

Validity of Fitbit Charge 2 in Controlled College Physical Education Settings

Xiaofen D. Keating¹ (corresponding author), Jingwen Liu², Xiaolu Liu³, Rulan Shangguan⁴, Jianmin Guan⁵, & Li Chen⁶

¹Associate Professor, Department of Curriculum and Instruction at the University of Texas at Austin, USA, xk93@austin.utexas.edu

²Doctoral Candidate, Department of Curriculum and Instruction at the University of Texas at Austin, USA, jliu5@utexas.edu

³Doctoral Student, Department of Curriculum and Instruction at the University of Texas at Austin, USA, xiaolu.liu@utexas.edu

⁴Doctoral Candidate, Department of Curriculum and Instruction at the University of Texas at Austin, USA, gulan@utexas.edu

⁵Associate Professor, Department of Health and Kinesiology at the University of Texas at San Antonio, USA, Jianmin.guan@utsa.edu

⁶Professor, Department of Sports Management at Delaware State University, USA, lchen@desu.edu

Abstract

Purpose: The study aimed to determine the validity of the Fitbit Charge 2 (FC2) for measuring PA steps in college physical education settings using the Yamax SW-200 (YX) as the criterion device. **Methods:** Undergraduate students ($n = 17$) worn both FC2 and YX in two conditions (i.e., controlled and college physical education settings). The absolute value of average percent error (APE) was calculated to evaluate the agreement between FC2 and YX in both conditions. **Results:** FC2 estimates of total steps in walking, jogging, and running in the controlled condition were within acceptable range (i.e., $APE < 5\%$) without gender difference. However, FC2 tended to overestimate total steps per class in physical education teaching football and baseball units, and gender difference was significant while no difference in sport units were shown. **Conclusions:** FC2 revealed promising validity evidence measuring total steps in a controlled setting and its validity may be questionable in college physical education settings where student physical activity patterns are irregular.

Key words: Fitbit Charge 2, emerging adults, physical education in higher education

Introduction

It is widely known that the risks of many lifestyle-related illnesses can be greatly reduced by simply increasing the physical activity (PA) level on a regular basis (Reiner, Niermann, Jekauc, & Woll, 2013, US Department of Health and Human Services [USDHHS], 2008), which can significantly reduce medical costs and prolong life span in the general population (Colberg et al., 2016; Humphreys, McLeod, & Ruseski, 2014). Surprisingly, with well documented positive health benefits of PA, students have become more sedentary than ever before, regardless of age, gender, and ethnicity (Deliens, Deforche, Bourdaudhuij, & Clarys, 2015; Hesketh, Lakshman, & van Sluijs, 2017). In addition, a steep decline in PA is found among adolescents and young adults (Corder et al., 2017). Of great importance, it has been found that there are long-

term carry-over effects of sedentary habits. Sedentary children and young adults are more likely to develop sedentary lifestyles throughout their adulthood, which will lead to increasing risks of a poor health (Cawley, Frisvoldc, & Meyerhoeferd, 2013; Cleland, Dwyer, & Venn, 2012). As a result, researchers have suggested that PA should be promoted as early as possible to accrue health benefits (Trost, Cliff, Ahmadi, Van Tuc, & Hagenbuchner, 2017; Vella et al., 2016).

Among a variety of approaches of PA promotion, a number of studies have indicated that quality physical education has great potentials to help students increase their PA levels and maintain a physically active lifestyle (Dauenhauer, Keating, & Lambdin, 2017; Kohl & Cook, 2013; Sallis et al., 2012). In fact, promoting health-related fitness and PA has become one of the key components in National Association for Sport and Physical Education (NASPE) (2002) standards for physical education, along with other important components such as movement competency, knowledge about movement performance, personal and social responsibilities and recognition of the value of PA for personal well-being (Sallis et al., 2012). In addition, the general educational reform trend that emphasizes data-driven pedagogy in physical education (Dauenhauer et al., 2017; Soland, 2017) warrants the necessity of using valid methods to accurately assess students' PA, in order to plan and implement high quality physical education programs (Dauenhauer & Keating, 2011; Scruggs, 2013). However, the complexity of assessing PA has also been well documented in the literature (Kolt, 2013; Skender et al., 2016). In essence, in order to generate accurate PA data, the four elements in FITT principle (i.e., frequency, intensity, time, and type) need to be considered simultaneously when quantifying PA (Keating, Shangguan, Huang, Liu, & Chen, 2017). Looking at one element of PA at a time is likely to result in misleading or incomplete outcomes (Skender et al., 2016). The interactions among the four elements involved in FITT principle make it very difficult to objectively measure PA in a setting which involves a large group of individuals moving and interacting simultaneously within a relatively short period of time (Hurvitz, Moudon, Kang, Saelens, & Duncan, 2014; Keating et al., 2017). Researchers have noted that it is particularly more challenging to measure students' PA in a physical education setting due to the dynamic movement patterns (e.g., short and intermittent bouts of PA with inconsistent patterns across bouts) (Clemes & Biddle, 2013).

Therefore, special attention must be given to PA measurement in physical education focusing on the settings (i.e., indoor vs. outdoor), types of PA, the number and age of students, and total time needed (Kang, Mahar, & Morrow, 2016). Specifically, devices and methods used for measuring PA in physical education settings not only need to be reliable and valid, but also be relatively easy to administrate in a large number of students (Bort-Roig, Gilson, Puig-Ribera, Contreras, & Trost, 2014; Van Camp & Hayes, 2017). As suggested by Kang and colleagues (2016) the measuring and recording of student PA levels in class must not take a long time.

With the attempt of seeking feasible and practical means of

assessing students' PA in physical education, PA wearables have caught the attention of professionals in the field of PA measurement and assessment, and physical education (Dauenhauer et al., 2017; Lee, Drake, & Williamson, 2015; Melton, Buman, Vogel, Harris, & Bigham, 2016; Rote, 2017). As a matter of fact, PA wearables have gained a great deal of popularity in educational settings and hence are suggested to be a solution for assessing PA in physical education (Lee et al., 2015; Lindberg, Seo, & Laine, 2016; Rotich, 2016). Previous research indicated that PA wearables can produce objective PA data for teachers to enhance teaching effectiveness in physical education (Dauenhauer, Keating, & Lambdin, 2016; Kang et al., 2016; Rote, 2017; Sgrò et al., 2017) and promote PA during recess in elementary schools (Van Camp & Hayes, 2017). It was found that K-12 students could better learn how to analyze their own PA patterns using data collected by their PA wearables (Lee et al., 2015). Sgro and associates (Sgrò et al., 2017) used a PA wearable to assess elementary students' standing long jump performance and quantified key elements in skill acquisition. Studies using wearables measuring elementary student PA also found that attending physical education class had a positive carry-over effects on student in- and outside school PA (Dauenhauer & Keating, 2011; Silva et al., 2017). However, research on the motivational effects of PA wearables yielded inconsistent findings. A study done by Kerner and Goodyear (2017) indicated the negative influence of wearables on adolescents' PA motivation due to the peer competitions. Using Fitbit, resoundingly positive experiences were reported by both K-12 (Lehman, 2017) and college students (Rote, 2017). In Rote's (2017) study, a significant increase in PA were found among the experimental group, in which college students wore Fitbit devices throughout a semester. Regarding gender differences in PA data accuracy measured by wearables, it is surprising that this research question has not been fully examined in the literatures considering that the most recently published systematic literature reviews on the topic have not addressed the variable of gender (e.g., Evenson, Goto, & Furberg, 2015; Lyons, Lewis, Mayrsohn, & Rowland, 2014; Muntaner, Vidal-Conti, & Palou, 2016; Yang & Hsu, 2010). In spite of the lack of evidence concerning the accuracy of measuring PA in males vs. females, a handful studies have used PA wearables for both genders (Huang, Xu, Yu, & Shull, 2016; Melton et al., 2016; Rote, 2017).

Many studies have examined a variety of popular PA wearables and comparisons between different devices were made (Hickey et al., 2016; Kim, Barry, & Kang, 2015). It is important to note that examining concurrent validity seems to be the common approach to validate PA wearables in the existing literature (e.g., Montoye, Mitrzyk, & Molesky, 2017; Reid et al., 2017; Rosenberger, Buman, Haskell, McConnell, & Carstensen, 2016). In general, PA data measured by a target wearable(s), such as step counts, is compared against PA data reported by a formerly validated criterion device. The absolute percent error (APE) is usually calculated to evaluate agreement between the target and criterion device (Brian & Haegele, 2017), which is calculated based on the formula of $(\text{measured step} - \text{actual step})/\text{actual step} * 100$. To date, Yamax SW-200 has been served as the criterion device in measuring step counts in many studies (e.g., Clemes & Biddle, 2013; McCullagh, Brady, Dillon, Horgan, & Suzanne, 2016; Vanroy et al., 2014) due to its high reliability and validity in controlled settings at various

speeds.

Among a number of PA wearables that have been used in physical education teaching, Fitbit (Fitbit Inc., San Francisco, CA) is one of the widespread use PA wearables (Kerner & Goodyear, 2017; Rosenberger et al., 2016; Rote, 2017). As one of the commercially available PA trackers (e.g. Fitbit®, iWatch®, Jawborn, Garmin®, etc.), Fitbit is a PA self-monitoring device that has the capacity to track 24-hr PA in a free living setting and provide individualized feedback against PA goals set by the users (www.fitbit.com). Accompanied by smart phone Apps and/or web-based portals, Fitbit can wirelessly download PA data online or on users' smart phone to track and summarize PA, sleep, and weight data at any given time (www.fitbit.com). Moreover, the gamification system is also available for users to form virtual social groups, which allows self-invited group members to access individual rankings of any of the above variables (i.e., PA, sleep, and weight changes). Thus, instead of only focusing on measuring data related to individual healthy behaviors such as pedometers (Vallance, Eurich, Gardiner, Taylor, & Johnson, 2016), Fitbit also allows educators to use it for educational purpose as well by synching group data (Chu et al., 2017; Van Camp & Hayes, 2017). In addition, Fitbit Apps permits users to create a screen ID and choose what information can be shared in a group to protect user privacy. As such, Fitbit has been utilized for teaching and research on student PA behaviors (Cadmus-Bertram, Marcus, Patterson, Parker, & Morey, 2015; Kerner & Goodyear, 2017; Rote, 2017).

A handful of studies have tested the reliability and validity of Fitbits in both controlled environment and free-living conditions (Chu et al., 2017; Evenson, Goto, & Furberg, 2015; Montoye et al., 2017; Reid et al., 2017; Van Camp & Hayes, 2017). Previous studies on Fitbit indicated that there was high inter-device reliability for steps, distance, energy expenditure, and sleep among certain Fitbit models (Evenson et al., 2015). For example, through reviewing previously published projects on Fitbit, Evenson and associates (2015) reported that walking- and running-based Fitbit trials had consistently shown great inter-device reliability (steps: intraclass CC 0.76-1.00; distance: intraclass CC 0.90-0.99, and energy expenditure in Kcals: intraclass CC 0.71-0.97). Similar results were also reported by the study done by Takacs and associates (2014).

As noted earlier, PA wearables have been constantly changing in order to increasingly advance PA wearables. The same trend is also observed in Fitbit as a series of Fitbit products are available as time passes (i.e., Fitbit One, Fitbit Charge 2, etc.) (www.fitbits.com). Therefore, the way forward for enriching our knowledge on Fitbit and physical education is to continuously test and compare the reliability and validity of the newly released Fitbits in order to have the acceptable level of confidence for using these wearables (Montoye et al., 2017; Rosenberger et al., 2016). Overall, researchers have suggested that continuous efforts on examining the reliability and validity of new PA wearables are needed to keep up with the rapid and endless changes in PA trackers in various settings (Chu et al., 2017; Evenson et al., 2015; Van Camp & Hayes, 2017).

To date, to the best of our knowledge, there has been minimal research on the validity and reliability of FC2 because it is a relatively new product which was released in 2016 (www.fitbit.

com). Furthermore, no published studies have examined the accuracy of PA data collected by FC2 in physical education lessons in higher education settings. As such, it is unclear if FC2 is valid among college students. Given that researchers and practitioners usually choose PA wearables based on their primary PA outcomes of interest and costs (Rosenberger et al., 2016), it seems that FC2 may be continuously used in physical education settings due to its reasonably low costs and popularity (Van Camp & Hayes, 2017). Information about the validity of FC2 could increase our confidence in using FC2 to accurately measure college students' PA and enhance their learning in planning and living healthy lifestyles. Therefore, the purpose of this study was to examine the accuracy of FC2 measuring total steps in two different conditions: controlled and physical education settings. The reason for focusing on the accuracy of total steps was that it is one of the most commonly measured PA variables by PA wearables (Chu et al., 2017; Evenson et al., 2015). In addition, gender and sport unit differences (i.e., football vs. baseball) in the PA data accuracy measured by FC2 were also explored. It was hypothesized that in the controlled setting, FC2 mean APEs were within acceptable range (i.e., less than 5%). The hypotheses in physical education setting were: (a) the mean APE (i.e., the average APE across all classes) of FC2 was significantly higher among female than their male counterparts; and (b) the mean APE of FC2 was significantly higher in American football units than that in baseball units because football consists of more different and changes in types of PA (i.e., throwing, jumping, running, pushing, etc.) than those in baseball.

Methods

This study was approved by the university IRB committee affiliated with the senior author. Before data collection started informed consent forms were obtained from participants. To protect participants' privacy, each student was assigned a number, which was the same number as his/her Fitbit. No student names were used during the data collection.

Research Design

Concurrent validity was tested using Yamax SW-200 (YX) as the criterion device in two conditions (controlled vs. physical education settings). The controlled setting was defined as the environment where participants were able to repeatedly perform three types of PA (i.e., walking, jogging, and running) as required by the instructors and FC2 step count output could be compared to participant's actual counts as well as the counts generated by YX. Physical education setting was referred to the traditional physical education teaching condition with instruction, feedback, student practices, etc. Participants simultaneously wore the two trackers performing walking, jogging, and running three times in the controlled condition. Participants' PA patterns in the physical education consisted of standing, walking, running, and jumping, which are typical types of PA performed in physical education. Both football and baseball units consisted of warm-up, skill learning, skill-related drills, game play, and cool down. Each physical education class lasted 3 hours with a half of the class time in a multi-activity gym for learning how to play football and baseball. There was enough equipment for all students so that they did not need to take turns for practice. The leading instructor had

more than 30 years of teaching experience in higher education. All students fully participated in class and both male and female students were treated the same because it was their first time learning American sports (i.e., football and baseball) and they were at the same skill level.

Participants

In total, 17 Chinese college students participated in this study with 4 males and 13 females ($M_{age} = 20.35 \pm 0.86$, $BMI_{female} = 20.54 \pm 1.21$, $BMI_{male} = 23.36 \pm 1.49$). All the participants were enrolled in a summer program at a large university in the U.S. where they learned how to play football and baseball and the sport culture associated with these sports. All the participants were reportedly healthy without major health concerns during their program study in the US, and had no experience related to football and baseball prior to the program.

Instruments

Similar to other Fitbit products (www.fitbit.com), FC2 uses a 3-axis accelerometer and altimeter. Users wear FC2 on the non-dominant wrist like a watch. It provides various estimates of daily PA levels, including step counts, distance traveled, active minutes, floors climbed, and hourly activity. After connecting FC2 to a smartphone via Bluetooth, the users can access the above information using the Fitbit app. In addition, FC2 allows users to time a period of PA and track real-time statistics such as steps, active time, etc.

YX was selected as a criterion step counter for the physical education setting study. It is a pendulum-based pedometer worn on the hip, which provides estimates for step count taken since reset to zero. As noted earlier, the accuracy of YX in step counting has been supported by previous studies (Clemes & Biddle, 2013; Huang et al., 2016). Some studies reported a mean APE less than 5% among healthy adults (De Cocker, De Meyer, De Bourdeaudhuij, & Cardon, 2012).

Procedure

Four researchers, who had intensive training and experience with using YX and FC2, administrated the two conditions. On the first day of the 4-week program, the leading researcher introduced YX and FC2 to participants, including how to properly wear the devices, how to reset YX to zero, and how to read step counts on both devices. The same indoor multi-purpose gymnasium facility was used for both conditions. Participants had worn the two trackers for one week before the data collection began to allow participants to get used to wearing them.

In the controlled setting, the participants wore FC2 on their non-dominant wrist while performing three different types of PA: walking, jogging, and running for 100 steps. In order to differentiate running from jogging, participants were instructed to run as fast as they could while jogging was defined as running slowly without sweating. Because FC2 can also measure heart rate, students were instructed to make sure that their heart rate was at 70-90% of their maximum heart rate when running. For each movement, the participants took three trials, and performed at their own pace in the same gym. Before each trial, the participants were asked to start the timer function on FC2 so that the data during

the specified time period were separately recorded with the other time (Van Camp & Hayes, 2017). After each trial was completed, the participants ceased all movements. Using participant's own smartphone, they recorded the real-time step count readings on FC2 display in an online data collection survey on Qualtrics survey tools. Participants then stopped the timer on FC2.

The physical education setting study was conducted during the second to fourth week of the summer program in the same gymnasium. Participants took two sport units (i.e., baseball and flag football) taught by two qualified instructors from the physical education teacher education program of the institution that hosted the summer program. All participants took part in the same PA in class. In total, 10 classes (i.e., four and six classes for the baseball unit and football unit, respectively) had usable data. The first two classes for each unit were excluded because they were used for participants to get used to wear the wearables and the other classes had more than 50% of missing values and thereby were eliminated from the data analyses. An equal sample size across subgroups is not required for linear mixed-model analysis and the unequal sample size in subgroups does not cast bias in linear mixed-model analysis (West, Welch, & Galecki, 2014). In each class, the participants took a 90-minute class to acquire basic movement skills, practice game tactics, and participate in skill learning centered practices. Participants were instructed to wear both devices properly and simultaneously throughout the 10 classes. Before each class started, the participants reset YX reading to zero and started the timer on FC2 at the same time while remaining seated. At the end of each class, the participants ceased all movements and recorded the step readings of each device through an online data collection survey using their smartphones. Four researchers were at the gym to ensure all data were recorded correctly.

Data Analyses

For the data collected in the controlled setting, the APE for each trial was calculated using the formula of $(FC2 \text{ step counts} - 100)/100 * 100\%$, and an absolute value was obtained (Rosenberger et al., 2016). The mean APE, which was the average of three APEs, was calculated for walking, jogging, and running, respectively. In addition, a one sample *t*-test was performed to compare the differences between the number of step counts captured by FC2 and the designated step count of 100. Also, Mann-Whitney *U* test was used to test gender difference in mean APE in all conditions. This is due to the unequal distribution of mean APE between male and female participants and unequal sample size in male and female subgroups (Gravetter & Wallnau, 2016; West et al., 2014).

For the physical education setting, the APE of FC2 was calculated for each participant in each class using the formula as follows:

$APE = [(FC2 - YX)/YX] * 100$ where FC2 refers to the step count recorded by FC2, and YX refers to the step count recorded by YX. An average APE of FC2 was calculated for the entire sample throughout the 16 classes.

In order to explore the gender and sport unit difference in the accuracy of FC2, a linear mixed-model analysis was performed. The dependent variable was the APE of FC2 of each participant in a particular class. Gender, sport unit, and the interactions between the two variables were analyzed as fixed effects, whereas each

individual participant difference was treated as random effects (Talbot, Brede, Price, & Metter, 2017). All statistical analyses were performed using SPSS 21.0.

Results

FC2 Validity in Controlled Setting

The descriptive statistics of step counts and APEs of FC2 were displayed in Table 1. One sample *t*-test only showed that FC2 significantly under-estimated step counts in walking condition. There was no significant difference between FC2 measured steps and the actual step count (i.e., 100 steps) in the other three conditions. The APEs of FC2 in the 100-step walking, jogging, and running conditions were less than 5%, which were within the acceptable range. Mann-Whitney *U* test showed no gender differences in mean APE across the three conditions.

| Condition | Steps (SD) | Mean APE (SD) | <i>t</i> (df) |
|-----------|---------------|---------------|---------------|
| Walking | | | |
| Total | 97.21 (4.87)* | 4.54% (3.16%) | -2.30 (15) |
| Female | 96.17 (4.99) | 5.17% (3.44%) | |
| Male | 100.33 (3.09) | 2.67% (.47%) | |
| Jogging | | | |
| Total | 101.64 (3.61) | 3.26% (2.14%) | 1.81 (15) |
| Female | 101.22 (3.84) | 3.22% (2.24%) | |
| Male | 102.88 (2.93) | 3.38% (2.10%) | |
| Running | | | |
| Total | 102.58 (5.08) | 4.42% (3.48%) | 2.04 (15) |
| Female | 101.72 (4.67) | 3.94% (2.84%) | |
| Male | 105.17 (6.09) | 5.83% (5.22%) | |

Note: APE = absolute percent error; SD = standard deviation; * $p < .05$ level

Physical Education Setting

The overall mean APE of FC2 was 36.57% across the 10 classes, which was greater than the acceptable value (i.e., 5%) (Smith, Egercic, Bramble, & Secich, 2017). Linear mixed-model analysis results showed that there was a significant gender difference in mean APE [$F(1, 117) = 6.60, p = .011$]. The mean APE of FC2 was significantly higher in males than that for females (see Table 2).

Using linear mixed-model analysis, we conducted further analysis about gender and sport unit differences in steps counts across 10 classes. Step counts measured by YX did not yield gender difference [$F(1, 118) = .063, p = .80$] or sport unit difference [$F(1, 118) = .001, p = .98$]; whereas step counts measured by FC2 yielded significant gender difference [$F(1, 151) = 5.13, p = .03$], but no sport unit difference [$F(1, 151) = 1.40, p = .24$] or interaction between gender and unit. FC2 data suggested that male students took an average of 3370.42 1522.03 steps, which were significantly more than that for female students (i.e., 2822.17 ± 1160.86 steps) (see Table 2).

However, no significant mean APE of FC2 in gender by sport unit effect was found [$F(1, 117) = .39, p = .53$]. The difference in

research on the topic is urgently needed to inform the designer and manufacturers about the needs of teachers and learners in using PA wearables (Lee et al., 2015).

The current study examined the accuracy of FC2 in the controlled and physical education settings. To our knowledge, our study marks the first attempt to assess the concurrent validity of FC2 against previous validated device (i.e., YX) in physical education settings, even though other Fitbit products have been validated in both lab and free-living settings (Alharbi, Bauman, Neubeck, & Gallagher, 2016; Chu et al., 2017). Previous studies have indicated that in non-controlled settings, the measurement of PA domains (i.e., FITT) varied among wearable devices (Rosenberger et al., 2016). The findings generated by this study enriched our knowledge about the validity of FC2 in controlled and college physical education settings. Overall, in line with what has been reported in the literature (Evenson et al., 2015), the data from the current study suggested that FC2 could produce accurate data concerning walking, jogging, and running in a controlled setting. For combined PA in college physical education settings, however, FC2 tended to have an unacceptable APE (i.e., greater than 5%) and overestimated total steps, regardless of gender and sport units. Gender difference in mean APE of FC2 was also significant, overestimating more total steps in males than their female counterparts while YX did not reveal gender difference. Each of the highlighted results was further discussed below.

Mean APE of FC2 by Teaching Content (Football vs. Baseball)

Based on the data presented in Table 2, it is obvious that FC2 had more total steps than YX by gender and sport units. The possible explanation could be related to the different body parts where the wearables were worn. Similar to the research design used in the study by Chu and associates (2017), the current study also used a waist-worn wearable to compare the step counts with that generated by a wrist-worn wearable. In particular, FC2 was worn on the wrist, capturing upper body movements, even when the lower body parts are not moving. On the other hand, YX is worn around the waist area, resulting relatively lower sensitivity to upper body movements. Given that the physical education settings in the current study consisted of sport skill learning that concentrated on upper body movements (i.e. throwing and catching in football; pitching, hitting, and catching in baseball) as well as lower body movements (i.e. jogging, running, and jumping), it was highly possible that FC2 captured "steps" (arm movement) in addition to the actual steps taken by feet whereas YX only recorded the feet movement. Cautions need to be exercised when interpreting the result, however. Our data may also suggest that FC2 might be a better wearable for measuring PA in physical education settings, because it measures both upper body and lower body movements in comparison with YX. Due to the lack of valid data to verify the above contention in the current study, future research on the topic is warranted.

It was unexpected that there was not a significant difference in mean APE of FC2 between sport units. In our classes, football sessions consisted of more upper body movement (i.e., blocking, throwing, catching, snapping, etc.) than baseball (i.e., running, batting and throwing). We hypothesized that the APE of FC2 would

| Table 2 Average Total Step Counts per Class, Mean Absolute Percent Error of Fitbit Charge 2, and Yamax SW-200 in the Physical Education Settings | | | |
|---|-----------------|--------------------------|-------------------|
| Subgroup | Mean APE of FC2 | Average Total Steps (SD) | |
| | (SD) | FC2 | YX |
| Gender | | | |
| Male | 46.94% (25.38%) | 3329.36 (1342.80)* | 2389.92 (1076.92) |
| Female | 33.73% (23.56%) | 2806.50 (1121.49)* | 2301.89 (1019.73) |
| Sport Unit | | | |
| Baseball unit | 37.84% (24.13%) | 3098.44 (1175.88) | 2376.55 (824.38) |
| Football unit | 35.68% (24.83%) | 2817.29 (1196.73) | 2280.49 (1156.87) |
| Gender by Sport Units | | | |
| Female | | | |
| Baseball | 34.35% (24.25%) | 2996.35 (1045.32) | 2395.78 (761.16) |
| Football | 33.29% (23.25%) | 2678.14 (1159.78) | 2231.89 (1178.35) |
| Male | | | |
| Baseball | 51.78% (18.76%) | 3475.38 (1561.46) | 2297.70(1091.80) |
| Football | 43.91% (28.93%) | 3246.83 (1232.46) | 2447.56 (1099.32) |

Note: FC2 = Fitbit Charge 2; YX = Yamax SW-200; APE = absolute percent error; SD = standard deviation; asteps counts measured by FC2 compared between genders; **p* < .05.

mean APE of FC2 between sport units was also not significant [F(1, 117) = .67, *p* = .42]. Overall, the APE of FC2 in the physical education settings of this study exceeded the acceptable range for total step counting devices (i.e., lower than 5%) (Smith et al., 2017).

Discussion

Valid measures of PA wearables are pivotal for both research and teaching physical education. This is not only because an adequate amount of PA on a regular basis is directly related to the reduced risk level for poor health (Cleland et al. 2012), but also students PA is an important aspect to be considered in planning and implementation of quality physical education (Sallis et al., 2012), which aims to physically educate all students, regardless of age, gender, ethnicity, SES, etc. (Kohl & Cook, 2013; Trost et al., 2017).

A number of studies have suggested that PA wearables can greatly improve the accuracy and efficiency in PA assessment in physical education, with great potentials to facilitate quality physical education for all students (Bower & Sturman, 2015; Kolt, 2013; Lee et al., 2015). However, the unique PA patterns in skill acquisition (i.e., short bouts and constant changing in PA types among walking, running, and jumping), funding limit, and time restrictions for measuring a large number of students in physical education call for a group solution that is relatively cheap and is sensitive to inconsistent movements (Rosenberger et al., 2016; Van Camp & Hayes, 2017). With the possibility of producing cheaper and more reliable and feasible PA wearables in the near future, it seems that the use of PA wearable will be increased in school settings (Lee et al., 2015; Lindberg et al., 2016). As a result, more

be higher in football sessions compared to baseball sessions, as more upper body movement may be falsely recorded as "steps" by FC2 in football. The non-significant mean APE of FC2 may suggest that FC2 was robust regardless of sport units. However, it is also possible that the non-significant finding was resulted from students' skill level. As noted earlier, all the participants were introduced to the two sports for the first time because they never learned American sports in China. In baseball sessions, participants ended up moving upper body frequently when they practiced batting and throwing. Also in game plays, participants tended to miss hitting the ball often and ended up running around in the gym, which was different from typical baseball plays. These may result in high APE in the baseball sessions. More experimental studies are needed to further validate FC2 involving participants with higher levels of sport skills in various sport units in physical education settings.

Gender Differences in Mean APE of FC2 in Controlled vs. Physical Education Settings

Like what was reported by Van Camp & Hayes (2017), gender differences were observed, regardless of sport units. Interestingly, Chu and associates (2017) conducted a study on comparison of measuring steps between wrist-worn Fitbit Flex and waist-worn ActiGraph for adults in a free-living setting. The authors also found a similar gender difference that males had more steps. However, it is critical to point out that YX data did not show any gender difference (see Table 2) and no different drills/practices were given to male students in class as the instructors made sure that both genders were treated equally in class. Therefore, it is likely that FC2 demonstrated more errors in measuring PA in males in physical education settings while such a gender difference was not shown in the controlled setting. What makes more errors of FC2 measuring male students in our physical education setting, however, is unknown and warrants more research on the topic in the future.

Limitations

Some limitations need to be addressed and cautions should be exercised when generalizing the results of the current student to its population. First, this study used a convenience sample of participants and the sample size is relatively small, which may limit the generalizability or repeatability of results. Studies with larger samples are needed to replicate the results of our study in order to use FC2 with confidence to measure PA in physical education settings. Secondly, there was an unproportioned ratio of females and males. The much larger number of female students may have skewed the data. Thirdly, only YX was used as the criterion device, generating limited inter-device validity, focusing only on total steps per class. Other important PA variables such as total PA time, moderate and vigorous PA (MVPA) time, distance travelled, and Kcals consumed should also be validated in the future. Fourthly, participants never learned football and baseball before. Therefore, participants' skill level was very low, which may be different from that with highly skilled participants. Future research is needed to examine if skill levels would influence the PA data accuracy measured by FC2. And lastly, the type of sports taught in the physical education settings were limited to basic skill learning

concerning football and baseball, that mainly involves upper body movements. This is most likely to cause the overestimation of step counts by FC2 in comparison with YX. Other activities that involves movements of different body parts (e.g. soccer, tag games, etc.) should be considered for further research.

References

- Alharbi, M., Bauman, A., Neubeck, L., & Gallagher, R. (2016). Validation of Fitbit-Flex as a measure of free-living physical activity in a community-based phase III cardiac rehabilitation population. *European Journal of Preventive Cardiology*, *23*, 1476-1485.
- Bort-Roig, J., Gilson, N. D., Puig-Ribera, A., Contreras, R. S., & Trost, S. G. (2014). Measuring and influencing physical activity with smartphone technology: A systematic review. *Sports Medicine*, *44*(5), 671-686.
- Bower, M., & Sturman, D. (2015). What are the educational affordances of wearable technologies? *Computers & Education*, *88*, 343-353.
- Brian, A., & Haegele, J. A. (2017). Gopher FITStep Pro accuracy when measuring steps and moderate-to-vigorous physical activity. *Acta Gymnica*, *47*(1), 3-10.
- Cadmus-Bertram, L. A., Marcus, B. H., Patterson, R. E., Parker, B. A., & Morey, B. L. (2015). Randomized trial of a Fitbit-based physical activity intervention for women. *American Journal of Preventive Medicine*, *49*(3), 414-418.
- Cawley, J., Frisvold, D., & Meyerhoefer, C. (2013). The impact of PE on obesity among elementary school children. *Journal of Health Economics*, *32*, 743-755.
- Cleland, V., Dwyer, T., & Venn, A. (2012). Which domains of childhood physical activity predict physical activity in adulthood? A 20-year prospective tracking study. *British Journal of Sports Medicine*, *46*(8), 595-602.
- Clemes, S. A., & Biddle, S. J. H. (2013). The use of pedometers for monitoring physical activity in children and adolescents: Measurement considerations. *Journal of Physical Activity and Health*, *10*(2), 249-262.
- Chu, A. H., Ng, S. H., Paknezhad, M., Gauterin, A., Koh, D., Brown, M. S., & Müller-Riemenschneider, F. (2017). Comparison of wrist-worn Fitbit Flex and waist-worn ActiGraph for measuring steps in free-living adults. *PLoS One*, *12*(2), 1-13.
- Colberg, S. R., Sigal, R. J., Yardley, J. E., Riddell, M. C., Dunstan, D. W., Dempsey, P. C., ... Tate, D. F. (2016). Physical activity/exercise and diabetes: A position statement of the American Diabetes Association. *Diabetes Care*, *39*(11), 2065-2079.
- Corder, K., Winpenney, E., Love, R., Brown, H. E., White, M., & van Sluijs, E. (2017). Change in physical activity from adolescence to early adulthood: A systematic review and meta-analysis of longitudinal cohort studies. *British Journal of Sports Medicine*, *0*, 1-9.
- Dauenhauer, B. D., & Keating, X. D. (2011). The influence of physical education on physical activity levels of urban elementary students. *Research Quarterly for Exercise and Sport*, *82*, 512-520.
- Dauenhauer, B. D., Keating, X. D., & Lambdin, D. (2016). Effects of a three-tiered intervention model on physical activity and fitness levels of elementary school children. *Journal of Primary*

- Prevention*, 37, 313-327.
- Dauenhauer, B., Keating, X. D., & Lambdin, D. (2017). An examination of physical education data sources and collection procedures during a federally funded grant. *Journal of Teaching in Physical Education*, 37, 46-58.
- De Cocker, K. A., De Meyer, J., De Bourdeaudhuij, I. M., & Cardon, G. M. (2012). Non-traditional wearing positions of pedometers: Validity and reliability of the Omron HJ-203-ED pedometer under controlled and free-living conditions. *Journal of Science and Medicine in Sport*, 15(5), 418-424.
- Deliens, T., Deforche, B., Bourdeaudhuij, I. D., & Clarys, P. (2015). Determinants of physical activity and sedentary behavior in university students: a qualitative study using focus group discussions. *BMC Public Health*, 15, 201.
- Evenson, K. R., Goto, M. M., & Furberg, R. D. (2015). Systematic review of the validity and reliability of consumer wearable activity trackers. *International Journal of Behavioral Nutrition and Physical Activity*, 12, 1091-1101.
- Gravetter, F. J., & Wallnau, L. B. (2016). *Statistics for the behavioral sciences*. Boston, MA: Cengage Learning.
- Hesketh, K. R., Lakshman, R., & van Sluijs, E. M. F. (2017). Barriers and facilitators to young children's physical activity and sedentary behavior: A systematic review and synthesis of qualitative literature. *Obesity Reviews*, 18(9), 987-1017.
- Hickey, A., Newham, J., Slawinska, M. M., Kwasnicka, D., McDonald, S., Del Din, S., ... Godfrey, A. (2016). Estimating cut points: A simple method for new wearables. *Maturitas*, 83, 78-82.
- Huang, Y., Xu, J., Yu, B., & Shull, P. B. (2016). Validity of FitBit, Jawbone UP, Nike+ and other wearable devices for level and stair walking. *Gait & Posture*, 48(Supplement C), 36-41.
- Humphreys, B. R., McLeod, L., & Ruseski, J. E. (2014). Physical activity and health outcomes: Evidence from Canada. *Health Economics*, 23, 33-54.
- Hurvitz, P. M., Moudon, A. V., Kang, B., Saelens, B. E., & Duncan, G. E. (2014). Emerging technologies for assessing physical activity behaviors in space and time. *Frontiers in Public Health*, 2, 1-15.
- Kang, M., Mahar, M. T., & Morrow Jr., J. R. (2016). Issues in the assessment of physical activity in children. *Journal of Physical Education, Recreation & Dance*, 87, 6, 35-43.
- Keating, X. D., Shangguan, R., Huang, Y., Liu, X., & Chen, L. (2017). Relating Chinese university administrators' physical activity to self-rated health and body mass index. *ICHPER-SD Journal of Research*, 9(1), 3-10.
- Kerner, C., & Goodyear, V. A. (2017). The motivational impact of wearable healthy lifestyle technologies: A self-determination perspective on fitbits with adolescents. *American Journal of Health Education*, 48(5), 287-297.
- Kim, Y., Barry, V. W., & Kang, M. (2015). Validation of the ActiGraph GT3X and activPAL accelerometers for the assessment of sedentary behavior. *Measurement in Physical Education & Exercise Science*, 19(3), 125-137.
- Kohl, H. W. III, & Cook, H. D. (2013). Educating student body: Taking physical activity physical education school. Retrieved from <http://www.nationalacademies.org/hmd/Reports/2013/Educating-the-Student-Body-Taking-Physical-Activity-and-Physical-Education-to-School.aspx>
- Kolt, G. S. (2013). The complexity of measuring physical activity and sedentary behavior. *Journal of Science and Medicine in Sport*, 16, 1.
- Lee, V. R., Drake, J., & Williamson, K. (2015). Let's get physical: K-12 students using wearable devices to obtain and learn about data from physical activities. *TechTrends: Linking Research and Practice to Improve Learning*, 59(4), 46-53.
- Lehman, D. (2017). Wearable fitness trackers help Manitoba high school reach goals. *Physical & Health Education Journal*, 83(1), 1-1.
- Lindberg, R., Seo, J., & Laine, T. H. (2016). Enhancing physical education with exergames and wearable technology. *IEEE Transactions on Learning Technologies*, 9(4), 328-341.
- Lyons, E. J., Lewis, Z. H., Mayrsohn, B. G., & Rowland, J. L. (2014). Behavior change techniques implemented in electronic lifestyle activity monitors: A systematic content analysis. *Journal of Medical Internet Research*, 16(8), e192. <https://doi.org/doi:10.2196/jmir.3469>
- McMinn, D., Rowe, D. A., Stark, M., & Nicol, L. (2010). Validity of the new lifestyles NL-1000 accelerometer for measuring time spent in moderate-to-vigorous physical activity in school settings. *Measurement in Physical Education & Exercise Science*, 14(2), 67-78. <https://doi.org/10.1080/10913671003715516>
- Melton, B. F., Buman, M. P., Vogel, R. L., Harris, B. S., & Bigham, L. E. (2016a). Wearable devices to improve physical activity and sleep: A randomized controlled trial of college-aged African American women. *Journal of Black Studies*, 47(6), 610-625.
- Melton, B. F., Buman, M. P., Vogel, R. L., Harris, B. S., & Bigham, L. E. (2016b). Wearable devices to improve physical activity and sleep: A randomized controlled trial of college-aged African American women. *Journal of Black Studies*, 47(6), 610-625. <https://doi.org/10.1177/0021934716653349>
- Montoye, A. H. K., Mitzryk, J. R., & Molesky, M. J. (2017). Comparative accuracy of a wrist-worn activity tracker and a smart shirt for physical activity assessment. *Measurement in Physical Education & Exercise Science*, 21(4), 201-211.
- Muntaner, A., Vidal-Conti, J., & Palou, P. (2016). Increasing physical activity through mobile device interventions: A systematic review. *Health Informatics Journal*, 22(3), 451-469.
- Reid, R. E. R., Insogna, J. A., Carver, T. E., Comptour, A. M., Bewski, N. A., Sciortino, C., & Andersen, R. E. (2017). Validity and reliability of Fitbit activity monitors compared to ActiGraph GT3X+ with female adults in a free-living environment. *Journal of Science & Medicine in Sport*, 20(6), 578-582.
- Rosenberger, M. E., Buman, M. P., Haskell, M. L., McConnell, M. V., & Carstensen, L. L. (2016). Twenty-four hours of sleep, sedentary behavior, and physical activity with nine wearable devices. *Medicine & Science in Sports & Exercise*, 48(3), 457-465.
- Rote, A. E. (2017a). Physical activity intervention using Fitbits in an introductory college health course. *Health Education Journal*, 76(3), 337-348.
- Rote, A. E. (2017b). Physical activity intervention using Fitbits in an introductory college health course. *Health Education*

- Journal*, 76(3), 337-348. <https://doi.org/10.1177/0017896916674505>
- Rotich, W. K. (2016). Leveraging the habit-forming aspects of technology to increase levels of physical activity. *Strategies*, 29(6), 15-19.
- Sasaki, J. E., John, D., & Freedson, P. S. (2011). Validation and comparison of ActiGraph activity monitors. *Journal of Science and Medicine in Sport*, 14, 411-416.
- Scruggs, P. W. (2013). Pedometer steps/min in physical education: Does the pedometer matter? *Journal of Science and Medicine in Sport*, 16(1), 36-39. <https://doi.org/10.1016/j.jsams.2012.05.011>
- Sgrò, F., francesco. sgro@unikore. i., Mango, P., Pignato, S., Schembri, R., Licari, D., & Lipoma, M. (2017). Assessing standing long jump developmental levels using an inertial measurement unit. *Perceptual & Motor Skills*, 124(1), 21-38. <https://doi.org/10.1177/0031512516682649>
- Silva, D. A. S., Chaput, J.-P., Katzmarzyk, P. T., Fogelholm, M., Hu, G., Maher, C., ... Tremblay, M. S. (2017). Physical education classes, physical activity, and sedentary behavior in children. *Medicine & Science in Sports & Exercise, Publish Ahead of Print*. <https://doi.org/10.1249/MSS.0000000000001524>
- Singh, A. K., Rehab, M. C., Farmer, C., van Ben Berg, M. L. E. V. D., Killington, M., & Barr, C. J. (2016). Accuracy of the FitBit at walking speeds and cadences relevant to clinical rehabilitation populations. *Disability and Health Journal*, 9, 320-323.
- Skender, S., Ose, J., Chang-Claude, J., Paskow, M., Brühmann, B., Siegel, E. M., ... Ulrich, C. M. (2016). Accelerometry and physical activity questionnaires - a systematic review. *BMC Public Health*, 16, 515. <https://doi.org/10.1186/s12889-016-3172-0>
- Soland, J. (2017). Combining academic, noncognitive, and college knowledge measures to identify students not on track for college: A data-driven approach. *Research & Practice in Assessment*, 12, 5-19.
- Takacs, J., Pollock, C. L., Guenther, J. R., Bahar, M., Napier, C., & Hunt, M. A. (2014). Validation of the Fitbit One activity monitor device during treadmill walking. *Journal of Science and Medicine in Sport*, 17, 496-500.
- Thompson, C. J. (2012). Improving instructional decision-making by observing technology integration in the classroom: A case study. *Journal of Technology Integration in the Classroom*, 4(1), 5-13.
- Trost, S., Cliff, D., Ahmadi, M., Van Tuc, N., & Hagenbuchner, M. (2017). Sensor-enabled activity class recognition in preschoolers: Hip versus wrist data. *Medicine & Science in Sports & Exercise, Publish Ahead of Print*. <https://doi.org/10.1249/MSS.0000000000001460>
- Van Camp, C. M., & Hayes, L. B. (2017). Identifying beneficial physical activity during school recess: Utility and feasibility of the Fitbit. *Journal of Behavioral Education*, 27, 394-409.
- Vanroy, C. (2014). Physical activity monitoring in stroke: SenseWear Pro2 activity accelerometer versus Yamax Digi-Walker SW-200 pedometer. *Disability and Rehabilitation: An International, Multidisciplinary Journal*, 36(20), 1695-1703.
- Yang, C.-C., & Hsu, Y.-L. (2010). A review of accelerometry-based wearable motion detectors for physical activity monitoring. *Sensors*, 10(8), 7772-7788. <https://doi.org/10.3390/s100807772>
- Vanroy, C., Vissers, D., Cras, P., Beyne, S., Feys, H., Vanlandewijck, Y., & Truijten, S. (2014). Physical activity monitoring in stroke: Sense Wear Pro2 Activity accelerometer versus Yamax Digi-Walker SW-200 Pedometer. *Disability and Rehabilitation*, 36(20), 1695-1703.
- Vella, S. A., Schranz, N. K., Davern, M., Hardy, L. L., Hills, A. P., Morgan, P. J., ... Tomkinson, G. (2016). The contribution of organized sports to physical activity in Australia: Results and directions from the Active Healthy Kids Australia 2014 Report Card on physical activity for children and young people. *Journal of Science and Medicine in Sport*, 19(5), 407-412.
- West, B. T., Welch, K. B., & Galecki, A. T. (2014). *Linear mixed models: a practical guide using statistical software*. CRC Press.
- Yang, C. C., & Hsu, Y. L. (2010). A review of accelerometry-based wearable motion detectors for physical activity monitoring. *Sensors*, 10(8), 7772-7788. <https://doi.org/10.3390/s100807772>

