Making Lifelogging Usable: Design Guidelines for Activity Trackers

Jochen Meyer[1], Jutta Fortmann[2], Merlin Wasmann[1], and Wilko Heuten[1]

[1] OFFIS Institute for Information Technology, Germany
[2] University of Oldenburg, Germany

Abstract. Of all lifelogging tools, activity trackers are probably among the most widely used ones receiving most public attention. However, when used on a long-term basis e.g. for prevention and wellbeing, the devices' acceptance by the user and its usability become critical issues. In a user study we explored how activity trackers are used and experienced in daily life. We identified critical issues with regard not just to the HCI topics wearability, appearance of the device, and display and interaction, but also to aspects of modeling and describing the measured and presented data. We suggest four guidelines for the design of future activity trackers. Ideally, activity tracking would be fulfilled by a modular concept of building blocks for sensing, interaction and feedback that the user can freely combine, distribute and wear according to personal preferences and situations.

1 Introduction

The evolution of wearable devices for data capture in the last few years has changed the typical user of such devices from the researchers and early adaptors to normal persons using them in their daily lifes. While multimedia devices with video, audio and image capture such as Google Glass are spectacular and represent the technological state of the art, it is the fairly lo-fi activity trackers that are currently probably most widely being used and receiving the most public attention.

Products such as the Fitbit One[1], Nike Fuelband[2], or Jawbone UP[3] are attractive, easy to use and fairly low-cost consumer products for monitoring daily activity. They are not just a valuable source of data on their own right, e.g. for health monitoring and behavior change [5], but also provide metadata for other applications such as activity recognition in life logging.

Using these trackers as part of lifelogging or for prevention and wellbeing, however, is a long-term effort possibly covering many years. In these cases, user acceptance and usability of monitoring devices in users' daily lives also in an ergonomical and aesthetic sense become important aspects.

[1] `https://www.fitbit.com/one`
[2] `http://www.nike.com/us/en_us/lp/nikeplus-fuelband`
[3] `https://jawbone.com/up`

X. He et al. (Eds.): MMM 2015, Part II, LNCS 8936, pp. 323–334, 2015.

We therefore explored how activity trackers are used in daily life. We researched user experiences and examined usability and perceived comfort. From these observations we identified factors that influence the use of activity tracking devices in daily life. Based on these we discuss design implications and a vision for the future design of activity trackers.

2 Related Work

While lifelogging was initially understood as capturing primarily images, nowadays it is more and more seen that lifelogging must in fact be understood broader. [2] presents six classes of data that are relevant for lifelogging, namely passively as well as actively captured media, mobile context and activity data, computer activity, and biometric information, where the latter also includes behaviors such as physical activity. Ryoo et al. differentiate between wearable lifelogging devices with high data rate (e.g. video) and low data rate (e.g. GPS signal) [16]. Lifelogs have been used to capture images [9], track context information [10], or monitor actions taken at a computer [7].

[2] also discusses the main requirements for lifelogging devices. He particularly addresses technical issues, including stability, reliability, battery life, and storage, but also identifies unobtrusive wearability as a necessity and mentions the intrusive nature of multiple wearable devices, taking into account issues such as direct skin contact needed by some devices. We continue this discussion by focussing particularly on state-of-the-art activity trackers as specific lifelogging devices.

Activity trackers are routinely used in interventions, e.g. to encourage physical activity, as well as in public health studies [17] or to monitor elderly persons' behaviors (e.g. [15], [8]). In this context the feasibility and usability of activity trackers are frequently discussed. Many studies have been done about the precision and expressive value of the measured data (e.g. [3]). However, end-users ask more differentiated questions about their data [13], therefore precision alone is not the key. A positive attitude of the user towards monitoring (e.g. [6]) is helpful to ensure compliance to interventions, and it's a prerequisite for unsolicited use of activity trackers as part of lifestyle and wellbeing management. Equally important are the design requirements for mobile interventions, such as [4], and there are a number of evaluations, e.g. of Android based pedometer apps [11], or comparison of two approaches for personal health research [18]. In previous work [14] we gained first insights into the usability of activity trackers in daily life. We continue this work by systematically exploring how users interact and work with recent activity trackers.

3 Activity Trackers

Activity trackers vary e.g. in form factor and size, types of data measured, display type and size, interaction design, connectivity to portals and mobile phones, price, and motivational and persuasive measures. Also numerous activity tracking apps for smartphones are available. In this study we focussed on the investigation of form factor, input methods and the presentation of feedback, taking into account both, dedicated activity tracking devices, and smartphone apps.

Form factor: Clip-shaped activity trackers can be attached to clothes, e.g. a trouser pocket or a blouse, or can be worn invisibly inside a pocket. Bracelet-shaped activity trackers are worn on the wrist like a watch.

Input: A single button is used to navigate through the display of different types of information or to switch between different modes, such as the sleep monitoring mode and the daily activity mode. Various patterns of tapping on the device are used to initiate the display of information or to switch between different modes.

Feedback: A numeric display on the device shows numbers such as the steps done or the calories burnt so far. A graphic display on the device uses abstract visualizations to display results, such as a colored progress bar to show the progress towards a pre-defined goal, or a motivational image. Vibration may be used to represent e.g. the achievement of a goal or to confirm certain inputs.

Apps use a mobile phone's sensors such as the accelerometer or the GPS receiver to measure activity. Smartphones are normally carried in a pocket, attached to the arm or belt, or hold in the hand during activities. The phone's touch screen or physical buttons are used to make input. Feedback is typically presented via the phone's high-resolution display and can be complemented by sound and vibration signals. As apps are cheap or even free, and quick and easy to download and install, the inhibition threshold to use them is low.

4 Study Design

We aimed to understand which factors influence the use or non-use of activity trackers in daily life. Particularly we wanted to know:

- Which form factor do people prefer and why?
- Do the users understand the devices' measurements in the way they are presented?
- Which are motivating and hindering factors for using or not using the devices?

We therefore conducted an exploratory study in which we investigated how popular activity trackers are used and experienced under real-life circumstances.

We acquired 12 participants (7 males) from a participant database, through public announcements, and personal contacts. They were aged between 25 and 70 (M = 43.7, SD = 18.9). None of the participants suffered from serious health problems. All of them stated to be interested in health and in monitoring their behavior. All participants had access to and experience in operating a computer. Nine participants stated that they owned a smartphone.

We chose four activity trackers that prototypically represent today's typical products capturing biometric data including, but not limited to physical activity

Table 1. Activity trackers used in the study

Activity Tracker	Form Factor	Input	Feedback	Synchronization
Fitbit Ultra / One	Clip	One-Button	Numeric and simple graphic display	Wireless through PC docking station to fitbit.com
Fitbit Flex	Wristband (can be worn in pocket)	Tapping	5 point progress bar; vibration	Wireless through PC dongle or Bluetooth to fitbit.com
Nike FuelBand	Wristband	One-button	Numeric and simple graphic display	USB connection via PC to nike.com
Runtastic Pedometer	Smartphone app	Touch, button	Hi-res graphic display, sound, vibration	Automatically to runtastic.com (optional)

(see Table 1). They vary in form factor, input method, and feedback display to cover a broad range of the design space of current activity trackers. To keep the technical barriers as low as possible, we avoided trackers that required specific mobile phones for synchronization.

Fig. 1. From left to right: Fitbit One[1], Fitbit Ultra[4], Fitbit Flex[5], Nike FuelBand[2], Runtastic Pedometer App[6]

Participants took part in an initial meeting in which we explained the procedure of the study and introduced all activity trackers that were used in the study. Each participant received a Fitbit Ultra[4] or One, a Fitbit Flex[5] and a Nike FuelBand (see Figure 1). In the first, four-day phase the participants were asked to install and use each of the activity trackers for one day in their daily life. Owners of an iOS or Android phone were also asked to install and use the Runtastic Pedometer app[6] for a day (see Figure 1). This phase aimed at making

[4] `https://help.fitbit.com/customer/de/portal/articles#product_ultra`
[5] `http://www.fitbit.com/us/flex`
[6] `https://www.runtastic.com/en`

the participants familiar with all devices. During this phase, participants documented their initial experiences with each activity tracker in a short protocol. Afterwards, in the second, 10-day phase the participants were asked to use at least one of the activity trackers as continuously as possible in their daily life, as long as they felt comfortable. Participants were free to choose the activity tracker(s) they liked most from the three/four we gave them. During this phase, participants filled out a daily diary in which they briefly documented their experiences.

After the second phase, participants took part in individual meetings that began with the completion of the System Usability Scale (SUS) [1] and the Comfort Rating Scales (CRS) [12] for each activity tracker they had used continuously in the second phase. Afterwards, we conducted a semi-structured interview. We asked about the participants' choice of device, about their use of the measured data and whether they found the data sufficient and understandable. We asked about the data presentation on the device, the use of the portal, and whether they believed they would use the activity tracker over a longer period of time. The interview notes and the participants' diaries were coded by three experts and jointly clustered.

5 Study Results

5.1 Quantitative Results

For the 10-day phase, 9 participants chose to use a Fitbit Ultra or One, 6 used a Nike Fuelband, 5 chose a Fitbit Flex, and 1 used the Runtastic app. Several participants chose to use two different devices simultaneously. The usability of the devices was in general assessed very well (Median SUS for Fitbit Flex: 92.5; Fitbit Ultra/One: 92.5; Fuelband: 85; Runtastic: 72.5). The Fuelband and the Runtastic app received considerably lower values than the Fitbit trackers (see

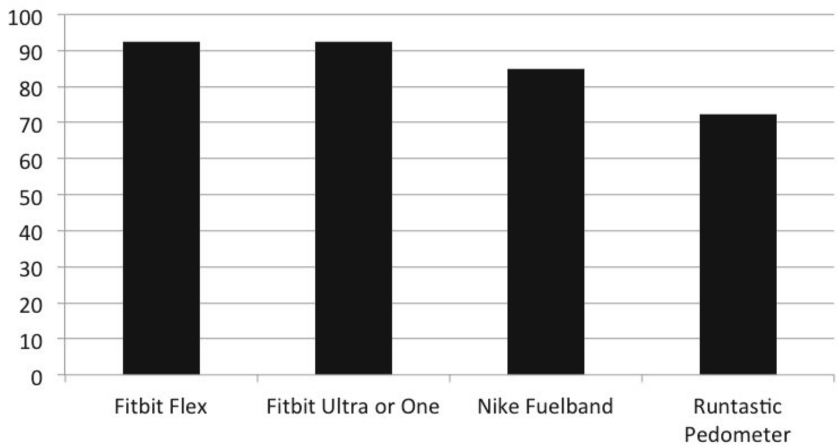

Fig. 2. Median SUS scores for the activity trackers

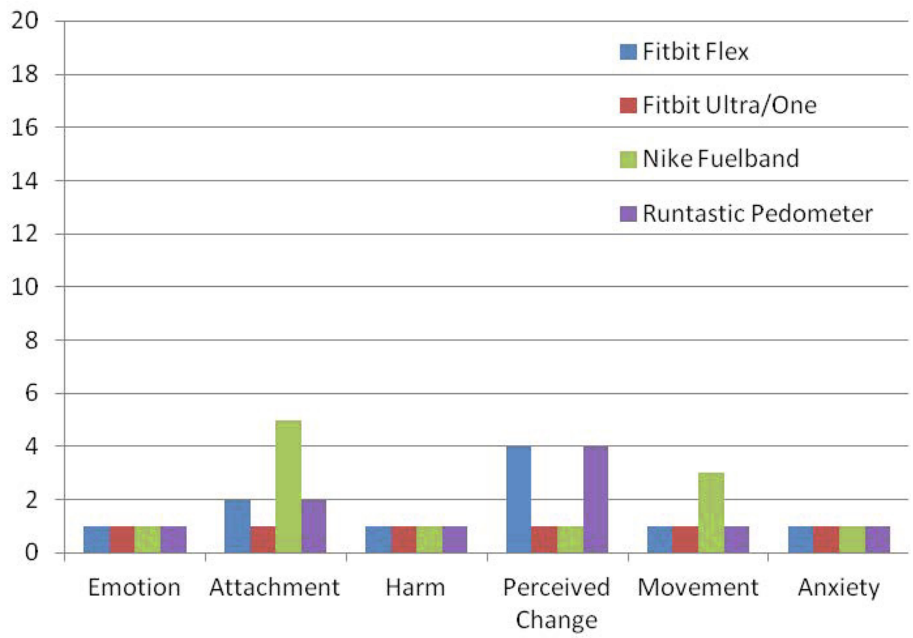

Fig. 3. Median CRS scores for the activity trackers

Figure 2). The Comfort Rating Scales show that in general all activity trackers were rated well with regard to comfort (see Figure 3). The Fuelband performed slightly weaker in attachment (Mdn = 5) and movement (Mdn = 3). These scales reflect how the users felt the device on their body and how it affected the way they moved. Fitbit Flex and the Runtastic app received slightly worse rates in perceived change (Mdn = 4). This scale describes in how far the users felt that the device made them feel physically different.

5.2 Qualitative Results

From the interviews and diaries we identified four concepts as relevant aspects for the design and use of activity trackers: perception by other persons and aesthetic appearance, wearability in daily live, interaction and visualization, and activity measures and the validity of the measured data.

Perception and Appearance. With activity trackers being highly personal devices the visual and aesthetic appearance and its perception by third persons was important.

Many participants appreciated when devices were *inconspicuous* and were not or hardly noticed by other persons. They liked it being invisible and didn't want to draw attention to the fact that they track their activities.

However, some participants liked that wristbands were noticed as a *fashionable accessory*. The Fuelband's design was mentioned as a positive example.

When trackers, both clips and wristbands, would be visible, the *aesthetic appearance* became important. The devices would need to be fashionable and fit to the clothes. This included different and customizable colors for the wristbands as well as the clips.

Wearability. Wearability means the practical consequences of having to wear and use a device 24/7 and in a person's different contexts throughout daily live.

Regarding the *form factor* wrist bands were often felt to interfere with daily activity, whereas clips were in general better accepted, also because they could be worn in a pocket. This, however, was also dependent on the clothes worn: Not all clothes have appropriate pockets. And attaching the clip e.g. to the belt was also a matter of aesthetics. A modular concept allowing use as both a wristband and a "bit", as implemented by the Flex, was appreciated for its flexibility; but having to handle many parts was found to be annoying.

Robustness of the trackers was an important property. Nine participants were concerned about damaging the device. They took it off during gardening work, while swimming or taking a shower, even if it was said to be waterproof. The smaller clip devices that were appraised for inconspicuousness were felt to be easily lost. The stiff design of the Fuelband was positively perceived as being robust and safe. On the other hand it was also reported to be disturbing e.g. when working at a computer: *"It felt like having an elephant on the arm"* (P11).

Four participants considered *smartphone apps* not to be as practical as dedicated activity tracking devices because they would not wear a smartphone continuously and close enough to the body which resulted in incorrect and incomplete data.

Intensity of Interaction. The presentation of tracked data on the device, and how to interact with the device e.g. to change tracking modes was another important topic.

Most participants preferred the *attractive, possibly colorful data visualization* with text, graphics, or animations e.g. on the Fuelband over more minimalistic displays such as the 5-dot-display of Fitbit Flex. They liked getting detailed data, and some liked to use a wristband tracker as a watch replacement.

This demand for visualization might, however, also *change over time*: One participant mentioned that after a longer period of use, a less detailed display would be enough.

Overall, participants found *interacting with the devices* easy. They could easily manage the one-button interface as implemented by the Fitbit One/Ultra and Fuelband. A button-less tap interface like on the Fitbit Flex caused problems. Six participants said it was not intuitively usable and they sometimes made misentries. Sometimes taps were not recognized as such, and sometimes taps were misleadingly recognized from normal arm movements in daily life.

Participants reported they would regularly monitor their progress throughout the day and considered the *continuous feedback at-a-glance* that they received from the devices' displays an important feature.

Direct feedback on the *achievement of a daily goal* was considered very effective: *"Five points on the Flex!", "Vibration of the wristband is motivating"* (P9).

Two participants wanted to be more flexible in defining daily goals or more complex goals such as fitness beyond daily activity.

Activity Measures and Validity of Data. The measured activity data as the primary outcome of using a tracker, raised important questions and issues for the participants.

Seven participants confirmed that monitoring in itself is already motivating and increases the awareness of one's own activity behavior.

An appropriate *level of abstraction* was important to ensure that the data was understandable for the user. Most users instantly understood the concept of step counts that were naturally accepted as the primary measure. However, they were also basically aware that step count as such was already an abstract measure for different types of activities and was not applicable e.g. to cycling. In general, they therefore appreciated alternative activity measures beyond step count. Calories burnt that are calculated by virtually all trackers were found to be interesting, but confusion was caused when it was not clear whether the calories count includes the resting metabolic rate or not. Participants had difficulties in *understanding more abstract measures* for activity such as the Nike Fuelpoints or the flower of the Fitbit One/Ultra.

The *validity of data* was a concern for many participants. They reviewed the results of different activities and compared multiple devices' measurements. They criticized that non-walking activities, such as swimming or biking were not adequately reflected. Incomplete data such as when forgetting to wear or to activate the device was an important aspect that annoyed the participants.

Participants appreciated getting *detailed views* and analyses of their activity data. For most participants the web portals which they would check every couple of days were appropriate. Some preferred an app over a portal, few used the on-device display only.

Beyond the scope of this study were general observations on technical problems with regard to the installation of and daily interaction with the device. In general, users were able to operate the devices, but we didn't research that in detail.

6 Guidelines for the Design of Activity Trackers

Our qualitative results show a broad range of users' preferences. Little surprisingly and like in most HCI designs there is no "one size fits all" device that everybody likes. Rather the appropriate design needs to take into account differing preferences, contexts and sometimes conflicting requirements. For activity trackers we identify four guidelines that should be taken into account:

Make it either invisible, or make it fashionable: Most users liked the trackers to be not noticed by others. The best tracker would therefore be worn invisibly. However, pocket-worn or clip-on devices may not be appropriate to the user's dressing habits. Trackers that are worn visibly are perceived as a fashionable accessory, much like a watch or jewelry. In these cases, designs must match the user's taste and therefore considerably vary between different users and period of use. Adding familiar functions beyond activity tracking such as a watch might mitigate acceptance problems.

Intervention defines interaction, and intervention changes over time: There is a trade-off between unobtrusiveness of wearing and rich interaction. Most users were at first exited about the detailed data they could instantly see on the devices, and they accepted more obtrusive trackers when they provided richer information. But, after some time of use they were satisfied with more limited instant feedback if this attended a device that was easier to use in daily life. Ultimately, the users' decision on how they want to use the device defines what the appropriate interaction method is. A tracker's interaction design may therefore either be intentionally limited to a given intervention and would probably not be used beyond, or it must be adaptable to the changing needs.

Perceived robustness is as important as actual robustness: Trackers should ideally be worn 24/7, and particularly in phases of increased physical activity such as labor, household and gardening, or transportation. There it is wet or dirty, physical force may act on the device, and there is the risk of loss. Although today's devices are in fact quite robust and water-resistant, users were very eager to protect their devices and avoid loss or damage. Consequently they occasionally put off the devices in such contexts, leading to incomplete data. It is therefore not enough to just *build* the device as robust. Rather the users must also *perceive* the device as being reliable and not easily broken or lost. Alternatively, trackers could be cheap single-use devices similar to e.g. contact lenses that are disposed of after use.

Fuzzy but reliable data is better than pseudo-precise but possibly wrong data: Users were highly interested in the data they collected, and they were highly annoyed if the data didn't reflect their behaviors correctly. This is not a simple matter of technically precise measurements: We found too simplistic measures such as step count inherently insufficient, as these measures were *per se* not applicable when accounting for e.g. cycling. Also such measures may make different assumptions about when exactly a certain movement counts as a step and when it's just quivering with a leg. Users notice such deviations between the trackers' measurements and their own perception and will lose confidence in the trackers. The trackers therefore need alternative measures that must be both, easy to understand, and perceived correct for the user. This is primarily a challenge of data modeling. The "'traditional"' step count may be too fine-grained and not appropriate for all activities, but active minutes a day may on the other be too coarse and simplistic. It seems likely that measures must be user-specific to account for a user's individual behaviours and preferences. This requires modeling the user's needs with respect to activity tracking, and matching them with what can be measured.

Taking into account these conflicting requirements it seems unlikely that one single device will be able to fulfill them. Not just do preferences vary between users, but also a single user has different needs over time and even throughout the day. An approach could be to understand an activity tracker not as one single and monolithic system, but a combination of multiple sensing and feedback devices of varying precision, interactivity, and obtrusiveness. For long term monitoring the user might choose a very small sensor with no feedback at all, whereas for a specific health behavior change intervention more precision and a rich interaction would make a more obtrusive system acceptable for a limited period of time. The user would also have the choice to change devices throughout the day, using an invisible one when working, and wearing a sophisticated wristband in the evening. This approach calls for an ecosystem of interconnected low-cost devices. Specific parts of today's devices may in the future be integrated into smart clothes. In the long term, even implanted devices that are injected under the skin or integrated into a tooth would be thinkable.

7 Conclusion

In a qualitative field study we identified four concepts as relevant for the design of activity trackers: perception and appearance, wearability, intensity of interaction, and activity measures and validity of data. Of these we derived four concrete design guidelines, where the first three are closely related to HCI, and the fourth is a data modeling challenge. Our guidelines extend existing work such as [2] that identifies general requirements for lifelogging devices, with usability being just one of multiple aspects. While some points identified there relate well to our results, our focus on usability allows us to be much more concrete and derive concrete recommendations.

Our results are based on feedback from 12 participants. We believe that this sample size is large enough for sound qualitative results. However we obtained the results in an unsupervised diary-based approach; therefore we had to rely on the participants' reports and were unable to double-check the results.

We conclude that "the universal" all-purpose activity tracking device does not exist. Rather there will be a need for different designs, concepts, and data models, depending on the personal preferences of the user and the intended use, taking into account our design implications. While initially our results relate to activity trackers only, most of the rationale behind the guidelines e.g. on appearance, wearability, or robustness apply to general lifelogging devices, too. Therefore an important next step would be to examine the usability of other lifelogging devices such as cameras in order to validate and possibly adapt the results. In future work the guidelines could be applied in the design process of a new lifelogging device to prove their validity.

References

1. Brooke, J.: SUS - A quick and dirty usability scale. In: Jordan, P.W., Thomas, B., McClelland, I.L., Weerdmeester, B. (eds.) Usability Evaluation in Industry. CRC Press (1996)
2. Byrne, D., Kelly, L., Jones, G.J.F.: Multiple multimodal mobile devices: Lessons learned from engineering lifelog solutions. In: Handbook of Research on Mobile Software Engineering: Design, Implementation and Emergent Applications (2012)
3. De Cocker, K.A., De Meyer, J., De Bourdeaudhuij, I.M., Cardon, G.M.: 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 (2012)
4. Consolvo, S., Everitt, K., Smith, I., Landay, J.A.: Design requirements for technologies that encourage physical activity. In: Proc. of CHI 2006, pp. 457–466 (2006)
5. Consolvo, S., McDonald, D.W., Toscos, T., Chen, M.Y., Froehlich, J., Harrison, B., Klasnja, P., LaMarca, A., LeGrand, L., Libby, R., Smith, I., Landay, J.A.: Activity sensing in the wild: A field trial of ubifit garden. In: Proc. of CHI 2008, pp. 1797–1806. ACM (2008)
6. Fausset, C.B., Mitzner, T.L., Price, C.E., Jones, B.D., Fain, B.W., Rogers, W.A.: Older adults use of and attitudes toward activity monitoring technologies. Proc. of the Human Factors and Ergonomics Society Annual Meeting 57(1), 1683–1687 (2013)
7. Gemmell, J., Bell, G., Lueder, R., Drucker, S., Wong, C.: Mylifebits: Fulfilling the memex vision. In: Proceedings of the Tenth ACM International Conference on Multimedia, MULTIMEDIA 2002, pp. 235–238. ACM, New York (2002)
8. Giehoff, C., Wallbaum, T., Rahn, A., Rolf, C., Boerner, R.J., Neitzel, L., Link, I.: Support of dementia patients by home tele monitoring (Unterstützung von Demenz-Betroffenen durch Home-Tele-Monitoring). In: 58. Annual Symposium of the German Society for Medical Informatics, Biometrics and Epidemiology (58. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS) e.V.) (2013)
9. Hodges, S., Williams, L., Berry, E., Izadi, S., Srinivasan, J., Butler, A., Smyth, G., Kapur, N., Wood, K.: Sensecam: A retrospective memory aid. In: Dourish, P., Friday, A. (eds.) UbiComp 2006. LNCS, vol. 4206, pp. 177–193. Springer, Heidelberg (2006)
10. Joki, A., Burke, J., Estrin, D.: Compaignr: A Framework for Participatory Data Collection on Mobile Phones. Technical report, Centre for Embedded Network Sensing, University of California, Los Angeles (2007)
11. Khan, D.U., Ananthanarayan, S., Le, A.T., Siek, K.A.: Evaluation of Physical Activity Monitoring Applications for Android. In: UbiComp 2012 Workshop on Evaluating Off- the-Shelf Technologies for Personal Health Monitoring (2012)
12. Knight, J.F., Baber, C., Schwirtz, A., Bristow, H.W.: The Comfort Assessment of Wearable Computers. In: Proc. of ISWC 2002 (2002)
13. Li, I., Dey, A.K., Forlizzi, J.: Understanding my data, myself: Supporting self-reflection with ubicomp technologies. In: Proc. of UbiComp 2011, pp. 405–414 (2011)
14. Meyer, J., Hein, A.: Live Long and Prosper: Potentials of Low-Cost Consumer Devices for the Prevention of Cardiovascular Diseases. J. Med. Internet Res. 2(2) (2013)

15. Philllips, L.J., Markis, N.: Using the Fitbit Motion Tracker to measure physical activity in residential care/assisted living residents. In: State of the Science Congress on Nursing Research (2012)
16. Ryoo, D.-W., Bae, C.: Design of The Wearable Gadgets for Life-Log Services based on UTC. IEEE Transactions on Consumer Electronics 53(4), 1477–1482 (2007)
17. Schmidt, M.D., Blizzard, L.C., Venn, A.J., Cochrane, J.A., Dwyer, T.: Practical considerations when using pedometers to assess physical activity in population studies. Research Quarterly for Exercise and Sport 78(3), 162–170 (2007)
18. Wilcox, L., Feiner, S.: Evaluating Physical Activity Monitors for Use in Personal Health Research. In: UbiComp 2012 Workshop on Evaluating Off- the-Shelf Technologies for Personal Health Monitoring (2012)