University of Twente

Faculty of Behavioural, Management & Social Science

Department Psychology, Health & Technology

A subjective evaluation of the fitness application RunKeeper® using the Perceived Persuasiveness Questionnaire (PPQ)

Bachelor Thesis, 2019

By

Philipp Sebastian Kill

Supervisor: Dr. N. Köhle

Second Supervisor: Dr. N. Beerlage-de Jong

Abstract

Background: The usage of persuasive eHealth technologies has been shown to be an effective way of tackling health problems in various areas. However, to develop efficient eHealth technologies the end user must be addressed as well. Not involving end users might result in a technology that is not tailored towards the needs and skills of that user and does not pay attention to user characteristics. For this reason, the Perceived Persuasiveness Questionnaire (PPQ) was developed based on the Persuasive Systems Design Model (PSD). It is a tool which measures how well users of an eHealth technology perceive its persuasive elements. However, the questionnaire itself has not been adequately tested for reliability and validity yet and thus must be used with caution.

Aim: The aim of this study is to find out more about the reliability of the constructs of the Perceived Persuasiveness Questionnaire. Additionally, it will be tested whether the user characteristic of physical activity has an effect on later use continuance of the eHealth application RunKeeper®.

Methods: This study uses a cross-sectional quantitative approach by making use of questions about demographics, physical activity, as well as the Perceived Persuasiveness Questionnaire (PPQ). Using convenience- and snowball sampling a total sample of 35 participants was recruited. Based on their answers, a reliability analysis for the construct of the PPQ was conducted and it was investigated whether a difference in use continuance could be observed between individuals with low/moderate physical activity and individuals with high physical activity.

Results: The results of the reliability analysis for the PPQ showed that internal consistency was low, since four out of nine constructs failed to meet acceptance criteria ($\alpha < 0.7$). Furthermore, no difference could be observed between low/moderate physically active users and high physically active ones ($p = .137$).

Conclusion: The development of an evaluative tool for eHealth technologies is imperative. This study has shown that the PPQ could be used as such tool, but its constructs lack internal consistency. These will need to be revised, for example by combining them or changing certain items. Additionally, the fact that physical activity had no impact on later use continuance, designers of future eHealth technologies must not pay attention to this particular user characteristic. However, it is mandatory that more characteristics will be investigated to optimally tailor future interventions.

Table of contents

Introduction

Nowadays, technology is a constant part of daily life, in virtually all domains ranging from entertainment purposes to health assistance. In the health care sector, eHealth technologies have become highly influential. These include technologies that track, monitor, and inform the user; facilitate communicative encounters between health stakeholders; and in general are used to improve personal health and health services (Shaw et al., 2017). These technologies are in demand because the worldwide numbers of non-communicable diseases are projected to rise (Mathers & Loncar, 2005) but could be largely prevented through the use of eHealth technologies (WHO, 2013). According to WHO (2013), around 75% of heart disease, stroke, and type II diabetes, as well as roughly 40% of cancer cases could be prevented with a healthy behaviour.

Studies on eHealth have shown that it can be used to facilitate such healthy behaviour by tackling issues like obesity (Cadmus-Bertram et al., 2013; Hutchesson et al., 2013), alcohol intake (Hester, Delaney, & Campbell, 2012; Simmons, Heckman, Fink, Small, & Brandon, 2013), low physical activity (Schwerdtfeger, Schmitz, & Warken, 2012; Sriramatr, Berry, & Spence, 2014), cancer-related fatigue (Seiler, Klaas, Tröster, & Fagundes, 2017), and mental health problems (Stratton et al., 2017). eHealth is not limited to the clinical environment but can also be used at home through means of personal computers, laptops, tablets, the internet, wearables, or smart phones (Cunningham, Wake, Waller, & Morris, 2014). One characteristic of eHealth is that it is often personalized towards the user and delivers tailored information, which yields better outcome results (Ijsselsteijn, de Kort, Midden, Eggen, & van den Hoven, 2006; Kaptein, Markopoulos, de Ruyter, & Aarts, 2015). Next to the benefits of the actual treatment, several studies have revealed that eHealth interventions are more cost-effective than traditional health care (Elbert et al., 2014; van Keulen et al., 2010). This is promising, as the demand for outcome- and cost-effective eHealth interventions increases (Murray, 2014). One possibility to meet that demand comes in the form of smartphone applications.

Smartphones already offer a variety of apps aimed at improving health in various ways (Edwards et al., 2016). However, one major problem in the context of eHealth is the fact that users often struggle with use continuance, meaning that they fail to use their eHealth technology (e.g. an app) for a long period of time (Bol, Helberger, & Weert, 2018; Kelders, Kok, & Gemert-Pijnen, 2011). Lacking use continuance is a problem for eHealth because better results can be achieved when people engage in an intervention for a longer period of time (Ryan, Patrick, Deci, & Williams, 2008). There are two keys aspects to consider in the design process of an

eHealth technology if it is supposed to engage the user over a longer period of time. The first aspect includes the characteristics of the potential user, because factors such as age, gender, and education were already determined to have an influence on use and use continuance (Reinwand, Schulz, Crutzen, Kremers, & de Vries, 2015).

Looking at eHealth technologies, multiple factors play a role for determining its later effectiveness. As already mentioned, use continuance is important for eHealth technologies to achieve their full potential and achieve good behaviour change. In addition, one important factor that needs to be considered when looking at this are the user characteristics of eHealth technologies. Campbell, MacAuley, McCrum, and Evans (2001) have shown in their paper that age has a great influence on the use of an eHealth technology. In their study, older adults showed different motivating factors for exercise and this highlights that a "one-fits-all" approach is not the right solution. Additionally, Reinwand et al. (2015) have shown that females use eHealth interventions more than males. They have also found out differences for user who are ill and for people who are in a relationship. Next to these influencing factors, level of education, quality of life, or income did not have any influence on the use of eHealth (Reinwand et al., 2015). One further factor that influences the use of eHealth application for health promotion is work-related stress (Bregenzer, Wagner-Hartl, & Jiménez, 2017). Individuals with more stress have been shown to seek health promotion applications more than people with less stress. One further factor that has been shown to influence continued use of an eHealth technology is whether individuals smoke or do not smoke. The results of a study done by van Keulen et al. (2010) indicate that non-smokers use eHealth technology more than smokers. Along with the previously mentioned studies these characteristics show that eHealth interventions can not be universally used for every type of user. Often times, interventions reach the users more effectively that are not the main target group (van Keulen et al., 2010). This is a problem because it means that the people suffering from a health problem, e.g. smoking or low physical activity, do not seek the interventions that do already exist. However, knowledge about the users alone is not sufficient and eHealth technologies have to be designed in a specific way to act as persuasive technology that changes the user´s attitude or behaviour (Oinas-Kukkonen & Harjumaa, 2009).

In this context, Fogg (1999) describes persuasion as "an attempt to shape, reinforce, or change behaviors, feelings, or thoughts about an issue, object, or action". For most cases, persuasive technologies are built up on Behaviour Change Support Systems (BCSS) (Oinas-Kukkonen, 2013). Oinas-Kukkonen (2013) states that "A behavior change support system (BCSS) is a sociotechnical information system with psychological and behavioral outcomes

designed to form, alter or reinforce attitudes, behaviors or an act of complying without using coercion or deception.". This basically means that any designed system, e.g. a health promoting app, uses some form of persuasion to achieve its goal to promote healthy behaviour, e.g. more physical activity. These systems are usually designed based on underlying theoretical models.

The Persuasive Systems Design model (PSD)

One such model to design a BCSS is the PSD model, which can primarily be used as a framework for the design and development process of persuasive systems, e.g. health promoting apps (Oinas-Kukkonen & Harjumaa, 2009). In their model Oinas-Kukkonen and Harjumaa (2009) describe that it is important to include certain qualities in a persuasive technology to make it effective.

These qualities are defined into the four categories: Primary Task Support, Dialogue Support, Credibility Support, and Social Support (Oinas-Kukkonen & Harjumaa, 2009). *Primary Task Support* aids the user by breaking down complex behaviour into simple steps, tailoring and personalizing the presented information, providing performance monitoring, and enabling behaviour rehearsal. All these principles increase the persuasive abilities of a system and help the users to reach their primary goal. Kelders, Kok, Ossebaard, and Van Gemert-Pijnen (2012) found significant differences in aspects such as intended usage, frequency of interaction with the system, or duration of usage when looking at the number of primary task support elements. This shows the importance of primary task support elements in the design of a health application.

Dialogue Support describes design principles that deal with computer-human dialogues, and includes feedback in terms of praise, reminders, rewards, or suggestions. Dialogue support encourages the user to carry out the primary task and is crucial for increasing the user´s positive affect towards the technology (Lehto, Oinas-Kukkonen, & Drozd, 2012). It includes features such as argumentation, prompts, and feedback given to the user at an appropriate frequency level.

Credibility Support includes principles that increase the technology´s credibility which in turn will lead to an increase in persuasion. These principles include the degree of expertise presented, the level of trustworthiness, surface-credibility, and verifiability. Sillence, Briggs, Harris, and Fishwick (2006) found that perceived credibility could positively predict the selection of health advice sites. This shows that a high credibility is necessary for an effective

health technology to positively influence future users. The last category deals with the amount of social support included in a persuasive technology.

Social Support increase the persuasiveness of a technology by giving the users the opportunity to compare themselves to others, learn, compare, or compete with peers, or by receiving public recognition for individual achievements (Oinas-Kukkonen & Harjumaa, 2009). Social support incorporates aspects such as social learning and social comparison which increase the user´s motivation to use the system (Oinas-Kukkonen & Harjumaa, 2009; Orji, 2017).

Despite the fact that the PSD offers guidelines for the design process of persuasive technologies, the need for evaluative tools for eHealth technologies is still existent (Oinas-Kukkonen, 2013). Greenhalgh and Russell (2010) have shown that while some tools for eHealth evaluation exist, these often fail to deliver the insights expected by them. This means that currently it is difficult for designers, or any other party for that matter, to evaluate how well the eHealth technology works. However, improvement in eHealth is only possible when one has sufficient knowledge about which aspects of it work and which do not work as intended. One such tool that could possibly be used to effectively evaluate eHealth persuasive technologies is the Perceived Persuasiveness Questionnaire developed by Lehto et al. (2012).

The Perceived Persuasiveness Questionnaire (PPQ)

The main aim of the PPQ is the evaluation of how well persuasive elements can be recognized in an eHealth technology. Participants are asked to indicate their subjective feeling about the tested eHealth technology, which then can be used to evaluate how these participants perceive the persuasive elements. Many existing evaluation tools focus on the most effective design during the development of a product/application. However, it is necessary to also evaluate a technology after it has been finished and is released. Catwell and Sheikh (2009) stress that it is important to continuously evaluate any eHealth technology as long as it is being used. This stresses the importance of the PPQ as it can be used to evaluate the effectiveness of a finished product, such as a mobile app, a website, or other health related technology.

The PPQ consists out of nine constructs, which combined can give an indication of how well the persuasiveness of the technology is perceived. Each construct is measured using a Likert-scale ranging from "strongly disagree" to "strongly agree". Since the PPQ has not been validated yet, there are no standards for scoring. Currently, the most feasible way of scoring is

to calculate the mean for a construct using it´s corresponding items. For the purpose of the current study the PPQ has been adapted so that it´s items are in line with the evaluation of the mobile application RunKeeper®(ASICS Digital, 2019).

The first construct of the PPQ is called "*Primary Task Support*" which is also the central construct of the PSD model. In the PPQ, "*Primary Task Support*" is measured with three items, for example item 23: "*Runkeeper application helps me change my exercising habits*" which are all aimed at the primary goal of the user.

Another construct of the PPQ which plays an important role is "*Perceived Dialogue Support*" as mentioned in the PSD model. In the PPQ it is measured with three items including item 17: "*Runkeeper application provides me with appropriate counselling.*". This construct should not be confused with "*Perceived Social Support*" which refers to the options that a system provides to share information or experiences with peers. One item, in the to this study adapted version of the PPQ which measures Social support, is item 30: "*Learning from my peers' actions is beneficial for me.*". In total, *Perceived Social Support* is measured with three items.

Another construct of the PPQ is "*Perceived Credibility*" which is measured with five items such as item 19: "*Runkeeper application is clearly made by health professionals.*". These four constructs can all be found in the PSD model. However, the PPQ added five different constructs, the first one being "*Unobtrusiveness*".

Unobtrusiveness refers to the applicability of the technology in the daily life of the user. The better the unobtrusiveness of a system the easier it is for the user to implement it and also to use it on a regular basis (Lehto et al., 2012). This is crucial to any health technology as most interventions are not one-time interventions but instead try to change the behaviour of the user over a longer period of time. In order to measure unobtrusiveness, the PPQ uses four items, for example item 14: "*Using Runkeeper application disrupts my daily routines.*". A construct which is similar to unobtrusiveness is the construct of "*Perceived Effort*". Perceived Effort describes the effort which a user must make while using a technology and includes three items. For instance, item 31: "*Using Runkeeper application is difficult*.".

The next construct of the PPQ is called "*Perceived Persuasiveness*" and refers to how well the system manages to persuade the user to start using it, continue using it, and how strong the impact of the system is on the user. A higher perceived persuasion is achieved through attitude change that is elicited if a message is successfully delivered to and processed by the

user. The PPQ measures perceived persuasiveness with three items, for example item 20: "*Runkeeper application has an influence on me.*". Similar to this construct the construct of "*Perceived Effectiveness*" measures how effective the user evaluates the intervention to be. A technology that scores high for this construct manages to successfully achieve behaviour change and is overall effective. In the PPQ this construct is measured with three items in total including item 12: "*In my opinion, using RunKeeper application has an effect on my willingness to exercise.*".

The final construct of the PPQ is "*Use Continuance*". Use continuance refers to the intention of the test subject to continue using a certain health technology. High use continuance is often achieved through a successful combination of the constructs mentioned earlier and is important in the prolonging implementation of a health application. This construct is measured using four items in total, including item 22: "*I am going to continue using Runkeeper application.*".

Due to the fact that the PPQ is still in its development, no clear scoring standards have been set yet. As of now, the most feasible way to compute scores for the constructs is to take an overall mean of all items for one construct respectively. Because the PPQ works with a Likert-scale, the mean scores can have a range from one to five. If done in that way, a high score on all constructs, except for *Perceived Effort*, means that this construct is reflected in a good way. Due to the fact that *Perceived Effort* is a negatively stated construct, a low score will indicate low effort, which is desirable.

This study

Above, it has been shown that models to guide the design process of eHealth technologies exist, for example the PSD model. However, tools are still needed to evaluate the finished technologies and gain insights into user experience. These tools are needed because continuous evaluation of eHealth technologies are necessary and existing tools are scarce (Catwell & Sheikh, 2009). Next to theoretical model like the PSD model, Rabin and Bock (2011) have also shown that users of health promoting application expect certain features. It is therefore necessary to come up with a tool that can measure the effectiveness of an eHealth technology. The PPQ represents such a potential tool but has not been sufficiently tested on reliability and validity yet. Beerlage-de-Jong et al. (2016) have found that the constructs of the PPQ might be revised and have also shown that internal consistency is relatively low. Despite

these first results, the need for an effective evaluative tool is present and therefore the PPQ needs to be investigated further than just one study.

Hence, the main goal of this study is to *test the reliability of the constructs that make up the PPQ and evaluate whether the internal consistency is acceptable*. To do this, a subjective evaluation of the fitness application "RunKeeper®" will be conducted using the Perceived Persuasiveness Questionnaire. A similar evaluation was done by Beerlage-de-Jong et al. (2016) and it showed that five out nine constructs did not reach the threshold to be seen as reliable. Since the PPQ has not been validated and tested extensively apart from that study, it can be hypothesized that "internal consistency among the constructs of the Perceived Persuasiveness Questionnaire will be generally low and not all constructs will meet acceptance criteria".

Due to the fact that eHealth technologies struggle with use continuance, it is important to investigate which factors influence this construct. As already mentioned, different studies have already tried to highlight some user characteristics that influence the continuous use of a technology. It was also introduced that some interventions are most successful for people who need them least (e.g. smoking interventions for non-smokers). Whilst these insights are already valuable, the need for more research on the topic still exists and it is important to find out which characteristics influence what type of eHealth technology. Therefore, the second aim of this study is to examine whether physical activity has an effect on use continuance in the context of using the fitness application RunKeeper®. Since an increase of physical activity is often a goal of eHealth technologies, knowledge about its influence before starting such a technology could help to better tailor future interventions. The research question for this will be formulated as: "Can a difference in use continuance of the RunKeeper® app be observed when comparing individuals who score low/moderate on physical activity versus those who score high?"

Methods

Design

For this study, a quantitative cross-sectional design was applied. A participant sample was provided with a questionnaire as mean to collect data. All participants received the same instructions and went through the same procedure. The study was approved by the Ethical Committee of the University of Twente (request number 190129).

Participants

For the current study participants were recruited using convenience and snowball sampling. Most participants were gathered by asking friends or family, and by asking roommates of those people. After meetings with friends, their peers were contacted via phone and asked whether they had time to participate in the study. Additionally, the internal system "Sona system" by the University of Twente also provided some participants. The Sona System is an internal system by the University of Twente which rewards participants for taking part in studies by granting them a certain amount of credits. Students of the track Psychology and Communication Science have to obtain 15 SONA credits during their Bachelor´s program in order to get a degree. The sampling period lasted about five weeks starting April 2019. Since this study complements the study by Beerlage-de-Jong et al. (2016), German participants were targeted, but nationality was not set as a general inclusion criteria for the current study because it had no influence on the hypotheses/research questions.

In total, the sample consisted of 35 participants *(60% female, 40% male)* who came from different educational backgrounds. The actual sample included 33 participants of German nationality and 2 participants who were raised but not born in Germany. These two participants were also included as nationality was not of concern for this research. The participants were aged 19 to 54 with a mean of 23,20 (*SD=5.61*). More detailed information about participant characteristics can be found in Table 1.

Table 1.

Characteristics of the participants (N=35)

	N	Mean (Min-Max)	Std. Dev.
Nationality			
German	33 (94%)		
Other	2 (6%)		
Gender			
Male	14 (40%)		
Female	21 (60%)		
Age		23.20 (19-54)	5.61
Scientific Background			
Behavioural, Management & Social Science	27 (77%)		
Information & Communication Technology Science	1 (3%)		
Science & Technology	2 (6%)		
Other	5 (14%)		
eHealth familiarity			
Very	6 (17%)		
Somewhat	20 (57%)		
Not at all	9 (26%)		
Level of physical activity			
High	18 (51%)		
Moderate	15 (43%)		
Low	2 (6%)		

Materials

RunKeeper® application

For the main purpose of this study, participants were asked to evaluate the fitness application RunKeeper® using the PPQ. The app was presented on the researcher´s phone which used the to that point latest android version of it (9.9.1 (5900)). This was done firstly because it made the process of the study faster and easier for participants, and secondly to ensure that all participants used the same version and interface of RunKeeper®. The app has multiple features such as setting personal goals, training for a certain event, or sharing personal achievements with friends. After choosing a certain activity, such as running, the app uses GPS to track the user and then gives feedback on performance. Illustrations of the actual interface of RunKeeper® can be found in Figure 1.
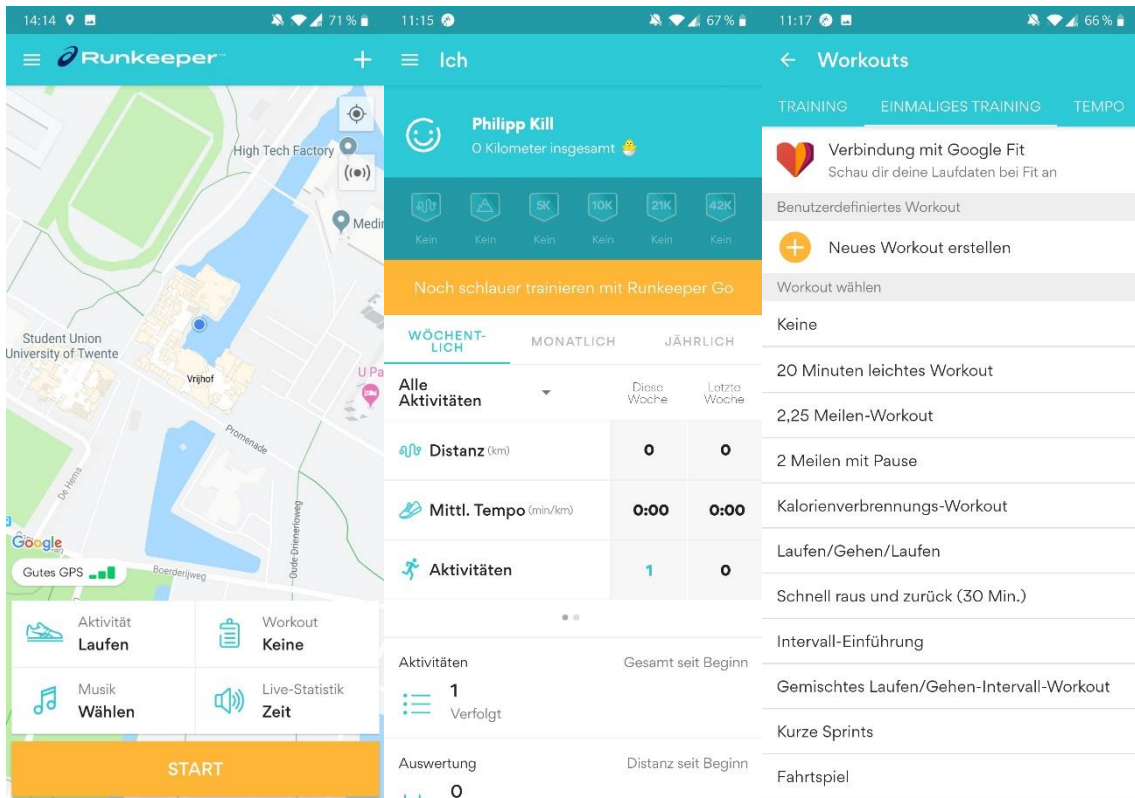
Figure 1. Screenshots of RunKeeper® application.

In addition to the app, participants were given a pen-and-paper version of the PPQ. Since the PPQ was already described in the introduction part, it will not be further discussed in this section. In addition, participants had to answer four demographic questions asking about age, gender, nationality, and familiarity with e-health technology.

Measurement of Physical activity (PA)

In order to obtain information about physical activity of the participants, the short version of the International Physical Activity Questionnaire (IPAQ) was used (IPAQ, 2005). The questionnaire identifies the level of activity an individual engages in, by taking into consideration the past seven days. The IPAQ consists out of seven questions which target activity levels for the four areas of sitting, walking, moderate-activity, and vigorous-activity. Open questions such as "During the last 7 days, on how many days did you do vigorous physical activities like heavy lifting, digging, aerobics, or fast bicycling?" ask for an indication in either hours and/or minutes for the specific area of activity. The questionnaire measures intensity using MET-values, that is the ratio of the rate of energy expended during an activity to the rate of energy expended at rest. For example, a 4-MET activity expends four times more energy

10

than the body at rest. In the case of the IPAQ, vigorous-intensity activity is rated as an 8-MET activity, moderate-intensity activity is rated as 4-MET activity, and walking is rated as a 3.3-MET activity. The individual values for each area of PA can be computed by multiplying the indicated minutes by the number of days per week and the MET-factor (3.3, 4, or 8) for each activity. Additionally, one value for Total Physical Activity can be computed by adding up the three MET-values for walking, moderate-intensity activity, and vigorous-intensity activity respectively. The four above mentioned values can then be used to categorize each participant in one of three categories of physical activity (low, moderate, and high). Multiple factors have to be taken into account before placing a participant in one of the categories. The criteria for each category can be found in Table 2. Craig et al. (2003) have found that the IPAQ has good reliability and validity which is why it will be used as measure for physical activity.

Table 2.

Physical activity categories of the IPAQ

	Criteria
Low	Not meeting any criteria for either of the categories "Moderate" or "High"
Moderate	a) 3 or more days of vigorous-intensity activity of at least 20 minutes per day b) 5 or more days of moderate-intensity activity and/or walking of at least 30 minutes per day c) 5 or more days of any combination of walking, moderate-intensity or vigorous intensity activities achieving a minimum Total PA of at least 600 MET-minutes/week.
High	a) vigorous-intensity activity on at least 3 days achieving a minimum Total PA of at least 1500 MET-minutes/week b) 7 or more days of any combination of walking, moderate-intensity or vigorous-intensity activities achieving a minimum Total PA of at least 3000 MET-minutes/week.

Procedure

At the start of the study, all participants were given a short introduction about the purpose of the study using a PowerPoint presentation and an information sheet. Here, the purpose of the study and the Runkeeper® application was presented and explained. All participants were asked if they had sufficiently understood what they had to do and that, if they

had any questions or concerns, they could communicate this with the researcher at any time. All participants participated voluntarily and were given a written informed consent prior to the start of the study. They were also informed that they could withdraw from the study at any point without having to give any specific reason.

Following the introduction and information part, the RunKeeper® application was presented on the researcher's phone to ensure equal conditions. Participants were instructed to set the app to the activity "running" and then had to choose a 2,25-mile workout. Afterwards, participants were asked to freely explore the app for a minimum of at least two minutes time. They were instructed to stop exploring the app once they had the feeling to have a sufficient overview of its functionality but not before the two minutes time gap was reached. Next, they were given a pen-and-paper version of the PPQ adapted to the RunKeeper® application, complemented by demographic questions as well as a questionnaire asking about physical activity.

Data analysis

At first the demographics of all participants were analysed using descriptive statistics with the program IBM SPSS (Version 24).

For the main goal of this research, all participants were analysed when looking at the PPQ results. To find out internal consistency and reliability of the constructs of the PPQ, Cronbach´s alpha values were calculated using the answers to the Runkeeper® evaluation. To gain further insight, Cronbach´s alpha values if one item was removed were also calculated. If necessary, findings of interest within the Cronbach´s alpha analysis were further analysed using an inter-item correlational analysis. This was done for all constructs with an alpha value below 0.7.

Since the PPQ is not fully developed yet, it was assumed that its constructs would not show an overall high internal consistency. Thus, the prior work done by Beerlage-de-Jong et al. (2016) will be taken into account for a second reliability analysis. Her study find that the original constructs might not be appropriate and instead has chosen to combine the constructs of "Perceived Primary Task Support", "Perceived Persuasiveness", "Perceived Effectiveness", as well as item 16, originally belonging to the construct "Perceived Credibility". Additionally, item 1 was removed from the original construct "Perceived Effort" and was placed with

"Unobtrusiveness". A second reliability analysis was conducted based on these newly arranged constructs.

The second research question analyses the difference between two groups (low/moderate and high levels of physical activity) and will therefore, use a between-groups approach. To test whether significant differences between levels of physical activity for the two groups can be found, an independent samples t-test was run to check for a significant difference in means ($p < .05$). For this t-test the participants that scored low on physical activity and the participants that scored moderate were merged into one group. This was done because the "Low" category only featured two participants. Thus, the t-test compared the group of participants that were classified as high in physical activity against those who did not fall into that category.

Results

Reliability analysis of PPQ constructs

To test the reliability of each construct, a reliability analysis was conducted. The Cronbach´s alpha values for this analysis can be found in Table 3. Notably, 5 out of the 9 constructs (or 55%) show a high value for internal consistency with alpha values higher than 0.7. The constructs "Perceived Primary Task Support (TASK)", "Perceived Persuasiveness (PERS)", "Unobtrusiveness (UNOB)", "Use Continuance (CONT)", and "Perceived Effectiveness (EFFE)" all show good internal consistency and the constructs PERS and CONT can be scantly improved by removing item 25 and item 7 respectively. The four constructs with an alpha value smaller than 0.7 all remain below that threshold even if one item is removed indicating that their items are not fully consistent or that some items measure a different construct than assumed.

Table 3.

Cronbach´s alpha values for original PPQ constructs

Construct	Cronbach´s Alpha	Items	Mean	Std. Dev.	Cronbach´s alpha if item deleted
TASK	**.805**	15	3.14	1.115	.622
		23	2.94	1.136	.777
		26	3.43	1.195	.793
DIAL	.433	11	3.37	1.031	.177
		17	3.03	.923	.414
		18	3.49	.818	.380
CRED	.575	4	3.59	.857	.466
		10	3.85	.657	.505
		16	3.44	.746	.505
		19	3.15	.925	.644
		27	3.50	.663	.467
PERS	**.799**	5	2.76	1.251	.727
		20	2.79	.992	.630
		25	3.15	1.149	**.819**
UNOB	**.851**	1	3.40	1.063	.840
		6	2.91	1.121	.750
		14	3.57	.917	.812
		24	3.31	1.157	.835
CONT	**.920**	7	2.71	1.226	**.922**
		8	3.00	.970	.898
		13	2.66	1.187	.874
		22	2.63	1.165	.890
EFFO	.343	2	3.77	.973	**-.254**
		9	3.69	.932	.119
		31	4.31	.676	.576
EFFE	**.789**	3	3.14	1.141	.789
		12	3.23	1.140	.620
		29	3.09	1.121	.720
SOCI	.526	21	3.23	1.003	.613
		28	4.06	1.027	.409
		30	3.34	1.056	.184

Looking at Table 3, it is notable that the constructs *Perceived Dialogue Support, Perceived Credibility*, *Perceived Effort*, and *Perceived Social Support* score below the threshold of acceptance for internal consistency ($\alpha < 0.7$).

The first construct analysed was Perceived Effort (EFFO) which scored lowest on reliability. Additionally, if item 2 is removed from that construct the Cronbach´s alpha becomes negative. To further investigate this, a correlational analysis between the items 2, 9, and 31 was performed. The result of this analysis shows that item 31 is barely positively correlated to item 2 and is negatively correlated to item 9. It can also be seen that only item 2 and 9 are significantly positively correlated (Table 4). Meaning that item 31 might belong to a different construct.

Table 4.

Inter-Item correlation for the items of the construct "Perceived Effort"

	Item2	Item9	Item31
Item2	1	.405*	.068
Item9	.405*	1	-.119
Item31	.068	-.119	1

*. Correlation is significant at the 0.05 level (2-tailed).

The next construct scoring below the threshold for acceptance of reliability was *Perceived Dialogue Support*. Looking at Table 5 it becomes obvious that all three items are only weakly or moderately correlated to each other. This shows that an increase in one item only leads to a slight increase in the other two. However, for items measuring the same construct a stronger relationship should be present.

Table 5.

Inter-Item correlation for the items of the construct "Perceived Dialogue Support"

	Item11	Item17	Item18
Item11	1	.236	.268
Item17	.236	1	.098
Item18	.268	.098	1

The next construct that was analysed further was *Perceived Credibility* (Table 6). The inter-item correlation for the five items shows that items 10, 16, and 27 are all moderately (0.2 – 0.4) or strongly (>0.4) correlated to each other with significant p-values ($p < 0.05$). However, they are only weakly correlated to item 4 and 19 which themselves are significantly correlated to each other. This shows that items 10, 16, and 27 could belong to one construct while items 4 and 19 belong to a different one.

Table 6.

Inter-Item correlation for the items of the construct "Perceived Credibility"

	Item4	Item10	Item16	Item19	Item27
Item4	1	.161	.159	.391*	.320
Item10	.161	1	.385*	-.056	.591**
Item16	.159	.385*	1	.097	.398*
Item19	.391*	-.056	.097	1	-.074
Item27	.320	.591**	.398*	-.074	1

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

The last construct analysed for inter-item correlations was *Perceived Social Support* (Table 7). It can be seen that only items 28 and 30 correlated strongly and significantly with each other. Item 21 does not show a strong correlation to either of the two. This means that an increase in item 21 does hardly show an increase in the other two items.

Table 7.

Inter-Item correlation for the items of the construct "Perceived Social Support"

	Item21	Item28	Item30
Item21	1	.101	.257
Item28	.101	1	.442*
Item30	.257	.442*	1

*. Correlation is significant at the 0.05 level (2-tailed).

After the initial reliability analysis, a second one was performed based on the adjusted constructs. It can be seen that the new construct TASK + PERS + EFFE + item16 scores higher ($\alpha = .922$) than before ($\alpha = .805$, $\alpha = .799$, or $\alpha = .789$). Additionally, adding item 1 to the construct *Perceived Effort* led to an increase in reliability from $\alpha = .343$ to a new value of $\alpha = .534$. However, this new value is still inadequate because it is still lower than 0.7. Removing item 1 from the construct *Unobtrusiveness* leads to a decreased alpha value but it remains acceptable, as its new value is .840. The rest of the constructs remained unchanged and thus their alpha values did not change either. Detailed information about this as well as Cronbach´s alpha if one item was deleted can be found in Table 8.

Table 8.

Cronbach´s alpha values for alternative PPQ constructs

Construct	Cronbach´s Alpha	Items	Mean	Std. Dev.	Cronbach´s alpha if item deleted
TASK +	**.922**	3	3.12	1.166	.922
PERS +		5	2.76	1.251	.916
EFFE (+16)		12	3.18	1.158	.906
		15	3.09	1.128	.906
		16	3.42	.751	**.926**
		20	2.79	.992	.910
		23	2.88	1.139	.916
		25	3.15	1.149	.913
		26	3.36	1.194	.919
		29	3.03	1.132	.910
DIAL	.433	11	3.37	1.031	.177
		17	3.03	.923	.414
		18	3.49	.818	.380
CRED (-16)	.505	4	3.59	.857	.251
		10	3.85	.657	.454
		19	3.15	.925	**.595**
		27	3.50	.663	.390
UNOB (-1)	**.840**	6	2.91	1.121	.717
		14	3.57	.917	.792
		24	3.31	1.157	.819
CONT	**.920**	7	2.71	1.226	**.922**
		8	3.00	.970	.898
		13	2.66	1.187	.874
		22	2.63	1.165	.890
EFFO (+1)	.534	1	3.40	1.063	.343
		2	3.77	.973	.345
		9	3.69	.932	.412
		31	4.31	.676	**.638**
SOCI	.526	21	3.23	1.003	**.613**
		28	4.06	1.027	.409
		30	3.34	1.056	.184

Differences in Use Continuance

To obtain an answer for the second research question, the relationship between physical activity (PA) and use continuance was explored. As seen in Table 9, the low/moderate and high physical activity distributions were sufficiently normal for the purpose of conducting a t-test, i.e., skew $<|2.0|$ and kurtosis $<|9.0|$ (Schmider, Ziegler, Danay, Beyer, & Bühner, 2010). First, the assumption of homogeneity of variances was tested and satisfied via Levene´s F test, $F(33)$ = 2.53, p = .122. The independent samples t-test was not associated with a statistically significant effect, $t(33)$ = -1.52, p = .137. Thus, there was no difference in use continuance between participants who scored high on physical activity and those who scored low/moderate. Based on these results, the answer to the second research question is that physical activity has no significant effect on the use continuance of the tested application RunKeeper®.

Table 9.

Descriptive Statistics associated with Use Continuance (N=35)

Physical Activity	N	M	SD	Skew	Kurtosis
Low/moderate	17	2.68	.73	-.08	-1.07
High	18	3.15	1.08	.07	-1.08

Discussion

The main goal of this study was to gain insights into the internal consistency of the PPQ constructs. The results of this reliability analysis for the original PPQ constructs showed that five of the nine constructs had acceptable reliability. This confirms the first hypothesis, that internal consistency will generally be low, because four out of nine constructs (44%) failed to meet acceptance criteria to be accepted as internally consistent and therefore reliable. These four constructs were *Perceived Dialogue Support (DIAL)*, *Perceived Credibility (CRED)*, *Perceived Effort (EFFO)*, and *Perceived Social Support (SOCI)*. Even though the absolute majority of constructs (56%) managed to meet acceptance criteria, the first hypothesis is seen as confirmed because this value is not high enough for a questionnaire to be seen as generally

internally consistent or reliable. Therefore, it can be advised that the PPQ should be used with caution, since it is not clear if it currently measures the constructs of the PSD model accurately.

The second reliability analysis based on the adjusted PPQ constructs (Beerlage-de-Jong et al., 2016) would also confirm this hypothesis. It showed that three out of the now total seven constructs showed acceptable reliability values. The four constructs that were not internally consistent in the first analysis remained that way for the second. Interestingly, when comparing the results of the reliability analysis to the results found by Beerlage-de-Jong et al. (2016), it can be observed that only the constructs of "*Use Continuance*" and "*Perceived Effectiveness*" show the same results with values greater than 0.7 and therefore, being accepted as internally consistent. This implies that these two constructs might already be measured in a reliable way. Additionally, the new construct which is made up out of "Perceived Primary Task Support", "Perceived Persuasiveness", "Perceived Effectiveness", and item 16 stands out with a reliability value of 0.922. This shows that the new grouping of these constructs already proposes a valuable step towards general reliability of the PPQ.

It is notable, that the original constructs "*Perceived Primary Task Support*", "*Perceived Persuasiveness*", and "*Unobtrusiveness*" scored high on internal consistency in this study but failed to do so in the work done by Beerlage-de-Jong et al. (2016). Additionally, "*Perceived Credibility*" and "*Perceived Effort*" scored high in the study by Beerlage-de-Jong et al. (2016), but failed to do so in the current study. This is an interesting finding since the samples of the two studies show similar characteristics. In both studies the majority of participants was familiar with eHealth, had a similar mean age (23.20 and 24.71), and were recruited from similar backgrounds. The only difference that can be observed are the nationalities of the two samples. In this study the majority of participants was either German or of German background and in the study by Beerlage-de-Jong et al. (2016) the sample consisted out of Finish and Dutch students. It could be that the nationalities interpreted the questions differently. However, this is not a clear finding and more attention must be paid to these construct in future studies involving the PPQ.

These results show that the PPQ in its current state covers some constructs of the PSD model reliably well, whereas some others fail to be internally consistent. This was highlighted with the additional inter-item analysis for the construct *Perceived Effort* as item 9 and 31 proved to be negatively correlated with each other. The Cronbach´s alpha value if item 31 was to be deleted showed that this would improve reliability notably. This could mean that item 31 measures a different construct and special attention will have to be paid to that construct in

future development of the PPQ. In addition, the construct Perceived Social Support also showed only one strong, significant positive correlation between items 28 and 30. Item 21 failed to show these results implying that this item also might not adequately measure the construct. Especially interesting was that for the construct Perceived Credibility three items (10, 16, and 27) correlated with each other and the remaining two items (4, 19) also correlated with each other. However, between these two "groups" of no strong and significant correlation could be observed. This is interesting because it might indicate that the first chunk of items (10, 16, 27) measures one construct, and the other two items measure a different one. These findings highlight the fact that the PPQ in its current form might not adequately measures its constructs as intended. Therefore, it is important to come up with new ways of measuring these constructs. One recommendation is to come up with new questions for a construct with low reliability and then replacing the ones with low reliability with the new questions. If this is to be done a new reliability analysis must be conducted to check for internal consistency again.

Another interesting finding of this study was that the two constructs *Perceived Dialogue Support* and *Perceived Social Support* both failed to meet acceptance criteria for their alpha values. A possible explanation for this could be that both constructs deal with direct interaction between the user and the persuasive technology. However, the main difference is that dialogue support focuses on human-to-technology interaction, whereas social support deals with human-to-human communication (Lehto et al., 2012). These seemingly similar constructs might cause confusion in users because they are not able to distinguish between the two. This in turn might be the reason why both constructs fail to be reliable because users confuse the items for each construct respectively. This same result can be found in the study done by Beerlage-de-Jong et al. (2016) and emphasizes that the items of the two constructs should either be re-worked or that the PPQ should only focus on one of the two. This would reduce confusion for the construct and might be a solution to reliably measure at least one of the constructs. It is important, to pay special attention to the construct Social Support because several studies have pointed out its importance for persuasive technologies (Lehto et al., 2012; Orji, 2017). In order to be an effective evaluative tool for such technologies the PPQ should be able to measure this construct reliably. All of these observations highlight that the PPQ is still in its infancy and will have to be revised to fit validity and reliability standards in the future.

Because use continuance of a persuasive technology is influenced by the characteristics of the user (Campbell et al., 2001; Reinwand et al., 2015), the second research question of this paper aimed at finding out whether differences in physical activity were associated with differences in use continuance of the RunKeeper® application. The result of this analysis did

not show a significant difference in use continuance between subjects with high physical activity and subjects with moderate/low physical activity. Therefore, it can be assumed that the user characteristic of physical activity level has no influence on the later use continuance of an eHealth technology. This is contradictory to a study done by Bregenzer et al. (2017) who found that lower physical health, including lower physical activity, was positively associated with eHealth app usage. One possible explanation for these differing findings could be that the current study only measured intended use continuance while Bregenzer et al. (2017) measured actual use continuance. Further studies targeting eHealth are necessary to investigate which user characteristics influence use continuance. One possibility would be to target a specific health problem, e.g. low physical activity, and come up with a questionnaire asking about as many characteristics as possible. This in combination with longitudinal study using a fitness application could then give insights into which character traits influence use continuance. A detailed picture about these influencing characteristics would help designers of future eHealth interventions to create more effective technologies that are better tailored for their potential users.

Strengths, Limitations, and Recommendations

Some characteristics of this study stand out as strong points and are worthy to be mentioned here. The first strong point is that because the researcher was physically present during the study, the participants were able to ask questions in case of confusion. This is a strong point because if participants were only instructed to download RunKeeper® and use it, the exact procedure might not be clear to them. This could have resulted in an unnecessarily negative picture of the application which later biases the answers to the PPQ.

Because there were restrictions on where to conduct the study (e.g. in a lab), participants had the opportunity to complete the study at a place of their choice. This means that the participants were most likely not influenced by the environment and felt comfortable while conducting the task.

However, next to the strengths there are certain limitations which must be addressed. Looking at the PPQ, it has to be mentioned that, in its current state it does not include all design principles of the PSD model. The PSD model not only features overall categories but also includes 28 design principles which offer more detailed guidelines for the optimal design process of a persuasive technology. Because of the compact design of the PPQ it is not able to accurately measure all 28 design principles. This might also cause the effect that some questions

target multiple constructs because they are framed rather broad. For the future it would be recommended to come up with a way to measure all design principles accurately. While this would result in a longer version of the PPQ it would ensure that all constructs, including their design principles, are measured adequately.

Additionally, the way in which participants were categorized for physical activity has to be seen as limitation. Even though the IPAQ has been shown to have good validity and reliability (Craig et al., 2003), one problem with it is that participants tend to over-report their physical activity (Rzewnicki, Auweele, & Bourdeaudhuij, 2007). This is a problem because it leads to fewer participants falling into the category "Low" which means that this category is not adequately represented in the data. A similar pattern could be observed in the current study as only two participants were categorized as being low in physical activity. Additionally, the short form of the IPAQ which was used in this study only asks for physical activity within the past seven days. This does not give a stable and enduring overview of participants physical activity level. An additional limitation which is linked to the previous one is the fact that most participants were students at the University of Twente. Many students use their bicycle to get to different places (RWS, 2018) and this leads to a higher score on the IPAQ. This is a limitation because it also means that the "Low" category is not adequately represented. Therefore, future studies should make sure that their sample is more representative and that it is possible to properly investigate the differences between low and high physical activity. These limitations must be taken into account when interpreting the results of this study and further research is necessary to deepen the understanding about user characteristics of eHealth and the PPQ.

Future work on the PPQ might also consider to look at comparable research and adaptions to the PPQ could be made with the input of similar questionnaires such as O'Brien and Toms (2010) who developed a questionnaire measuring user engagement. This could be used as input for the PPQ, as active engagement in an eHealth app is necessary to achieve positive results and behaviour change. While this questionnaire is different in its focus, its main goal is to find out information about how users perceive certain elements of a technology. This is similar to the PPQ since it also focuses on how well users perceive the persuasion of an eHealth technology. Even though persuasion and engagement are different, it can be said that higher engagement will increase the likelihood of persuading the user. This is why it is recommended for the improvement of the PPQ to take similar questionnaires into account.

Conclusion

The Perceived Persuasiveness Questionnaire proposes an imperative tool to evaluate user perception of persuasive elements in eHealth technology. However, the subjective evaluation of RunKeeper® has shown that it currently fails to meet reliability standards and therefore needs to be revised. Additionally, the results of the second part of this study have shown that physical activity seems to play no role in later use continuance of eHealth technology. For future developments in eHealth technologies it is recommended to find out more information about user characteristics such as physical activity. As for the PPQ, it already offers a promising evaluative tool but needs to better cover the design principles of the PSD model before it can be used in standard practice.

References

ASICS Digital, I. (2019). RunKeeper.   Retrieved May 25, 2019 from https://runkeeper.com/

Beerlage-de-Jong, N., Kulyk, O., Kuonanoja, L., Wentzel, J., Oinas-Kukkonen, H., & Gemert-Pijnen, J. V. (2016). Evaluation of the Perceived Persuasiveness Questionnaire. *International Journal of Human-Computer Studies*.

Bol, N., Helberger, N., & Weert, J. C. M. (2018). Differences in mobile health app use: A source of new digital inequalities? *The Information Society, 34*(3), 183-193. doi:10.1080/01972243.2018.1438550

Bregenzer, A., Wagner-Hartl, V., & Jiménez, P. (2017). Who uses apps in health promotion? A target group analysis of leaders. *Health Informatics Journal*, 1460458217738121. doi:10.1177/1460458217738121

Cadmus-Bertram, L., Wang, J. B., Patterson, R. E., Newman, V. A., Parker, B. A., & Pierce, J. P. (2013). Web-based self-monitoring for weight loss among overweight/obese women at increased risk for breast cancer: the HELP pilot study. *Psycho-Oncology, 22*(8), 1821-1828. doi:10.1002/pon.3219

Campbell, P. G., MacAuley, D., McCrum, E., & Evans, A. (2001). Age Differences in the Motivating Factors for Exercise. *Journal of Sport and Exercise Psychology, 23*(3), 191-199. doi:10.1123/jsep.23.3.191

Catwell, L., & Sheikh, A. (2009). Evaluating eHealth Interventions: The Need for Continuous Systemic Evaluation. *PLOS Medicine, 6*(8), e1000126. doi:10.1371/journal.pmed.1000126

Craig, C. L., Marshall, A. L., Sjöström, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., . . . Sallis, J. F. (2003). International physical activity questionnaire: 12-country reliability and validity. *Medicine & science in sports & exercise, 35*(8), 1381-1395.

Cunningham, S. G., Wake, D. J., Waller, A., & Morris, A. D. (2014). Definitions of eHealth. In A. Gaddi, F. Capello, & M. Manca (Eds.), *eHealth, Care and Quality of Life* (pp. 15-30). Milano: Springer Milan.

Edwards, E. A., Lumsden, J., Rivas, C., Steed, L., Edwards, L. A., Thiyagarajan, A., . . . Walton, R. T. (2016). Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. *BMJ Open, 6*(10), e012447. doi:10.1136/bmjopen-2016-012447

Elbert, N. J., van Os-Medendorp, H., van Renselaar, W., Ekeland, A. G., Hakkaart-van Roijen, L., Raat, H., . . . Pasmans, S. G. M. A. (2014). Effectiveness and Cost-Effectiveness of eHealth Interventions in Somatic Diseases: A Systematic Review of Systematic Reviews and Meta-Analyses. *Journal of medical Internet research, 16*(4), e110. doi:10.2196/jmir.2790

Fogg, B. J. (1999). Persuasive technologies. *Communications of the ACM, 42*(5), 27-29.

Greenhalgh, T., & Russell, J. (2010). Why Do Evaluations of eHealth Programs Fail? An Alternative Set of Guiding Principles. *PLOS Medicine, 7*(11), e1000360. doi:10.1371/journal.pmed.1000360

Hester, R. K., Delaney, H. D., & Campbell, W. (2012). The College Drinker's Check-Up: Outcomes of two randomized clinical trials of a computer-delivered intervention. *Psychology of Addictive Behaviors, 26*(1), 1-12. doi:10.1037/a0024753

Hutchesson, M. J., Collins, C. E., Morgan, P. J., Watson, J. F., Guest, M., & Callister, R. (2013). Changes to dietary intake during a 12-week commercial web-based weight loss program: a randomized controlled trial. *European Journal Of Clinical Nutrition, 68*, 64. doi:10.1038/ejcn.2013.194

Ijsselsteijn, W., de Kort, Y., Midden, C., Eggen, B., & van den Hoven, E. (2006, 2006//). *Persuasive Technology for Human Well-Being: Setting the Scene.* Paper presented at the Persuasive Technology, Berlin, Heidelberg.

IPAQ. (2005). Guidelines for Data Processing and Analysis of the International Physical Activity Questionnaire (IPAQ).   Retrieved from www.ipaq.ki.se

Kaptein, M., Markopoulos, P., de Ruyter, B., & Aarts, E. (2015). Personalizing persuasive technologies: Explicit and implicit personalization using persuasion profiles. *International Journal of Human-Computer Studies, 77*, 38-51. doi:https://doi.org/10.1016/j.ijhcs.2015.01.004

Kelders, S. M., Kok, R. N., & Gemert-Pijnen, J. E. W. C. V. (2011). *Technology and adherence in web-based interventions for weight control: a systematic review*. Paper presented at the

Proceedings of the 6th International Conference on Persuasive Technology: Persuasive Technology and Design: Enhancing Sustainability and Health, Columbus, Ohio, USA.

Kelders, S. M., Kok, R. N., Ossebaard, H. C., & Van Gemert-Pijnen, J. E. W. C. (2012). Persuasive system design does matter: a systematic review of adherence to web-based interventions. *Journal of medical Internet research, 14*(6), e152-e152. doi:10.2196/jmir.2104

Lehto, T., Oinas-Kukkonen, H., & Drozd, F. (2012). Factors affecting perceived persuasiveness of a behavior change support system.

Mathers, C. D., & Loncar, D. (2005). Updated projections of global mortality and burden of disease, 2002-2030: data sources, methods and results. *Geneva: World Health Organization*.

Murray, E. (2014). eHealth: where next? *British Journal of General Practice, 64*(624), 325. doi:10.3399/bjgp14X680365

O'Brien, H. L., & Toms, E. G. (2010). The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology, 61*(1), 50-69. doi:10.1002/asi.21229

Oinas-Kukkonen, H. (2013). *A foundation for the study of behavior change support systems* (Vol. 17).

Oinas-Kukkonen, H., & Harjumaa, M. (2009). *Persuasive Systems Design: Key Issues, Process Model, and System Features* (Vol. 24).

Orji, R. (2017). *Why are persuasive strategies effective? Exploring the strengths and weaknesses of socially-oriented persuasive strategies.* Paper presented at the International Conference on Persuasive Technology.

Rabin, C., & Bock, B. (2011). Desired Features of Smartphone Applications Promoting Physical Activity. *Telemedicine and e-Health, 17*(10), 801-803. doi:10.1089/tmj.2011.0055

Reinwand, D. A., Schulz, D. N., Crutzen, R., Kremers, S. P. J., & de Vries, H. (2015). Who Follows eHealth Interventions as Recommended? A Study of Participants' Personal Characteristics From the Experimental Arm of a Randomized Controlled Trial. *Journal of medical Internet research, 17*(5), e115. doi:10.2196/jmir.3932

RWS. (2018). Cycling Facts 2018. Retrieved May 30, 2019 from https://www.government.nl/documents/reports/2018/04/01/cycling-facts-2018

Ryan, R. M., Patrick, H., Deci, E. L., & Williams, G. C. (2008). Facilitating health behaviour change and its maintenance: Interventions based on self-determination theory. *The European Health Psychologist, 10*(1), 2-5.

Rzewnicki, R., Auweele, Y. V., & Bourdeaudhuij, I. D. (2007). Addressing overreporting on the International Physical Activity Questionnaire (IPAQ) telephone survey with a population sample. *Public Health Nutrition, 6*(3), 299-305. doi:10.1079/PHN2002427

Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust? *Methodology*.

Schwerdtfeger, A., Schmitz, C., & Warken, M. (2012). Using Text Messages to Bridge the Intention-Behavior Gap? A Pilot Study on the Use of Text Message Reminders to Increase Objectively Assessed Physical Activity in Daily Life. *Frontiers in Psychology, 3*(270). doi:10.3389/fpsyg.2012.00270

Seiler, A., Klaas, V., Tröster, G., & Fagundes, C. P. (2017). eHealth and mHealth interventions in the treatment of fatigued cancer survivors: A systematic review and meta-analysis. *Psycho-Oncology, 26*(9), 1239-1253. doi:10.1002/pon.4489

Shaw, T., McGregor, D., Brunner, M., Keep, M., Janssen, A., & Barnet, S. (2017). What is eHealth (6)? Development of a Conceptual Model for eHealth: Qualitative Study with Key Informants. *Journal of medical Internet research, 19*(10), e324. doi:10.2196/jmir.8106

Sillence, E., Briggs, P., Harris, P., & Fishwick, L. (2006). A framework for understanding trust factors in web-based health advice. *International Journal of Human-Computer Studies, 64*(8), 697-713. doi:https://doi.org/10.1016/j.ijhcs.2006.02.007

Simmons, V. N., Heckman, B. W., Fink, A. C., Small, B. J., & Brandon, T. H. (2013). Efficacy of an experiential, dissonance-based smoking intervention for college students delivered via the Internet. *Journal of Consulting and Clinical Psychology, 81*(5), 810-820. doi:10.1037/a0032952

Sriramatr, S., Berry, T. R., & Spence, J. C. (2014). An Internet-Based Intervention for Promoting and Maintaining Physical Activity: A Randomized Controlled Trial. *American Journal of Health Behavior, 38*(3), 430-439. doi:10.5993/AJHB.38.3.12

Stratton, E., Lampit, A., Choi, I., Calvo, R. A., Harvey, S. B., & Glozier, N. (2017). Effectiveness of eHealth interventions for reducing mental health conditions in employees: A systematic review and meta-analysis. *PLOS ONE, 12*(12), e0189904. doi:10.1371/journal.pone.0189904

van Keulen, H. M., Bosmans, J. E., van Tulder, M. W., Severens, J. L., de Vries, H., Brug, J., & Mesters, I. (2010). Cost-effectiveness of tailored print communication, telephone motivational interviewing, and a combination of the two: results of an economic evaluation alongside the Vitalum randomized controlled trial. *International Journal of Behavioral Nutrition and Physical Activity, 7*(1), 64. doi:10.1186/1479-5868-7-64

WHO. (2013). 10 facts on noncommunicable diseases Retrieved May 10, 2019 from https://www.who.int/features/factfiles/noncommunicable_diseases/en/

Appendices

Appendix A. Study information sheet

Introduction

You are being invited to participate in a research study titled "The Evaluation of the Perceived Persuasiveness Questionnaire". This study is being done by Jessica Bormann and Philipp Kill from the Faculty of Behavioural, Management and Social Sciences at the University of Twente.

This research aims to validate the Perceived Persuasiveness Questionnaire (PPQ), which is a questionnaire designed to evaluate the subjective persuasiveness of technologies.

During the study we will run through two phases.

In the first phase you will perform a closed card sort. Card sorting is a method aimed at evaluating or developing a system's information structure. In this method, participants are asked to sort cards with excerpts of the system's content on it, into (for them) meaningful groups. In this study, you will first get introduced to the nine constructs of the PPQ. The researcher will explain these constructs to you in a short PowerPoint presentation. Then the closed card sorting task follows. We ask you to group the cards with the PPQ items into predefined categories (the nine PPQ constructs). Lastly, a picture is taken to document the result of the card sorting, which does not include the participant.

In phase two you will evaluate the Runkeeper application. You will receive a scenario-based-task which simulates a real-life-user-experience with this application. After this task you will get the opportunity to freely explore the application. Finally, you are asked to fill in the PPQ based on the experiences you had with the Runkeeper application.

Your participation in this study is entirely voluntary and you can pause or withdraw at any time. To the best of our ability your answers in this study will remain confidential.

Study contact details for further information:

Philipp Kill, p.s.kill@student.utwente.nl

Jessica Bormann, j.bormann@student.utwente.nl

Contact Information for Questions about Your Rights as a Research Participant

If you have questions about your rights as a research participant, or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the Secretary of the Ethics Committee of the Faculty of Behavioural, Management and Social Sciences at the University of Twente by ethicscommittee-bms@utwente.nl

Appendix B. Informed Consent Sheet

Title research: Evaluation of the Perceived Persuasiveness Questionnaire

I hereby declare that I have been informed in a manner which is clear to me about the nature and method of the research. My questions have been answered to my satisfaction. I agree of my own free will to participate in this research. I reserve the right to withdraw this consent without the need to give any reason and I am aware that I may withdraw from the experiment at any time. If my research results are to be used in scientific publications or made public in any other manner, then they will be made completely anonymous. My personal data will not be disclosed to third parties without my express permission.

If you have any complaints about this research, please direct them to the secretary of the Ethics Committee of the Faculty of Behavioural Sciences at the University of Twente, Drs. L. Kamphuis-Blikman P.O. Box 217, 7500 AE Enschede (NL), telephone: +31 (0)53 489 3399; email: l.j.m.blikman@utwente.nl).

Signatures

_____   _____   _____

Name of participant   Signature   Date

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

_____   _____   _____

Researcher name   Signature   Date

Study contact details for further information:

Philipp Kill, p.s.kill@student.utwente.nl

Jessica Bormann, j.bormann@student.utwente.nl

Appendix C. Full Questionnaire including demographic questions, IPAQ, and PPQ.

Demographic Questions

1. What is your age?
2. What is your gender?
   - Male
   - Female
   - Other
   - I prefer not to answer
3. What is your Nationality?
   - German
   - · Other
4. What is your Scientific background?
   - Behavioural, Management & Social Science
   - Information & Communication Technology Science
   - Science & Technology
   - Electrical Engineering, Mathematics & Computer Science
   - Other
5. How familiar are you with eHealth technology?
   - Very
   - Somewhat
   - Not at all

Questions about physical activity

We are interested in finding out about the kinds of physical activities that people do as part of their everyday lives. The questions will ask you about the time you spent being physically active in the **last 7 days**. Please answer each question even if you do not consider yourself to be an active person. Please think about the activities you do at work, as part of your house and yard work, to get from place to place, and in your spare time for recreation, exercise or sport.

Think about all the **vigorous** activities that you did in the last **7 days**. **Vigorous** physical activities refer to activities that take hard physical effort and make you breathe much harder than normal. Think only about those physical activities that you did for at least 10 minutes at a time.

1. During **the last 7 days**, on how many days did you do **vigorous** physical activities like heavy lifting, digging, aerobics, or fast bicycling?

   _____ **days per week**

   [ ] **No** vigorous physical activities → **Skip to question 3**

2. How much time did you usually spend doing vigorous physical activities on one of those days?

   _____ **hours per day**

   _____ **minutes per day**

$\boxed{}$ Don't know/Not sure

Think about all the **moderate** activities that you did in the **last 7 days**. **Moderate** activities refer to activities that take moderate physical effort and make you breathe somewhat harder than normal. Think only about those physical activities that you did for <u>at least 10 minutes</u> at a time.

3.   During the **last 7 days**, on how many days did you do **moderate** physical activities like carrying light loads, bicycling at a regular pace, or doubles tennis? Do not include walking.

_____ **days per week**

$\boxed{}$ **No** moderate physical activities → **Skip to question 5**

4.   How much time did you usually spend doing moderate physical activities on one of those days?

_____ **hours per day**
_____ **minutes per day**

$\boxed{}$ Don't know/Not sure

Think about the time you spent **walking** in **the last 7 days**. This includes at work and at home, walking to travel from place to place, and any other walking that you have done solely for recreation, sport, exercise, or leisure.

5.   During the **last 7 days**, on how many days did you walk for at least 10 minutes at a time?

_____ **days per week**

$\boxed{}$ No walking → **Skip to question 7**

6.   How much time did you usually spend **walking** on one of those days?

_____ **hours per day**

_____ **minutes per day**

$\boxed{}$ Don't know/Not sure

The last question is about the time you spent **sitting** on weekdays during the **last 7 days**. Include time spent at work, at home, while doing course work and during leisure time. This may include time spent sitting at a desk, visiting friends, reading, or sitting or lying down to watch television.

7.   During **the last 7 days**, how much time did you spend **sitting** on a week day?

_____ **hours per day**

_____ **minutes per day**

$\boxed{}$ Don't know/Not sure

Perceived Persuasiveness Questionnaire

The following questionnaire is the Perceived Persuasiveness Questionnaire which will be used to evaluate the previously used application "RunKeeper". Please use an **X** to indicate which statement of "Strongly disagree", "Disagree", "Neither agree nor disagree", "Agree", and "Strongly agree" best fits each question.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
1. Using Runkeeper application is practical / convenient for me.					
2. Using Runkeeper application does not require a lot of effort from me.					
3. My chances of starting with exercising improve by using Runkeeper application.					
4. Runkeeper application shows expertise.					
5. Runkeeper application is not personally relevant for me.					
6. Using Runkeeper application fits into my daily life.					
7. I am not going to use Runkeeper application from now on.					
8. I am considering discontinuing using Runkeeper application.					
9. Using Runkeeper application is straightforward for me.					
10. Runkeeper application is unreliable.					
11. Runkeeper application encourages me.					
12. In my opinion, using Runkeeper application has an effect on my willingness to exercise.					

13. I will be using Runkeeper application in the future.					
14. Using Runkeeper application disrupts my daily routines.					
15. Runkeeper application does not help me to start with exercising.					
16. Runkeeper application does not provide confidence.					
17. Runkeeper application provides me with appropriate counselling.					
18. Runkeeper application provides me with appropriate feedback.					
19. Runkeeper application is clearly made by health professionals.					
20. Runkeeper application has an influence on me.					
21. I get support from my peers through Runkeeper application when I need it.					
22. I am going to continue using Runkeeper application.					
23. Runkeeper application helps me change my exercising habits.					
24. Finding the time to use Runkeeper application is a problem for me.					
25. Runkeeper application makes me reconsider my exercising habits.					
26. Runkeeper application provides me with means to exercise.					
27. Runkeeper application is trustworthy.					
28. Through Runkeeper application, I cannot					

share my experiences with my peers.					
29. In my opinion, Runkeeper application has no effect on my exercising habits / weight.					
30. Learning from my peers' actions is beneficial for me.					
31. Using Runkeeper application is difficult.					

Full overview of the IPAQ (short form)

Questions about physical activity

We are interested in finding out about the kinds of physical activities that people do as part of their everyday lives. The questions will ask you about the time you spent being physically active in the **last 7 days**. Please answer each question even if you do not consider yourself to be an active person. Please think about the activities you do at work, as part of your house and yard work, to get from place to place, and in your spare time for recreation, exercise or sport.

Think about all the **vigorous** activities that you did in the last **7 days**. **Vigorous** physical activities refer to activities that take hard physical effort and make you breathe much harder than normal. Think only about those physical activities that you did for at least 10 minutes at a time.

1. During **the last 7 days**, on how many days did you do **vigorous** physical activities like heavy lifting, digging, aerobics, or fast bicycling?

    _____ **days per week**

    ☐ **No** vigorous physical activities → **Skip to question 3**

2. How much time did you usually spend doing vigorous physical activities on one of those days?

    _____ **hours per day**

    _____ **minutes per day**

    ☐ Don't know/Not sure

Think about all the **moderate** activities that you did in the **last 7 days**. **Moderate** activities refer to activities that take moderate physical effort and make you breathe somewhat harder than normal. Think only about those physical activities that you did for at least 10 minutes at a time.

3. During the **last 7 days**, on how many days did you do **moderate** physical activities like carrying light loads, bicycling at a regular pace, or doubles tennis? Do not include walking.

    _____ **days per week**

☐ **No** moderate physical activities → **Skip to question 5**

4. How much time did you usually spend doing moderate physical activities on one of those days?

_____ **hours per day**
_____ **minutes per day**

☐ Don't know/Not sure

Think about the time you spent **walking** in **the last 7 days**. This includes at work and at home, walking to travel from place to place, and any other walking that you have done solely for recreation, sport, exercise, or leisure.

5. During the **last 7 days**, on how many days did you walk for at least 10 minutes at a time?

_____ **days per week**

☐ No walking → **Skip to question 7**

6. How much time did you usually spend **walking** on one of those days?

_____ **hours per day**

_____ **minutes per day**

☐ Don't know/Not sure

The last question is about the time you spent **sitting** on weekdays during the **last 7 days**. Include time spent at work, at home, while doing course work and during leisure time. This may include time spent sitting at a desk, visiting friends, reading, or sitting or lying down to watch television.

7. During **the last 7 days**, how much time did you spend **sitting** on a weekday?

_____ **hours per day**

_____ **minutes per day**

☐ Don't know/Not sure