risk in children with a family history of obesity. *PLoS One*, *8*(11), e79143.

Sirard, J. R., & Pate, R. R. (2001). Physical activity assessment in children and adolescents. *Sports Medicine*, *31*(6), 439–454. https://doi.org/10.2165/00007256-200131060-00004.

Troiano, R. P., Berrigan, D., Dodd, K. W., Masse, L. C., Tilert, T., & McDowell, M. (2008). Physical activity in the United States measured by accelerometer. *Medicine and Science in Sports and Exercise*, *40*(1), 181.

Tudor-Locke, C., Craig, C. L., Beets, M. W., Belton, S., Cardon, G. M., Duncan, S., … Hatano, Y. (2011). How many steps/day are enough? Children and adolescent. *International Journal of Behavioral Nutrition and Physical Activity*, *8*, 78–91.

Van Camp, C. M., & Hayes, L. B. (2012). Assessing and increasing physical activity. *Journal of Applied Behavior Analysis*, *45*(4), 871–875. https://doi.org/10.1901/jaba.2012.45-871.

Washington, W. D., Banna, K. M., & Gibson, A. L. (2014). Preliminary efficacy of prize-based contingency management to increase activity levels in healthy adults. *Journal of Applied Behavior Analysis*, *47*, 231–245. https://doi.org/10.1002/jaba.119.

World Health Organization (2011). Global recommendations on physical activity for health (5–17 years old). Retrieved from http://www.who.int/dietphysicalactivity/physical-activity-recommendations-5-17 years.pdf.

Zerger, H. M., Normand, M. P., Boga, V., & Patel, R. R. (2016). Adult attention and interaction can increase moderate-to-vigorous physical activity in young children. *Journal of Applied Behavior Analysis*, *49*, 449–459. https://doi.org/10.1002/jaba.317.

How to cite this article: Van Camp CM, Berth D. Further evaluation of observational and mechanical measures of physical activity. *Behavioral Interventions*. 2018;1–13. https://doi.org/10.1002/bin.1518