A Comparison of a Smartphone App and a Wrist-Worn Fitness Tracker for Self-monitoring of Physical Activity by Older and Younger Users

Rebekka Kupffer, Melanie Wutzler, Josef F. Krems, and Georg Jahn(✉)

Institute of Psychology, Chemnitz University of Technology, Wilhelm-Raabe-Str. 43, 09120 Chemnitz, Germany
{rebekka.kupffer,melanie.wutzler}@s2014.tu-chemnitz.de, {josef.krems,georg.jahn}@psychologie.tu-chemnitz.de

Abstract. Wearables capable of monitoring steps are important elements of behavior change interventions to increase physical activity. For intervention studies, there is currently the choice between smartphone apps and fitness trackers for self-monitoring daily steps. We report results from a pilot study, in which younger and older participants experienced both types of devices, rated usability, and performed usability tasks. The fitness tracker, which was operated with a single touch-sensitive button, proved advantageous in subjective and objective usability. The discussion includes further aspects of the choice between smartphone apps and fitness trackers for use in interventions.

Keywords: Mobile health · Wearables · Quantifying self · Behavior change Pedometer app · Accelerometry

1 Introduction

Increasing physical activity is of foremost importance in improving public health. Self-monitoring is a main component of behavior change interventions to increase physical activity [1] and wearable devices that record physical activity support self-monitoring very effectively. They capture activity more accurately than subjective estimates, which require effortful awareness and are distorted by memory lapses, imprecision, and motivated judgment. Before smartphones and fitness trackers became widely available, pedometers have proven effective in increasing daily physical activity quantified by step counts. In studies, in which participants recorded their daily step counts in diaries and had been assigned a goal of daily steps to strive for, the increase achieved by pedometers, diaries, and goal setting in groups of sedentary adults was about 2000 steps per day [2]. The effect of such an intervention is likely attenuated if

R. Kupffer and M. Wutzler have contributed equally.

B. Guidi et al. (Eds.): GOODTECHS 2017, LNICST 233, pp. 331–341, 2018.
https://doi.org/10.1007/978-3-319-76111-4_33

individuals are not participating in a scientific study and long-term effects are still unclear. Moreover, the achievable increase in physical activity is likely smaller for older adults with impaired mobility. Nonetheless, technological support for self-monitoring is an important element of behavior change interventions even in older age groups.

Today, fitness trackers and smartphone apps are available for self-monitoring of physical activity. They afford potentially powerful additional functionality for achieving behavior change (e.g., adaptive goal setting, individualized feedback, prompts, rewards, suggestions for activity, social networking, and social support), however, particularly for older target groups, usability will critically impact acceptance and consequently effectiveness. Several aspects of usability differ between the two types of wearables and need to be considered when deciding which type should be employed in a behavior change intervention with older adults.

First of all, individuals who are already using a smartphone are likely familiar with its basic functions, can power it on and off, can activate the display, can charge the smartphone, can access apps, and can switch between apps. All these necessary operations can be serious obstacles for less experienced individuals. For older users who do not own a smartphone and lack experience with computers and digital wearables, a standalone fitness tracker may be better suited for simple self-monitoring of physical activity if it does not pose serious usability problems. Presumably, a simple standalone fitness tracker will be better accepted and used with less problems than a smartphone app.

For self-monitoring, it is important that the devices can be used without assistance. Many fitness trackers require to be used with a smartphone or a computer for being set up and for accessing activity data as statistics over multiple days. Setting up the fitness tracker could be performed by a second person, but day to day use should not require assistance by a second person. Putting the fitness tracker on and off, charging, and accessing recorded activity data all have to be managed by the person who uses the device for self-monitoring. We test for both device types how capable older users are of these tasks that have to be performed in daily use.

For exploring, whether to employ a smartphone app or a fitness tracker in behavior change interventions with older individuals, and to study the usability of examples of both device types, we enrolled participants in a study that consisted of three phases each lasting 5 days. Accelerometry was used in all three phases for a reference measurement of physical activity. In the first phase, participants just wore the accelerometer, estimated their daily steps, and noted their estimate in a diary. For the second and third phase, they were asked to strive for a daily step goal of 7000 steps and used either a smartphone app or a wrist-worn fitness tracker for self-monitoring. They kept a diary of the daily step counts. Usability ratings and usability tests were scheduled before and after the participants had gained experience in using the devices for self-monitoring.

2 Method

2.1 Participants

The study included 20 participants in an older and a younger age group. The 13 older participants (7 female, 6 male) were between 65 and 81 years of age and were recruited in a lecture for senior citizens, on a public square in the city of Chemnitz, and from a participant database. The 7 younger participants (4 female, 3 male) were between 21 and 30 years of age and were recruited via a student mailing list or at Chemnitz University of Technology. Demographic information about the participant groups is shown in Table 1. Participants received 40 Euros as compensation. Three students chose to collect participant hours as part of a curricular requirement instead.

Table 1. Demographic information on the older (>65 years) and younger (<30 years) participant groups.

	Older (N = 13)	Younger (N = 7)
Age in years	70.03 (4.48)	24.87 (3.38)
BMI in kg/m^2	26.59 (2.71)	21.87 (1.64)
CPQ-score	18.76 (6.15)	26.99 (2.86)
SPQ-score	15.27 (5.29)	23.08 (4.49)

Note. Means and standard deviations (in brackets). BMI: body mass index; CPQ: Computer Proficiency Questionnaire; SPQ: Smartphone Proficiency Questionnaire

2.2 Devices

Smartphone and Pedometer App. When physical activity was self-monitored with a smartphone app, participants used Google Nexus 5 Android smartphones with the "Pacer" pedometer app. The app was installed and set up before the participants received the smartphones. In March 2017, Pacer was the best-rated pedometer app (4.6/5 stars) in the German Android app-store. The app is always running in the background. Therefore, it is not necessary to start a measurement before going for a walk. The step count for the current day is shown on the Android homescreen. Additional information can be accessed when the app is opened (see Fig. 1). GPS tracking has to be activated in order to pace the steps made.

Fig. 1. A screenshot from the pedometer smartphone app (Pacer), the fitness tracker (Mi Band 2), and the toggling cycle for display content on the fitness tracker.

The smartphone was carried in a waistbelt that also contained the accelerometer (see Fig. 2). Participants were instructed to wear the smartphone on the right hip.

Fig. 2. Waistbelt and movisens Move II accelerometer.

Wrist-Worn Fitness Tracker. When physical activity was self-monitored with a fitness tracker, participants used the Mi Band 2 (Xiaomi). The fitness tracker is usually worn in a wristband and shows information on an OLED display that is activated and toggled by a single touch-sensitive button (see Fig. 2). By default, three screens can be cycled through: The current time is shown when the display is activated. The second touch switches to the step count, which is automatically reset at midnight. A third touch switches to measuring the heart rate. A pounding heart symbol is shown while the heart rate is measured with an optical sensor, then the heart rate is shown. The display turns off automatically and starts with showing the time when activated again.

The Mi Band 2 has to be set up with a smartphone app that is available for iOS and Android. The fitness tracker was set up and configured before the participants received it. They did not use the fitness tracker together with a smartphone and thus, the additional features available in the app such as a history of daily step counts, energy consumption, covered distance, sleep quality etc. were not accessed. Participants were instructed to wear the fitness tracker on the left wrist.

Accelerometer. For baseline and reference measurements of step counts, each participant wore a movisens Move II activity tracker (movisens, Karlsruhe, Germany) (see Fig. 2). It combines a three-axis accelerometer, a barometric altimeter, and a temperature sensor. The raw data can be processed by the DataAnalyzer software from movisens to generate activity reports including step counts. The participants were instructed to wear the accelerometer in the waistbelt on the right hip.

2.3 Materials and Tasks

A questionnaire on computer proficiency, a questionnaire on smartphone proficiency, and a brief usability scale to collect subjective ratings were employed. The Computer Proficiency Questionnaire (CPQ) [3] asks for subjective ratings on a 5-point scale of the ability to perform tasks with a computer (basic tasks, printing, communication, calendar functions, using the internet, entertainment). The CPQ-score varies between 6 and 30 and higher values indicate higher proficiency. We used a German translation of the CPQ and prepared a smartphone proficiency questionnaire (SPQ) consisting of 13 items by rephrasing appropriate items of the CPQ. The SPQ-score varies between 5 and 25.

For subjective ratings of usability, the System Usability Scale (SUS) [4] was used. It collects ratings on a 5-point scale for 10 items, which are then converted to a score between 0 and 100. Higher scores indicate better usability.

A different set of tasks was put together for each device to test usability (see Table 2). Participants' hands and the devices were recorded on video while the tasks were performed. Task times (efficiency) and error frequencies (effectiveness) were coded from the videos.

Table 2. Usability tasks performed with the smartphone app and with the fitness tracker.

Smartphone app	Fitness tracker
1. Switch the display on	1. Put the fitness tracker in the wristband
2. Search for the pedometer app and open the app	2. Put the wristband on
3. Walk a round in the room and check the step count (repeated two times)	3. Switch the display on
4. Search for a screen that shows yesterday's step count	4. Search for the step count
5. Search for a screen that shows the total distance recorded in the app	5. Walk a round in the room and check the step count (repeated two times)
	6. Check your heart rate
	7. Take off the wristband

2.4 Procedure

The study encompassed four appointments at the university and three 5-day intervals of activity tracking as shown in Fig. 3. Before the first appointment, participants were informed about the purpose and procedure of the study and provided informed consent.

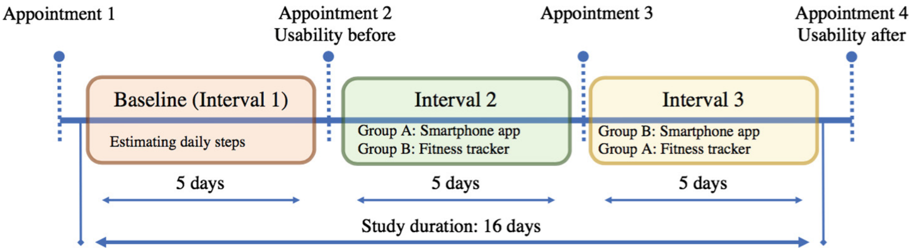

Fig. 3. Procedure with scheduling of appointments and intervals of activity tracking. Usability tests and usability ratings for both device types took place at appointments 2 (before using the devices for self-monitoring) and 4 (after self-monitoring).

At the first appointment, the tests of physical activity were conducted, participants filled in a questionnaire about demographic information, the CPQ, and the SPQ. They were given a diary and were instructed to fill in at the end of each day the number of steps they had taken, information about the weather (good, neutral, bad), and their mood (good, neutral, bad). The participants also received a short manual on how to use and charge the devices.

For the first interval of activity tracking (baseline), the participants received the movisens accelerometer and the waistbelt. In the first five days of activity tracking, they were supposed to estimate how many steps they had taken and to note the estimated step count in the diary every night before going to bed.

At the second appointment, participants performed the usability tasks first with the fitness tracker and then with the smartphone. They were instructed to try to find the solutions by their own, without any help. When they indicated that they needed help or when they could not complete a task, the experimenter demonstrated the task and the participant tried again. Then, they filled in the SUS whereby each question was asked for the fitness tracker at first, and then for the smartphone app. This was followed by some open questions.

For the second interval of activity tracking, participants received an activity goal of a minimum of 7000 steps per day and their first activity tracker. Half of the participants (group A) started with the smartphone app and the remaining half (group B) with the fitness tracker. Participants were instructed to document the step counts indicated by the respective device in their diaries each day before going to bed together with information on the weather and their mood for this day.

At the third appointment, participants in group B returned the fitness trackers, which were checked, charged, reset, and then given to participants in group A. Participants in group B used the smartphone app in the subsequent activity tracking interval. The third interval of activity tracking proceeded as the second interval. The instructions (activity goal of 7000 steps per day minimum and diary notes of step count, weather, and mood) remained the same. At the fourth and final appointment participants returned the devices and the usability-tests from the second appointment were repeated with the same instructions, followed by the SUS ratings and an open interview.

3 Results

3.1 Step Counts

All 20 participants completed the study resulting in 300 days of activity tracking, of which accelerometer recordings of eight days (2.7%) are missing because of technical failure. Participants' mean estimates of daily step counts can be compared to the mean daily step count recorded by the accelerometer for the five days in the first interval of activity tracking (baseline). Both means are shown separately for each participant in Fig. 4. For about half of the participants the mean daily estimate was clearly off the actual mean daily step count and more participants underestimated than overestimated their mean step count.

Figure 5 shows boxplots of mean daily step counts by type of tracker and age group, and by weather and gender. Mean step counts were higher for males than for females, higher at days with positive than negative weather ratings, and higher at days with positive than negative mood ratings. The type of fitness tracker and the age of participants

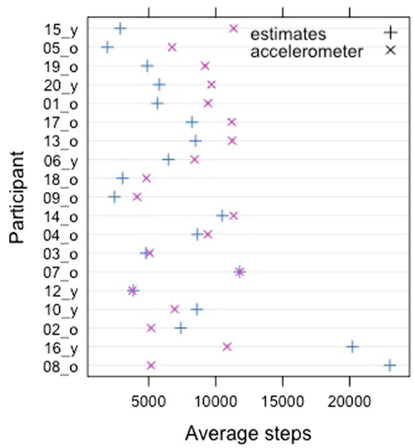

Fig. 4. Means of estimated and accelerometer daily step counts in the first interval of activity tracking (baseline) separately for each of 19 participants (missing data for 11_y) ordered by the difference between actual and estimated mean step counts.

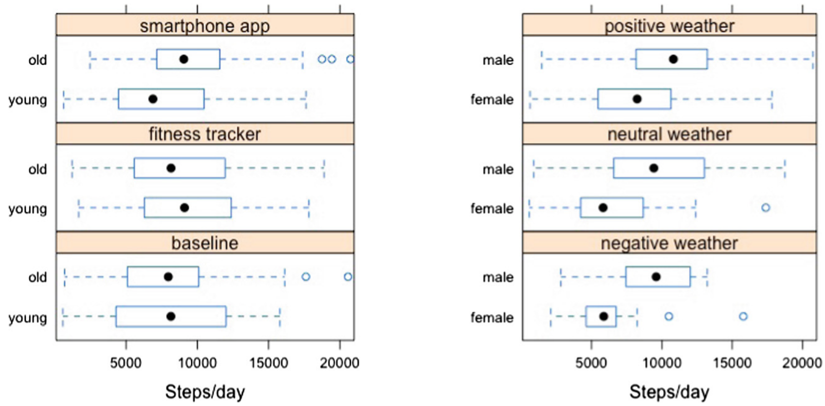

Fig. 5. Boxplots of mean accelerometer step counts by type of tracker and age group (left), and by weather rating and sex (right).

had no consistent effects on mean step counts. Linear mixed effects modeling of step count data from all three intervals of activity tracking confirmed that sex, weather, and mood but neither device type nor age were significant predictors of step counts.

3.2 Usability Tasks and Usability Ratings

All tasks and the SUS ratings were performed twice by all participants, once *before* they used the devices for self-monitoring (at appointment 2) and once *after* they had used the devices for self-monitoring. The tasks differed for the two devices, thus performance data were analyzed and are reported separately for each device.

Task effectiveness was quantified by the average error frequency for each device, which indicates how often participants could not complete tasks or asked for help. Task efficiency was quantified by the total time needed to perform the tasks.

Table 3 presents mean error frequencies and mean total times for the smartphone app and for the fitness tracker, before and after self-monitoring, and separately for the younger and the older age group. Both variables indicate better performance at the second test session (after) and for the younger age group reflecting practice and age effects. There was no difference between the age groups in the mean error frequency with the fitness tracker at the second session (both at zero) (Table 3).

Table 3. Mean error frequencies and mean total times for the usability tasks performed with the smartphone app (left) and the different usability tasks performed with the fitness tracker (right) before and after self-monitoring and separately for the younger and the older age group.

	Smartphone app		Fitness tracker	
	Young	Old	Young	Old
Error frequency				
Before	4.42 (2.86)	19.27 (3.59)	1.59 (1.58)	4.27 (2.00)
After	0.00 (0.00)	13.19 (5.23)	0.00 (0.00)	0.00 (0.00)
Time (sec)				
Before	155.86 (18.40)	339.14 (49.64)	77.14 (10.29)	129.21 (17.42)
After	65.62 (11.59)	268.33 (68.07)	40.71 (3.67)	60.82 (4.40)

Note. Means and standard errors (in brackets). Because different tasks were performed, data cannot be compared directly between devices.

Figure 6 shows mean SUS-scores with standard errors for the fitness tracker and the smartphone app before and after using the devices for self-monitoring, separately for

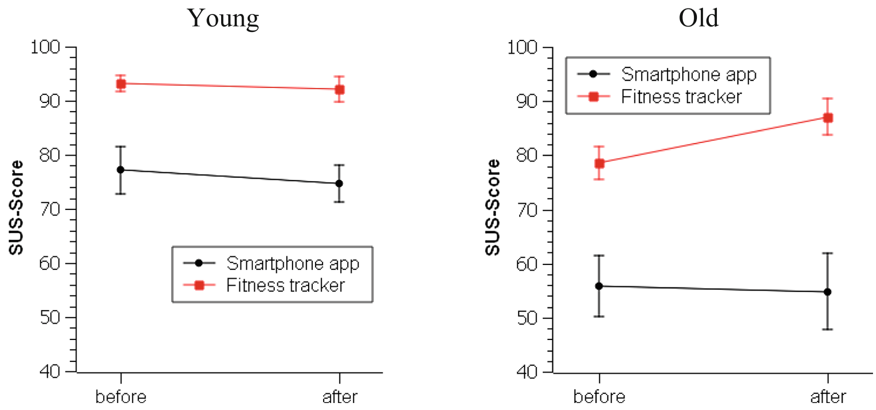

Fig. 6. Mean SUS-scores with standard errors for the fitness tracker and the smartphone app before and after using the devices for self-monitoring and separately for the younger (left) and the older (right) age group.

the younger and the older age group. Usability was rated higher for the fitness tracker at both sessions in both age groups and the difference in mean SUS-scores slightly increased from before to after in the older age group.

Furthermore, the influence of computer proficiency, smartphone proficiency, and age on task time was examined by separate linear regressions. These regressions were computed only for the older group (see Table 4) because there was hardly any variance in the CPQ and SPQ-scores in the younger group. Self-rated smartphone proficiency (SPQ) was a better predictor of task time than self-rated computer proficiency with both devices and importantly, smartphone proficiency was less influential for the time required for fitness tracker tasks. For the smartphone app tasks, smartphone proficiency was a better predictor of task time than age in the older group.

Table 4. Linear regressions predicting the total time required for performing usability tasks in the older group by self-rated computer proficiency, smartphone proficiency, and age.

	Smartphone app			Fitness tracker		
	β	*SE*	R^2	β	*SE*	R^2
CPQ	−0.06	8.75	.00	−0.25	2.97	.06
SPQ	−0.57	9.04	.33	−0.36	3.43	.13
Age	0.44	10.58	.19	0.40	3.79	.16

Note. No direct comparisons are possible between the results for the smartphone app and the fitness tracker because different usability tasks were performed.

4 Discussion

Two devices for self-monitoring daily steps were compared in a pilot study for an informed decision about which to employ in an intervention to increase the physical activity of older adults. Younger and older adults used both a smartphone app and a wrist-worn fitness tracker. The fitness tracker was rated more favorably than the smartphone app by older and younger participants. This usability advantage of a simple wrist-worn fitness-tracker over a smartphone pedometer app was also confirmed by a smaller age effect on usability task performance particularly after some practice.

The participants used the devices while striving for a goal of 7000 steps per day after they had already estimated daily step counts during a baseline interval. Comparisons and analyses of step counts in the present study should be interpreted cautiously because the sample size was small and activity was recorded for a few days only. In the present study, there was no increase in physical activity from the baseline but results from larger previous studies suggest that both devices could increase physical activity in behavior change interventions [1]. Baseline step counts could have been increased already by self-monitoring in the form of estimating daily steps and noting the estimates in a diary. The comparison of estimated with accelerometer step counts clearly and unsurprisingly shows that any technical support improves self-monitoring considerably.

The clear usability advantage of the fitness tracker once again demonstrates that reduced complexity (one button, no choice reaction) increases accessibility and acceptance. Notably, this result is in contrast with the conclusion in a previous study [5] comparing a smartphone app and a fitness tracker, in which participants had experienced only their preferred device and the smartphone app was rated more favorably. In the present study, 19 of the 20 participants stated in the final interview that they would prefer the fitness tracker for everyday use. This needs to be qualified a bit because some also noted that it was inconvenient to carry the smartphone for the sole purpose of tracking steps or as a second smartphone. Of course, this disadvantage would dissolve if an app would be used on a smartphone that is carried anyway.

Several differences between the device types that were not the focus of the present study deserve to be mentioned because they may be important if devices are chosen for interventions. If a person already uses a smartphone, self-monitoring with an app on one's own device is possible and experience with the smartphone and its operating system may help in using the app. Ideally, the device for self-monitoring would be worn continuously while activity occurs. A smartphone is likely carried less continuously than a wrist-worn fitness tracker. To really carry it continuously, the smartphone has to be charged over night or during other intervals without activity. Smartphones have to be charged more often than simple fitness trackers. On the other hand, putting on the fitness tracker may be forgotten. With a smartphone that is used for various purposes it is less likely to miss extended intervals of activity. How a smartphone is carried varies more than how a wrist-worn fitness tracker is carried. Certain ways of carrying the smartphone may result in more accurate step counts than others, however, studies examining the accuracy of step counts for recent smartphone apps arrived at quite positive accuracy evaluations. In the present study, the smartphone was carried in a belt. A wrist-worn fitness tracker more easily than a smartphone can be equipped with sensors that can pick up physiological signals. Measuring the heart rate may be a valuable functionality in interventions targeting physical activity of older individuals because the interventions may aim at activity of a certain intensity and the heart rate may be used as feedback for ensuring activity at the intended intensity level (e.g., walking at the right speed to be physically active with individually moderate intensity). Overall, a simple fitness tracker seems to be the better choice in intervention studies that include smartphone novices.

References

1. Sullivan, A.N., Lachman, M.E.: Behavior change with fitness technology in sedentary adults: a review of the evidence for increasing physical activity. Front. Public Health **4**, 289 (2017)
2. Bravata, D.M., Smith-Spangler, C., Sundaram, V., Gienger, A.L., Lin, N., Lewis, R., Stave, C.D., Olkin, I., Sirard, J.R.: Using pedometers to increase physical activity and improve health: a systematic review. JAMA **298**(19), 2296–2304 (2007)

3. Boot, W.R., Charness, N., Czaja, S.J., Sharit, J., Rogers, W.A., Fisk, A.D., Mitzner, T., Lee, C.C., Nair, S.: Computer proficiency questionnaire: assessing low and high computer proficient seniors. Gerontologist **55**(3), 404–411 (2015)
4. Brooke, J.: SUS: a quick and dirty usability scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, I.L. (eds.) Usability Evaluation in Industry, pp. 189–194. Taylor & Francis, London (1996)
5. Fong, S.S.M., Ng, S.S.M., Cheng, Y.T.Y., Zhang, J., Chung, L.M.Y., Chow, G.C.C., Chat, Y.T.C., Chan, I.K.Y., MacFarlane, D.J.: Comparison between smartphone pedometer applications and traditional pedometers for improving physical activity and body mass index in community-dwelling older adults. J. Phys. Ther. Sci. **28**(5), 1651–1656 (2016)