https://doi.org/10.54141/psbd.1357839

PAMUKKALE JOURNAL OF SPORT SCIENCES
ISSN: 1309-0356
dergipark.org.tr/en/pub/psbd

Pamukkale J Sport Sci, 14(3), 279-296, 2023 Research Article

Investigation of the Validity and Reliability of Two Smart Bands Selected That Count Steps at Different Walking Speed

Mehmet Emin YILDIZ[1*] İlker GÜNEL[2]

[1]*Faculty of Sport Sciences, Department of Recreation, Uşak University, Uşak, Türkiye*
[2]*Faculty of Sport Sciences, Department of Sport Management, Uşak University, Uşak, Türkiye*

Keywords
Reliability,
Smart Band,
Step Counting,
Validity,
Walking

Article History
Received 09 September 2023
Revised 03 December 2023
Accepted 23 December 2023
Available Online 29 December 2023

* **Corresponding Author**:
Mehmet Emin YILDIZ
E-mail Address:
m.yildiz@usak.edu.tr

ABSTRACT

Whether activity monitor smart wristbands that have become widespread can accurately estimate the step count while walking is a matter of curiosity. The current study aims to analyze the validity and reliability of step count (SC) estimation at normal walking and fast walking pace under controlled conditions of two selected smart wristbands of the leading wearable device vendors, Xiaomi (Mi4) and Huawei (H4). Twenty healthy adult male and twenty healthy adult female were included in the study and analyzed separately. The mean age of male and female participants was 22.25 and 21.62 years, with BMI values of 24.22 kg/m² and 21.42 kg/m², respectively. The above-ground walking protocol consisted of four separate five-minute tests: Normal Walking Test, Normal Walking Retest, Fast Walking Test, and Fast Walking Retest. In the study, the analyses were performed by using activity monitor measurements and criterion measurements (the number of steps determined from video recordings), compatibility of test-retest measurement values, error indicators (MPE and MAPE), Intraclass Correlation Coefficients (ICC), and Bland-Altman limits of agreement. According to the current study results, it was revealed that the MAPE values recorded for Mi4 and H4 smart wristbands for both normal and fast walking pace in male and female participants were <5%, which was deemed excellent. According to all analyses, the H4 device was found to be valid and reliable, but according to ICC and Pearson Correlation analyses, the Mi4 device was not found to be valid and reliable at fast walking pace.

INTRODUCTION

The number of wearable devices like smart watches and wristbands has more than doubled worldwide in three years, increasing to 325 million in 2016 and 722 million in 2019. The number of wearable devices, whose commerce market has been growing rapidly, is estimated to have exceeded one billion by 2022 (Statista, 2022). Without the need for expensive laboratory-level equipment, electronic devices that can be worn or carried on the body enable individuals to measure and monitor such physical activity (PA) as step count (SC) and distance, kilocalories, sleep, and other health-related measures quickly, inexpensively, and objectively through powerful microchips using unique algorithms and smart sensors (An et al., 2017; Dobkin & Martinez, 2018; Evenson et al., 2015). Among these outputs, SC and distance measurements while walking are considered the most popular and transformable outputs used today (Bassett et al., 2017; Carlin & Vuillerme, 2021).

Activity tracker smart wristbands are less expensive than a smartwatch due to less expensive hardware and often fewer sensors. Therefore, they usually have better battery life but a limited interface to view monitoring results (Henriksen et al., 2018). The subset of consumer wearable devices used to monitor physical activity and fitness-related measurements are called "activity monitors" or "fitness monitors" (Evenson et al., 2015). Activity monitors can be easily synchronized via a smartphone or computer application and thus can provide continuous PA measurements over weeks, months, or years (Hartung et al., 2020).

Exercise and health researchers also use these popular electronic devices to monitor and facilitate PA behavior changes (Topalidis et al., 2021). Despite the widespread sales and popularity of these wearable fitness monitors over the past decade, which monitor PA and provide feedback at appropriate intervals, the assessments in terms of their use, accuracy, or consistency remain insufficient (Bunn et al., 2018; Dooley et al., 2017).

The most common brands whose validities are most commonly analyzed are Fitbit, Apple, Jawbone, and Garmin, which are quite expensive and beyond the financial means of a significant segment of society. Although the number of published studies on assessing well-known activity monitors has increased (Bunn et al., 2018), only a few studies have addressed low-cost activity monitors (Carlin & Vuillerme, 2021; Xie et al., 2018). These devices are cheaper than $50 (Amazon, 2023). Some recent analyses have revealed that some metrics for activity monitors are accurate enough to measure PA in research settings (Shin et al., 2019; Straiton et al., 2018).

In most studies, there is a lack of consistency among the published validation protocols. This limits valid comparisons among devices. Current articles recommend comprehensive and transparent verification of the devices to ensure that wearable technology can be used safely and to its full potential (Johnston et al., 2020; Silfee et al., 2018). It was found that some of the studies on new, popular, and inexpensively available activity monitors lacked such procedures as gender-separate validity and test-retest reliability, which are commonly recommended in the relevant research protocol guidelines (Johnston et al., 2020).

This study tries to eliminate what the related literature lacks. It aims to analyze the validity and reliability of SC estimation at normal walking and fast walking pace under controlled conditions of two selected smart wristbands of the leading wearable device vendors, Xiaomi (Mi4) and Huawei (H4). This analysis was performed separately with adult male and female data. This research method and its results are necessary to use these devices more safely in scientific research PA monitors. The data obtained in the current study are deemed significant in terms of allowing comparing with similar studies to be conducted. With the current study's data, it is estimated that the use of these technologies in PA research will increase, and the accessibility of PA data of large populations with more accurate methods will be enabled.

METHODS

The participants were given an oral explanation of the protocol before signing written informed consent to participate in the study, and they were presented with the opportunity to ask questions (Kastelic et al., 2021). While preparing the study method and applying the study protocol, the guidelines of the Consumer Technology Association (CTA) and the best practice protocols for the validation of wearable step counters recommended by Towards Intelligent Health and Well-Being Network of Physical Activity Assessment (INTERLIVE) were taken as a basis (CTA, 2016; Johnston et al., 2020). This study was approved by the decision of Clinical Research Ethics Committee in Uşak University Faculty of Medicine dated 25.05.2022 and numbered 84-84-10, and it was conducted in accordance with Helsinki Declaration.

Study Group

Based on previous studies on the subject, we determined that at least twenty (20) male and twenty (20) female participants were sufficient in order to collect valid data (Carlin & Vuillerme, 2021). All the participants completed the walking tests. Healthy adult male and female students studying at a university in the Western region of Türkiye were invited to participate in the study through direct invitation. All participants were students studying in the Coaching Education Department and Sports Management Department. A few were licensed athletes, but no tests were applied to determine their physical activity levels. Several criteria were determined as the inclusion and exclusion criteria per the recommendations of previous studies. The inclusion criteria were having Body Mass Index (BMI) values less than obesity (<30 kg/m^2), being an adult between the ages of 18-64, and being physically healthy. Obesity, old age, and physical problems were assumed to affect walking movement. These differences should be the subject of research in other studies (Johnston et al., 2020).

The exclusion criteria that negatively affected the participants' walking exercise were pregnancy and having an implanted electromagnetic device (An et al., 2017), neurological diseases and cognitive problems reported by the individual, any musculoskeletal injuries and/or surgeries that may have affected walking, and other problems that may affect the individual's walking ability (Svarre et al., 2020).

The participants that the researchers determined through interviews gave informed written consent before starting the walking tests and completed and signed the standard Physical Activity Readiness Questionnaire (PAR Q). After that, the following demographic and anthropometric data of the participants who were deemed suitable for walking tests were obtained by the researcher through the interview method, and they were written down in the Demographic Information section of the Case Report Form: The number, gender, age, height, body weight, and dominant arm information of the participant (Carlin & Vuillerme, 2021). During the following days, their walking tests were performed on flat ground by making appointments with each of them. The researchers calculated their BMI based on the self-reported heights and body weights of all the participants. The participants participated in the exercises by wearing non-high-heeled shoes and seasonal casual clothes that would not negatively affect their walking patterns. All the participants were physically, cognitively, and spiritually healthy. All but one participant were dominantly right-handed participants. All the participants were Turkish-origin individuals. The data of the study were collected in May 2022. The characteristics of the participants are presented in Table 1.

Data Collection Tools

Activity Monitors: According to the information shared by IDC, the companies that sell the most intelligent wristbands in the Middle East, Türkiye, and Africa are Huawei and Xiaomi (IDC, 2022). For the research, we chose two consumer activity monitor devices that were widely used due to their affordable costs. During the walking tests, an interface was selected on which only the day, time, and step count information would be displayed on the screens of Mi4 and H4, which are the activity monitors.

Mi Band 4 (Mi4; Model: XMSH07HM, Xiaomi Comminations Co., Ltd.) is an activity monitor smart wristband considered a low-cost wearable device. It measures the number of steps through the 3D gyroscope and the 3D accelerometers. The Chinese technology giant Xiaomi announced through its official Twitter account on 09 Jul 2020 that Mi4 had become the world's best-selling smart wristband (Xiaomi, 2022). Mi4 was activated by connecting to the Mi Fit (version: 3.6.0) mobile application installed on an Android phone via Bluetooth.

Huawei Band 4 (H4; Model: ADS-B29, Huawei Technologies Co., Ltd.) is an activity tracker smart wristband considered a low-cost wearable device. It includes a pedometer function that measures the number of steps through the 3D accelerometer. It works with an Android 4.4 and higher operating system or iOS 9.0 and higher (Huawei, 2022). H4 was activated by connecting to the Huawei Health (version: 10.1.1.312) mobile application installed on an Android phone via Bluetooth.

Data Collection Process

Considering the guidelines of the Consumer Technology Association (CTA) and the recommendations of related scientific studies, the exercise duration was determined as five minutes to allow the participants to reach steady-state measurements (CTA, 2016). In order to allow comparison, two different exercise intensities were determined: 'normal walking' and 'fast walking'. Besides, test-retest walking was performed at each walking speed for device consistency specified in related research recommendations (Johnston et al., 2020). Therefore, the on-ground walking protocol included four separate tests. These were the Normal Walking Test, Normal Walking Retest, Fast Walking Test, and Fast Walking Retest.

During each test, the walking was performed on a 32-meter long and 2-meter-wide line with smooth turning points to avoid sharp turns. Turns from right to left did not have a negative effect on the walking pace. During the Normal Walking test and Normal Walking Retest, one of the researchers reminded the participants that they should walk at their standard daily walking pace. During the Fast-Walking Test and Fast-Walking Retest, one of the researchers also reminded the participants that they should walk faster than their average daily walking pace. Another researcher checked the duration of the walking tests with a chronometer (Slx 7061, Selex). The researcher counted aloud the last 10 seconds of the five-minute duration to enable the participant to slow down and stop at the last second.

During all the walking tests, the participants were made to wear the two smart wristbands on their non-dominant wrists simultaneously, just as in similar studies. The two smart wristbands were placed proximal to the radial and ulnar bones and dorsal to the wrist (Carlin & Vuillerme, 2021). While the order of the wristbands was Mi4 and H4 from proximal to distal for half of the male and female participants, the order of the wristbands was H4 and Mi4 from proximal to distal for the other half of the participants.

Before initiating the test applications, all the protocol details were explained to the participants to avoid any interruptions during the protocol, and their questions were responded. To check whether the wristbands were working before testing, the researcher asked the participant to walk towards the starting point. The researcher observed the screen to make sure the activity monitors were active. Before the participants began walking, the number of steps on the Mi4 and H4 smart wristband screens was recorded as the start-up number, and the screens of the activity monitors were photographed at the beginning of the walking line on flat ground. At the end of the five-minute walk, the current step count on the smart wristband screen was recorded as the final number. At the same time, the participants stood in an upright and stable position, and the screens of the activity monitors were photographed. Then, the data saved in the Excel file on the computer were checked by comparing the relevant photos. The difference between the start-up and final numbers obtained in each walking test (the final number minus the start-up number) was calculated as the estimation score of the smart wristband’s walking steps. This procedure was repeated for each participant and the four walking tests separately. At the end of the walking, the total number of laps that each participant walked (one lap is 68 meters) and the distance of the last step in the last lap were calculated, and the total distance was recorded in meters. Later on, the walking speed was calculated as km/hour. After each test, the participants were given a two-minute rest period to complete the procedures.

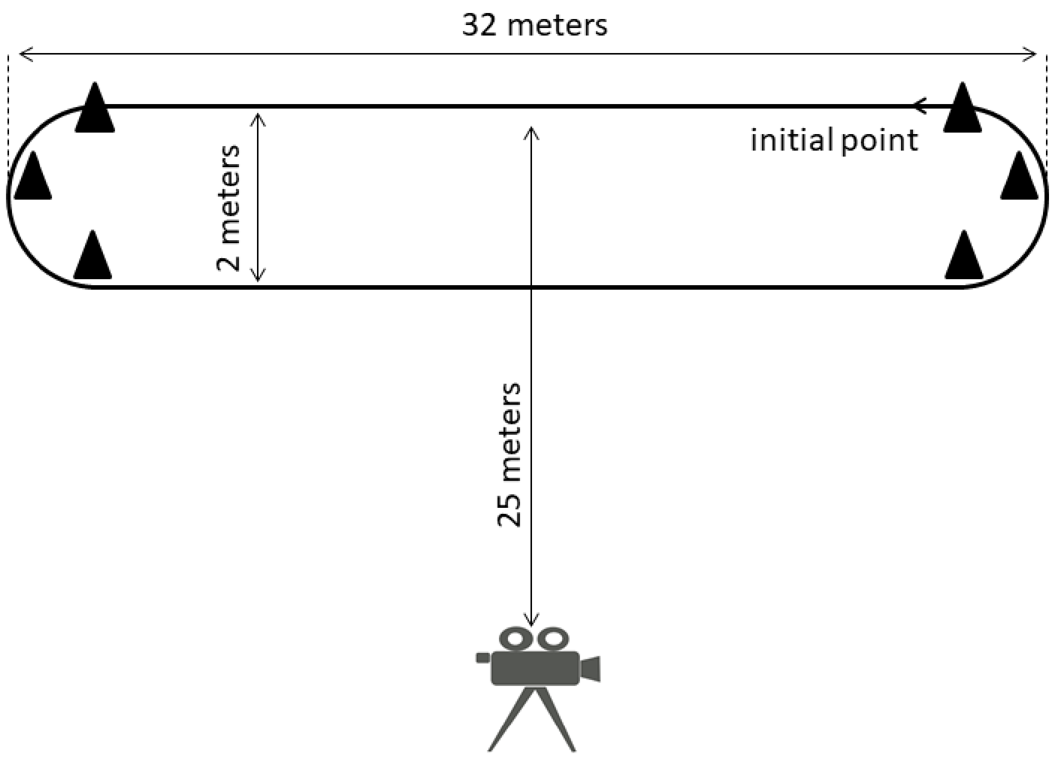
Criterion Measure

As used in previous similar studies, video-based step count was used as the gold standard for actual step count (Johnston et al., 2020; Carlin & Vuillerme, 2021). A video camera (Samsung S7 Edge mobile phone) was determined as a criterion and convergent measurement. Video recordings were made in HD quality with a frame width of 1280, frame height of 720, data speed of 12014 kbps, frame rate of 30 fps, and in mp4 format. The recording device was used by fixing it on a tripod at a distance of 25 meters so that the entire 32-meter walking track would fit on the recording screen (Figure 1).

Video recordings of all the walking tests were numbered separately and transferred to the computer. When the video recordings were opened on the computer monitor (15 inches), it was observed that all the step movements of the participants were clear throughout the entire track. In situations of uncertainty, when it was unclear whether any steps were performed or not, and when it was challenging to follow fast steps, the video was viewed in slow motion. The step definition was accepted as "the act of raising one foot and placing it elsewhere, which results in altering the body mass center" (Johnston et al., 2020). In our study, following this definition, the participant who was ready with both feet parallel to each other before starting to walk, started walking by taking the first step forward with the researcher's "start" instruction. At the end of the time, the participant took the last step forward, transferred his weight, and released the other foot behind. Thus, a repetitive body load transfer was prevented.

Video-based step count was conducted independently by two observers; in case of inconsistency between the step counts obtained by the observers from the video recordings, the two observers repeated the step count, and after reaching a consensus, the actual number of steps (criterion) was recorded in an Excel file on the computer.

Figure 1
Walking Track and Video Recording



Data Analysis

Descriptive statistics were determined, and the step count errors recorded in each activity monitor were calculated as follows:

- Step count error = [(measured steps – actual steps) / actual steps] x 100
- Other definitions are as follows:
- Measured steps are the value of the steps provided by the activity monitor.
- Actual steps are the number of steps counted manually from the video.
- Walk distance is measured using a measuring tape on the walking track (this was only used with duration to determine walking speed).
- Walking time is the time measured by the chronometer and video recording.

Statistical analyses were performed using Microsoft Excel and SPSS 24.0 (SPSS Inc., Chicago, IL, USA). In this study, previously published recommendations were considered to assess the validity of the activity monitors in the male and female participants separately (Johnston et al., 2020; Kastelic et al., 2021). In the study, the measurements of the activity monitors and their criteria measurement fit (step counts determined by the video recordings) were performed by using test and retest measurement values fit, error indicators, Intraclass Correlation Coefficients (ICC), and Bland-Altman agreement analysis limits. Mean Percent Error (MPE) was calculated to investigate group-level agreement, step count estimation, and error direction. Mean Absolute Percent Error (MAPE) was used to investigate the agreement at the individual level. ICC was used to measure the extent of the agreement between the activity monitors and the criteria measure. Since the duration in test and retest measurements is constant, but the speed is relative, albeit slightly, differences in the number of steps may negatively affect the fit, so Pearson Correlation analysis was performed with the error indicator MPE data. For reliability, by taking the Pearson Correlation Coefficient into account, the fit value was accepted as ≥0.75 perfect, 0.60-0.74 good, 0.40-0.59 moderate, and <0.40 poor in fit analyses (Cicchetti, 1994). Parameters for creating Bland-Altman plots were calculated to investigate the systematic and random error. The normal distribution of mean and percent error data was confirmed by the Shapiro-Wilk test. The statistical significance was accepted as $\alpha<0.05$.

RESULTS

All 20 male and 20 female participants in the study completed the procedures. Table 1 describes the characteristics of male and female participants. Previous studies on the subject used the walking speed of 4.8 km/h, assuming that it represented the normal walking speed (Clemes, et al., 2010; Steeves et al., 2011). In our study, the mean normal walking speed was 3.9 ± 0.4 km/h in male participants and 3.6 ± 0.4 km/h in female participants. The mean fast/paced walking speed was 6.6 ± 0.5 km/h in male participants and 6.1 ± 0.4 km/h in female participants. The test results obtained represent these realized speed ranges.

As could be seen in Table 1, the mean age of male participants was 22.25 years (Sd ± 4.56), their mean height was 1.77 meters (Sd ± 0.04), and their mean body weight was 76.42 kg (Sd ± 8.18), with a BMI value of 24.22 kg/m² (Sd ± 2.48). The mean age of female participants was 21.62 years (Sd ± 2.57), their mean height was 1.64 meters (Sd ± 0.57), and mean body weight was 58.00 kg (Sd ± 6.62), with a BMI value of 21.42 kg/m² (Sd ± 2.68).

Table 1
The Characteristics of the Participants

Gender	Characteristics	N	$\bar{X}$	Sd	Min.	Max.
Male	Age (years)	20	22.25	4.56	19	40
	Height (m)	20	1.77	0.04	1.70	1.85
	Weight (kg)	20	76.42	8.18	57.40	89.30
	BMI (kg/m²)	20	24.22	2.48	18.74	28.33
Female	Age (years)	20	21.62	2.57	19	29
	Height (m)	20	1.64	0.57	1.55	1.76
	Weight (kg)	20	58.00	6.62	45.50	71.00
	BMI (kg/m²)	20	21.42	2.68	16.03	27.39

Note. BMI: Bady Mass Index

As could be seen in Table 2, it was found that the compatibility of the Mi4 device according to the criteria step count at normal walking pace was excellent with 83.1% in male participants, and good with 71.6% in female participants. It was also found that the compatibility of the H4 device was good with 70.5% in male participants and good with 70.2% in female participants according to the criteria at normal walking pace. Besides, according to the criteria at fast walking pace, the compatibility of the Mi4 device was found to be poor with 44.2% in male participants and poor with 39.7% in female participants. In contrast, the compatibility of the H4 device was found to be moderate with 59.2% in male participants, and good with 63.1% in female participants according to the criteria at fast walking pace.

Table 2
MAPE Values and Intraclass Correlation Coefficients (ICC) Analysis of the Activity Monitors in Male and Female Participants According to Their Walking Paces

Walking Paces	Devices	Gender	MAPE	Intraclass Correlation Coefficients	ICC Value	ICC P
Normal Walking	Mi4	Male	2.18	.831	5.926	**.000****
		Female	0.11	.716	3.521	**.004****
	H4	Male	1.20	.705	3.386	**.005****
		Female	0.85	.702	3.361	**.006****
Fast Walking	Mi4	Male	0.86	.442	1.791	.107
		Female	2.47	.397	1.659	.139
	H4	Male	3.19	.592	2.450	**.029***
		Female	3.55	.631	2.708	**.018***

Note. MAPE: Mean Absolute Percentage Error
*$p < .05$
**$p < .01$

The MAPE values of the Mi4 device were recorded perfectly, with an error of less than 3%, at both normal walking and fast walking pace in both male and female participants. The MAPE values of the H4 device were also recorded perfectly, with an error of less than 3%, at normal walking pace in male and female participants. The MAPE values of the H4 device were recorded with an error of less than 5% at fast walking pace in both male and female participants.

As could be seen in Table 3, when the test-retest compatibility of the MPE values according to the criterion step count measurement of the Mi4 device was examined, it was found that this fit was high in male participants (r = .921) at normal walking pace. In contrast, it was moderate in female participants (r = .670). When the test-retest compatibility of the H4 device was examined, it was revealed that this fit was moderate in male and female participants (r = .523 and r = .543, respectively) at normal walking pace. It was also revealed that the test-retest compatibility of the Mi4 device was low both in male and female participants (r = .289 and r = .371, respectively) at fast walking pace. Besides, the test-retest compatibility of the H4 device was found to be moderate both in male and female participants (r = .523 and r = .543, respectively) at fast walking pace.

Table 3
Test-Retest Fit (Pearson Correlation) Analysis Regarding the Mean Percent Errors (MPE) of the Devices According to Criteria Measurement at Two Walking Paces

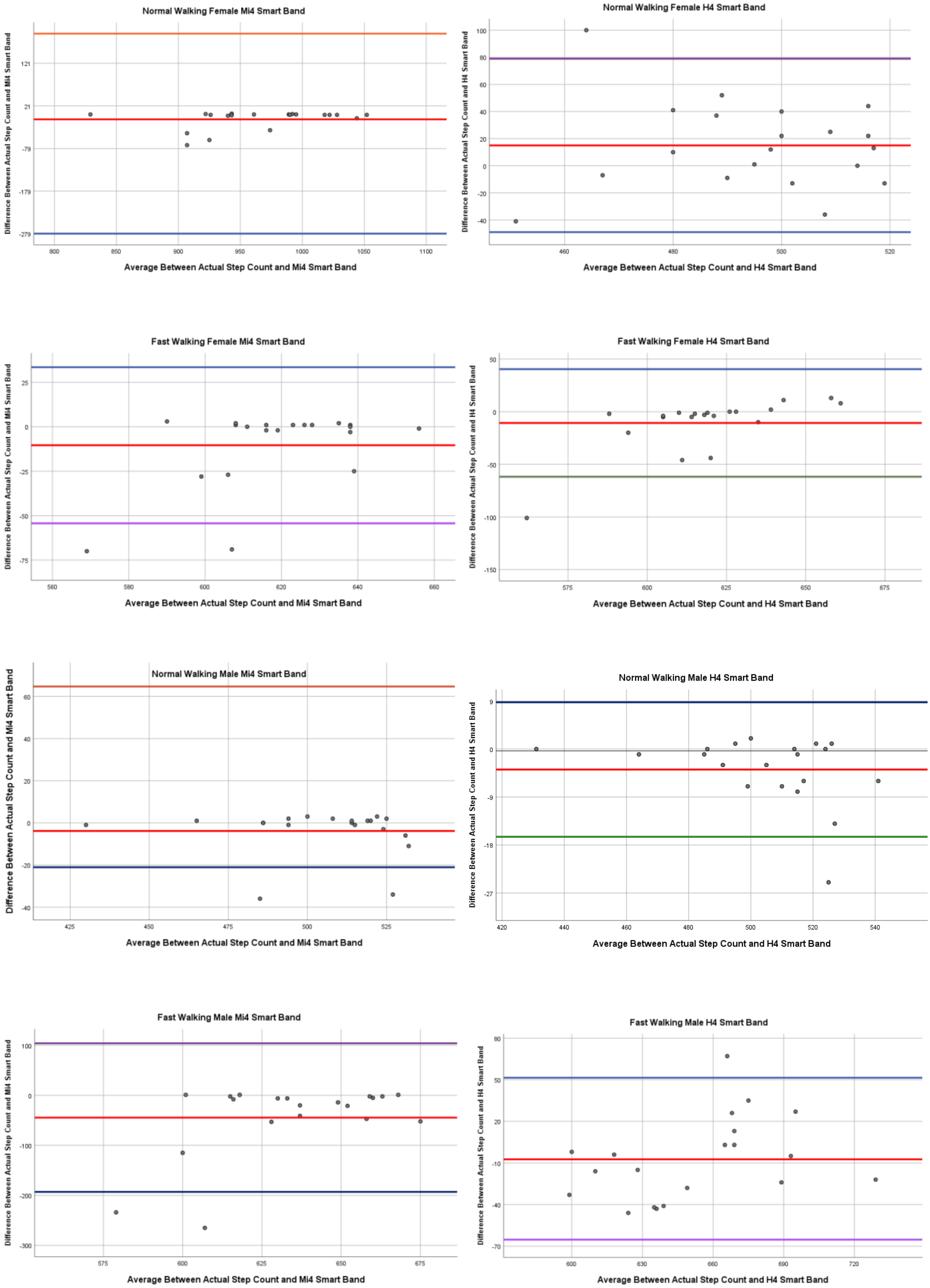
Walking Paces	Devices	Gender	Tests	$\overline{X}$	Sd	r	P
Normal Walking	Mi4	Male	Test	-.731	2.138	.921**	**.000**
			Retest	-1.723	4.526		
		Female	Test	-2.126	4.591	.670**	**.001**
			Retest	-3.203	8.579		
	H4	Male	Test	-.733	1.205	.523**	**.009**
			Retest	-1.138	1.773		
		Female	Test	-1.047	1.729	.543**	**.007**
			Retest	-0.645	1.604		
Fast Walking	Mi4	Male	Test	-6.416	10.588	.289	.108
			Retest	-10.407	12.878		
		Female	Test	-1.712	3.604	.371	.054
			Retest	-6.310	9.419		
	H4	Male	Test	-1.109	4.615	.523**	**.009**
			Retest	2.174	9.177		
		Female	Test	-1.731	4.204	.543**	**.007**
			Retest	-3.043	7.581		

*p < .05
**p < .01

The lower and upper colored lines in the Bland-Altman plot represent the 95% CI limits of the mean difference. The red straight line in the middle represents the equity point. In other words, it is where the difference between the criteria and the device measurement values is equal to 0. When the Bland–Altman plots were examined, it could be said that the compatibility between the criterion number of steps of Mi4 and H4 device was within the acceptable limits in both normal walking and fast walking pace in male and female participants.

Figure 2

Bland-Altman Plots Calculating the Agreement Between Activity Monitors and Actual/Criteria Step Counts in Different Walking Paces



DISCUSSION

It is recommended that future studies should use all of the validity indicators to enable consumers and researchers to make conscious decisions regarding both the validity and reliability of activity monitor devices and to facilitate comparison between devices (Bunn et al., 2018). Previous Xiaomi Smart Band 4 studies evaluated step count validity during the walking/running protocol. However, all these studies had heterogeneous adult groups. Together with the fact that there are studies on the Huawei Talk Band B3 version, no research could be found in the literature on the validity of the step count for the Huawei Smart Band 4 device.

The current study tries to eliminate what the related literature lacks. It aims to analyze the validity and reliability of SC estimation at normal walking and fast walking pace under controlled conditions of two selected smart wristbands of the leading wearable device vendors, Xiaomi (Mi4) and Huawei (H4). Furthermore, it is thought that this study will make significant contributions to the reliability procedures of devices such as gender-based validity and test-retest, which are considered deficiencies in the studies on cheap and accessible activity monitors and which are also recommended in the relevant research protocol guidelines (Silfee et al., 2018; Johnston et al., 2020).

Today, there is no consensus on definite standards for validity, reliability, and sensitivity metrics that will indicate that using a particular calculated metric is adequate for a given situation (Kastelic et al., 2021). Besides, suppose an activity monitor is to be used as an outcome measure in a clinical experiment or as an alternative gold standard measurement tool for step count. In that case, it is recommended that the device display a deficient level of measurement error (MAPE ≤ 5%). However, it is suggested that a higher level of measurement error (MAPE ≤ 10%–15%) may be acceptable if the device is being validated for use by the general population (Johnston et al., 2020). According to this criterion, the MAPE values recorded for Mi4 and H4 smart wristbands for both normal and fast walking pace in male and female participants in the current study were found to be excellent, with a value of <5%.

In a study conducted in laboratory conditions and with an adult population, it was found that there was high accuracy regarding the number of steps recorded with Mi Band 4 and the number of steps recorded with a video recorder in both comfortable (r = 0,665) and fast walking (r = 0,759; de la Casa Pérez et al., 2022). In another study, the validity rate was found to be high for Xiaomi Mi Band 2, with a mean error ratio of <5% (St Fleur et al., 2021). In the current study, the compatibility of the Mi4 device according to the criteria step count at normal walking pace was excellent with 83.1% in male participants and good with 71.6% in female participants. According to the criteria at fast walking pace, the compatibility of the Mi4 device was found to be poor with 44.2% in male participants and poor with 39.7% in female participants. When the steps were not accurately estimated, the Mi4 device tended to underestimate the values. These results differ from the results of the above research. This difference may be due to the walking styles of individuals from different societies included in the study. The reason for this is worth further investigation.

In a study conducted, it was reported that Huawei Talk Band 3 performed very well in terms of the accuracy and stability of SC measurement (Xie et al., 2018). In the current study, the compatibility of the H4 device was good with 70.5% in male participants and good with 70.2% in female participants according to the criteria at normal walking pace. The compatibility of the H4 device was found to be moderate with 59.2% in male participants and good with 63.1% in female participants according to the criteria at fast walking pace. This result is compatible with the results of research conducted on previous Huawei models.

However, in the current study, when the reliability of the devices was calculated, the test-retest compatibility of the Mi4 device was low in both male and female participants (r = .289 and r= .371) at fast walking pace. According to ICC and Pearson Correlation analysis, it was seen that the Mi4 device was not valid and reliable at fast walking pace. The highest compatibility of the Mi4 device was found in male participants at normal walking pace (r = .921). The test-retest compatibility of the H4 device was moderate in male and female participants both at normal walking pace (r = .523 and r = .543, respectively) and at fast walking pace (r = .523 and r = .543, respectively). According to all analyses, the H4 device gave both valid and reliable results.

Limitations

Compared with previous studies, this study is thought to have various strengths. First of all, it involves reliability analysis as well as validity analysis. Secondly, the protocol used is accepted in laboratory conditions to analyze the accuracy of smart wearable devices. Thirdly, the study has an adequate sample of male and female participants, and the sample is homogeneous.

However, there are some limitations in our study. First, although we evaluated the SC at different intensities, we did not measure the step count in clinical populations such as individuals with obesity or disability who might have walking differences due to pathology.

Secondly, this study was conducted in a laboratory setting and, therefore, cannot be generalized to leisure activities.

CONCLUSION

According to the current study results, it was revealed that the MAPE values recorded for Mi4 and H4 smart wristbands for both normal and fast walking pace in male and female participants were <5%, which was deemed excellent. The validity and reliability of the H4 device were determined to be at an acceptable level. However, according to ICC and Pearson Correlation analysis, the Mi4 device was not found to be valid and reliable at fast walking pace. In this case, the Mi4 device manufacturers should review their algorithms for SC measurement stability at fast walking pace. Nevertheless, if device evaluations based solely on MAPE values and Bland-Altman plot analyses are considered sufficient, Mi4 and H4 devices can be used in research on the PA levels of populations. Currently, mainstream devices can reliably measure the number of steps used as effective health assessment indicators. Future research should further investigate why there are differences between devices and how activity states affect accuracy; therefore, they should guide and help activity monitor manufacturers improve their algorithms. The interventions targeting physical activity through wearable devices should consider these results when selecting a wearable device as an objective measure of physical activity. The validity of the activity monitors used to measure steps should be analyzed not only for continuous walking in laboratory settings but also for activities focusing on arm movements, intermittent walking, and daily mobility.

Acknowledgements

The authors would like to express their gratitude to the participants who voluntarily participated in this research.

Authors' contributions

The first author took responsibility for the study design, conceptualization, and references. All authors contributed to the study implementation, data collection, data analysis, writing and editing processes.

Declaration of conflict interest

No conflict of interest is declared by the authors.

REFERENCES

Amazon, (2023 Nov 29). "*Smart Band*" *Results.* https://www.amazon.com

An, H. S., Jones, G. C., Kang, S. K., Welk, G. J., & Lee, J. M. (2017). How valid are wearable physical activity trackers for measuring steps?. *European journal of sport science*, *17*(3), 360–368. https://doi.org/10.1080/17461391.2016.1255261

Bassett, D. R., Jr., Toth, L. P., LaMunion, S. R., & Crouter, S. E. (2017). Step Counting: A Review of Measurement Considerations and Health-Related Applications. *Sports medicine (Auckland, N.Z.)*, *47*(7), 1303–1315. https://doi.org/10.1007/s40279-016-0663-1

Bunn, J. A., Navalta, J. W., Fountaine, C. J., & Reece, J. D. (2018). Current State of Commercial Wearable Technology in Physical Activity Monitoring 2015-2017. *International journal of exercise science*, *11*(7), 503–515. PMC5841672

Carlin, T., & Vuillerme, N. (2021). Step and Distance Measurement From a Low-Cost Consumer-Based Hip and Wrist Activity Monitor: Protocol for a Validity and Reliability Assessment. *JMIR research protocols*, *10*(1), e21262. https://doi.org/10.2196/21262

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284-290. https://doi.org/10.1037/1040-3590.6.4.284

Clemes, S. A., O'Connell, S., Rogan, L. M., & Griffiths, P. L. (2010). Evaluation of a commercially available pedometer used to promote physical activity as part of a national programme. *British journal of sports medicine*, *44*(16), 1178–1183. https://doi.org/10.1136/bjsm.2009.061085

CTA (Consumer Technology Association), (2016). *Physical Activity Monitoring for Step Counting*. (2023 Nov 29) https://www.cta.tech/

de la Casa Pérez, A., Latorre Román, P. Á., Muñoz Jiménez, M., Lucena Zurita, M., Laredo Aguilera, J. A., Párraga Montilla, J. A., & Cabrera Linares, J. C. (2022). Is the Xiaomi Mi Band 4 an Accuracy Tool for Measuring Health-Related Parameters in Adults and Older People? An Original Validation Study. *International journal of environmental research and public health*, *19*(3), 1593. https://doi.org/10.3390/ijerph19031593

Dobkin, B. H., & Martinez, C. (2018). Wearable Sensors to Monitor, Enable Feedback, and Measure Outcomes of Activity and Practice. *Current neurology and neuroscience reports*, *18*(12), 87. https://doi.org/10.1007/s11910-018-0896-5

Dooley, E. E., Golaszewski, N. M., & Bartholomew, J. B. (2017). Estimating Accuracy at Exercise Intensities: A Comparative Study of Self-Monitoring Heart Rate and Physical Activity Wearable Devices. *JMIR mHealth and uHealth*, *5*(3), e34. https://doi.org/10.2196/mhealth.7043

Evenson, K. R., Goto, M. M., & Furberg, R. D. (2015). Systematic review of the validity and reliability of consumer-wearable activity trackers. *The international journal of behavioral nutrition and physical activity*, *12*, 159. https://doi.org/10.1186/s12966-015-0314-1

Hartung, V., Sarshar, M., Karle, V., Shammas, L., Rashid, A., Roullier, P., Eilers, C., Mäurer, M., Flachenecker, P., Pfeifer, K., & Tallner, A. (2020). Validity of Consumer Activity Monitors and an Algorithm Using Smartphone Data for Measuring Steps during

Different Activity Types. *International journal of environmental research and public health*, *17*(24), 9314. https://doi.org/10.3390/ijerph17249314

Henriksen, A., Haugen Mikalsen, M., Woldaregay, A. Z., Muzny, M., Hartvigsen, G., Hopstock, L. A., & Grimsgaard, S. (2018). Using Fitness Trackers and Smartwatches to Measure Physical Activity in Research: Analysis of Consumer Wrist-Worn Wearables. *Journal of medical Internet research*, *20*(3), e110. https://doi.org/10.2196/jmir.9157

Huawei (2022, Aug 20). *Wearables band 4*. https://consumer.huawei.com/tr/wearables/band4/

IDC (2022, Aug 20). *Wearable Devices in the Middle East, Turkey, and Africa: Market Analysis and Vendor Highlights, 2021.* https://www.idc.com/research/viewtoc.jsp?containerId=META49036222

Johnston, W., Judice, P. B., Molina García, P., Mühlen, J. M., Lykke Skovgaard, E., Stang, J., Schumann, M., Cheng, S., Bloch, W., Brønd, J. C., Ekelund, U., Grøntved, A., Caulfield, B., Ortega, F. B., & Sardinha, L. B. (2020). Recommendations for determining the validity of consumer wearable and smartphone step count: expert statement and checklist of the INTERLIVE network. *British journal of sports medicine*, *55*(14), 780–793. https://doi.org/10.1136/bjsports-2020-103147

Kastelic, K., Dobnik, M., Löfler, S., Hofer, C., & Šarabon, N. (2021). Validity, Reliability and Sensitivity to Change of Three Consumer-Grade Activity Trackers in Controlled and Free-Living Conditions among Older Adults. *Sensors (Basel, Switzerland)*, *21*(18), 6245. https://doi.org/10.3390/s21186245

Shin, G., Jarrahi, M. H., Fei, Y., Karami, A., Gafinowitz, N., Byun, A., & Lu, X. (2019). Wearable activity trackers, accuracy, adoption, acceptance and health impact: A systematic literature review. *Journal of biomedical informatics*, *93*, 103153. https://doi.org/10.1016/j.jbi.2019.103153

Silfee, V. J., Haughton, C. F., Jake-Schoffman, D. E., Lopez-Cepero, A., May, C. N., Sreedhara, M., Rosal, M. C., & Lemon, S. C. (2018). Objective measurement of physical activity outcomes in lifestyle interventions among adults: A systematic review. *Preventive medicine reports*, *11*, 74–80. https://doi.org/10.1016/j.pmedr.2018.05.003

St Fleur, R. G., St George, S. M., Leite, R., Kobayashi, M., Agosto, Y., & Jake-Schoffman, D. E. (2021). Use of Fitbit Devices in Physical Activity Intervention Studies Across the Life Course: Narrative Review. *JMIR mHealth and uHealth*, *9*(5), e23411. https://doi.org/10.2196/23411

Statista (2022, Aug 20). *Connected wearable devices worldwide 2016-2022*. Published by Federica Laricchia, Feb 14, 2022. https://www.statista.com/statistics/487291/global-connected-wearable-devices/#statisticContainer

Steeves, J. A., Tyo, B. M., Connolly, C. P., Gregory, D. A., Stark, N. A., & Bassett, D. R. (2011). Validity and reliability of the Omron HJ-303 tri-axial accelerometer-based pedometer. *Journal of physical activity & health*, *8*(7), 1014–1020. https://doi.org/10.1123/jpah.8.7.1014

Straiton, N., Alharbi, M., Bauman, A., Neubeck, L., Gullick, J., Bhindi, R., & Gallagher, R. (2018). The validity and reliability of consumer-grade activity trackers in older, community-dwelling adults: A systematic review. *Maturitas*, *112*, 85–93. https://doi.org/10.1016/j.maturitas.2018.03.016

Svarre, F. R., Jensen, M. M., Nielsen, J., & Villumsen, M. (2020). The validity of activity trackers is affected by walking speed: the criterion validity of Garmin Vivosmart® HR and StepWatch™ 3 for measuring steps at various walking speeds under controlled conditions. *PeerJ*, *8*, e9381. https://doi.org/10.7717/peerj.9381

Topalidis, P., Florea, C., Eigl, E. S., Kurapov, A., Leon, C. A. B., & Schabus, M. (2021). Evaluation of a Low-Cost Commercial Actigraph and Its Potential Use in Detecting Cultural Variations in Physical Activity and Sleep. *Sensors (Basel, Switzerland)*, *21*(11), 3774. https://doi.org/10.3390/s21113774

Xiaomi (2022, Aug 20). *Mi Smart Band 4*. https://www.mi.com/tr/mi-smart-band-4/

Xie, J., Wen, D., Liang, L., Jia, Y., Gao, L., & Lei, J. (2018). Evaluating the Validity of Current Mainstream Wearable Devices in Fitness Tracking Under Various Physical Activities: Comparative Study. *JMIR mHealth and uHealth*, *6*(4), e94. https://doi.org/10.2196/mhealth.9754