12 Years of Self-tracking for Promoting Physical Activity from a User Diversity Perspective: Taking Stock & Thinking Ahead

SOFIA YFANTIDOU, PAVLOS SERMPEZIS, and ATHENA VAKALI, Aristotle University of Thessaloniki, Greece

Despite the indisputable personal and societal benefits of regular physical activity, a large portion of the population does not follow the recommended guidelines, harming their health and wellness. The World Health Organization has called upon governments, practitioners, and researchers to accelerate action to address the global prevalence of physical inactivity. To this end, an emerging wave of research in ubiquitous computing has been exploring the potential of interactive self-tracking technology in encouraging positive health behavior change. Numerous findings indicate the benefits of personalization and inclusive design regarding increasing the motivational appeal and overall effectiveness of behavior change systems, with the ultimate goal of empowering and facilitating people to achieve their goals. However, most interventions still adopt a "one-size-fits-all" approach to their design, assuming equal effectiveness for all system features in spite of individual and collective user differences. To this end, we analyze a corpus of 12 years of research in self-tracking technology for health behavior change, focusing on physical activity, to identify those design elements that have proven most effective in inciting desirable behavior across diverse population segments. We then provide actionable recommendations for designing and evaluating behavior change self-tracking technology based on age, gender, occupation, fitness, and health condition. Finally, we engage in a critical commentary on the diversity of the domain and discuss ethical concerns surrounding tailored interventions and directions for moving forward.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing systems and tools**; **HCI design and evaluation methods**; *HCI theory, concepts and models*; • **Applied computing** → *Consumer health.*

Additional Key Words and Phrases: metadata analysis, persuasive systems design, health behavior change, mHealth, personal informatics

ACM Reference Format:
Sofia Yfantidou, Pavlos Sermpezis, and Athena Vakali. 2022. 12 Years of Self-tracking for Promoting Physical Activity from a User Diversity Perspective: Taking Stock & Thinking Ahead. In *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '22 Adjunct), July 4–7, 2022, Barcelona, Spain.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3511047.3538029

1 INTRODUCTION

According to the World Health Organization (WHO), regular physical activity is amongst the determinants of good population health, generating significant personal and societal benefits. However, worldwide, 1 in 4 adults and 3 in 4 adolescents do not engage in regular physical activity, contributing to what has been coined as the "inactivity pandemic" [42]. To join the fight against the global prevalence of physical inactivity, technological innovations need to motivate individuals who are not sufficiently active to incorporate regular exercise into their daily routine. Research has shown that interactive technology can incite desirable health behavior change (HBC) in terms of healthy nutrition, regular physical activity, and disease management, among others [33, 47]. Ultimately though, physical activity interventions are the most prevalent among all HBC technological interventions, accounting for more than 1 out of 3 publications in the related literature [17].

On the same note, emerging, ubiquitous self-tracking technology (ST) has been a game-changer for health and wellness promotion, focusing on physical activity. ST refers to "the practice of gathering data about oneself on a regular basis and then recording and analysing the data to produce statistics and other data (such as images) relating to regular habits, behaviours and feelings" [32, §1]. The philosophy behind it is that continuous monitoring can raise awareness regarding the user's behavioral patterns, whether beneficial or detrimental for their health, and assist them with improving their overall health outcomes. In recent years, a substantial amount of HCI research has been directed towards the design of such technologies, demonstrating the benefits and the persuasive power of strategically-designed ST for HBC [17]. However, despite the abundance of research studies in the field, prior work still has limitations in terms of system inclusive design, scattered domain knowledge, and technology abandonment.

Specifically: (i) The **"one-size-fits-all" mentality** is still prevalent in the majority of ST interventions' design; however, it might not always be fair across user segments. According to Monteiro-Guerra et al. [37], interactive technology has not yet reached its full potential in terms of personalization and inclusive design, lacking context awareness, despite the great potential of personalized approaches in inciting positive HBC. (ii) There is **scattered knowledge concerning ST interventions adopting a diversity perspective**. It is unclear which kinds of personalization strategies exist, how to implement them in practice, and which of those strategies hold the most persuasive power for diverse user segments. In other words, given the large number of available system features (e.g., goal-setting, group challenges, reminders) and related literature, it is unclear whether different features have different motivational appeal and, consequently, effectiveness in promoting an increase in physical activity or not. (iii) The adoption of ST might be widespread, but **technology abandonment is far from a rare phenomenon**. White papers report varied attrition rates for ST, ranging from 30% to 70% [20, 43]. While these numbers are not conclusive, they highlight the need for continuous evaluation of the user experience. However, there exist limited guidelines on personalizing the different performance metrics for diverse user groups.

The limitations above highlight the importance of viewing the HBC domain through the lens of diversity and inclusive design, a perspective rarely explored in related literature. This paper aims to address this gap by conducting an exhaustive analysis and synthesis of prior knowledge accumulated in and expressed through two primary datasets (metadata analysis). While these data capture a comprehensive picture of the knowledge in the domain, they have never been analyzed through a diversity perspective, leading to the identification of personalization guidelines for HBC interventions. Specifically, our contributions are as follows:

- **C1 - Diversity analysis of feature effectiveness for persuasive ST design:** We identify the most effective persuasive elements of ST, as per the Persuasive Systems Design (PSD) framework [41], based on the reported results of 117 prior interventions, totaling 6 million participants, stretching over 12 years (2008-2020). To achieve this, we conduct a previously unexplored personalized metadata analysis of the "PAST SELF" open corpus [52, 53], where we compare the effectiveness of certain persuasive strategies and feature groups between diverse user segments, as well as the general population. Among others, we unveil that rewards barely hold any persuasive power for the elderly, opposite to the general population, or that social support strategies have a much stronger positive effect on gender-targeted interventions than mixed-gender interventions. We also identify unexplored dimensions in existing studies, which can serve as guidelines for future research in inclusive HBC interventions, focusing on physical activity. Through this contribution, we surface biases in ST design and move beyond the "one-size-fits-all" approach, comparing the persuasive power of ST features for diverse user segments by synthesizing metadata extracted from prior research.

C2 - Recommendations for inclusive ST design & evaluation: We compare and contrast feature effectiveness and persuasive power among user groups of varying demographics, in terms of age, gender, occupation, fitness level, and health condition. The findings of this analysis reveal potential design and evaluation recommendations for fairer, more inclusive ST. Among others, our findings unveil the benefits of integrating human conversational agents or avatars in elderly-targeted ST interventions, given the high effectiveness of the "Similarity" persuasive strategy for this user segment and the exploitation of discussion forums and chatting functionality for gender-targeted interventions due to the high effectiveness of the "Social Learning" persuasive strategy in prior work. We accompany the aforementioned design recommendations with different sets of evaluation metrics that are better suited for certain feature and user groups. Following prior literature [31], we encourage a multi-faceted evaluation beyond quantitative data analysis by incorporating self-reported and qualitative metrics, as well as contextual factors and metadata, capturing multiple aspects of user engagement. Through this contribution, we facilitate researchers and practitioners in designing and evaluating the ST user experience with the ultimate goal of providing inclusive functionality to avoid increased attrition rates.

We structure the remaining of this paper as follows: Section 2 discusses HBC theories and their application for ST design. Section 3 introduces the methodology of our metadata analysis, while Section 4 presents and discusses our findings regarding feature effectiveness and evaluation for diverse user groups. Finally, Section 5 concludes our work and touches upon ethical implications and privacy concerns of personalization in personal informatics.

2 RELATED WORK

This section presents related work in the field of ST design for HBC. Section 2.1 discusses behavior change theories within the ubiquitous computing domain, focusing on the PSD framework, which is crucial for our work. Section 2.2 presents literature reviews and meta-analyses, revealing current challenges and the novelty of the presented work.

2.1 Behavior Change Theories

Ubiquitous computing and HCI researchers and practitioners have long understood the importance of theoretically founded design, designing, and implementing technological interventions inspired by numerous psychology, behavioral science, sports science, and behavioral economics theories. Behavior change theories can positively affect ST by informing design, guiding evaluation, and inspiring alternative experimental designs [23]. Some of the most widely-used theories in the domain [53] include the Social Cognitive Theory [46], the Behaviour Change Technique Taxonomy [36], the Transtheoretical Model [44] and the Self-Determination Theory [15].

While behavior change theories provide indispensable high-level guidance, they do not describe how their theoretical concepts could be translated into real-world ST, leaving the interpretation to the respective researcher or practitioner. To bridge this gap, Oinas-Kukkonen and Harjumaa [41], among others, introduced the widely adopted PSD framework. PSD describes the content and functionality required in a behavior change product or service to increase its persuasive power. In particular, the PSD framework defines four persuasive strategies categories, i.e., *primary task support*, *dialogue support*, *credibility*, and *social support*; each having seven sub-groups within. However, personalization is out of the scope of the PSD framework, which adopts an "one-size-fits-all" approach. To address this limitation, we build upon the PSD taxonomy to conduct our metadata analysis, as described in Section 3, from a user diversity perspective.

2.2 ST Reviews and Meta-analyses: Design Space, Limitations & Open Questions

The exploitation of ubiquitous ST for increasing physical activity or decreasing sedentary behavior is an emerging field of study gathering growing scientific interest [17]. Thus, it is no surprise that numerous mapping and literature reviews and meta-analyses have attempted the synthesis of the aforementioned primary study results. Nevertheless, the majority do not adopt a diversity perspective.

A wave of research occupies itself with interventions targeting a specific user segment or PSD strategy. For instance, there exist reviews targeting users with specific traits, such as age [1, 16, 45], race [38], occupation [6], or physical and mental health conditions [39, 40, 49]. Others assess the effectiveness of specific behavior change techniques or ST features, such as gamification [22], chatting functionality [10], or virtual and physical rewards [50], social sharing [18], or the application of machine learning [51] on the activity levels of individuals. However, contrary to the comprehensive nature of our work, the exact targeting of such works does not allow for comparative analysis between different population segments of persuasive strategies.

Adopting a comparative approach, a number of studies consider broader inclusion and exclusion criteria for their primary studies, incorporating multiple user segments and functionalities. For instance, Epstein et al. [17] published an exhaustive mapping review of the personal informatics literature, seeking to answer questions related, but not limited, to identifying areas of interest within personal informatics, tracking motivations, challenges, ethical concerns, and scientific contributions. While these reviews significantly contribute to the domain, their scope is orthogonal to our analysis, as they rely on empirical data to provide high-level guidelines. In contrast, our work is entirely data-driven, analyzing the outcomes of 12 years of related literature to unveil inclusive design and evaluation recommendations. On a similar note, Matthews et al. [34] and Aldenaini et al. [2] published systematic reviews assessing the effectiveness of mobile-application-based interventions in motivating physical activity identifying novel research directions. However, they neither assess the effectiveness of individual persuasive strategies nor personalize their recommendations to specific user groups. To address the first limitation, Aldenaini et al. [3] published a follow-up systematic review, evaluating the effectiveness of individual persuasive strategies in promoting physical activity. Similar to our work, they categorize intervention components under the PSD framework and report success rates per technique. However, they do not adopt a diversity perspective assuming equal effectiveness across diverse users. Closer to our work, Yfantidou et al. [53] conducted a literature review of HBC ST interventions for increasing physical activity. Among others, they presented an assessment of the persuasive power of the PSD strategies based on the reported results of prior literature through the so-called "PAST Score", an effectiveness indicator. By exploiting the open-access corpus of this work [52], we attempt the diversification of the "PAST Score", through our research methodology presented in Section 3, and the presentation of guidelines for the design and evaluation of inclusive ST systems that are tailored to the needs and the goals of their users.

3 METHODOLOGY

This section presents the methodology of our metadata analysis of two primary research studies from a user diversity perspective. Specifically, we utilize the open access "PAST SELF" corpus [52] and the PSD framework [41] to assess ST feature effectiveness, and provide guidelines for personalized system design and evaluation.

The PSD Framework. As mentioned in Section 2.1, the PSD framework introduces 28 persuasion strategies grouped into four categories, namely, *primary task support*, *dialogue support*, *credibility*, and *social support*. PSD is a widely-adopted framework making it easier to compare our results with related meta-analyses and reviews [3], while it also enables us to compare and contrast primary studies integrating different theoretical elements by utilizing common constructs, i.e., the PSD strategies. We utilize an extended version of PSD to capture all persuasive strategies encountered in the studied literature. Specifically, we consider strategies, such as goal-setting, general information provision, punishment, and variability, which are not present in the original taxonomy (category "Other" from here onward). Goal-setting refers to system- or user-defined physical activity goals for the users to achieve; general information provision refers to sharing information with the users regards physical activity, health and general well-being; punishment refers to negative treatment for under-performance or failing to achieve preset goals; and variability refers to the system's ability to provide a variable experience to the user through variable rewards, game elements, interfaces, and hidden tasks. For definitions for the PSD framework's original strategies, see [41].

The "PAST SELF" Corpus. On a similar note, the "PAST SELF" Corpus is a rich source of metadata for 117 primary studies, containing information about the PSD strategies encountered in the included literature, as well as intervention characteristics and sample demographics for all studies (35 encoded fields in total). Note here that the corpus provides a complete overview of the literature in the domain for more than a decade. The corpus definition follows a systematic methodology to ensure the quality of included studies, based on the guidelines introduced by Kitchenham's [27] widely recognized protocol for conducting systematic reviews. To locate the primary studies, the authors perform a broad search in Google Scholar, Scopus, IEEE Xplore, and Web of Science digital libraries utilizing a well-defined boolean search query fine-tuned according to the guidelines of Spanos and Angelis [48]. Overall, the authors screened 16774 articles after duplicate elimination, of which they removed 374 based on date criteria, 15112 based on the title, and 802 based on the abstract. After a full-text read of the remaining 546 articles, they excluded 429 articles based on predefined inclusion/exclusion criteria. Hence, 117 articles synthesized the final corpus. Inclusion criteria required a peer-review process, English language, a clearly-defined user intervention utilizing a ubiquitous device, and a quantitative assessment of the intervention's results. For our metadata analysis, we utilize the following subset of fields:

- Intervention Duration (in weeks): The intervention duration in number of weeks;
- Sample Size: The number of participants in the study;
- Male: The number of male participants in the study;
- Special Criteria: Any special criteria of the participants, e.g., health condition;
- PSD - Primary Task Support: The Primary Task Support elements present in the intervention;
- PSD - Dialogue: The Dialogue elements present in the intervention;
- PSD - System Credibility: The System Credibility elements present in the intervention;
- PSD - Social Support: The Social Support elements present in the intervention;
- Independent Variable: All PSD strategies utilized in the intervention;
- Positive Result: A number from the set [-1, -0.5, 0, 0.5, 1], depending on the success of the intervention (-1 for statistically significant negative result and 1 for statistically significant positive result).

Table 1. Notation for the score generation functions of the personalized "PAST Score".

Notation	Explanation
t_i	Technique i, $t_i \in \{1, 2, \ldots, 32\}$
n_{papers}	Total number of papers
p_{i,j_u}	Paper j of intervention with user sample u with technique i, 1 if technique i appears in paper j, 0 otherwise
r_{i,j_u}	Result of paper j of intervention with user sample u with technique i, $r_{i,j} \in \{-1, -0.5, 0, 0.5, 1\}$
w	The preferred weight for the PAST_score, $w \in [0, 1]$

The Personalized PAST Score. Based on the encoded metadata, [53] calculates the so-called "PAST Score", an indicator of the effectiveness of each PSD strategy for the general population. However, we tweak the original formulas by introducing a personalization criterion moving beyond the "one-size-fits-all" approach. PAST score defines the efficacy and frequency of a technique t_i (based on literature results) for the general population. We extend this, and define the efficacy of the technique t_i for the user segment u, as the sum of the coded results of the u-sampled interventions the technique appears in, divided by the number of these papers, and the frequency of the technique t_i, as the number of u-sampled interventions that the technique appears divided by the total number of papers for this segment (for notations see Table 1):

$$\text{efficacy}(t_i, u) = \frac{\sum_{j_u} r_{i,j_u}}{\sum_{j_u} p_{i,j_u}} \quad \text{and} \quad \text{frequency}(t_i, u) = \frac{\sum_{j_u} p_{i,j_u}}{\sum_{j_u} p_{j_u}} \tag{1}$$

The scores are normalized in the $[+1, +5]$ interval for comparison purposes. Note here that each metric has its shortcomings, hence we report the metrics combined (i.e., PAST Score) but also separately in Section 4. Specifically, efficacy alone can be misleading, since a rare strategy that appears in a single study with positive results would get a maximum score, ignoring its generalizability capacity. To overcome this shortcoming, we also utilize a strategy's frequency, as an indicator of confidence in the strategy's persuasiveness capacity. However, frequency alone would punish rare but potentially ground-breaking results, while over-rewarding commonly used strategies. The final score for a technique is a combination of its reported personalized efficacy and usage frequency in the investigated papers. For our analysis, in equation 2, we assume a weight of $w = 0.5$, i.e., equal importance for efficacy and frequency. Specifically:

$$\text{Personalized\_PAST\_Score}(t_i, u) = w * \text{efficacy}(t_i, u) + (1 - w) * \text{frequency}(t_i, u) \tag{2}$$

Based on the personalized version of the "PAST Score", we then assess the effectiveness of different PSD strategies for diverse segments of the population. For each segment, we also compare PSD effectiveness with the general population. Additionally, we provide a comparison between large-scale interventions (sample size > 50 or intervention duration > 10 weeks) versus all included interventions to explore potential similarities and differences between them. Specifically, we choose to analyze population segments defined by: age (children, adolescents, elderly, mixed), gender (male, female), occupation (office workers, university students), fitness and physique (inactive population, overweight and obese population), and health condition (patients with cancer, diabetes type 2 or heart disease).

The Evaluation Dimensions. To provide comprehensive guidance for HCI and ubiquitous computing researchers and practitioners, we accompany the PSD effectiveness with indicative features and evaluation metrics. We consider a spectrum of evaluation dimensions, measuring different qualities of the user experience, inspired by the work of Lalmas et al. [31]. Based on the findings of our analysis, we recommend a combination of the following evaluation dimensions to capture the various aspects of user experience with ST. Specifically, we consider four distinct aspects:

- **Perceived Self Aspect:** the user's self-reported image of their everyday experiences, as well as psychological, technological, social, and health factors, usually measured through qualitative evaluation methods;
- **Physical Self Aspect:** the user's physical reaction to the interaction with the system, which can be interpreted as the physical activity performed in response to the system's intervention
- **Behavioral Self Aspect:** the user's behavioral response to the system, calculated via user-system interaction metrics, such as wear-time and session duration
- **Environmental Aspect:** the external factors that affect the user's interaction with the system and the execution of the desired behavior, such as weather or location.

4 ANALYSIS RESULTS & DISCUSSION: INCLUSIVE FEATURE EFFECTIVENESS & EVALUATION

This section discusses the most effective PSD strategies per population segment based on the personalized "PAST Score". It also unveils the unexplored "corners" of the ST design space for diverse segments, identifying research gaps to motivate future work in the field (Section 4.1). In the meantime, it provides a set of recommendations of indicative system features and matching evaluation metrics that are best suited for each use case and user segment (Section 4.2).

4.1 PSD Strategies Effectiveness from a Personalized Perspective: Across Population Segments

According to [53], interventions that have utilized self-monitoring, goal-setting, and rewards tend to have higher success rates, indicating the effectiveness of such persuasive strategies for the general population. As a first extension, we explore the correlation between the extended PSD framework strategies' effectiveness (expressed through the personalized "PAST Score") and the different population segments (Figure 1). Naturally, the aforementioned PSD strategies show higher effectiveness (red-like colors) across most population segments. Specifically, self-monitoring has a high or very high personalized "PAST Score" (+4 or +5) for 75% of user segments and goal-setting and rewards for almost 70%. In Section 4.2 we will discuss possible exceptions to the rule. Overall, high scoring strategies (+4 and +5) are rare accounting for 14% of the heat map, with medium (+3), and low or very low (+2 and +1) scoring strategies accounting for 13% and 74%, respectively. High-scoring strategies also tend to belong to certain categories. We notice that the "Dialogue" and "Primary Task Support" strategies show overall higher effectiveness ($\mu = 2.33$, $\sigma = 1.39$ and $\mu = 2.13$, $\sigma = 1.39$) compared to "Other" ($\mu = 2.00$, $\sigma = 1.51$), "Social Support" ($\mu = 1.84$, $\sigma = 0.98$) and "System Credibility" ($\mu = 1.13$, $\sigma = 0.38$) across multiple segments.

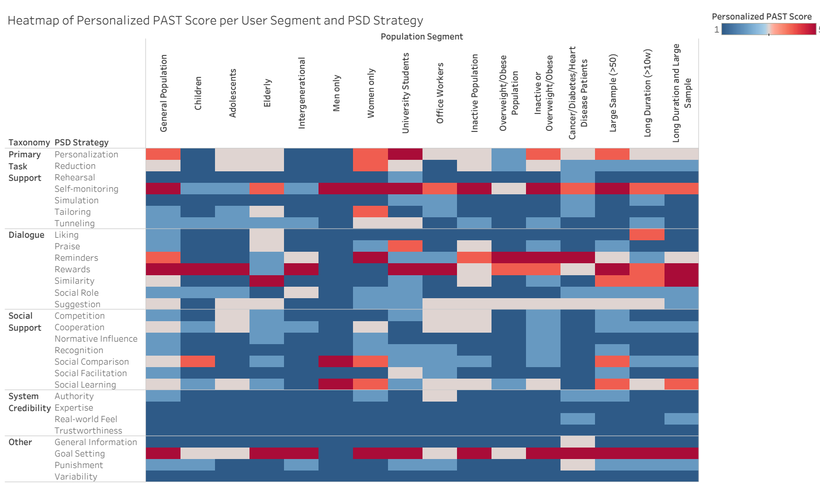

Fig. 1. A heat map of persuasive strategies effectiveness (in terms of PAST Score) versus population segments. Blue color indicates lower effectiveness (1-2), while beige and red colors indicate medium (3) and high (4-5) effectiveness, respectively. The "Other" category indicates strategies of the extended PSD taxonomy.

It is important to note here that lower scores are not necessarily indicators of low effectiveness but also unexplored research domains. This is because the "PAST Score" incorporates a strategy's frequency into the scoring function; hence novel or unexplored strategies might appear with lower scores. While this might seem like a disadvantage at first, incorporating frequency is essential to avoid over-rewarding infrequent techniques for individual positive results. However, researchers and practitioners are encouraged to adapt the weight parameter, w, to their application's needs. To this end, in Figure 2, we present an alternative version of the heat map presenting the frequency scores of each PSD

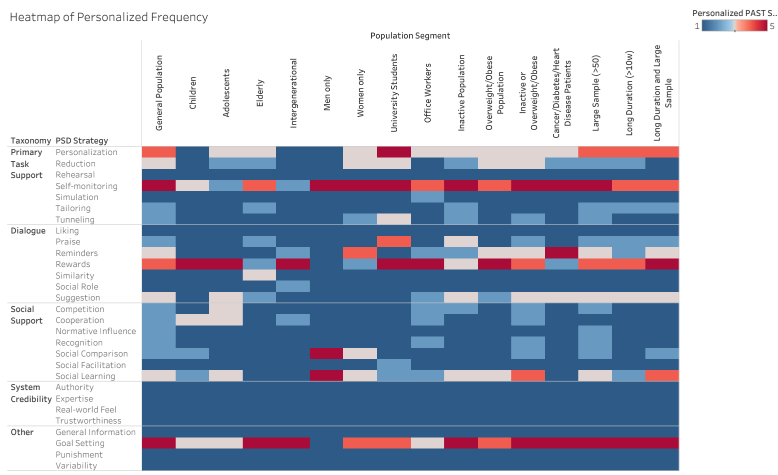

Fig. 2. A heat map of persuasive strategies frequency score versus population segments. Blue color indicates lower frequency (1-2), while beige and red colors indicate medium (3) and high (4-5) frequency, respectively.

strategy across population segments ($w = 0$). Immediately, we can see that all "System Credibility" strategies without exception have a meager frequency score (+1). Similarly, 3 out of 4 "Other" strategies (except for goal-setting) also have a very low frequency. Even in the popular categories, there are PSD strategies that have been barely experimented with. For instance, social facilitation (6% of population segments above very low frequency), and normative influence (12%) from the "Social Support" category, social role (6%), similarity (6%), and liking (0%) from "Dialogue", and finally, simulation (6%) and rehearsal (0%) from "Primary Task Support". Note that a percentage of 0% strategies above very low frequency does not necessarily indicate a full absence of related interventions but potentially minimal experimentation. Similarly, there exists an imbalance in the number of interventions exploiting various PSD strategies between different population segments. For example, women-targeted interventions are more common and diverse than men-targeted ones (31% of PSD strategies above very low frequency vs. 10%). Also, popular samples include inactive or overweight and obese populations (48% of PSD strategies above very low frequency) and office workers (38%). On the contrary, intergenerational and children-based interventions are rarer and more monotonous, as only 20% of PSD strategies have above very low frequency score for these segments. Given this intuition, we can deduce that certain PSD strategies, such as those under "System Credibility" and "Other", or user segments, such as men and children (see note on ethics in Section 5) offer fertile ground for further exploration. On the contrary, other strategies, such as goal-setting and rewards, or user groups, such as women and office workers, have received ample scientific attention.

Figure 3 presents the opposite heat map, where we completely exclude the frequency factor from the personalized "PAST Score" ($w = 1$), accounting only for reported efficacy. We notice that despite the significant lack of "System Credibility" strategies in the related literature, some show very high efficacy scores (+5) for certain population segments, e.g., real-world feel for adolescents or authority for office workers. This suggests that future self-tracking interventions incorporate these previously considered secondary features. Similarly, we notice that even techniques with low "PAST Score" for the general population show high or very high efficacy (+4 or +5) for certain segments, such as rehearsal for university students (+4 in efficacy compared to +1 generic "PAST Score"), liking for the elderly, and intergenerational interventions (+5 and +4 from +2), or rehearsal for university students (+4 from +2). Such visualization provides us with

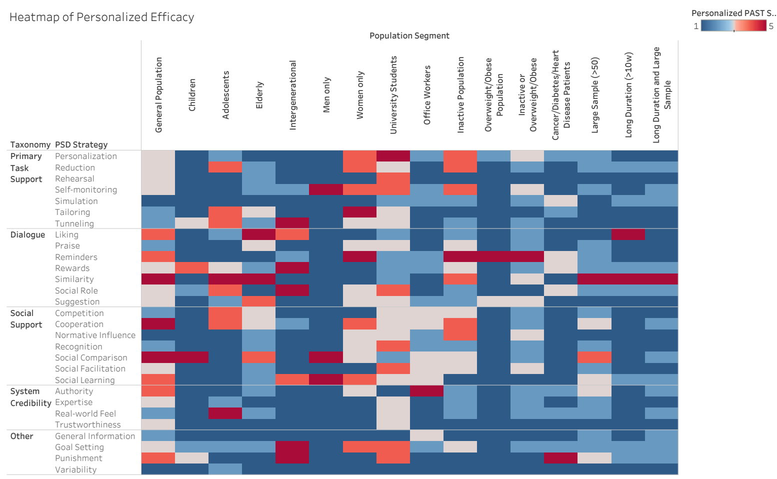

Fig. 3. A heat map of persuasive strategies efficacy versus population segments. Blue color indicates lower efficacy (1-2), while beige and red colors indicate medium (3) and high (4-5) efficacy, respectively.

incites about less tested but promising and potentially ground-breaking PSD strategies that pave the ground for future work in inclusive HBC interventions focusing on physical activity.

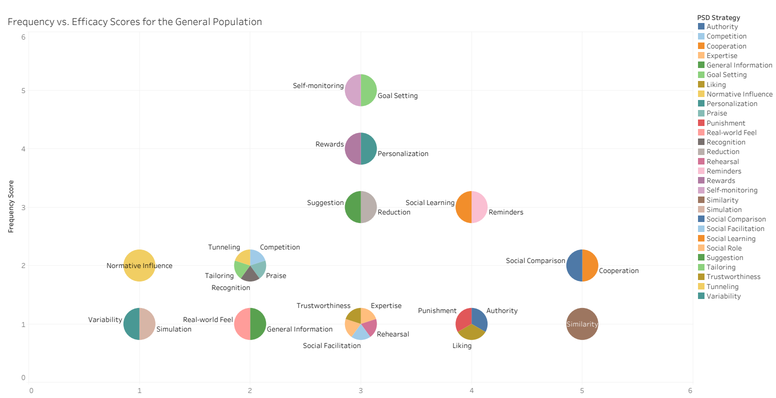

Fig. 4. A scatter plot of efficacy scores (x-axis) versus frequency scores (y-axis) for each PSD strategy for the general population.

Summing up, Figure 4 gives us an overview of the comparison between PSD strategies' frequency (y-axis) versus efficacy (x-axis) in prior literature for the general population. Each circle in the figure represents one or more PSD strategies (overlapping strategies are shown as pie chart pieces). We notice that the most popular techniques, e.g., self-monitoring and goal-setting, ultimately show overall medium efficacy. At the same time, some popular techniques in commercial self-tracking devices, i.e., praise or competition, show overall low efficacy. Some others (see bottom left part of the plot) appear to be inefficient; however, this may be due to insufficient evidence, and we need more data to conclude their effectiveness. Nevertheless, the most interesting part of the plot -in terms of potential for future work-is the bottom-right corner, which presents the least tried techniques, i.e., similarity, social comparison, cooperation, liking, punishment, and authority, that have led to successful results in prior HBC interventions.

This section has given an overview of the persuasive power of the PSD strategies for distinct population segments while we further our discussion on a case-to-case basis below.

4.2 PSD Strategies Effectiveness from a Diversity Perspective: Use Cases

This section presents a series of engaging, indicative use cases where we compare the effectiveness of each PSD strategy across diverse segments of the population through radar plots. PSD strategies are placed on the circle's circumference, whereas the inner polygons indicate the personalized "PAST score" for each strategy. The blue polygon refers to the general population, while the red (or green) to specific user segments.

4.2.1 Age-group Differences. Initially, we explore differences in persuasive power across various age groups compared to the general population. The radar plots in Figure 5 present a comparison of PSD strategies' effectiveness between young users (children and adolescents) and the general population, while the radar plot in Figure 6 focuses on the elderly population.

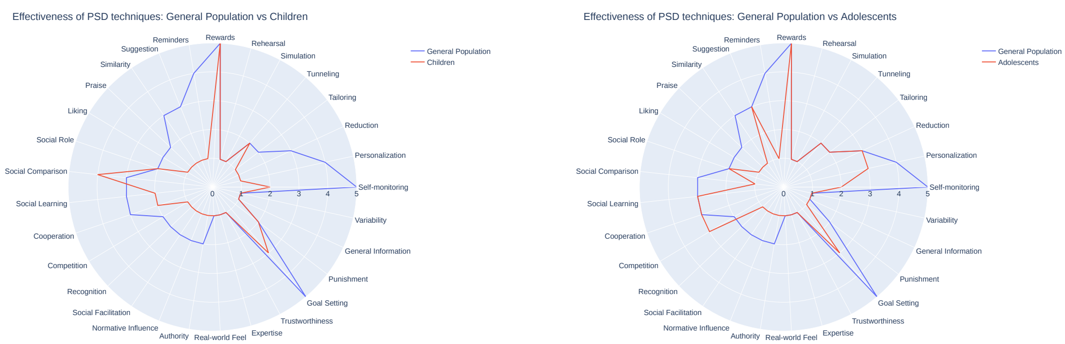

Fig. 5. PSD Strategies effectiveness comparison between the general population, children (left), and adolescents (right), respectively.

We notice that the most powerful strategy, based on reported results, for inciting positive HBC in children through ST is *rewards* (+5), followed by *social comparison* (+4), which shows more persuasive power for children compared to the general population. Rewards are usually tied to gamification in children-oriented interventions, where real-world physical activity is translated into virtual commodities. Examples of such rewards include game currency or extra playtime [7, 21]. With regards to social comparison, prior research has utilized decoy opponents used for real-life comparisons [7]. On the contrary, there is a large drop in "Primary Task Support" effectiveness (e.g., self-monitoring +2 vs. +5) and goal-setting (+3 vs. +5). We conjecture that such features, designed and fine-tuned on adult populations, might not be directly applicable to children. Hence, further research is required towards designing kid-friendly primary task support for positive HBC. Given the complexity of self-reporting, we could recommend a combination of indicative physical and interaction factors (Physical Self Aspect, Behavioral Self Aspect; see Section 3) for evaluating the effectiveness of features in interventions with children populations. Specifically, the adoption of desired behavior can be measured through tracked data, such as activity duration, or type of activities performed. At the same time, the interaction success can be evaluated via feature access, number of sessions, activity performed between sessions, or number of acquired commodities and rewards.

Children

Recommended strategies: Virtual Rewards, Social Comparison
Evaluation metrics: Activity between sessions, Acquired rewards, Age-based activity duration comparison
Strategies needing further exploration: Self-monitoring, Goal-setting, Personalization

Similarly, interventions on adolescents (Figure 5-right) utilizing *rewards* also show high effectiveness (+5), but "Social Support" features, such as *social learning*, *cooperation* and *competition* offer more promising results compared to Social Comparison (+3 vs. +1). Rewards (material or virtual) in adolescent interventions slightly diverge from gamification practices. Material rewards take the form of monetary prizes and gifts, while virtual rewards move away from in-game commodities to badges and point systems [9, 13, 28, 35]. Additionally, discussion groups in social media platforms have been integrated in HBC interventions as a means of social facilitation [13, 35], similarly to individual and group-based competitions [9, 13]. Note that the effectiveness of "Primary Task Support" strategies, such as personalization and reduction, seems to gradually increase with age. Evaluation metrics here should capture the social nature of adolescent-based interventions. Apart from tracked data, e.g., activity duration, time between sessions, energy expenditure (Physical Self Aspect), it is important to collect interaction metrics (Behavioral Self Aspect), such as content views (e.g., messages, posts, e-mails), notifications, and content response time, content shares, comments, and likes, as well as user content generation. Similarly, environmental factors (Environmental Aspect) such as social network and interactions can offer a more comprehensive view of the intervention's success.

Adolescents

Recommended strategies: Virtual and Physical Rewards, Social Facilitation, Cooperation, Competition
Evaluation metrics: User content generation, Comments, Shares, Interactions, User Network
Strategies needing further exploration: Self-monitoring, Goal-setting, Personalization

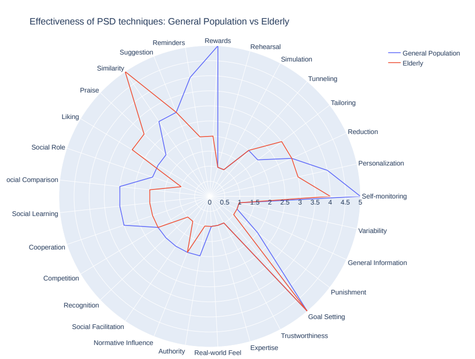

Fig. 6. PSD Strategies effectiveness comparison between the general population and the elderly population.

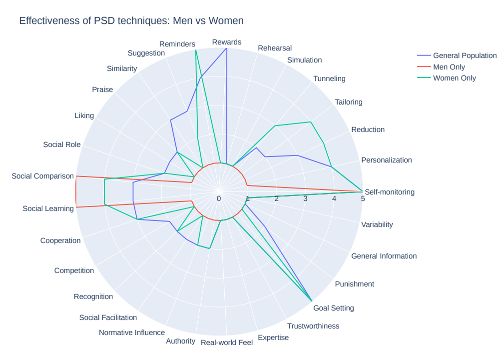

Fig. 7. PSD Strategies effectiveness comparison between the general population, men and women.

When it comes to the elderly population (Figure 6), the radar plot shows a different picture. "Social Support" strategies, i.e., social comparison, social learning, and cooperation, show decreased effectiveness (+1 or +2), giving ground to "Dialogue" strategies, such as *liking*, *praise* (+3) and especially *similarity* (+5). Similarity usually takes the form of

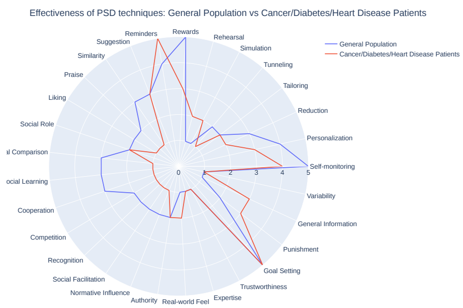

Fig. 8. PSD Strategies effectiveness comparison between the general population, and cancer, diabetes, and heart disease patients.

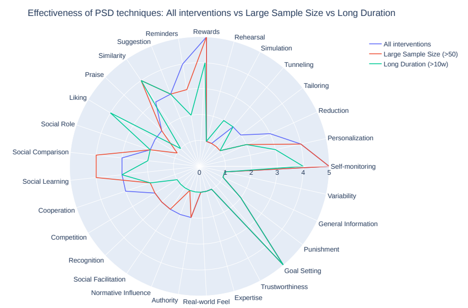

Fig. 9. PSD Strategies effectiveness comparison between all interventions, and interventions with a larger sample (>50 users) or a longer duration (>10 weeks).

embodied conversational agents and human avatars [4, 8]. Liking is implemented via voice-controlled personal assistants [25], while praise can range from simple positive feedback for goal attainment to virtual pets exhibiting positive or negative emotions depending on the user behavior [8]. The common denominator amongst these features is that they provide a sense of human connection to elderly users who might experience solitude. On the contrary, rewards show a significant decrease in effectiveness for this population segment, contrary to younger samples. For this segment of the population, relevant evaluation metrics include sedentariness duration, or type of activities performed, measuring the direct effect of the intervention (Physical Self Aspect), as well as feature access, content views, interactions with conversational agents, and wear time (Behavioral Self Aspect). Also, self-reported physical and mental health or physical activity competency are relevant factors when tailoring interventions for older users (Perceived Self Aspect).

Elderly

Recommended strategies: Liking, Praise, Similarity
Evaluation metrics: Sedentariness duration, feature access, conversational agent interactions, health
Strategies needing further exploration: Rewards, Tailoring

4.2.2 Gender Differences. With regards to gender differences, Figure 7 presents a comparison of the effectiveness of PSD strategies for men and women versus the general population. We notice that "Social Support" strategies, i.e., *social comparison* and *social learning*, show higher effectiveness for gender-specific interventions compared to generic interventions, especially for male samples (+5 vs. +4 for females vs. +3 for general population). Also, rewards have surprisingly low scores (+1) compared to the general population. However, as seen from the limited red spikes in the radar plot, interventions targeted to males are rare in the related literature, offering grounds for future work.

With regards to female-based interventions, mainly *tailoring* (+2 increment), and *tunneling*, *reduction*, and *reminders* to a smaller extend (+1 increment), show promising results with regards to effectiveness, exceeding that of the general population. Reduction can be translated into system features through goal-setting guidance and progress bars for supporting partial goal accomplishment [5, 54]. Tailoring can take the form of contextualized suggestions for short active breaks [19], and tunneling has been implemented as information provision accompanied by data-driven suggestions, or gradually increasing goals [5, 14]. With regards to social features, gender-based interventions utilize more closed social platforms compared to adolescent-based interventions, such as private Fitbit communities or team chats to create a collective sense of trust for progress sharing [5, 54]. Given the high importance of social features across genders, evaluation metrics should incorporate social factors, such as self-reported social support, group cohesion, social comparison tendencies (Perceived Self Aspect), as well as interaction analytics, such as content shares and views, and time between content view and activity (Behavioral Self Aspect). At the same time, evaluation metrics targeted to "Primary Task Support" strategies are also suitable, especially for female-based interventions, given their high effectiveness for this segment. Specifically, reduction, tailoring, and tunneling feature access should be evaluated by the activity performed as a result (Physical Self Aspect).

Females

Recommended strategies: Goal-setting, Reduction, Tailoring, Social Learning, Social Comparison, Reminders
Evaluation metrics: Self-reported social support, Group cohesion, Social comparison tendencies, Interaction Analytics, Feature Access, Time between access and physical activity
Strategies needing further exploration: Rewards

Males

Recommended strategies: Self-monitoring, Social Learning, Social Comparison
Evaluation metrics: Social support, Group cohesion, Social comparison tendencies, Interaction Analytics
Strategies needing further exploration: Rewards, Goal-setting, Reminders, Personalization

4.2.3 Health-related Differences. Concerning patients suffering from a physical health condition, such as diabetes type 2, heart disease, or cancer (Figure 8), we directly notice that "Social Support" strategies have a lower "PAST Score" compared to the general population, while *reminders* and *goal-setting* share the highest "PAST Score" (+5). Low scores, though, can be partly attributed to the low frequency of these PSD strategies for this segment (See Figure 2), indicating a direction of future work. On the other hand, *reminders*, *punishment*, *real-world feel*, *rehearsal*, and *simulation* show a slight increase in effectiveness compared to the general population (+1), while provision of *general information* shows the largest increase (+2). We assume that this is because this segment has a higher need for information about which forms of physical activity are suitable for their condition. Punishment can take the form of monetary punishment for failure to reach goals or negative feedback [11, 26], while real-world feel can refer to app store responsiveness [12] or contact form availability for support [29]. With regards to primary-task support, rehearsal has been implemented as short exercise demonstration videos [28], and simulation as a linkage between real-world physical activity and virtual world experience [24]. To evaluate the user experience for this segment, one needs to incorporate self-reported metrics of health condition, and goals and expectations (Perceived Self Aspect), tracked metrics (Physical Self Aspect), as well as interaction metrics, such as content views (Behavioral Self Aspect) to evaluate for example information utility.

Cancer/Diabetes/Heart Disease Patients

Recommended strategies: Goal-setting, Reminders, Information provision, Punishment
Evaluation metrics: Health condition, Goals & Expectations, Goal accomplishment rate, Content views
Strategies needing further exploration: Social comparison, Social learning, Cooperation

4.2.4 Observed Differences based on Sample Size and Intervention Duration.

This section presents the reported results of large-scale and long-term interventions to indicate the longevity of the PSD strategies' effectiveness. We define large-scale interventions as those with a sample size larger than 50 participants and long-term interventions as those with a duration longer than 10 weeks. Amongst the first things we can notice is that *goal-setting* sustains its maximum "PAST Score" both for large-scale and long-term interventions, whereas *self-monitoring*, and *rewards*' effectiveness whines with time (+4 from +5). This is in accordance with prior literature [30], which reports diminished long-term effectiveness of incentives for inciting and maintaining HBC. On the contrary, "Social Support" features, such as *social comparison* and *social learning*, show improved effectiveness for large-scale interventions (+4 from +3). Another interesting point is that despite similarity and liking being quite rare in the included literature, they are correlated with better effectiveness (+1 and +2 increment, respectively in "PAST Score") for long-term interventions.

5 CONCLUSION

This paper presents the results and related discussion of a metadata analysis performed on top of the PSD framework and an open corpus of 117 HBC interventions utilizing ST. Through our exploration, we debunk the "one-size-fits-all" mentality in the design space of personal informatics, highlighting -yet again- the significance of inclusive design, diversity and personalization. Specifically, we analyze the effectiveness of various persuasive strategies across diverse population segments and showcase the differences and similarities between them and the general population (C1). Additionally, we accompany the analysis with data-driven recommendations on system features and evaluation metrics suitable for each segment (C2) and provide directions for future work.

While we focus on physical activity interventions, future work may validate our results in different HBC domains, such as smoking cessation, or diet monitoring, to investigate possible changes in the perceived effectiveness of PSD strategies. It is also important to note that while design guidelines can facilitate the development of ST, researchers and practitioners need to be mindful of their epistemic status. While we strongly believe in the value of empirical data for generating design guidelines, we suggest that our recommendations be treated similarly to "design hypotheses", which require additional testing. As a final note, our results illustrate a lack of research on how different population segments (e.g., based on age, gender, and health status) interact with ST, which presents an opportunity for future work. For instance, the limited number of interventions targeting racial minorities did not allow us to perform an analysis based on race. Future research should focus on equitable access to ST for HBC, taking into account the socioeconomic status, the variable health, and the technological literacy of diverse population segments while being sensitive to their cultural and linguistic needs. On a similar note, privacy was rarely touched upon in our corpus. However, data privacy (e.g., data sharing, retention policy, third party involvement) is fundamental for trusted and secure use of ST, especially for sensitive groups such as minors or minorities, whose data sharing might raise ethical concerns.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813162. The content of this paper reflects only the authors' view and the Agency and the Commission are not responsible for any use that may be made of the information it contains. The authors would also like to thank the web developers, T. Valk and S. Karamanidis, for their contribution to the development of the online exploration tool.

REFERENCES

[1] 2020. Critical Reflections on Technology to Support Physical Activity among Older Adults. *ACM Transactions on Accessible Computing* 13, 1 (2020). https://doi.org/10.1145/3374660

[2] Noora Aldenaini, Oladapo Oyebode, Rita Orji, and Srinivas Sampalli. 2020. Mobile Phone-Based Persuasive Technology for Physical Activity and Sedentary Behavior: A Systematic Review. *Frontiers in Computer Science* 2 (2020), 19. https://doi.org/10.3389/fcomp.2020.00019

[3] Noora Fahad Aldenaini, Felwah Alqahtani, Rita Orji, and Srinivas Sampalli. 2020. Trends in Persuasive Technologies for Physical Activity and Sedentary Behavior: A Systematic Review. *Frontiers in Artificial Intelligence* 3 (2020), 7.

[4] Mohammed Alsaqer and Samir Chatterjee. 2017. Helping the elderly with physical exercise: Development of persuasive mobile intervention sensitive to elderly cognitive decline. In *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom).* IEEE, 1–6.

[5] Danielle Arigo, Leah M Schumacher, Emilie Pinkasavage, and Meghan L Butryn. 2015. Addressing barriers to physical activity among women: A feasibility study using social networking-enabled technology. *Digital health* 1 (2015), 2055207615583564.

[6] Yash Babar, Jason Chan, and Ben Choi. 2018. "Run forrest run!": Measuring the impact of App-enabled performance and social feedback on running performance. In *International Conference on Information Systems 2018, ICIS 2018*, Vol. 2023. Association for Information Systems (AIS), San Francisco, California, United States, 1–17.

[7] Shlomo Berkovsky, Mac Coombe, Jill Freyne, Dipak Bhandari, and Nilufar Baghaei. 2010. Physical activity motivating games: virtual rewards for real activity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 243–252.

[8] Timothy W Bickmore, Rebecca A Silliman, Kerrie Nelson, Debbie M Cheng, Michael Winter, Lori Henault, and Michael K Paasche-Orlow. 2013. A randomized controlled trial of an automated exercise coach for older adults. *Journal of the American Geriatrics Society* 61, 10 (2013), 1676–1683.

[9] Kacie CA Blackman, Jamie Zoellner, Adil Kadir, Brandon Dockery, Sallie Beth Johnson, Fabio A Almeida, D Scott McCrickard, Jennie L Hill, Wen You, and Paul A Estabrooks. 2015. Examining the feasibility of smartphone game applications for physical activity promotion in middle school students. *Games for health journal* 4, 5 (2015), 409–419.

[10] Susan Weber Buchholz, Joellen Wilbur, Diana Ingram, and Louis Fogg. 2013. Physical activity text messaging interventions in adults: a systematic review. *Worldviews on Evidence-Based Nursing* 10, 3 (2013), 163–173.

[11] Neel P Chokshi, Srinath Adusumalli, Dylan S Small, Alexander Morris, Jordyn Feingold, Yoonhee P Ha, Marta D Lynch, Charles AL Rareshide, Victoria Hilbert, and Mitesh S Patel. 2018. Loss-framed financial incentives and personalized goal-setting to increase physical activity among ischemic heart disease patients using wearable devices: the ACTIVE REWARD randomized trial. *Journal of the American Heart Association* 7, 12 (2018), e009173.

[12] Fabio Ciravegna, Jie Gao, Neil Ireson, Robert Copeland, Joe Walsh, and Vitaveska Lanfranchi. 2019. Active 10: Brisk walking to support regular physical activity. In *Proceedings of the 13th EAI International Conference on pervasive computing technologies for healthcare.* 11–20.

[13] Rekesh Corepal, Paul Best, Roisin O'Neill, Frank Kee, Jennifer Badham, Laura Dunne, Sarah Miller, Paul Connolly, Margaret E Cupples, Esther MF Van Sluijs, et al. 2019. A feasibility study of 'The StepSmart Challenge'to promote physical activity in adolescents. *Pilot and feasibility studies* 5, 1 (2019), 1–15.

[14] Prabu David, Janet Buckworth, Michael L Pennell, Mira L Katz, Cecilia R DeGraffinreid, and Electra D Paskett. 2012. A walking intervention for postmenopausal women using mobile phones and interactive voice response. *Journal of telemedicine and telecare* 18, 1 (2012), 20–25.

[15] Edward L Deci and Richard M Ryan. 2008. Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian psychology/Psychologie canadienne* 49, 3 (2008), 182.

[16] Steriani Elavsky, Lenka Knapova, Adam Klocek, and David Smahel. 2019. Mobile health interventions for physical activity, sedentary behavior, and sleep in adults aged 50 years and older: a systematic literature review. *Journal of aging and physical activity* 27, 4 (2019), 565–593.

[17] Daniel A. Epstein, Clara Caldeira, Mayara Costa Figueiredo, Xi Lu, Lucas M. Silva, Lucretia Williams, Jong Ho Lee, Qingyang Li, Simran Ahuja, Qiuer Chen, Payam Dowlatyari, Craig Hilby, Sazeda Sultana, Elizabeth V. Eikey, and Yunan Chen. 2020. Mapping and Taking Stock of the Personal Informatics Literature. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020). https://doi.org/10.1145/3432231

[18] Daniel A. Epstein, Bradley H. Jacobson, Elizabeth Bales, David W. McDonald, and Sean A. Munson. 2015. From "nobody cares" to "way to go!": A design framework for social sharing in personal informatics. *CSCW 2015 - Proceedings of the 2015 ACM International Conference on Computer-Supported Cooperative Work and Social Computing* (2015), 1622–1636. https://doi.org/10.1145/2675133.2675135

[19] Joseph Finkelstein, McKenzie Bedra, Xuan Li, Jeffrey Wood, and Pamela Ouyang. 2015. Mobile App to Reduce Inactivity in Sedentary Overweight Women.. In *MedInfo.* 89–92.

[20] Malay Gandhi. 2015. *Deconstructing the Fitbit IPO and S-1.* Rock Health. Retrieved July 28, 2020 from https://rockhealth.com/deconstructing-fitbit-s-1/

[21] Ainara Garde, Aryannah Umedaly, S Mazdak Abulnaga, Leah Robertson, Anne Junker, Jean Pierre Chanoine, J Mark Ansermino, and Guy A Dumont. 2015. Assessment of a mobile game ("MobileKids Monster Manor") to promote physical activity among children. *Games for health journal* 4, 2 (2015), 149–158.

[22] Juho Hamari, Jonna Koivisto, and Harri Sarsa. 2014. Does gamification work?–a literature review of empirical studies on gamification. In *2014 47th Hawaii international conference on system sciences.* IEEE, Washington, DC, United States, 3025–3034.

[23] Eric B. Hekler, Predrag Klasnja, Jon E. Froehlich, and Matthew P. Buman. 2013. Mind the theoretical gap: Interpreting, using, and developing behavioral theory in HCI research. *Conference on Human Factors in Computing Systems - Proceedings* (2013), 3307–3316. https://doi.org/10.1145/2470654.2466452

[24] Christoph Höchsmann, Denis Infanger, Christopher Klenk, Karsten Königstein, Steffen P Walz, and Arno Schmidt-Trucksäss. 2019. Effectiveness of a behavior change technique–based smartphone game to improve intrinsic motivation and physical activity adherence in patients with type 2 diabetes: randomized controlled trial. *JMIR serious games* 7, 1 (2019), e11444.

[25] Bang Hyun Kim and Karen Glanz. 2013. Text messaging to motivate walking in older African Americans: a randomized controlled trial. *American Journal of Preventive Medicine* 44, 1 (2013), 71–75.

[26] SangBin Kim, SinJae Lee, and JungHyun Han. 2018. StretchArms: Promoting stretching exercise with a smartwatch. *International Journal of Human–Computer Interaction* 34, 3 (2018), 218–225.

[27] Barbara Kitchenham and Stuart Charters. 2007. *Guidelines for performing systematic literature reviews in software engineering.* Technical Report. Keele University and Durham University.

[28] Susanne Hwiid Klausen, Lars L Andersen, Lars Søndergaard, Janus Christian Jakobsen, Vibeke Zoffmann, Kasper Dideriksen, Anne Kruse, Ulla Ramer Mikkelsen, and Jørn Wetterslev. 2016. Effects of eHealth physical activity encouragement in adolescents with complex congenital heart disease: the PReVaiL randomized clinical trial. *International Journal of Cardiology* 221 (2016), 1100–1106.

[29] Thea JM Kooiman, Martijn De Groot, Klaas Hoogenberg, Wim P Krijnen, Cees P Van Der Schans, and Adriaan Kooy. 2018. Self-tracking of physical activity in people with type 2 diabetes: a randomized controlled trial. *CIN: Computers, Informatics, Nursing* 36, 7 (2018), 340–349.

[30] Jan-Niklas Kramer, Peter Tinschert, Urte Scholz, Elgar Fleisch, and Tobias Kowatsch. 2019. A cluster-randomized trial on small incentives to promote physical activity. *American journal of preventive medicine* 56, 2 (2019), e45–e54.

[31] Mounia Lalmas, Heather O'Brien, and Elad Yom-Tov. 2014. Measuring user engagement. *Synthesis Lectures on Information Concepts, Retrieval, and Services* 6, 4 (2014), 1–132.

[32] Deborah Lupton. 2014. Self-tracking cultures: towards a sociology of personal informatics. In *Proceedings of the 26th Australian computer-human interaction conference on designing futures: The future of design.* 77–86.

[33] Mia Liza A Lustria, Seth M Noar, Juliann Cortese, Stephanie K Van Stee, Robert L Glueckauf, and Junga Lee. 2013. A meta-analysis of web-delivered tailored health behavior change interventions. *Journal of health communication* 18, 9 (2013), 1039–1069.

[34] John Matthews, Khin Than Win, Harri Oinas-Kukkonen, and Mark Freeman. 2016. Persuasive technology in mobile applications promoting physical activity: a systematic review. *Journal of medical systems* 40, 3 (2016), 72.

[35] Jason A Mendoza, K Scott Baker, Megan A Moreno, Kathryn Whitlock, Mark Abbey-Lambertz, Alan Waite, Trina Colburn, and Eric J Chow. 2017. A Fitbit and Facebook mHealth intervention for promoting physical activity among adolescent and young adult childhood cancer survivors: A pilot study. *Pediatric blood & cancer* 64, 12 (2017), e26660.

[36] Susan Michie, Michelle Richardson, Marie Johnston, Charles Abraham, Jill Francis, Wendy Hardeman, Martin P Eccles, James Cane, and Caroline E Wood. 2013. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Annals of behavioral medicine* 46, 1 (2013), 81–95.

[37] Francisco Monteiro-Guerra, Octavio Rivera-Romero, Luis Fernandez-Luque, and Brian Caulfield. 2019. Personalization in Real-Time Physical Activity Coaching using Mobile Applications: A Scoping Review. *IEEE Journal of Biomedical and Health Informatics* 24, 6 (2019), 1738–1751.

[38] Andre Matthias Müller, Stephanie Alley, Stephanie Schoeppe, and Corneel Vandelanotte. 2016. The effectiveness of e-& mHealth interventions to promote physical activity and healthy diets in developing countries: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity* 13, 1 (2016), 109.

[39] Elizabeth L. Murnane, Tara G. Walker, Beck Tench, Stephen Voida, and Jaime Snyder. 2018. Personal informatics in interpersonal contexts: Towards the design of technology that supports the social ecologies of long-term mental health management. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018). https://doi.org/10.1145/3274396

[40] Francisco Nunes, Nervo Verdezoto, Geraldine Fitzpatrick, Morten Kyng, Erik Grönvall, and Cristiano Storni. 2015. Self-care technologies in HCI: Trends, tensions, and opportunities. *ACM Transactions on Computer-Human Interaction (TOCHI)* 22, 6 (2015), 1–45.

[41] Harri Oinas-Kukkonen and Marja Harjumaa. 2009. Persuasive systems design: Key issues, process model, and system features. *Communications of the Association for Information Systems* 24, 1 (2009), 28.

[42] World Health Organization. 2019. *Global action plan on physical activity 2018-2030: more active people for a healthier world.* World Health Organization.

[43] Endeavour Partners. 2014. *Inside Wearables Part 1: How behavior change unlocks long-term engagement.* Endeavour Partners. Retrieved July 28, 2020 from https://medium.com/@endeavourprtnrs/inside-wearable-how-the-science-of-human-behavior-change-offers-the-secret-to-long-term-engagement-a15b3c7d4cf3

[44] James O Prochaska and Wayne F Velicer. 1997. The transtheoretical model of health behavior change. *American journal of health promotion* 12, 1 (1997), 38–48.

[45] Susan B Quelly, Anne E Norris, and Jessica L DiPietro. 2016. Impact of mobile apps to combat obesity in children and adolescents: a systematic literature review. *Journal for Specialists in Pediatric Nursing* 21, 1 (2016), 5–17.

[46] Vanessa Ratten and Hamish Ratten. 2007. Social cognitive theory in technological innovations. *European Journal of Innovation Management* 10, 1 (2007), 90–108.

[47] Anna L Roberts, Abigail Fisher, Lee Smith, Malgorzata Heinrich, and Henry WW Potts. 2017. Digital health behaviour change interventions targeting physical activity and diet in cancer survivors: a systematic review and meta-analysis. *Journal of Cancer Survivorship* 11, 6 (2017), 704–719.

[48] Georgios Spanos and Lefteris Angelis. 2016. The impact of information security events to the stock market: A systematic literature review. *Computers & Security* 58 (2016), 216–229.

[49] Katta Spiel, Christopher Frauenberger, Os Keyes, and Geraldine Fitzpatrick. 2019. Agency of autistic children in technology research—A critical literature review. *ACM Transactions on Computer-Human Interaction (TOCHI)* 26, 6 (2019), 1–40.

[50] Kelley Strohacker, Omar Galarraga, and David M Williams. 2014. The impact of incentives on exercise behavior: a systematic review of randomized controlled trials. *Annals of Behavioral Medicine* 48, 1 (2014), 92–99.

[51] Anja Thieme, Danielle Belgrave, and Gavin Doherty. 2020. Machine Learning in Mental Health: A Systematic Review of the HCI Literature to Support the Development of Effective and Implementable ML Systems. *ACM Transactions on Computer-Human Interaction (TOCHI)* 27, 5 (2020), 1–53.

[52] Sofia Yfantidou. 2020. *syfantid/PAST-SELF-Framework-Data: Paper Release.* https://doi.org/10.5281/zenodo.4063377

[53] Sofia Yfantidou, Pavlos Sermpezis, and Athena Vakali. 2021. Self-Tracking Technology for mHealth: A Systematic Review and the PAST SELF Framework. *arXiv preprint arXiv:2104.11483* (2021).

[54] Jingwen Zhang and John B Jemmott III. 2019. Mobile app-based small-group physical activity intervention for young African American women: a pilot randomized controlled trial. *Prevention Science* 20, 6 (2019), 863–872.