Measuring psychological constructs in computer-tailored interventions

Novel possibilities to reduce participant burden and increase engagement

Short, C.E.; Smit, E.S.; Crutzen, R.

Publication date
2022

Document Version
Final published version

Published in
The European Health Psychologist

Citation for published version (APA):
Short, C. E., Smit, E. S., & Crutzen, R. (2022). Measuring psychological constructs in computer-tailored interventions: Novel possibilities to reduce participant burden and increase engagement. *The European Health Psychologist*, *22*(4), 801-815. https://www.ehps.net/ehp/index.php/contents/article/view/3378

Original Article

Measuring psychological constructs in computer-tailored interventions: novel possibilities to reduce participant burden and increase engagement

Camille E. Short
University of Melbourne, Australia
Eline S. Smit
University of Amsterdam, The Netherlands
Rik Crutzen
Maastricht University, The Netherlands

Abstract

Within the field of health psychology, there has been an enormous increase in behaviour change interventions that use digital technology. Answering questions and providing tailored feedback based on the answers provided by participants is the key working mechanism when using computer-tailoring in behaviour change interventions. This behaviour change method has proven to be (cost-)effective and results in participants being exposed to material that is tailored to their social-cognitive profile. At the same time, answering questions to assess this profile increases participant burden, which might contribute to low levels of engagement and high attrition - two of the key challenges in digital health.

This article provides insight into how routinely collected data and novel self-assessment methods can be used in computer-tailoring to measure psychological constructs and address these key challenges. The examples presented suggest that the development of novel proxy measures for measuring psychological constructs relevant to computer-tailoring is indeed possible. However, the extent to which measures are valid and actually do reduce participant burden and have other potential benefits is speculative and needs further investigation. The recommendations provided for future research and practice are hoped to serve as a stimulance for driving further momentum in this area.

Introduction

In the World Health Organization's Global Strategy on Digital Health, digital health is described as “the field of knowledge and practice associated with any aspect of adopting digital technologies to improve health, from inception to operation” (WHO, 2020). Within the field of health psychology, there has been an enormous increase in behaviour change interventions that use digital technology (Crutzen et al., 2018). To change behaviour, it is crucial to be aware of Lewin’s formula indicating that behaviour (B) is a function of a person (P) and his or her environment (E): B=*f*(P,E) (Lewin, 1936). In other words, digital technology should not only be used to target behaviour directly, but should also take the person and the environment in which the behavior takes place into account. The first step in doing so, is by using opportunities provided by digital technology to measure all three elements of this formula.

Technological possibilities to measure behaviour (B) are improving constantly. For example, physical activity and sleeping are behaviours that can be measured unobtrusively by means of mobile phones and watches, and online behaviour can be easily tracked (e.g., how people navigate through the Internet and what content they pay attention to (Skinner et al., 2017)). There are also possibilities with existing technologies in terms of capturing aspects of the environment (E). Mobile phones, for example, can track location. Measuring the person (P) is much more complicated, because we cannot directly measure people's cognitions or other psychological constructs. For now, we have to rely on indirect measures, such as reaction times and answers to questions.

Answering questions and providing tailored feedback based on the answers provided by participants is the key working mechanism when using computer-tailoring in behaviour change interventions (De Vries & Brug, 1999; Hawkins et al., 2008). This behaviour change method has proven to be (cost-)effective and results in participants being exposed to material that is tailored to their social-cognitive profile (Krebs et al., 2010; Noar et al., 2007; Smit et al., 2013; Wolfenden et al., 2015). At the same time, answering questions to assess this profile increases participant burden, which might contribute to low levels of engagement and high attrition - key challenges in digital health (Kohl et al., 2013; Short et al., 2018).

The topic of this article concerns novel possibilities for measuring psychological constructs related to the person. Whereas items in questionnaires are commonly used operationalisations that utilise natural language, other proxies might be more appropriate for linguistically diverse test takers. Moreover, these other proxies might reduce participant burden and as a result improve engagement and lower attrition, because it is not needed anymore to complete lengthy questionnaires. This may ultimately increase the impact of digital behaviour change interventions using computer-tailoring (Glasgow et al., 2006; Yardley et al., 2016). Therefore, the aim of this article is to provide insight into how routinely collected data and novel self-assessment methods can be used in computer-tailoring to measure psychological constructs and address key challenges in digital health (e.g., participant burden, engagement, attrition).

Trends in assessment of psychological constructs in computer-tailored interventions

Over 360 computer-tailoring studies have been conducted to date by researchers across health and computer sciences (Ghalibaf et al., 2019). Psychological constructs have been measured in approximately 60% of these studies, predominantly via questionnaires (91%), diaries or other written records (8%) (Ghalibaf et al., 2019). These psychological constructs can then also be used for tailoring purposes.

Selection of tailoring variables covering psychological constructs has typically been based on underlying theories guiding the intervention development. The Transtheoretical Model (Prochaska & Velicer, 1997), Social Cognitive Theory (Bandura, 1986), the Health Belief Model (Rosenstock, 1974) and the Reasoned Action Approach (Fishbein & Ajzen, 2010) have been the most commonly used theories (Ghalibaf et al., 2019). Resultantly, constructs like stage of change, self-efficacy, perceived benefits and barriers, and goals have been some of the most commonly used tailoring variables (Broekhuizen et al., 2012; see supplementary material).

More recently, there has been increasing recognition of the need to expand the theoretical basis of behaviour change interventions to address a broader set of behaviour change determinants (e.g., habits, affect) (O'Carroll, 2020; Rhodes et al., 2019), as well as determinants likely to impact on how people process intervention content (e.g., need for cognition) (Nikoloudakis et al., 2018; Smit, Linn, et al., 2015). In addition to this, there is growing criticism of the static nature of the theories cited above, with critics arguing they do not apply as well to behaviours that require ongoing participation (e.g., physical activity and healthy eating) as they do for limited occurrence health behaviours (e.g., health screening; Dunton, 2017). This has led to growing advocacy for considering application of these theories in the context of how determinants of behaviour may vary over-time and across situations (Chevance et al., 2020; Duckworth et al., 2016; Millar, 2017).

While the types of tailoring variables that have been used in interventions have been generally well reported, detailed information about how tailoring variables have been measured (e.g., number and content of items, response scales, psychometric properties) has not been as transparent, or heavily scrutinized in the literature (compared to outcome assessments for example). Anecdotally, the use of shorter measures has become more common as interventionists have tried to provide iterative feedback over time (requiring multiple assessments), and have moved from print and web-based delivery modes to mobile phones. Completing long questionnaires on mobile phones presents usability issues, and fails to capitalise on the advantages of real-time assessments that mobile devices can provide. Regardless of the degree of iterativity and the delivery modes used, though, greater attention should be paid to measurement of psychological constructs in computer-tailored interventions. This concerns both commonly used approaches and novel possibilities. Without sufficient information about the input used for tailoring, it is hard to gain more insight into whether tailoring output consists of relevant and well tailored messages.

Measurement as a fundamental issue

The latter touches upon a fundamental issue in psychology and related fields, including health psychology, health communication and behavior change science, as measurement of psychological constructs suffers from severe problems. That is, validated questionnaires often violate conditions required for validity (Hussey & Hughes, 2020). Fried (2017), for example, shows how commonly used 'validated' depression scales measure different aspects of depression. Also, results from typical psychometric analyses are not informative regarding an instrument's validity (Maul, 2017). For example, use of such analyses may fall short of providing rigorous, potentially falsifying tests of relevant hypotheses. Some of these underlying problems, specifically applied to explaining behaviour in behaviour change science, have been explained elsewhere in more detail (Peters & Crutzen, 2017). In short, most theories in psychology are lax when it comes to accuracy and precision of their definitions and operationalizations. This causes problems such as those identified in the aptly named article "The confounded self-efficacy construct" (Williams & Rhodes, 2016) and in the article by Fried (2017) cited earlier: researchers use terms such as 'attitude', 'habit', and 'intrinsic motivation' without having a sufficiently accurate definition to accompany that label, let alone rigorous and comprehensively developed instructions for how to develop measurement instruments for those constructs.

On the one hand, this means that there is a fundamental issue that needs to be solved. On the other hand, psychology in general, and health psychology, health communication and behavior change science more specifically, is applied to target (health) problems that cannot wait for this. So, the science of behaviour change needs to move simultaneously 'slow' and 'fast' (cf. Armitage, 2015). 'Slow' in the sense of working towards solutions to address underlying problems of measurement in psychology, such as unequivocal definition and measurement of psychological constructs without the need for central curation and oversight (Peters, 2020). 'Fast' in the sense that behavior change interventions will be developed meanwhile, as there is a pressing need to reduce morbidity and mortality related to human behaviour (Ritchie & Roser, 2019). This article focuses on the latter; how can we reduce participant burden, and subsequently increase engagement and reduce attrition, in currently developed behavior change interventions. More specifically, how can we do this by using novel possibilities for measuring psychological constructs.

Novel possibilities for measuring psychological constructs in computer-tailored interventions

If we look at possibilities to measure psychological constructs, then they can be presented in a variety of dimensions that reflect an underlying continuity (Peters & Crutzen, 2017). Looking at the dimension of drivenness, for example, on the one hand of the spectrum there is the use of questionnaires to assess psychological constructs and provide feedback based on the pre-specified tailoring rules. These rules are expert-driven (e.g., informed by theory or another rationale intervention developers have in mind), meaning that the rules are specified in advance. In current practice, the participant burden of expert-driven questionnaires is high because of the need to complete relatively long questionnaires. On the other hand of the spectrum there is, for example, the possibility to infer psychological constructs based on routinely collected data on online behaviour (e.g., Likes on Facebook; Kosinski et al., 2014). With routinely collected data, inferences are made based on a data-driven approach. As a result, the participant burden is relatively low given that no active contribution from the participant is needed. In short, drivenness (expert-data), and participant burden (high-low) are two dimensions on which possibilities for measuring psychological constructs vary. In the following, we will describe the possibilities to reduce participant burden, both applying a data-driven approach using routinely collected data and applying an expert-driven approach using purposively sampled data, yet using novel methodologies to reduce the associated high participant burden.

Deriving information about psychological constructs using routinely collected data

Devices and sensors are increasingly used in all aspects of everyday life and the amount of data that is generated and available for profiling users is staggering. The International Data Corporation estimated that there will be more than 59 zettabytes of data created and captured in 2020, with current trends suggesting the amount of data will double every four years (IDC, 2020). Undeniably, it is already common practice to utilise this data for audience segmentation. Companies like Google and Facebook, for example, facilitate targeted advertising by tracking what articles people read, recent purchases they have made, and even the content of their private emails and messages. Implicit in this approach is that they can obtain proxy measures of the person in terms of interests, desires and needs, and thus increase advertising efficacy by targeting those most susceptible or likely to find the advertisement relevant (Bidargaddi et al., 2017).

Analogous efforts to derive information about the person using routinely collected data are underway in psychiatry and personality psychology (Azucar et al., 2018; Bidargaddi et al., 2017). As with computer-tailoring, measurement of mental health symptoms and personality have traditionally been collected using questionnaires. For the field of psychiatry, utilising routinely collected data offers the potential to collect more temporally valid assessments of mood and symptom severity, and thus potentially offer more timely and targeted interventions. For example, a pilot study that tracked patients with bipolar disorder over twelve months found that clinical symptoms were related to objective smartphone measurements. More specifically, cell tower movements and call logs, which were described as proxy measures for physical activity and social communication, respectively (Beiwinkel et al., 2016). For the field of personality psychology, assessments utilising routinely collected data may also have public health benefits (e.g., tailoring health interventions to increase adoption and user experience). Although this area of research is still relatively young, many studies have been conducted investigating associations between online social media behaviours (e.g., using digital footprints such as likes, language used, pictures) and personality. A recent meta-analysis of 16 studies suggested that the overall strength of association (i.e., meta-analytic correlations) between automatically collected social media data and the big five personality traits ranges from 0.29 (agreeableness) to 0.40 (extraversion), which is in line with standard "correlational upper limits" for behaviour to predict personality (Azucar et al., 2018). As the strength of the association was improved when multiple digital footprints were included versus the use of a single type of digital footprint, the authors were optimistic that precision would improve as the field progresses and access to large datasets evolves.

These examples raise the question of how routinely collected data could be used in the context of delivering computer-tailored behaviour change interventions. Given the widespread use of audience segmentation commercially, one obvious application could be the identification of people who could benefit from an intervention (e.g., those with low mood in case of a mental health intervention). Given the popularity of social media platforms, targeting interventions based on social media footprints could significantly increase the reach of computer-tailored interventions, including reaching those who are not yet contemplating behavioural changes but may benefit from doing so based on their digital footprint. It also seems possible that at least some constructs that are typically assessed in order to provide tailored information could be approximated from routinely collected data. For example, it may be possible to deduce exercise habits using a smartphone by combining automatically collected data on behaviour frequency (e.g., using accelerometry, gps or movements between cell towers) with data on contextual cues (e.g., location, time of day, interactions with specific people). Likewise, constructs like intentions, attitudes and need for cognition could possibly be assessed based on browser history, focusing not just on what people click on, but what they avoid or do not attend to. To illustrate, if people's browser history shows web pages that mainly consist of (text accompanied with) pictures to take a relatively greater share than web pages with text only, this might be indicative of a lower rather than greater need for cognition (which might also be associated with, for example, educational level or age (Bruinsma & Crutzen, 2018)). This type of behavioural data might be particularly amenable to assessing aspects of psychological profiles that are less reflective in nature, such as implicit attitudes - a construct that is now usually measured by Implicit Association Tests (O'Shea & Wiers, 2020). While data collection is now relatively straight-forward, the intellectual challenge lies in considering how to model such high-definition data and derive meaningful summary statistics. In the context of developing proxy measures for computer-tailoring, this should be driven, at least in part, by specific scientific questions and hypotheses. This is equally true for purposively sampled data.

Deriving information about psychological constructs using purposively sampled data

This section explores methods of purposively assessing tailoring variables that move beyond the traditional questionnaires by developing questionnaires that are individually tailored in terms of content and length (e.g., applying computer-adaptive testing (CAT) methodology) or move towards more interactive multimedia-based approaches that entirely replace questionnaires (e.g., the use of images and serious games [i.e., games designed for a primary purpose other than pure entertainment]). While tailoring rules based on purposively sampled data remain expert-driven, the associated participant burden is much lower; something we will further illustrate in the following section using the three examples mentioned.

To briefly talk about developing individually tailored questionnaires first. When applying CAT, each questionnaire item is dynamically selected from a pool of items based on a measurement model (Smits et al., 2011). This results in a shorter questionnaire that is optimized for a specific individual and contains only items most likely to be relevant (e.g., most salient beliefs) for this particular person. This way, the questionnaire that serves as input for computer-tailored feedback becomes tailored in both length and content for each individual user. When applying CAT to mental health measurement, it was found that questionnaires can be reduced in length to one-third of the initial number of items (Smits et al., 2011). To the best of our knowledge, however, CAT has not yet been used in the context of computer-tailoring.

Second, the interest in serious games as assessment tools has been steadily increasing over the last several years in the domains of education, health, government and industry (Kato & De Klerk, 2017). This is owing to the perception that serious games can both promote user engagement (e.g., through interaction and multisensory environments) and provide more ecologically valid assessments, especially of skills and competencies (e.g., by measuring game behaviours that represent reactions, planning and prioritisation in real-time and "real like" environments) (De Klerk & Kato, 2017). For example, the game CancerSpace presents players (i.e., healthcare professionals) with real-world situations in which they must make care decisions similar to as they would in clinical practice. The game includes a number of interactions with patients in which the player must try to educate the patient and persuade him or her to undertake screening, thus providing insight into their knowledge, communication and problem solving skills (Swarz et al., 2010). A rising number of serious games have also been designed to both assess and train a person's cognitive functioning. The product BrainTagger, for example, has been designed to screen for delirium in older emergency patients. Each game is linked to a standard psychological task and its associated cognitive function (e.g., response inhibition) (Zhang & Chignell, 2020). In a similar vein, games have also been used to deliver cognitive bias modification training and assessment tasks online, with several already evaluated in the behaviour change field (Jayasinghe et al., 2020) and some commercial products widely available via app stores (Zhang et al., 2018).

As with the use of routinely collected data for assessment, the expertise required for advancing purposively collected game-based data for assessments is advancing but is still under development. In the game CancerSpace for example, the player's conversation choices are evaluated using pre-programmed decision trees (Swarz et al., 2010). This is akin to the expert-driven rules used in traditional computer-tailoring interventions. Whereas in BrainTagger, machine learning is used to adjust cognitive assessment scores by comparing differences in game parameters across tasks and individuals applying a data-driven approach. Establishing validity and the cost-benefit of using these assessment methods are additional key challenges (see Discussion section).

A lower hanging fruit may be the adoption of game-based elements into more traditional forms of assessment. For example, the use of avatar selection may be an engaging way to examine user self-perceptions, both real and ideal. This method would also lend itself to tailoring to a user profile (i.e., considering how elements of the person cluster together). A simplified example of how this approach could be utilised in a low cost way is highlighted in Text box 1.

To inform avatar development in an evidence-based way, or really, any profile-based tailoring method, person-based data collected from previous computer-tailoring studies could be examined for clusters. Identified clusters could then form the basis for avatars. For example, cluster analysis with data from 753 smokers who participated in an effectiveness trial of a web-based, computer-tailored smoking cessation programme based on smokers' baseline scores for pros and cons of quitting and quitting self-efficacy showed that among smokers in the preparation stage of change (i.e. motivated to quit smoking within one month), four clusters could be identified; Classic, Unprepared, Progressing and Disengaged Preparers (Smit et al., 2018). These clusters significantly differed with respect to all clustering variables, their gender, cigarette dependence and educational level. Most importantly, results suggest that smoking cessation interventions tailored to the preparation stage of change, i.e. the set of cognitions usually present in preparers, are only appropriate for the subgroup we defined as Classic Preparers. The other clusters might need different interventions as they display a different cognitive profile. Similarly, in a computer-tailoring intervention targeting post-treatment breast cancer survivors (Short et al., 2017), over 400 participants completed baseline and follow-up measures of psychological constructs, demographics and health status information using standard questionnaire items. This data could be used to examine how these variables cluster together, and importantly if clusters are related to intervention responsiveness and unmet needs. If so, tailoring based on these clusters in a future iteration of the intervention could be advantageous. Importantly this would reduce the burden associated with developing hundreds of iterations of intervention messages and may reduce 'tailoring waste' - i.e., message permutations that are developed but rarely delivered, or do little to increase relevance of information. By allowing users to select an avatar that corresponds to an evidence-based cluster, the burden of assessment could also be greatly reduced. Avatar-based tailoring will necessitate examining the extent to which avatar self-identification relates to current or ideal self-perceptions, and the extent to which this can be manipulated with intervention instructions. If both are achievable, avatars might have the added advantage of providing insights into discrepancies of self that the user would like to change (Klimmt et al., 2009; Meijer et al., 2020). Future research examining the utility of an avatar-based approach is therefore highly encouraged.

A third approach that could be considered is the replacement of standard questionnaire items with visual representations. This method has already gained traction in personality assessment, owing predominantly to perceptions that this approach can enhance engagement, reduce test taker fatigue (by requiring less attention to process), and may result in shorter test batteries due to the ability of images to provoke stronger reactions than text (Leutner et al., 2017; Meissner & Rothermund, 2015). There is some evidence to support these perceptions (e.g., Leutner et al., 2017), though as with all of the discussed methods validity still needs to be established. Research into the perceptions of these measures will also be needed.

Assessment To generate your avatar, please select which of the following images in each category best represents you over the last month in relation to your physical activity?			
Construct	**Intention**	**Self-efficacy**	**Social support**
Image options	*Avatar proudly wears goal and commitment badge (high)*	*Avatar is bounding over medium size hurdle (high)*	*Avatar surrounded by entourage (high)*
	Avatar is chilling out and a badge is collecting dust on the floor (low)	*Avatar is looking dwarfed by the hurdle (low)*	*Avatar is standing on its own (low)*

Tailoring to profile

Based on the choices made, the user will end up with 1 out of 8 possible avatars. The avatars can be used to generate computer-tailored feedback based on the three psychological constructs measured by creating the avatar.

Intention	Self-efficacy	Social support	Avatar profile
High	High	High	'You are set for success!'
High	High	Low	'You are almost ready for success, but let's ensure some more social support!'
High	Low	High	'You are almost ready for success, but let's get you some more confidence!'
High	Low	Low	'Your motivation is great, so let's ensure some social support and increase your confidence too!'
Low	High	High	'You are almost ready for success, but let's ensure you're sufficiently motivated!'
Low	High	Low	'Your confidence is great, so let's ensure you're sufficiently motivated and will receive some social support too!'
Low	Low	High	'Your social support is great, so let's ensure you're sufficiently motivated and increase your confidence too!'
Low	Low	Low	'Before making any changes to your physical activity, let's work on your motivation, confidence and social support!'

It is possible that the measures discussed in this section may be perceived as less trustworthy or credible than standard questionnaire-based approaches. Based on models of user experience and engagement (Crutzen et al., 2011; Short et al., 2015), this would lead to an increased likelihood of non usage of the intervention. On the flip side, if the measures are experienced as fun, or assessments lead to a greater sense of novelty or being more deeply understood, greater engagement could be expected.

Discussion

This article provides insight into how routinely collected data and novel self-assessment methods may be used in computer-tailoring to address key challenges in digital health (e.g., high participant burden, low engagement, high attrition). The examples presented from the literature (e.g., Swarz et al., 2010; Zhang & Chignell, 2020), and from our own creative efforts suggest that the development of novel proxy measures for measuring psychological constructs relevant to computer-tailoring is indeed possible. However, the extent to which these measures are valid and actually do reduce participant burden, increase engagement and have other potential benefits (e.g., facilitating profile-based tailoring) is speculative and needs further investigation. It also needs to be acknowledged that both scientific reasoning and creative efforts are needed to develop novel possibilities of measuring psychological constructs in computer-tailored interventions. Based on what has been achieved to date, and our own efforts in developing examples, it seems some psychological constructs (e.g., mood, personality) may be easier to capture and distinguish from other constructs than others (e.g., self-efficacy, social support). Our avatar example is one attempt to address this issue. The extent to which this approach actually does capture aspects of the person in a meaningful way that can be used for computer-tailoring also needs further investigation. It is hoped that this article serves as a stimulance for driving further momentum in this area. To this end, we next discuss some additional challenges to consider and describe recommendations for future research and practice.

Challenges of using novel possibilities for measurement

One of the advantages of utilising standard self-report questionnaires to measure psychological constructs for computer-tailoring is the simplicity of assessment. The background knowledge and skills to administer and interpret these standard assessments are also typically well aligned with the discipline expertise of those developing behaviour change interventions. Whereas, simplicity of collection and having the required expertise is less likely to be the case when drawing on routinely collected data and moving beyond standard self-report questionnaires.

When it concerns routinely collected data, first of all, data compilation can be complicated. Services and apps that collect data of interest are often owned and operated by businesses and therefore sit outside of mainstream health care and research. Second, where mainstream health data are available they are often in multiple silos. To fully capitalise on routinely collected data the ability to aggregate personal data sets from these sources will be necessary (Bidargaddi et al., 2017). Advanced technical and modelling expertise will also likely be needed. While the formation of multi-disciplinary teams is generally considered a pro, especially in the context of solving complex problems, working in such teams presents new challenges (e.g., overcoming field specific jargon) and sufficient time and willingness is needed to build a productive working relationship. Moreover, the ethical, legal, and social landscape varies, depending upon the domain (e.g., clinical, research, government) in which routinely collected data are used. The businesses that collect data and have expertise in person-based assessment may have lower ethical standards than what would be accepted in health and medical research and service delivery (Bidargaddi et al., 2017). For example, personal characteristics intuited from social media data (i.e., characteristics not explicitly disclosed by individuals) have already been used to target political propaganda prior to elections (Cadwalladr, 2017), and the availability of strategies for identifying individuals based on vulnerable emotional states has already been communicated to advertisers (Levin, 2017). While the prospect of being able to target individuals who may benefit from a behaviour change intervention is exciting and could expand the reach of public health initiatives, the dangers associated with misuse should be carefully considered and managed. Therefore, across all domains, development and implementation of guidelines and best practices is helpful and we will elaborate upon this in the next section (using ethical guidelines for COVID-19 tracing apps as an example).

Similarly, the expertise that is required to purposively collect - and subsequently interpret - data through, for instance, game-based assessment methods (e.g., the avatar example we provided), is still under development. While current applications make use of both expert-driven decision trees (Swarz et al., 2010) and user-driven machine learning (Zhang & Chignell, 2020), the validity of these approaches may be compromised by engagement mechanics that are irrelevant to assessing the construct and thereby introduce additional error or noise (Kato & De Klerk, 2017). This has been proposed as a possible reason as to why gamified cognitive bias modification tasks have mixed findings (Boendermaker et al., 2016). Moreover, the development of game-based assessment methods as part of more traditional forms of assessment is also likely to bring about relatively high costs, which makes these novel assessment forms unlikely to be cost-effectiveness unless they are much more effective in reducing participant burden, increasing engagement, reducing attrition and as such ultimately increasing the effectiveness of computer-tailored health interventions, than the traditional assessment methods currently employed. With this issue of cost-effectiveness, however, come the challenges of defining the best outcome measure that can compare interventions across health behaviours, but is also sensitive to behaviour-specific changes resulting from the intervention, and of determining what increase in effects is required to justify the investments needed (Smit, De Vries, et al., 2015). This raises questions about whether metrics related to participant burden and intervention engagement and attrition would be sufficiently informative for the policy makers that are responsible for allocating limited funds and willingness to pay for each unit of effect when it concerns reducing participant burden, increasing engagement and/or decreasing attrition.

Recommendations for future research and practice

Whereas the examples presented in this article do suggest that routinely collected data and novel self-assessment methods may be useful for assessing psychological constructs relevant to computer-tailoring, the extent to which these measures are valid and actually do reduce participant burden, increase engagement, reduce attrition and have other potential benefits (e.g., facilitating profile-based tailoring) is speculative and needs further investigation. One of the most obvious steps to take is to compare the proposed assessment methods with traditional methods (e.g., a self-administered questionnaire). Whether such comparative attempts are, however, truly valuable is a concern the scientific community should be reflective about, as - as indicated before - even commonly-used questionnaires that would serve as comparison often violate conditions required for validity (Hussey & Hughes, 2020) and results from typical psychometric analyses may not be informative regarding an instrument's validity (Maul, 2017).

At the same time, however, there is the need to move 'fast' in the sense that digital behavior change interventions need to be developed with a reduced participant burden, increased engagement rates, reduced attrition and a wider reach, as there is a pressing need to reduce morbidity and mortality related to human behaviour (Ritchie & Roser, 2019). To establish whether the routinely collected data and novel self-assessment methods described in this article are able to respond to this need, future research efforts are required that focus on participants' perceived burden of completing the different measures and their engagement with the interventions that these measures are a part of. To illustrate this based on our avatar example, two versions of a computer-tailored intervention aimed at increasing physical activity may be created; one that includes the novel assessment method of self-efficacy and social support using the avatar and one that includes traditional questionnaire items pertaining to these two psychological constructs. Then, different approaches to research can be taken. For example, one may explore the time required to complete the different assessments and study participants' subjective experience regarding completion of the two assessments (e.g. in terms of perceived pleasantness and cognitive burden). Another example would be to assess intervention engagement after completion of various assessment methods. In both examples, specific attention could be paid to the linguistically diversity among test takers to provide evidence for the applicability of novel assessment methods across a broad range of possible intervention participants.

When it comes to recommendations for practice, one of the most pressing ones is the development and implementation of guidelines and best practices. A recent example are ethical guidelines for COVID-19 tracing apps (Morley et al., 2020). To be ethical, a contact-tracing app must abide by four principles: it must be necessary, proportional, scientifically valid and time-bound. These principles are derived from the European Convention on Human Rights, the International Covenant on Civil and Political Rights (ICCPR) and the United Nations Siracusa Principles, which specify the provisions in the ICCPR that limit how it can be applied. However, there are many ways in which an app can meet these principles. To address this gap, Morley et al. have synthesized 16 questions that designers, deployers and evaluators should answer. These questions are based on the principles mentioned above, but also how they translate into requirements (e.g., is it voluntary? does it require consent? is the purpose limited?). Transparency and informed consent are related to each other. When asking consent from participants in computer-tailoring, it should be explained that the intervention content provided to them (i.e. decision-making regarding content) depends on, for example, certain demographics and/or their social-cognitive profile (i.e. the logic behind it). In other words, the logic behind the decision-making should be explained (Crutzen et al., 2019). This raises questions about how to explain algorithm-based decisions to participants. We refer to Brkan (2018) for ways how to reconcile the potential recognition of the right to explanation with the transparency requirement. An important issue with data-driven approaches is that it can lead to new forms of discrimination in decision-making (e.g., based on gender or ethnicity). Such discriminatory consequences, however, can be mainly attributed to human bias and legal shortcomings. Therefore, suggested solutions include comprehensive auditing strategies, implementation of data protection legislation and transparency enhancing strategies (Favaretto et al., 2019).

Conclusion

Routinely collected data and novel self-assessment methods may be used in computer-tailoring to address key challenges in digital health (e.g., high participant burden, low engagement, high attrition), yet their application does not come without challenges. We have described how the proposed possibilities to measure psychological constructs may be used, as illustrated by concrete examples. The discussion of the challenges one may encounter when doing so and the recommendations for future research and practice are hoped to serve as a stimulance for driving further momentum in this area.

References

Armitage, C. J. (2015). Changing behaviour, slow and fast: Commentary on Peters, de Bruin and Crutzen. *Health Psychology Review, 9*, 30–33.

Azucar, D., Marengo, D., & Settanni, M. (2018). Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis. *Personality and Individual Differences, 124,* 150–159.

Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory. Prentice-Hall.

Beiwinkel, T., Kindermann, S., Maier, A., Kerl, C., Moock, J., Barbian, G., & Rössler, W. (2016). Using smartphones to monitor bipolar disorder symptoms: A pilot study. *JMIR Mental Health, 3,* e2.

Bidargaddi, N., Musiat, P., Makinen, V. P., Ermes, M., Schrader, G., & Licinio, J. (2017). Digital footprints: Facilitating large-scale environmental psychiatric research in naturalistic settings through data from everyday technologies. *Molecular Psychiatry, 22*, 164–169.

Boendermaker, W. J., Maceiras, S. S., Boffo, M., & Wiers, R. W. (2016). Attentional bias modification with serious game elements: Evaluating the shots game. *JMIR Serious Games, 4,* e20.

Brkan, M. (2018). Do algorithms rule the world? Algorithmic decision-making in the framework of the GDPR and beyond. Terminator or the Jetsons? *The Economics and Policy Implications of Artificial Intelligence.*

Broekhuizen, K., Kroeze, W., Poppel, M., Oenema, A., & Brug, J. (2012). A systematic review of randomized controlled trials on the effectiveness of computer-tailored physical activity and dietary behavior promotion programs: An update. *Annals of Behavioral Medicine, 44*, 259–286.

Bruinsma, J., & Crutzen, R. (2018). A longitudinal study on the stability of the need for cognition. *Personality and Individual Differences, 127,* 151–161.

Cadwalladr, C. (2017). *The great British Brexit robbery: How our democracy was hijacked.* The Guardian. www.theguardian.com/technology/2017/may/07/the-great-british-brexit-robbery-hijacked-democracy

Chevance, G., Baretta, D., Heino, M., Perski, O., & Klasnja, P. (2020). Characterizing and predicting person-specific, day-to-day, fluctuations in walking behavior. *SportRxiv, 10*.31236/osf.io/bzj6s.

Crutzen, R., Cyr, D., & De Vries, N. K. (2011). Bringing loyalty to e-health: Theory validation using three Internet-delivered interventions. *Journal of Medical Internet Research, 13*, e73.

Crutzen, R., Peters, G.-J. Y., & Mondschein, C. (2019). Why and how we should care about the General Data Protection Regulation. *Psychology & Health, 34*, 1347–1357.

Crutzen, R., Van der Vaart, R., Evers, A., & Bode, C. (2018). Public health, behavioural medicine, and eHealth technology. *In EHealth Research, Theory and Development: A Multidisciplinary Approach* (pp. 111–127). Routledge.

De Klerk, S., & Kato, P. M. (2017). The future value of serious games for assessment: Where do we go now? *Journal of Applied Testing Technology, 18,*

32–37.
De Vries, H., & Brug, J. (1999). Computer-tailored interventions motivating people to adopt health promoting behaviors: Introduction to a new approach. *Patient Education and Counseling, 36*, 99–105.
Duckworth, A. L., Gendler, T. S., & Gross, J. J. (2016). Situational strategies for self-control. *Perspectives on Psychological Science, 11*, 35–55.
Dunton, G. F. (2017). Ecological momentary assessment in physical activity research. *Exercise and Sport Sciences Reviews, 45*, 48–54.
Favaretto, M., De Clercq, E., & Elger, B. S. (2019). Big Data and discriminatino: Perils, promises and solutions. A systematic review. *Journal of Big Data, 6,* 12.
Fishbein, M., & Ajzen, I. (2010). Predicting and Changing Behavior: The Reasoned Action Approach. Taylor & Francis Group.
Fried, E. I. (2017). The 52 symptoms of major depression. Journal of Affective Disorders, 208, 191–197.
Ghalibaf, A. K., Nazari, E., Gholian-Aval, M., & Tara, M. (2019). Comprehensive overview of computer-based health information tailoring: A systematic scoping review. BMJ Open, 9, e021022.
Glasgow, R. E., Klesges, L. M., Dzewaltowski, D. A., Estabrooks, P. A., & Vogt, T. M. (2006). Evaluating the impact of health promotion programs: Using the RE-AIM framework to form summary measures for decision making involving complex issues. Health Education Research, 21, 688–694.
Hawkins, R. P., Kreuter, M. W., Resnicow, K., Fishbein, M., & Dijkstra, A. (2008). Understanding tailoring in communicating about health. Health Education Research, 23, 454–466.
Hussey, I., & Hughes, S. (2020). Hidden invalidity among 15 commonly used measures in social and personality psychology. Advances in Methods and Practices in Psychological Science, doi:10.1177/2515245919882903.
IDC. (2020). IDC's Global DataSphere Forecast Shows Continued Steady Growth in the Creation and Consumption of Data. https://www.idc.com/getdoc.jsp?containerId=prUS46286020
Jayasinghe, H., Short, C. E., Braunack-Mayer, A., Merkin, A., & Hume, C. (2020). Evidence regarding automatic processing computerized tasks designed for health interventions in real-world settings among adults: Systematic scoping review. Journal of Medical Internet Research, 22, e17915.
Kato, P. M., & De Klerk, S. (2017). Serious games for assesment: Welcome to the jungle. Journal of Applied Testing Technology, 18, 1–6.
Klimmt, C., Hefner, D., & Vorderer, P. (2009). The video game experience as 'true' identification: A theory of enjoyable alterations of players' self-perception. Communication Theory, 19, 351–373.
Kohl, L., Crutzen, R., & De Vries, N. K. (2013). Online prevention aimed at lifestyle behaviours: A systematic review of reviews. Journal of Medical Internet Research, 15, e146.
Kosinski, M., Bachrach, Y., Kohli, P., Stillwell, D., & Graepel, T. (2014). Manifestations of user personality in website choice and behaviour on online social networks. Machine Learning, 95, 357–380.
Krebs, P., Prochaska, J. O., & Rossi, J. S. (2010). A meta-analysis of computer-tailored interventions for health behavior change. Preventive Medicine, 51, 214–221.
Leutner, F., Yearsley, A., Codreanu, S.-C., Borenstein, Y., & Ahmetoglu, G. (2017). From Likert scales to images: Validating a novel creativity measure with image based response scales. Personality and Individual Differences, 106, 36–40.
Levin, S. (2017). Facebook told advertisers it can identify teens feeling 'insecure' and 'worthless'. The Guardian. www.theguardian.com/technology/2017/may/01/facebook-advertising-data-insecure-teens
Lewin, K. (1936). Principles of Topological Psychology. McGraw-Hill.
Maul, A. (2017). Rethinking traditional methods of

survey validation. Measurement: Interdisciplinary Research and Perspectives, 15, 51–69.
Meijer, E., Vangeli, E., Gebhardt, W. A., & Van Laar, C. (2020). Identity processes in smokers who want to quit smoking: A longitudinal interpretative phenomenological analysis. Health, 24, 493–517.
Meissner, F., & Rothermund, K. (2015). A thousand words are worth more than a picture? The effects of stimulus modality on the implicit association test. Social Psychological and Personality Science, 6, 740–748.
Millar, B. M. (2017). Clocking self-regulation: Why time of day matters for health psychology. Health Psychology Review, 11, 345–357.
Morley, J., Cowls, J., Taddeo, M., & Floridi, L. (2020). Ethical guidelines for COVID-19 tracing apps. Nature, 582, 29–31.
Nikoloudakis, I. A., Crutzen, R., Vandelanotte, C., Quester, P., Dry, M., Skuse, A., Rebar, A. L., Duncan, M. J., & Short, C. E. (2018). Can you elaborate on that? Addressing participants' need for cognition in computer-tailored health behavior interventions. Health Psychology Review, 12, 437–452.
Noar, S. M., Benac, C. N., & Harris, M. S. (2007). Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychological Bulletin, 133, 673–693.
O'Carroll, R. E. (2020). Self-regulation interventions – what do we know and where should we go? Health Psychology Review, 14, 159–164.
O'Shea, B. A., & Wiers, R. W. (2020). Moving beyond the relative assessment of implicit biases: Navigating the complexities of absolute measurement. Social Cognition, 38, s187–s207.
Peters, G.-J. Y. (2020). Decentralized construct taxonomies. PsyArXiv, doi:10.31234/osf.io/xebhn.
Peters, G.-J. Y., & Crutzen, R. (2017). Pragmatic nihilism: How a Theory of Nothing can help health psychology progress. Health Psychology Review, 11, 103–121.
Prochaska, J. O., & Velicer, W. F. (1997). The transtheoretical model of health behavior change. American Journal of Health Promotion, 12, 38–48.
Rhodes, R., McEwan, D., & Rebar, A. (2019). Theories of physical activity behaviour change: A history and synthesis of approaches. Psychology of Sport and Exercise, 42, 100–109.
Ritchie, H., & Roser, M. (2019). Causes of death. In Our World in Data. https://ourworldindata.org/causes-of-death
Rosenstock, I. M. (1974). The Health Belief Model and preventive health behavior. Health Education Monographs, 2, 354–386.
Short, C. E., DeSmet, A., Woods, C., Williams, S. L., Maher, C., Middelweerd, A., Müller, A. M., Wark, P. A., Vandelanotte, C., Poppe, L., Hingle, M. D., & Crutzen, R. (2018). Measuring engagement in e- & mHealth behaviour change interventions: Viewpoint of methodologies. Journal of Medical Internet Research, 20, e292.
Short, C. E., Rebar, A., James, E., Duncan, M., Courneya, K., Plotnikoff, R., Crutzen, R., & Vandelanotte, C. (2017). How do different delivery schedules of tailored web-based physical activity advice for breast cancer survivors influence intervention use and efficacy? Journal of Cancer Survivorship, 11, 80–91.
Short, C. E., Rebar, A. L., Plotnikoff, R. C., & Vandelanotte, C. (2015). Designing engaging online behaviour change interventions: A proposed model of user engagement. The European Health Psychologist, 17, 32–38.
Skinner, A. L., Attwood, A. S., Baddeley, R., Evans-Reeves, K., Bauld, L., & Munafò, M. R. (2017). Digital phenotyping and the development and delivery of health guidelines and behaviour change interventions. Addiction, 112, 1281–1285.
Smit, E. S., Brinkhues, S., De Vries, H., & Hoving, C. (2018). Subgroups among smokers in preparation: A cluster analysis using the I-Change model. Substance Use & Misuse, 53, 400–

411.
Smit, E. S., De Vries, H., Oberjé, E. J. M., & Evers, S. M. A. A. (2015). Easier said than done: Overcoming challenges in the economic evaluation of Internet-based lifestyle interventions. The European Health Psychologist, 17, 39–44.
Smit, E. S., Evers, S. M. A. A., De Vries, H., & Hoving, C. (2013). Cost-effectiveness and cost-utility of Internet-based computer tailoring for smoking cessation. Journal of Medical Internet Research, 15, e57.
Smit, E. S., Linn, A., & Van Weert, J. (2015). Taking online computer tailoring forward: The potential of tailoring message frame and delivery mode of online health behaviour change interventions. The European Health Psychologist, 17, 25–31.
Smits, N., Cuijpers, P., & Van Straten, A. (2011). Applying computerized adaptive testing to the CES-D scale: A simulation study. Psychiatry Research, 188, 147–155.
Swarz, J., Ousley, A., Magro, A., Rienzo, M., Burns, D., Lindsey, A. M., Wilburn, B., & Bolcar, S. (2010). CancerSpace: A simulation-based game for improving cancer-screening rates. IEEE Computer Graphics and Applications, 30, 90–94.
WHO. (2020). Global Strategy on Digital Health. https://www.who.int/DHStrategy
Williams, D. M., & Rhodes, R. E. (2016). The confounded self-efficacy construct: Review, conceptual analysis, and recommendations for future research. Health Psychology Review, 10, 113–128.
Wolfenden, L., Nathan, N., & Williams, C. M. (2015). Computer-tailored interventions to facilitate health behavioural change. British Journal of Sports Medicine, 49, 1478–1479.
Yardley, L., Spring, B. J., Riper, H., Morrison, L. G., Crane, D. H., Curtis, K., Merchant, G. C., Naughton, F., & Blandford, A. (2016). Understanding and promoting effective engagement with digital behavior change interventions. American Journal of Preventive Medicine, 51, 833–842.
Zhang, B., & Chignell, M. (2020). A framework for using cognitive assessment games for people living with dementia. IEEE International Conference on Serious Games and Applications for Health (SeGAH), 8, 10.1109/SeGAH49190.2020.9201813.
Zhang, M., Ying, J., Song, G., Fung, D. S., & Smith, H. (2018). Attention and cognitive bias modification apps: Review of the literature and of commercially available apps. JMIR Mhealth and Uhealth, 6, e10034.

Camille E. Short
Melbourne Centre for Behaviour Change, Melbourne School of Psychological Sciences,
University of Melbourne, Australia
Melbourne School of Health Sciences, University of Melbourne, Australia
camille.short@unimelb.edu.au

Eline S. Smit
Department of Communication Science, Amsterdam School of Communication Research/ASCoR,
University of Amsterdam, The Netherlands
E.S.Smit@uva.nl

Rik Crutzen
Department of Health Promotion, Care and Public Health Research Institute, Maastricht University,
The Netherlands
rik.crutzen@maastrichtuniversity.nl