



Setting Physical Activity Goals with a Virtual Coach: Vicarious Experiences, Personalization and Acceptance

Nele Albers¹ · Beyza Hizli¹ · Bouke L. Scheltinga² · Eline Meijer³ · Willem-Paul Brinkman¹

Received: 5 October 2022 / Accepted: 5 December 2022
© The Author(s) 2023

Abstract

Goal-setting is often used in eHealth applications for behavior change as it motivates and helps to stay focused on a desired outcome. However, for goals to be effective, they need to meet criteria such as being specific, measurable, attainable, relevant and time-bound (SMART). Moreover, people need to be confident to reach their goal. We thus created a goal-setting dialog in which the virtual coach Jody guided people in setting SMART goals. Thereby, Jody provided personalized vicarious experiences by showing examples from other people who reached a goal to increase people's confidence. These experiences were personalized, as it is helpful to observe a relatable other succeed. Data from an online study with a between-subjects with pre-post measurement design (n=39 participants) provide credible support that personalized experiences are seen as more motivating than generic ones. Motivational factors for participants included information about the goal, path to the goal, and the person who accomplished a goal, as well as the mere fact that a goal was reached. Participants also had a positive attitude toward Jody. We see these results as an indication that people are positive toward using a goal-setting dialog with a virtual coach in eHealth applications for behavior change. Moreover, contrary to hypothesized, our observed data give credible support that participants' self-efficacy was lower after the dialog than before. These results warrant further research on how such dialogs affect self-efficacy, especially whether these lower post-measurements of self-efficacy are associated with people's more realistic assessment of their abilities.

Keywords Goal-setting · Physical activity · Behavior change · Conversational agent · Self-efficacy · Motivation

Introduction

Lack of physical activity is one of the primary risk factors for cardiovascular disease, which causes an estimated 32% of deaths worldwide [1]. To support people in becoming more physically active, goal-setting is commonly used as it provides motivation and helps to stay focused on a desired outcome [2, 3]. As such, goal-setting is often part of eHealth applications (e.g., [4]), which have the potential to make behavior change support more effective and widely available [5–7]. However, for goals to be effective, they need to meet several criteria such as being aligned with other goals a person has [8] and being realistic and achievable [9].

We thus wanted to incorporate the setting of effective goals in eHealth applications for behavior change. To this end, we designed a goal-setting dialog for running or walking with the virtual coach Jody, as virtual coaches can provide guidance where traditionally therapists would have. Virtual coaches can also foster adherence to eHealth applications by improving engagement and connecting

✉ Nele Albers
n.albers@tudelft.nl

Beyza Hizli
beyzahizli@gmail.com

Bouke L. Scheltinga
b.l.scheltinga@utwente.nl

Eline Meijer
e.meijer@lumc.nl

Willem-Paul Brinkman
w.p.brinkman@tudelft.nl

¹ Intelligent Systems, Delft University of Technology, Delft, The Netherlands

² Biomedical Signals and Systems, University of Twente, Enschede, The Netherlands

³ Public Health and Primary Care, Leiden University Medical Center, Leiden, The Netherlands

with users [10]. In the goal-setting dialog, Jody asked users several questions to encourage the creation of SMART goals [11], which are goals that are specific, measurable, attainable, relevant, and time-bound. For example, Jody asked users why their goal was important to them based on a set of questions therapists use to find out what matters to their clients in their life [12]. Being helped to realize the reasons for wanting to achieve an outcome has previously been shown to lead to more personalized goals and better results [13].

Effective goals alone are, however, not sufficient for behavior change to be successful. This is because there is a variety of possible barriers relating to users' capability, opportunity, and motivation [14], such as a lack of self-efficacy [2, 15, 16]. One way to increase users' self-efficacy is to let them observe another person succeed [15], or, in other words, make vicarious experiences. According to social cognitive theory [17], there are four sources of self-efficacy for a behavior: mastery experiences, vicarious experiences, verbal persuasion, and somatic and affective states. Comparing these sources of self-efficacy, Warner et al [18] found that mastery experiences, vicarious experiences, and subjective perceptions of health had similar significant effects on exercise self-efficacy in a study with older adults. The meta-analysis by Ashford et al [19] further identified feedback on past or others' performance as producing the highest levels of self-efficacy in the context of promoting lifestyle and recreational physical activity, followed by vicarious experiences. Since goal-setting typically first takes place at the start of a behavior change intervention when people may not yet have any personal experience with the target behavior, using vicarious experiences to foster self-efficacy is intuitively appealing.

While it can be helpful to observe another person succeed, it matters how relatable this "other" is: a more relatable "other" is beneficial [15, 20]. Finding such a relatable "other" can be difficult in practice for underserved groups [21] such as older adults [18]. Given vicarious experiences from diverse people, behavior change applications can, however, make use of user data to find relatable others. In our goal-setting dialog, Jody, therefore, showed *personalized* examples of other people who had achieved a running or walking goal.

To evaluate this dialog, we conducted a study in which 39 people set a running or walking goal with Jody. Based on this study, we tested the following hypotheses:

- H1: People's self-efficacy is higher after the dialog with the virtual coach than before.
- H2: People's self-efficacy is higher after receiving personalized examples than after receiving generic examples.
- H3: The personalized examples are perceived as more motivating than generic examples.
- H4: People have a positive attitude toward the virtual coach.

In addition, we inquired what people found motivating about the examples from other people.

Materials & methods

We conducted our study in March 2022. The Human Research Ethics Committee of Delft University of Technology approved the study (Letter of Approval number: 1707), and we preregistered the study in the Open Science Framework (OSF) [22].

Study design

The study was set up as a mixed-design study with one between- and one within-subject factor. The between-subject factor was the type of examples shown (2 levels: personalized/generic), and the within-subject factor was time (2 levels: pre-/post-measurement).

Materials

We used the Qualtrics platform to host the online questionnaires and a Google Compute Engine to host the virtual coach.


Virtual coach

The virtual coach Jody was implemented in Rasa [23]. It introduced itself as being there to help users set a goal for becoming more physically active. After describing the benefits of physical activity and goal-setting, Jody asked users to provide an initial idea for their choice of either a running or a walking goal. This was to reduce the anchoring effect [24] of the examples from other people. Next, users were shown two examples from other people and subsequently asked to re-formulate their goal as specifically as possible. Afterward, users were asked to reflect on the relevance and attainability of their goal as well as to provide a deadline for reaching their goal. In case of very low or very high attainability, users were suggested to adapt their goal to make it less or more challenging. Finally, Jody summarized users' goals, upon which they could confirm their goal or change the behavior or its deadline. To increase the accessibility of the dialog, users could largely communicate by clicking on buttons with answer choices. A psychologist read through the dialog to ensure that the language and dialog structure were easy to follow. The implementation of the virtual coach [25] as well as a video of the dialog [26] can be found online.

Examples from other people

Each participant saw 2 out of 72 examples from other people. In a prior study, these examples were collected from 72 people, 4 each for 18 combinations of values for sex, age

Here is how the person from the first example introduces themselves:

 I'm 21 years old. I've practised volleyball for 7 years until I quit because of my school commitments. Then I've always kept myself trained: I run, walk, went to the gym for a while, but I realised that I prefer running through the streets over running on a treadmill.

This person achieved the following goal:


 I was able to run 4 km in a row. I achieved this by walking almost every day and started running slowly.

Fig. 1 Screenshot of part of the dialog showing an example of a person who achieved a running goal

range, and weekly exercise amount. These people from the prior study were asked to 1) introduce themselves to a new person they meet at a friend's gathering who is interested in their physical activity behavior, 2) describe a goal with regards to running or walking they achieved in the past year, and 3) describe how they achieved that goal. The latter was included because it can be motivating to read not only that somebody else achieved a goal but also how they managed to do so [27]. After the prior study, the collected examples were anonymized and corrected regarding spelling, punctuation, and grammar to allow them to be displayed in the goal-setting dialog. Fig. 1 shows how an example was presented in the dialog. The characteristics of the people who provided the examples are presented in Table 4 and all examples are available online [28].

Personalized Examples People in the “personalized” condition saw examples that were predicted to be most motivating by a linear regression model. To create this model, we conducted a second data collection study on Prolific in which 36 individuals each rated 18¹ examples on how similar to themselves they considered the corresponding person and how motivating they considered the goal and how the person achieved it. Afterward, the prediction model was set up to predict the motivation rating based on the absolute difference in individual characteristics between the person providing an example and the person seeing an example (Fig. 2). We chose 10 out of 22 variables associated with the predictors of behavior capability, opportunity, and motivation [29] using stepwise regression and correlation analysis (Table 1)². Furthermore, as not only *actual* but also *perceived* similarity can play a role [30], we used as three additional prediction variables the similarity ratings for three clusters of people from the examples based on two prototype examples per cluster. The three clusters were computed based on the similarity ratings and the prototypes were the two most centered

examples per cluster. These prototypes were not shown in the goal-setting dialog. The full model had a multiple R^2 of 0.23 and is shown in Table 1.

Generic Examples Using the same motivation ratings as in the “personalized” condition, people in the “generic” condition received a random selection of two of the three overall most motivating examples (Table 6).

Measures

Running or walking self-efficacy These were measured based on scales from 0 to 100, adapted from the Exercise Self-Efficacy Scale by McAuley [31] (see Online Resource 1).

Perceived motivational impact of examples Participants rated examples from other people on how motivating they perceived them on a scale from -3 to 3. 0 was labeled as “Neutral”.

Acceptance of the virtual coach We used an adaptation of the six questions by Provoost et al [32], each of which was rated on a scale from -3 to 3, with 0 labeled as “Neutral.”

Take-away from examples Participants provided a free-text response to the question “What can you take away from these examples for yourself?” after seeing the two examples from other people in the dialog.

Motivational factors from examples Participants gave a free-text response to the question “What do you find motivating about the running or walking goals that other people achieved?”

Transtheoretical Model (TTM)-stage for becoming physically active Using the World Health Organization's definition of physical activity [33] and guidelines on physical activity and sedentary behavior [34], we adapted the question by Norman et al [35] to physical activity to measure this stage of change.

¹ We randomly chose one example for each of the 18 combinations of values for sex (2 values), age range (3 values), and weekly exercise amount (3 values) from the people providing the examples.

² See Table 5 for how these individual characteristics were measured. Details on the individual items can be found in the provided dataset.

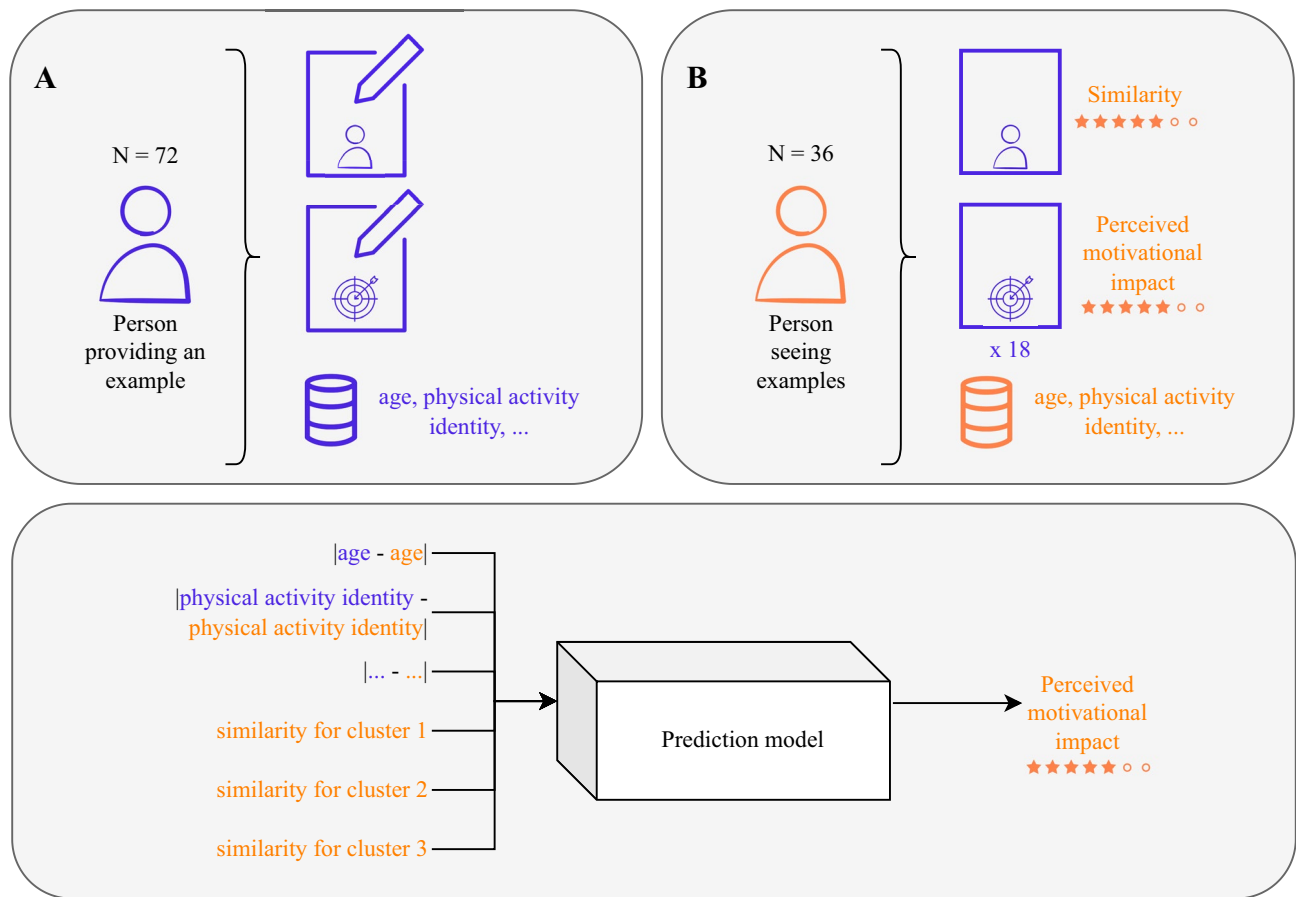


Fig. 2 Process of obtaining the prediction model used for choosing personalized examples. **A** 72 example people wrote about themselves and a goal they achieved. **B** 36 people each rated 18 examples on sim-

ilarity and perceived motivational impact. The results from **A** and **B** were used to obtain the prediction model

Table 1 Multiple linear regression model used to predict motivation ratings

Variable	Estimate	Beta	Std. Error	t value	Pr(> t)
Individual characteristics					
Age	-0.65	-0.08	0.28	-2.33	0.02*
Extraversion	-0.73	-0.08	0.33	-2.25	0.03*
Openness to experience	-0.79	-0.09	0.31	-2.58	0.01*
Household income	0.77	0.08	0.36	2.11	0.04*
Household size	0.61	0.07	0.33	1.84	0.07
Physical activity identity	-0.44	-0.05	0.39	-1.13	0.26
Sitting hours weekend day	0.57	0.06	0.37	1.55	0.12
TTM-stage	-0.53	-0.08	0.27	-2.00	0.05*
Godin leisure time activity level	-0.89	-0.20	0.18	-5.03	<0.001***
Running/walking self-efficacy	-0.82	-0.10	0.30	-2.72	<0.01**
Similarity ratings					
Similarity rating for cluster 1	0.13	0.09	0.05	2.43	0.02*
Similarity rating for cluster 2	-0.01	-0.01	0.05	-0.20	0.84
Similarity rating for cluster 3	0.40	0.25	0.06	6.79	<0.001***

TTM Transtheoretical Model

Signif. codes: 0 ***0.001 **0.01 *0.05

Participants

We aimed for a sample size of 36 based on 1) a power analysis using G*Power 3.1 [36] for ANOVA with repeated measures and within-between interaction leading to a sample size of 34 for a power of 0.8, an effect size of 0.25 (i.e., a medium effect size for ANOVA [37]), and an alpha of 0.05 and 2) wanting to recruit participants from 36 combinations of values for the variables smoking status³, sex, age, and weekly exercise amount. Eligible were people who were fluently English-speaking adults and who had not participated in the earlier data-gathering studies. 47 participants started the study and 39 were included in the analysis. Exclusion criteria are shown in Fig. 3. Participants were paid based on the minimum payment rules on Prolific (i.e., five GBP/hour).

Participant characteristics are shown in Table 2. Comparing Bayesian models with and without the condition as a predictor for each characteristic did not indicate systematic differences between the conditions for these characteristics.

Procedure

The study consisted of 1) a pre-questionnaire, 2) a goal-setting dialog with the virtual coach in which people saw either generic or personalized examples, and 3) a post-questionnaire (Fig. 3). Participants who successfully completed the pre-questionnaire were invited to the dialog about one week later. These participants were randomly assigned to the two conditions for the examples shown in the goal-setting dialog. Using an adaptation of the algorithm by Xiao et al [39], we aimed to balance the two conditions with regard to smoking status, self-efficacy for the preferred goal type, and two ratings of perceived similarity for clusters of people from the examples. The dialog lasted about seven minutes, after which participants were asked to complete the post-questionnaire.

Data preparation and analysis strategies

We created two index measures for the ratings of example types (i.e., personalized or generic) on perceived motivational impact as well as an index measure for the acceptance (Cronbach's α of 0.57, 0.56, and 0.87). Moreover, we reversed the scale of the TTM-stage for becoming physically active such that a higher value denotes a higher stage of change.

Instead of a frequentist analysis, we conducted a Bayesian analysis as it has been described as providing richer information [40]. We used Bayesian paired *t*-tests on the pre- and

Table 2 Participant characteristics for the two conditions

Characteristics	Generic	Personalized
Number		
- n	20	19
Age		
- mean (SD)	44.4 (17.7)	38.7 (14.6)
Sex		
- Female, n (%)	10 (50.0%)	10 (52.6%)
- Male, n (%)	10 (50.0%)	9 (47.4%)
Godin leisure time activity		
- mean (SD)	36.4 (31.7)	31.0 (22.3)
Running/walking self-efficacy*		
- mean (SD)	74.1 (30.5)	65.5 (24.0)
Sitting hours weekend day		
- mean (SD)	8.6 (3.7)	6.8 (4.3)
Smoking frequency		
- Less than once per day, n (%)	11 (55.0%)	13 (68.4%)
- At least once per day, n (%)	9 (45.0%)	6 (31.6%)
TTM-stage for becoming physically active		
- Precontemplation, n (%)	3 (15.0%)	0 (0.0%)
- Contemplation, n (%)	3 (15.0%)	8 (42.1%)
- Preparation, n (%)	6 (30.0%)	3 (15.8%)
- Action, n (%)	2 (10.0%)	3 (15.8%)
- Maintenance, n (%)	6 (30.0%)	5 (26.3%)
Weekly exercise amount		
- Less than 60 minutes per week, n (%)	8 (40.0%)	6 (31.6%)
- 60–150 minutes per week, n (%)	6 (30.0%)	6 (31.6%)
- More than 60 minutes per week, n (%)	6 (30.0%)	7 (36.8%)

SD Standard deviation, TTM Transtheoretical Model

*Running and walking self-efficacy were measured on scales from 0 to 100

post-measurement of self-efficacy for H1 and the perceived motivational impact ratings for the two types of examples for H3. For H2, we used a Bayesian two-sample *t*-test on the change in self-efficacy between the pre- and post-measurement, and for H4 a Bayesian one-sample *t*-test comparing the acceptance to a neutral value of zero. We evaluated the posterior probabilities that the hypotheses are true using the guidelines by Chechile [41] and their extension to probabilities below 0.5 by Andraszewicz et al [42]. These guidelines frame posterior probabilities as “bets” one can place with varying levels of confidence. The analyses were conducted in R with the Bayesian First Aid package [43].

For exploratory purposes, we additionally used the Bayesian First Aid package to compute the Pearson correlations between the TTM-stage for becoming physically active on the one hand and the change in self-efficacy between the pre- and post-measurement, motivational impact ratings for the two types of examples, and the acceptance on the other hand.

³ We used the smoking status as one balancing factor since Jody was developed as part of the Perfect Fit project [38], which aims to develop an eHealth application for quitting smoking and becoming more physically active.

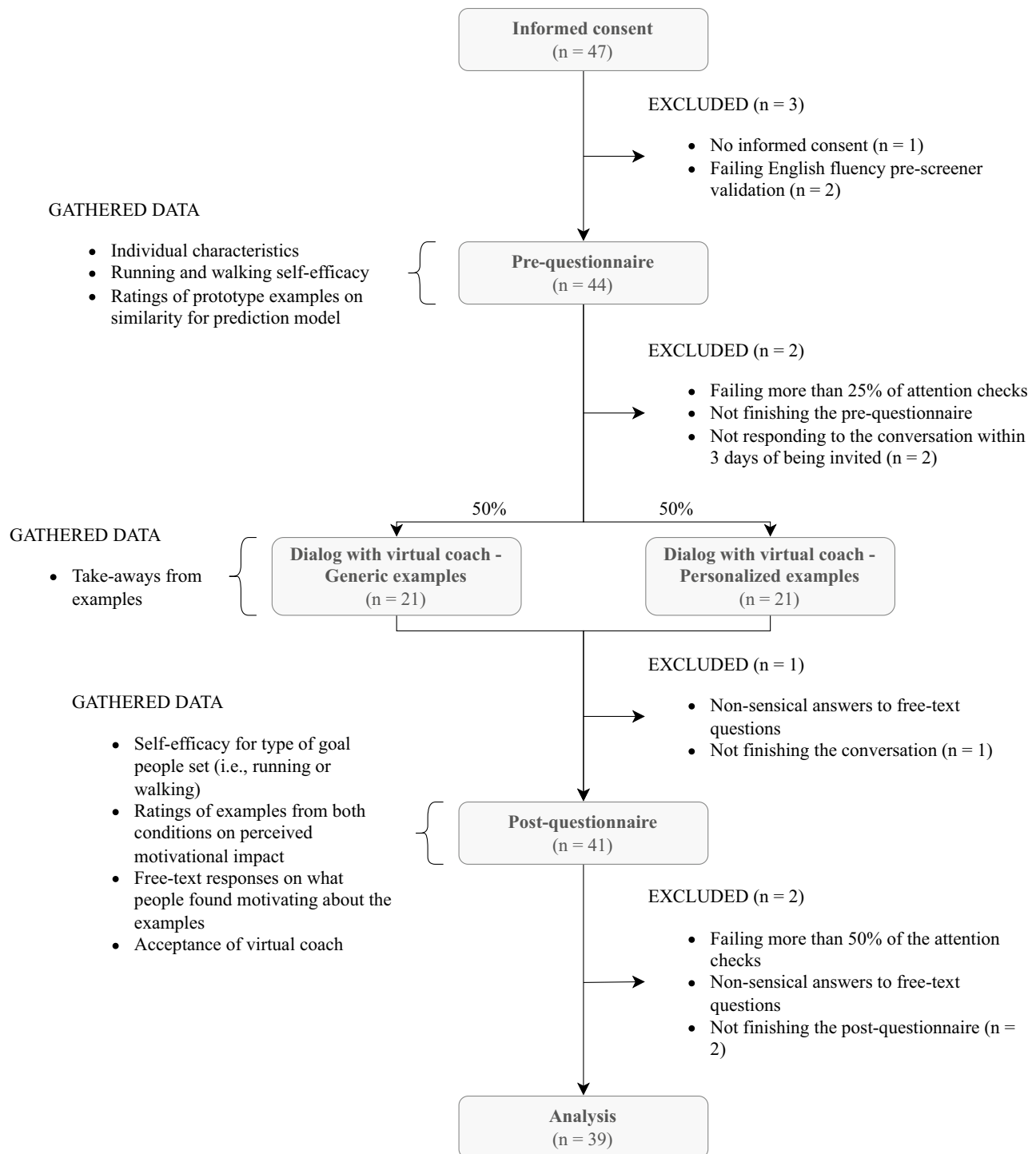
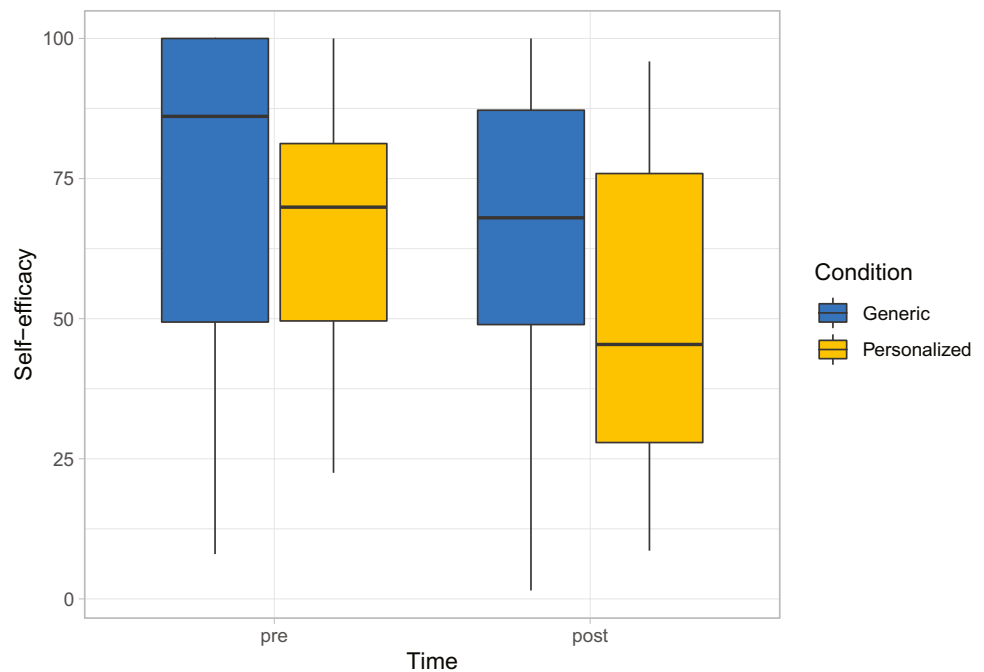


Fig. 3 Study design including study components, gathered data, and exclusion criteria

We further conducted a qualitative analysis of people's take-aways from the examples they saw as expressed in the dialog and what they found motivating about the examples as described in the post-questionnaire. Triangulating multiple sources serves to increase the validity of qualitative research [44]. Following the thematic analysis steps by Braun

and Clarke [45], BH first familiarized herself with the data before creating a draft coding scheme. She discussed the coding scheme with NA to arrive at a final coding scheme with nine codes. BH then coded all responses, using multiple codes if relevant. We found substantial agreement based on double-coding (Cohen's $\kappa = 0.79$) [46]. The responses from

Fig. 4 Self-efficacy before and after the dialog with the virtual coach for both conditions



one participant were afterward excluded from further analysis because the participant shared sensitive information. The codes assigned by BH were subsequently used to identify themes.

All data and analysis code are available in the online repository accompanying this article [47].

Results

Figure 4 shows that participants on average had a lower self-efficacy after than before the dialog. Quantifying this through our Bayesian analysis shows that the mean drop in self-efficacy was 12.38, which leads to a posterior probability of 0.002 that the self-efficacy is higher after the dialog than before (Table 3). This can be evaluated as a very strong bet against H1. Moreover, contrary to H2, Fig. 4 does not show a higher but rather a somewhat lower self-efficacy in the “personalized” condition compared to

the “generic” condition. Our Bayesian analysis suggests that it is not worth betting against H2 based on a posterior probability of 0.34 that H2 is true (Table 3). The perceived motivational impact of the personalized examples is, however, by on average 0.31 scale points higher than the one of the generic examples (Table 3). This leads to a posterior probability of 0.93 that H3 is true, which can be evaluated as a promising but risky bet. Regarding H4, the mean for people’s acceptance of Jody is 1.54, with the corresponding 95% credible interval ranging from 1.19 to 1.90 (Table 3). Based on a posterior probability of >0.99995, it is virtually certain that H4 is true.

Our exploratory analysis further provides no strong indication for correlations between the TTM-stage on the one hand and the change in self-efficacy and the motivational impact ratings for the two types of examples on the other hand (Table 7). For the acceptance, however, we see a small correlation [48] of 0.22 with a corresponding posterior probability of 0.92 that the correlation is greater than 0. Thus, we

Table 3 Results of Bayesian analyses for the four hypotheses

Parameter	Mean (SD)	95% CI	Post	Evaluation
H1: Self-efficacy				
Post - Pre	-12.38 (24.54)	[-20.36, -4.15]	0.002	Very strong bet against
H2: Change in self-efficacy				
Personalized - Generic	-3.57 (1.36)	[-21.03, 13.46]	0.34	Not worth betting against
H3: Perceived motivational impact				
Personalized - Generic	0.31 (1.19)	[-0.09, 0.72]	0.93	A promising but risky bet
H4: Acceptance				
Mean	1.54 (1.07)	[1.19, 1.90]	>0.99995	Virtually certain

SD Standard Deviation, *CI* Credible Interval, *Post* Posterior probability that the hypothesis is true

Fig. 5 The four themes for what people found motivating about examples from other people with sub-themes and participant quotes



obtain a promising but risky bet that people in higher stages of change have a more positive attitude toward Jody.

Figure 5 shows the four themes for what people found motivating about examples from other people with sub-themes and participant quotes. People were motivated by 1) specific, achievable, and challenging goals, 2) people who enjoyed their path to the goal, did not give up, achieved a goal step by step, and stayed consistent, 3) examples from people they could relate to, and 4) goal achievement itself. There are no clear differences between the conditions for the take-away responses (Fig. 6).

Discussion

The presented study examined the use of personalized vicarious experiences in a goal-setting dialog for physical activity with a virtual coach. Our results lead to a very strong bet against people's self-efficacy being higher after

than before the dialog. This is contrary to our expectations that the dialog and the vicarious experiences therein would increase self-efficacy [2, 15]. One possible explanation for these unexpected results is the Dunning-Kruger effect [49], according to which people with little experience or knowledge regarding a task tend to overestimate their competence. Thus, after the pre-measurement, thinking about their physical activity goal and ability to achieve it may have given people more knowledge about the task. This may have led people to judge their competence more realistically, in this case lower. Similar effects have been observed by Kang et al [50] and Rowland et al [51]. More precisely, participants had lower confidence to give lectures on elementary arithmetic and lower physical activity self-efficacy, respectively, after an intervention with vicarious experiences than before.

Moreover, mirroring the findings by Kang et al [50] that the drop in self-efficacy was stronger for a condition with more personalized vicarious experience, we also saw that it

is not worth betting against personalized examples leading to a higher increase in self-efficacy after the dialog than generic examples. This suggests that personalized experiences allowed people to get a more realistic picture of the task and their competence than generic ones. Interestingly, Kang et al [50] also found that the lower post-measurement of self-efficacy for the more personalized condition was more predictive of self-efficacy and perceived performance after a subsequent lecture than the post-measurement for the less personalized condition. As a more realistic assessment of one's abilities may help to set goals that are attainable, which is important for goals to be effective [11], these findings suggest that one should not only consider how high self-efficacy is, but also how realistic the assessment of abilities is that self-efficacy is based on. Whether personalized examples indeed lead to the setting of more attainable goals is an interesting question for future work.

Comparing personalized to generic vicarious experiences with regard to their perceived motivational impact, our data lead to a promising bet that personalized experiences are more motivating than generic ones. In light of our self-efficacy findings, the motivating effect was possibly not strong enough to also influence self-efficacy positively. To increase the effectiveness of the personalization, the prediction model with a multiple R^2 of only 0.23 might also need improvement. This might be done by incorporating other individual characteristics such as culture [52]. Notably, however, we have already tested 22 characteristics that are in general associated with the predictors of behavior capability, opportunity, and motivation. Hence, it may be worthwhile to focus on the content of the experiences rather than individual characteristics, which can be compared to taking a content-based rather than a collaborative filtering approach to recommender systems [53]. The results of our thematic analysis of what people found motivating about the experiences they saw can serve as a basis for this.

For our last hypothesis, we saw that people's attitude toward the virtual coach Jody is virtually certainly positive. This shows the potential of Jody for supporting people in setting effective goals, as the aspects covered in the attitude assessment such as a good relationship [54] and ease of use [55] are beneficial. However, given that people only had a single interaction with Jody, a novelty effect could have contributed to this positive attitude [56, 57].

Besides the low multiple R^2 of our prediction model used for the personalization of vicarious experiences, several further aspects warrant more research. First, some participants were in the precontemplation and maintenance stages of behavior change and thus either not yet interested in

becoming or already physically active (Table 2). We could hence speculate that the effect of the goal-setting dialog and the personalized vicarious experiences therein is stronger for a population who are either contemplating or preparing to change. The small positive correlation between the stage of change and the acceptance of Jody provides some support for this. However, we did not find much evidence concerning the change in self-efficacy and the motivational impact ratings.

Second, personalization effects might be stronger for people who are intrinsically motivated than for people who have a monetary incentive to participate. Specifically, since social judgment theory posits that a higher degree of involvement is associated with a larger latitude of rejection in which ideas are seen as objectionable [58], intrinsically motivated people may be more likely to "object" to generic examples. Third, we only assessed the effect of the dialog and the experiences on predictors of behavior, namely, self-efficacy and reflective motivation [29]. Yet, the dialog and experiences could also affect other predictors of behavior such as automatic motivation (e.g., impulses, reflex responses) [29]. It would, therefore, be interesting to assess how behavior itself is affected. Thereby, one needs to keep in mind that it may take some time before the setting of goals affects behavior as people think about how to reach their goal [59]. Adding support for such planning (e.g., [60, 61]) would also be valuable, as especially people with low self-efficacy may find it difficult to come up with an effective plan [2].

Conclusion

This study examined the effect of a goal-setting dialog and specifically personalized vicarious experiences in the context of a virtual coach for becoming more physically active. Vicarious experiences were provided by showing examples of how other people reached physical activity goals. The findings suggest that people see personalized examples as more motivating than generic ones and that people had a positive attitude toward the virtual coach. Moreover, contrary to what we hypothesized, the dialog negatively affected people's self-efficacy. Our data also provide some support that this negative effect was stronger for the personalized examples. These findings warrant further research on how self-efficacy can be improved in combination with goal-setting dialogs. It should thereby be taken into consideration whether these lower post-measurements of self-efficacy are associated with more accurate self-assessments of abilities.

Appendix

Characteristics of the people providing the examples

Table 4 Characteristics of the 72 people providing the examples

Characteristics	Value
Age	
- mean (SD)	42.83 (15.23)
- range	20–74
Sex	
- Female, n (%)	36 (50.00%)
- Male, n (%)	36 (50.00%)
Godin leisure time activity	
- mean (SD)	47.32 (32.14)
Household income	
- Less than £10,000, n (%)	12 (16.67%)
- £10,000–£15,999, n (%)	8 (11.11%)
- £16,000–£19,999, n (%)	6 (8.33%)
- £20,000–£29,999, n (%)	19 (26.39%)
- £30,000–£39,999, n (%)	8 (11.11%)
- £40,000–£49,999, n (%)	4 (5.56%)
- £50,000–£59,999, n (%)	5 (6.94%)
- £60,000–£69,999, n (%)	2 (2.78%)
- £70,000–£79,999, n (%)	4 (5.56%)
- £80,000–£89,999, n (%)	1 (1.39%)
- £90,000–£99,999, n (%)	1 (1.39%)
- £100,000–£149,999, n (%)	1 (1.39%)
- More than £150,000, n (%)	1 (1.39%)
Household size	
- 1, n (%)	10 (13.89%)
- 2, n (%)	11 (15.28%)
- 3, n (%)	20 (27.78%)
- 4, n (%)	21 (29.17%)
- 5, n (%)	5 (6.94%)
- 6, n (%)	4 (5.56%)
- 7, n (%)	1 (1.39%)
Personality	
- Extraversion, mean (SD)	4.17 (1.43)
- Openness to experience, mean (SD)	5.13 (1.13)
Running/walking self-efficacy*	
- mean (SD)	72.98 (23.72)
Sitting hours weekend day	
- mean (SD)	6.06 (3.43)
Smoking frequency	
- Less than once per day, n (%)	5 (6.94%)
- At least once per day, n (%)	67 (93.06%)
TTM-stage for becoming physically active	
- Precontemplation, n (%)	2 (2.78%)
- Contemplation, n (%)	16 (22.22%)
- Preparation, n (%)	11 (15.28%)
- Action, n (%)	18 (25.00%)

Table 4 (continued)

Characteristics	Value
- Maintenance, n (%)	25 (34.72%)
Weekly exercise amount	
- Less than 60 minutes per week, n (%)	24 (33.33%)
- 60–150 minutes per week, n (%)	24 (33.33%)
- More than 150 minutes per week, n (%)	24 (33.33%)

SD Standard Deviation, TTM Transtheoretical Model

*Running and walking self-efficacy were measured on scales from 0 to 100

Individual characteristics from prediction model

Table 5 Individual characteristics from the linear regression model used to predict motivation ratings and how they were measured

Characteristic	Measurement
Age	Computed by Prolific based on the date of birth
Extraversion	Big-Five personality dimension [63] measured with 7-point Likert scales
Openness to experience	Big-Five personality dimension [63] measured with 7-point Likert scales
Household income	Based on the question “What is your total household in-come per year, including all earners in your household (after tax) in GBP?” on Prolific
Household size	Based on the question “Including you, how many people live in your household?” on Prolific
Physical activity identity	Based on adapting the Exercise Identity Scale by [64] to physical activity (e.g., by replacing “exercise” with “physical activity”)
Sitting hours weekend day	Based on the International Physical Activity Questionnaire (IPAQ) [65]
TTM-stage	TTM-stage for becoming physically active based on adapting the question by Norman et al [35] to physical activity using the World Health Organization’s definition of physical activity [33] and guidelines on physical activity and sedentary behavior [34]
Godin leisure time activity level	Based on Godin [66]
Running/walking self-efficacy	Based on an adaptation of the Exercise Self-Efficacy Scale by citetmcauley1993 self (see Supplementary Information 1)

TTM Transtheoretical Model

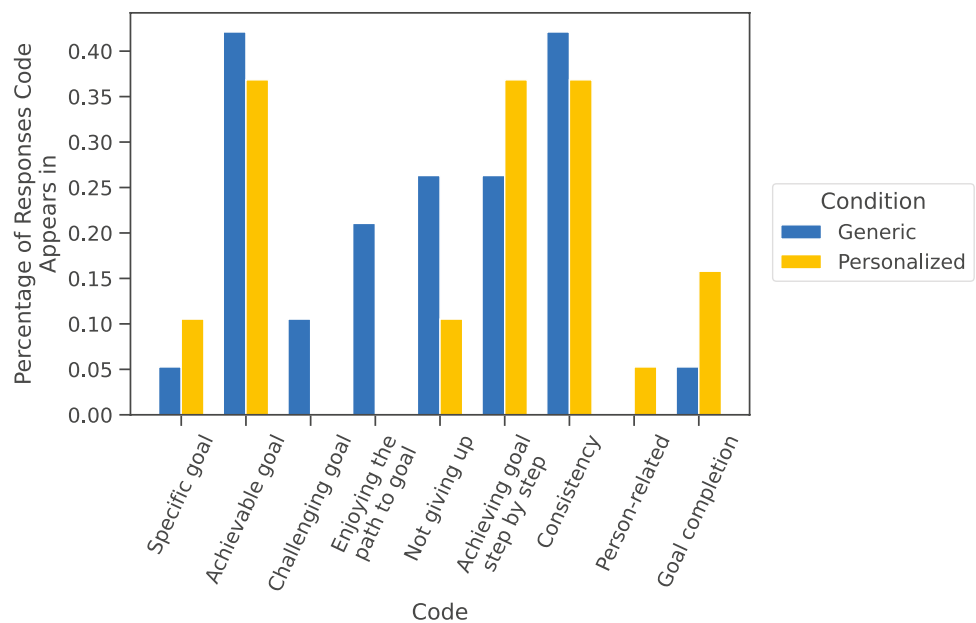
Generic examples

Table 6 The generic examples: People in the “generic” condition received two of the three overall most motivating examples. Each example consists of an introduction of the example person as well as a description of the goal and how the example person achieved the goal

Introduction	Goal and How They Achieved It
I enjoy the outdoors. My hobbies are building puzzles and riding horses. I’m easy to talk to and I listen well.	I have walked more than 10.000 steps in half a day. I achieved this by walking everywhere I needed to be.
Hello, I like to meet new people. One of my hobbies is watching tv. I am trying to stay in shape by walking.	I set a goal to walk every day regardless of weather factors for 1 straight month, and I achieved the goal. I achieved this by setting all my feelings aside and forced myself to ensure I met this goal.
I started having walks around country lanes which is nicer than walking alone on a road. I prefer to look at sheep and cows rather than cars and buses.	I increased the time that I am walking by walking further. I achieved this by building up my walks by doing a little bit more each month.

Code frequencies per condition

Fig. 6 Percentage of take-away responses each code from the coding scheme appears in per condition



Correlations between outcome measures and stage of change

Table 7 Results of Bayesian analyses of Pearson correlations between the TTM-stage for becoming physically active on the one hand and the outcome measures of change in self-efficacy between the pre- and

post-measurement, motivational impact ratings for the two types of examples, and the acceptance on the other hand

Outcome Measure	Mean (SD)	95% CI	Post > 0	Evaluation
Change in self-efficacy	0.00 (0.16)	[-0.31, 0.32]	0.49	Not worth betting against
Motivational impact generic examples	0.04 (0.16)	[-0.28, 0.36]	0.60	Not worth betting on
Motivational impact personalized examples	0.04 (0.17)	[-0.28, 0.37]	0.59	Not worth betting on
Acceptance	0.22 (0.16)	[-0.08, 0.53]	0.92	A promising but risky bet

TTM Transtheoretical Model, SD Standard deviation, CI Credible interval

Post > 0, Posterior probability that the correlation is greater than 0

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10916-022-01899-9>.

Acknowledgements This work is based on the study conducted for the publicly available Master's thesis by Hizli [62]. The authors would like to acknowledge the help they received from Kristell Penforinis in designing the goal-setting dialog, and from Millen van Osch in double-coding the free-text responses.

Author contributions All authors contributed to the study conception and design. Material preparation and data collection were performed by B.H. under the supervision of N.A. and W.B.; the data analysis was performed by B.H. and N.A.. B.S. and E.M. further checked all analysis code and data documentation. The first draft of the manuscript was written by N.A. and all authors commented on previous versions of the manuscript. All authors approved the final manuscript.

Funding This work is part of the multidisciplinary research project Perfect Fit, which is supported by several funders organized by the Netherlands Organization for Scientific Research (NWO), program Commit2Data - Big Data & Health (project number 628.011.211). Besides NWO, the funders include the Netherlands Organisation for Health Research and Development (ZonMw), Hartstichting, the Ministry of Health, Welfare and Sport (VWS), Health Holland, and the Netherlands eScience Center.

Declarations

Ethics approval The study was approved by the Human Research Ethics Committee of Delft University of Technology (Letter of Approval number: 1707).

Informed consent Informed consent was obtained from all participants included in the study.

Conflicts of interest The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes

were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. World Health Organization (2021) Cardiovascular diseases (cvds). [https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))
2. Locke EA, Latham GP (2002) Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American psychologist* 57(9):705
3. Epton T, Currie S, Armitage CJ (2017) Unique effects of setting goals on behavior change: Systematic review and meta-analysis. *Journal of consulting and clinical psychology* 85(12):1182
4. Romeo A, Edney S, Plotnikoff R, et al (2019) Can smartphone apps increase physical activity? systematic review and meta-analysis. *Journal of medical Internet research* 21(3):e12053
5. Alpay LL, Henkemans OB, Otten W, et al. (2010) E-health applications and services for patient empowerment: directions for best practices in the netherlands. *Telemedicine and e-Health* 16(7):787–791
6. Liao Y, Wu Q, Tang J, et al (2016) The efficacy of mobile phone-based text message interventions ('happy quit') for smoking cessation in china. *BMC Public Health* 16(1):1–11
7. Michie S, Yardley L, West R, et al (2017) Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. *Journal of medical Internet research* 19(6):e7126
8. O'Neil HF, Drillings M (1994) *Motivation: Theory and Research*. Routledge
9. Conzemius A, O'Neill J (2009) *The power of SMART goals: Using goals to improve student learning*. Solution Tree Press
10. Montenegro JLZ, da Costa CA, da Rosa Righi R (2019) Survey of conversational agents in health. *Expert Systems with Applications* 129:56–67

11. Doran GT, et al. (1981) There's smart way to write management's goals and objectives. *Management review* 70(11):35–36
12. Harteraad (????) Wat er toe doet. <https://watertoedoet.info/>
13. Lee MK, Kim J, Forlizzi J, et al (2015) Personalization revisited: a reflective approach helps people better personalize health services and motivates them to increase physical activity. In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp 743–754
14. Michie S, Van Stralen MM, West R (2011) The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implementation science* 6(1):1–12
15. Bandura A (1977) Self-efficacy: toward a unifying theory of behavioral change. *Psychological review* 84(2):191
16. Schunk DH, DiBenedetto MK (2021) Self-efficacy and human motivation. In: *Advances in motivation science*, vol 8, Elsevier, pp 153–179
17. Bandura A (1997) *Self-efficacy: The exercise of control*. Macmillan
18. Warner LM, Schüz B, Knittle K, Ziegelmann JP, Wurm S (2011) Sources of perceived self-efficacy as predictors of physical activity in older adults. *Applied Psychology: Health and Well-Being* 3(2):172–192
19. Ashford S, Edmunds J, French DP (2010) What is the best way to change self-efficacy to promote lifestyle and recreational physical activity? a systematic review with meta-analysis. *British journal of health psychology* 15(2):265–288
20. Ozer EM, Rowe J, Tebb KP, et al (2020) Fostering engagement in health behavior change: Iterative development of an interactive narrative environment to enhance adolescent preventive health services. *Journal of Adolescent Health* 67(2):S34–S44
21. Kariuki JK, Gibbs BB, Rockette-Wagner B, et al (2021) Vicarious experience in multi-ethnic study of atherosclerosis (mesa) is associated with greater odds of attaining the recommended leisure-time physical activity levels. *International Journal of Behavioral Medicine* 28(5):575–582
22. Hizli B, Albers N, Brinkman WP (2022b) Goal-setting dialogue for physical activity with a virtual coach. <https://doi.org/10.17605/OSF.IO/4DUWH>. <https://osf.io/4duwh>
23. Bocklisch T, Faulkner J, Pawlowski N, et al (2017) Rasa: Open source language understanding and dialogue management. *arXiv preprint* <http://arxiv.org/abs/1712.05181>
24. Furnham A, Boo HC (2011) A literature review of the anchoring effect. *The journal of socio-economics* 40(1):35–42
25. Hizli B (2022) Goal-setting dialogue for physical activity with a virtual coach: code. <https://doi.org/10.5281/zenodo.6647381>. https://github.com/PerfectFit-project/goal_setting_virtual_coach
26. Hizli B (2022) Example goal-setting dialog with the virtual coach jody. <https://youtu.be/7WJy7-H3QEQ>
27. Lockwood P, Jordan CH, Kunda Z (2002) Motivation by positive or negative role models: regulatory focus determines who will best inspire us. *Journal of personality and social psychology* 83(4):854
28. Hizli B, Albers N, Brinkman WP (2022) Data and code underlying the master thesis: Goal-setting dialogue for physical activity with a virtual coach. <https://doi.org/10.4121/20047328.v1>, https://data.4tu.nl/articles/dataset/Data_and_code_underlying_the_master_thesis_Goal-setting_dialogue_for_physical_activity_with_a_virtual_coach/20047328
29. Michie S, Atkins L, West R, et al. (2014) *The behaviour change wheel. A guide to designing interventions* 1st ed Great Britain: Silverback Publishing
30. Zhou S, Bickmore T, Paasche-Orlow M, et al (2014) Agent-user concordance and satisfaction with a virtual hospital discharge nurse. In: *International conference on intelligent virtual agents*, Springer, pp 528–541
31. McAuley E (1993) Self-efficacy and the maintenance of exercise participation in older adults. *Journal of behavioral medicine* 16(1):103–113
32. Provoost S, Kleiboer A, Ornelas J, et al (2020) Improving adherence to an online intervention for low mood with a virtual coach: study protocol of a pilot randomized controlled trial. *Trials* 21(1):1–12
33. World Health Organization (????) Physical activity. https://www.who.int/health-topics/physical-activity#tab=tab_1
34. World Health Organization (2020) *Who guidelines on physical activity and sedentary behaviour*. <https://apps.who.int/iris/bitstream/handle/10665/336656/9789240015128-eng.pdf?sequence=1&isAllowed=y>
35. Norman G, Benisovich S, Nigg C, et al (1998) Examining three exercise staging algorithms in two samples. In: *19th annual meeting of the Society of Behavioral Medicine*
36. Faul F, Erdfelder E, Buchner A, et al (2009) Statistical power analyses using g^* power 3.1: Tests for correlation and regression analyses. *Behavior research methods* 41(4):1149–1160
37. Cohen J (1988) *Statistical power analysis for the behavioral sciences*, second edition. Taylor & Francis Group, LLC
38. Meijer E, Penfornis K, Albers N, et al (2021) Perfect fit: een identiteitsgebaseerde virtuele coach die ondersteunt bij stoppen met roken en meer bewegen. *Nederlands Netwerk voor Tabaksonderzoek (NNvT) Congres*
39. Xiao L, Yank V, Ma J (2012) Algorithm for balancing both continuous and categorical covariates in randomized controlled trials. *Computer methods and programs in biomedicine* 108(3):1185–1190
40. Kruschke JK, Liddell TM (2018) *The bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a bayesian perspective*. *Psychonomic Bulletin & Review* 25(1):178–206
41. Chechile RA (2020) *Bayesian Statistics for Experimental Scientists: A General Introduction Using Distribution-Free Methods*. MIT Press
42. Andraszewicz S, Scheibehenne B, Rieskamp J, et al (2015) An introduction to bayesian hypothesis testing for management research. *Journal of Management* 41(2):521–543
43. Baath R (2014) Bayesian first aid: A package that implements bayesian alternatives to the classical *.test functions in r. In: *UseR! 2014 - the International R User Conference*
44. Nancy Carter R, Bryant-Lukosius D, Alba DiCenso R (2014) The use of triangulation in qualitative research. In: *Oncology nursing forum*, Oncology Nursing Society, vol 41, pp 545–7
45. Braun V, Clarke V (2006) Using thematic analysis in psychology. *Qualitative research in psychology* 3(2):77–101
46. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *biometrics* pp 159–174
47. Albers N, Hizli B, Scheltinga BL, Meijer E, Brinkman WP (2023) *Setting Physical Activity Goals with a Virtual Coach: Data and Analysis Code*, 4TU. Research Data, v1. <https://doi.org/10.4121/21806868.v1>
48. Cohen J (1992) A power primer. *Psychological bulletin* 112(1):155
49. Kruger J, Dunning D (1999) Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology* 77(6):1121
50. Kang N, Ding D, Van Riemsdijk MB, et al (2021) Self-identification with a virtual experience and its moderating effect on self-efficacy and presence. *International Journal of Human-Computer Interaction* 37(2):181–196
51. Rowland SA, Cohen MZ, Pullen CH, et al (2020) Vicarious experience to affect physical activity in women: A randomized control trial. *Western journal of nursing research* 42(4):286–292

52. Choi J, Lee HJ, Sajjad F, et al (2014) The influence of national culture on the attitude towards mobile recommender systems. *Technological Forecasting and Social Change* 86:65–79
53. Thorat PB, Goudar RM, Barve S (2015) Survey on collaborative filtering, content-based filtering and hybrid recommendation system. *International Journal of Computer Applications* 110(4):31–36
54. Zhang J, Oh YJ, Lange P, et al (2020) Artificial intelligence chatbot behavior change model for designing artificial intelligence chatbots to promote physical activity and a healthy diet. *Journal of medical Internet research* 22(9):e22,845
55. Venkatesh V, Morris MG, Davis GB, et al (2003) User acceptance of information technology: Toward a unified view. *MIS quarterly* pp 425–478
56. Croes EA, Antheunis ML (2021) Can we be friends with mit-suku? a longitudinal study on the process of relationship formation between humans and a social chatbot. *Journal of Social and Personal Relationships* 38(1):279–300
57. Sadeghi S, Gupta S, Gramatovici S, et al (2022) Novelty and primacy: a long-term estimator for online experiments. *Technometrics* pp 1–26
58. Sherif CW, Sherif M, Nebergall RE (1965) Attitude and attitude change: The social judgment-involvement approach. Saunders Philadelphia
59. Smith KG, Locke EA, Barry D (1990) Goal setting, planning, and organizational performance: An experimental simulation. *Organizational Behavior and Human Decision Processes* 46(1):118–134
60. Poppe L, Van der Mispel C, Crombez G, et al (2018) How users experience and use an ehealth intervention based on self-regulation: mixed-methods study. *Journal of medical Internet research* 20(10):e10,412
61. Beristain Iraola A, ´Alvarez S´anchez R, Petsani D, et al (2020) Virtual coaching for older adults at home using smart goal supported behavior change. In: *Companion Publication of the 2020 International Conference on Multimodal Interaction*, pp 251–255
62. Hizli B (2022) Goal-setting dialogue for physical activity with a virtual coach. Master’s thesis
63. Gosling S, Rentfrow P, Swann W (2003) A very brief measure of the big-five personality domains. *Journal of Research in Personality* 37:504–528
64. Anderson DF, Cychosz CM (1994) Development of an exercise identity scale. *Perceptual and motor skills* 78(3):747–751
65. Craig CL, Marshall AL, Sjstrm M, Bauman AE, Booth ML, Ainsworth BE, Pratt M, Ekelund U, Yngve A, Sallis JF, et al. (2003) International physical activity questionnaire: 12-country reliability and validity. *Medicine & science in sports & exercise* 35(8):1381–1395
66. Godin G (2011) The godin-shephard leisure-time physical activity questionnaire. *The Health & Fitness Journal of Canada* 4(1):18–22

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.