WHICH APP TO CHOOSE? AN ONLINE TOOL THAT SUPPORTS THE DECISION-MAKING PROCESS OF RECREATIONAL RUNNERS TO CHOOSE AN APP

Mark Janssen[1,2], Carine Lallemand[1], Kevin Hoes[2] and Steven Vos[1,2]

[1]Eindhoven University of Technology, Department of Industrial Design, Eindhoven, NL
[2]Fontys University of Applied Sciences, School of Sport Studies, Eindhoven, NL

Abstract

In recent years, there has been an exponential increase in the use of health and sports-related smartphone applications (apps). This is also reflected in App-stores, which are stacked with thousands of health- and sports-apps, with new apps launched each day. These apps have great potential to monitor and support people's physical activity and health. For users, however, it is difficult to know which app suits their needs. In this paper, we present an online tool that supports the decision-making process for choosing an appropriate app. We constructed and validated a screening instrument to assess app content quality, together with the assessment of users' needs. Both served as input for building the tool through various iterations with prototypes and user tests. This resulted in an online tool which relies on app content quality scores to match the users' needs with apps that score high in the screening instrument on those particular needs. Users can add new apps to the database via the screening instrument, making the tool self-supportive and future proof. A feedback loop allows users to give feedback on the recommended app and how well it meets their needs. This feedback is added to the database and used in future filtering and recommendations. The principles used can be applied to other areas of sports, physical activity and health to help users to select an app that suits their needs. Potentially increasing the long-term use of apps to monitor and to support physical activity and health.

Keywords: decision tool, sports apps, running, matching, multi-disciplinary

Introduction

Recent years have witnessed an exponential increase in the availability and use of sports-related apps (Janssen et al. 2017). This low-cost, mainstream technology to monitor sports performance is embedded in people's daily life. Especially among runners, research shows that about 50-75% of (event) runners use a running-related app (Janssen et al. 2017; Dallinga et al. 2015).

App-stores are stacked with thousands of sports-, fitness- and health-apps, with new apps launched every day. This comes with a significant challenge for users. There is an overload of available apps (Zhang, Zhang, and Halstead-Nussloch 2014), making it hard for users to decide which apps from the existing large inventory meet their needs, leading to frustration or doubts during the decision process and sometimes even resulting in not choosing any app at all (van Velsen, Beaujean, and van Gemert-Pijnen 2013). Thereby, it is also hard for users - often even impossible - to assess the qualities and limitations of an app before downloading it. The existing star-ratings in app stores can give users an idea of the quality of apps (Google 2016). However, assessing the app based on the number of stars-based user reviews can be unreliable (BinDhim and Trevena 2015). Acknowledging this problem, scholars developed instruments to review the quality of app content (e.g. Stoyanov et al. 2015). Yet, these tools are mostly domain-specific and are limited in scope. We present an online tool that supports the decision-making process to choose an app based on its content quality. We will describe the development of this tool and give insight into its three design principles. Which are (1) app content quality scores are matched to the users' needs with apps that score high in the screening instrument on those particular needs, (2) users can add new apps to the database via the screening instrument, making the tool self-supportive and future proof and (3) a feedback loop allows users to give feedback on the recommended app and how well it meets their needs. For this study, we focused on runners and running apps.

Development of the tool

We constructed and validated a screening instrument to assess app content quality. In parallel, we investigated the features runners need or wish in an app. Results of both methods served as the input for the built of the tool through various iterations with prototypes and user tests.

Construction and validation of the screening instrument to assess app quality

App-stores descriptions and ratings do not provide enough information to select the app that matches a user's need. To address this limitation, we relied on the construction of a screening instrument, developed to assess the qualities of apps from a multidisciplinary perspective. We combined a literature review and expert evaluations to gain insight into the qualities of apps. Then, we constructed an app quality screening instrument that was validated by researchers and end-users.

Developing an app quality screening instrument

To come up with features that are important to address app quality, we conducted a literature search. Combinations of search terms in different databases were used to identify relevant articles based on the content of the abstract and discussion section. The selected articles were used to construct a list of features related to app quality. In our case, aspects from existing screening instruments and taxonomies (e.g Stoyanov et al. 2015) were used together with empirical evidence from (1) health and behavioural science literature, such as exercise guidelines and behaviour change techniques (e.g. taxonomies of Behaviour Change (Abraham and Michie

2008), and insights from (2) design research literature, such as user-experience (e.g. Olla and Shimskey 2015), and (3) persuasive design (e.g. Fogg 2009). Our literature search resulted in just over a hundred features that were important to address app quality. Next to the literature review, eleven experts in the fields of Industrial Design and Engineering, Computer Science, Human Movement Science and Behavioural Sciences participated in expert panels. For full details of this study see (Dallinga et al. 2018). Results of the literature review and the expert panels were combined to construct the Sports App Screening Tool (SAST), encompassing 16 constructs (e.g. goal setting, monitoring, user experience), with a total 64 items scored on a 3-point-Likert scale.

Validation SAST

The SAST was tested and validated by researchers and end-users. First, five researchers (not study-related) independently and blindly screened the ten most downloaded running apps with the SAST. Interrater reliability was measured with Cohen's κ and was found to be sufficient ($\alpha \geq 0.669$). Second, the validity of the items together with the scores on the items was discussed in a group session with all five researchers. Small adjustments were made to the items, mostly of linguistic nature.

Third, we conducted a user study to determine the applicability of SAST. We used Participatory Action Research with 15 end-users (i.e. recreational runners). These participants used three apps with the highest scores in SAST (Nike + Run Club, Runkeeper and Strava) for three weeks. Via a questionnaire, the applicability of SAST for the selected runners was assessed. According to the runners, the items in the SAST were clear. However, the ease of use of SAST was dependent on the user's level of experience with apps. Less experienced or unexperienced app-users had more trouble using SAST. Again small, mostly linguistic, adjustments were made to the items.

Getting insight into users' needs

Parallel to the development of SAST, the user needs were investigated. Fifteen runners (the same sample who participated in the Participatory Action Research) filled in an open-ended questionnaire. Questions inquired about the feature's runners need or wish. Participants indicated that the user-friendliness of an app is the most important criteria: "often app builders try to fill up the app with as many features as possible, while it turns out that no matter how complete an app is, it will not be used when the usability is low" (participant 12). The survey also showed that (1) being able to monitor progress, (2) comparing current data with previous data and setting goals, (3) getting rewards, (4) getting feedback, and (5) sharing data with others are the functions mentioned as 'needed' by the runners.

We used this information (the screening of the ten most popular running apps and the obtained insight into users' needs) to build a prototype of the app decision tool.

From first prototype to stable release

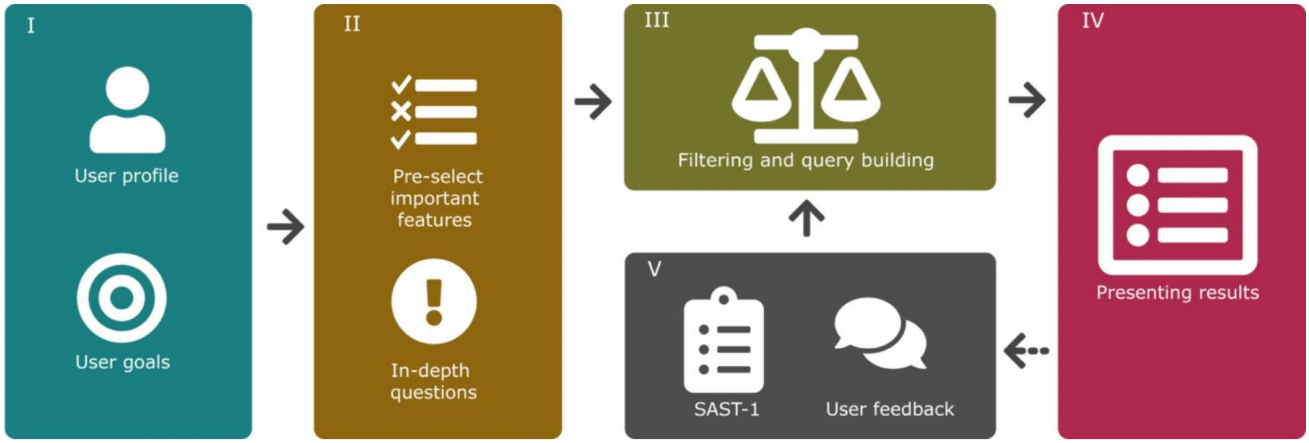

Figure 1: Primary building blocks of the decision tool: (i) user profile and goals, (ii) selection of important features, (iii) filtering, (iv) presenting results, and (v) feedback loop and adding new apps

A simple online platform with screen mock-ups was built (see figure 1): (1) user profile and user goals, (2) selection of important features, (3) filtering and matching, and (4) returning the results. A fifth building block was added subsequently and will be discussed later in the paper.

It was important to make the prototype tangible to get concrete feedback from users. Simultaneously with this prototyping phase, we conducted several sessions with end-users. We showed them the first prototype and asked them to talk us through everything that came up their mind. This think-aloud method (Someren, Barnard, and Sandberg 1994) not only gave us insight into the prototype itself but also into users' cognitive processes during the selection of an app.

We further developed the content, for example by reformulating the questions in the first building block. The question 'What are you looking for' was added (see figure 2, screen 1). Only if the answer turned out to be that the user was looking for an app with more functionalities than the current app, the step to fill in the user goals was skipped, the user is directly forwarded to 'selection of important features.' For the second building block, the functional requirements were initially derived from the survey on user needs. The user tests revealed additional categories such as 'music', 'giving general information', etc. The think-aloud thus also revealed that we should help users in their decision process by pre-selecting features based on their user goals. For example, if a runner chooses that (s)he is physically active for social reasons and to become fitter, in the third screen the functions: setting goals, work with training schedules and sharing results would be pre-selected (see figure 2, screen 3).

Besides information about which features are important according to the users, information was also needed about how advanced the feature should be. For this, we developed so-called, in-depth questions that provide more detail about an important function (see figure 2, screen 4). The in-depth questions corresponded to the items in SAST. For example, if a runner selects 'I can set goals', this corresponds to one of the 16 constructs of SAST, namely Goal Setting. Because this construct consists of six items, six matching in-depth questions are asked, for example 'set individual goals myself'. This connection between the SAST and the in-depth questions allows us to directly match user needs with the scores of the apps on SAST. This direct match between the scores on the SAST and the question asked in the tool is one of the three design principles.

At this moment in the design process, additional iterations with end-users were conducted. We decided to publicly release the decision tool to collect in-situ data on how it is used and what choices are made by users. Simultaneously with the release of the system, we continued iterating and developed a fifth building block.

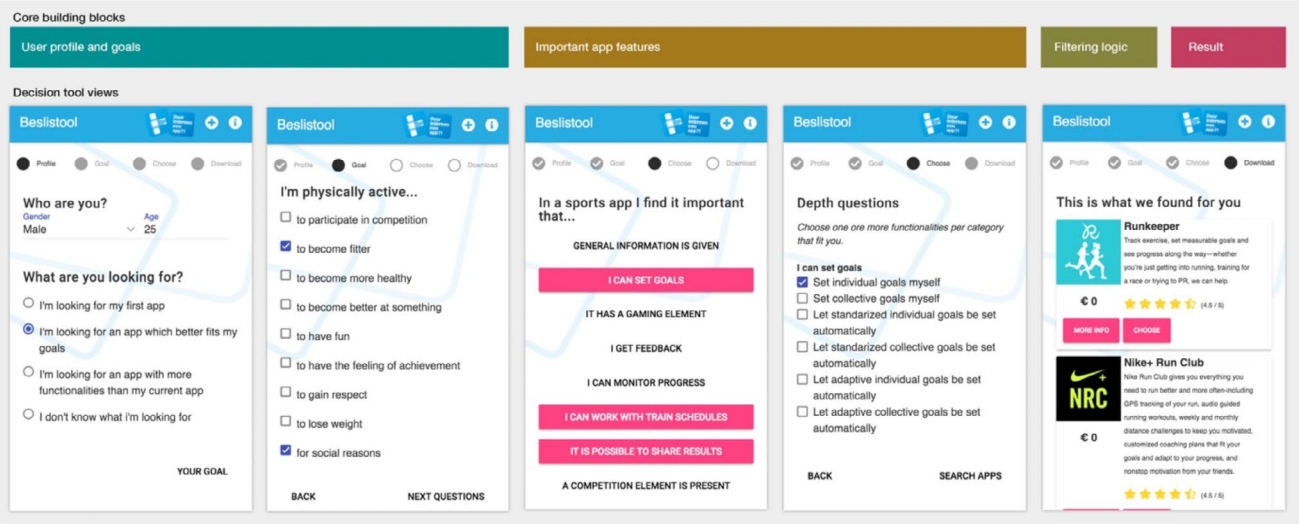

Figure 2: The sequence of screens of the decision tool, including an outline of the core building blocks at the top bar. In the first screen (left) the user profile is determined. The second screen enquires about the user goals, or 'why they want to be more physically active'. These two screens together form the first building block. Next, the user selects the features they find important. A pre-selection is made to help the user. Then they fill out the in-depth questions on the features they deem important. After the fourth screen, the decision tool filters and match apps that fulfil the runners' needs. The matching apps are presented on the fifth screen and the user can download the app of their choice.

Making the tool future-proof

To deal with the rapidly growing world of (smartphone) apps, we attempted to make the tool future proof by adding two principles (i.e. design principle 2 and 3): a screening function for end-users and a feedback loop. The screening principle for users allows new apps to be screened by users and added to the database. SAST is used to act as a neutral entry point to add apps to the database (see figure 1, block V). The final principle is the feedback loop, which invites end-users to give feedback on the decision process (only if they are willing to contribute to the validation process). After three weeks, the end-user receives an e-mail to rate the experience with the app. This user feedback is used to improve the filtering of the apps in the database in relation to the in-depth questions, important functions and goals.

Discussion and Conclusion

In this paper, we deployed a methodology to develop an online tool that supports the decision-making process to choose a smartphone application. We illustrated the method through a running-apps case study. Our approach included the development of a screening tool, the assessment of user needs and iterative prototyping based on user testing, which resulted in a decision-making tool relying on three design principles. First, the direct match between the scores on the SAST and the question asked to users in the tool. Second, a screening principle where users can add new apps to the database through the SAST, to make the tool self-supportive and future proof. Third, a feedback loop allows users who followed the tool's recommendation to give feedback on the recommended app. This feedback is added to the database and therefore directly used in future filtering and recommendations. Besides these principles, the multidisciplinary approach is an

essential asset in this context. This approach supports some practicalities like working on different aspects simultaneously, i.e. development of the screening tool was done simultaneously with the research on user needs and the iterations on the prototype were done parallel to the end-user testing. More importantly, we experienced that multidisciplinarity was required in several stages. For instance, the development and validation of the SAST where expertise of different disciplines were combined. But also, during the built of the first prototype where all disciplines were present to integrate the different expertise from the beginning of the process, making decisions that work for all the disciplines. Therefore, we managed to integrate all disciplines. Thereby, we argue that multidisciplinarity is not only required in this particular study, but that it applies to the whole domain of sports, health and design. Where approaches from several perspectives are necessary to design meaningful tools, services and practices.

Limitations and future work

First, we adjusted the screening tool to the needs of the end-users. In the future, we should consider making different versions of the screening tool, one that applies to end-user, but also a more extended version that could be used by researchers or experts. This could provide the decision tool with more detailed information. Second, we recommend fellow designers who want to follow our approach to integrate the feedback loop immediately in the first prototype. Adding it subsequently led to a limitation that the feedback loop was not user-tested. Finally, we used the same sample of runners in different steps. We would recommend broadening the view and feedback by recruiting new runners for each step. Finally, our three design principles should be applied to other contexts in the future, for instance, a different category of health apps or with different expertise within the teams.

Conclusion

The overall approach presented in this paper, as well as the underlying design principles, can be applied to other areas within sports, physical activity and health to help users to select an app that matches their respective needs. The overarching goal is to eventually increase more diversity and long-term use of apps to monitor and support physical activity and health.

Acknowledgements

This work was supported by SIA RAAK (Grant 2015-02-32P) and by Netherlands Organization for Scientific Research (Grant 023.007.056). Font Awesome icons are used (https://fontawesome.com/license)

References

Abraham, Charles, and Susan Michie. 2008. "A Taxonomy of Behavior Change Techniques Used in Interventions." *Health Psychology* 27 (3): 379–87. https://doi.org/10.1037/0278-6133.27.3.379.

BinDhim, Nasser, and Lyndal Trevena. 2015. "There's an App for That: A Guide for Healthcare Practitioners and Researchers on Smartphone Technology." *Online Journal of Public Health Informatics* 7 (2): 1–12. https://doi.org/10.5210/ojphi.v7i2.5522.

Dallinga, Joan, Mark Janssen, Jet Van Der Werf, Ruben Walravens, Steven Vos, and Marije Deutekom. 2018. "Analysis of the Features Important for the Effectiveness of Physical Activity–Related Apps for Recreational Sports: Expert Panel Approach." *Journal of Medical Internet Research* 20 (6). https://doi.org/10.2196/mhealth.9459.

Dallinga, Joan, Matthijs Mennes, Laurence Alpay, Harmen Bijwaard, and Marije Baart de la Faille-Deutekom. 2015. "App Use, Physical Activity and Healthy Lifestyle: A Cross Sectional Study." *BMC Public Health* 15: 833. https://doi.org/10.1186/s12889-015-2165-8.

Fogg, B J. 2009. "A Behavior Model for Persuasive Design." In *Persuasive '09 Proceedings of the 4th International Conference on Persuasive Technology*,. https://doi.org/10.1145/1541948.1541999.

Google. 2016. “How People Discover, Use, and Stay Engaged with Apps.” https://www.google.com/thinkwithgoogle.com/documents/how-users-discover-use-apps-google-research.pdf.

Janssen, Mark, Jeroen Scheerder, Erik Thibaut, Aarnout Brombacher, and Steven Vos. 2017. “Who Uses Running Apps and Sports Watches? Determinants and Consumer Profiles of Event Runners’ Usage of Running-Related Smartphone Applications and Sports Watches.” *PLOS ONE* 12 (7): e0181167. https://doi.org/10.1371/journal.pone.0181167.

Olla, Phillip, and Caley Shimskey. 2015. “MHealth Taxonomy : A Literature Survey of Mobile Health Applications,” no. 2014: 299–308. https://doi.org/10.1007/s12553-014-0093-8.

Someren, M., Y. Barnard, and J. Sandberg. 1994. *The Think Aloud Method; A Practical Guide to Modelling Cognitive Processes*. London: Academia Press.

Stoyanov, Stoyan, Leanne Hides, David Kavanagh, Oksana Zelenko, Dian Tjondronegoro, and Madhavan Mani. 2015. “Mobile App Rating Scale: A New Tool for Assessing the Quality of Health Mobile Apps.” *JMIR MHealth and UHealth* 3 (1): e27. https://doi.org/10.2196/mhealth.3422.

Velsen, Lex van, Desirée Beaujean, and Julia van Gemert-Pijnen. 2013. “Why Mobile Health App Overload Drives Us Crazy, and How to Restore the Sanity.” *BMC Medical Informatics and Decision Making* 13 (1): 23. https://doi.org/10.1186/1472-6947-13-23.

Vos, Steven, Mark Janssen, Jos Goudsmit, Coen Lauwerijssen, and Aarnout Brombacher. 2016. “From Problem to Solution: A Three-Step Approach to Design a Personalized Smartphone Application for Recreational Runners.” *Procedia Engineering* ISEA 00: 1–7.

Zhang, Chi, Xihui Zhang, and Richard Halstead-Nussloch. 2014. “Assessment Metrics , Challenges and Strategies for Mobile.” I*ssues in Information Systems* 15 (IIi): 59–66. http://iacis.org/iis/2014/88_iis_2014_59-66.pdf.