Experts Get Me Started, Peers Keep Me Going: Comparing Crowd- versus Expert-Designed Motivational Text Messages for Exercise Behavior Change

Roelof A. J. de Vries
Human Media Interaction
University of Twente
Enschede, The Netherlands
r.a.j.devries@utwente.nl

Cristina Zaga
Human Media Interaction
University of Twente
Enschede, The Netherlands
c.zaga@utwente.nl

Franciszka Bayer
Psychology, Health and Technology
University of Twente
Enschede, The Netherlands
f.bayer@student.utwente.nl

Constance H. C. Drossaert
Psychology, Health and Technology
University of Twente
Enschede, The Netherlands
c.h.c.drossaert@utwente.nl

Khiet P. Truong
Human Media Interaction
University of Twente
Enschede, The Netherlands
k.p.truong@utwente.nl

Vanessa Evers
Human Media Interaction
University of Twente
Enschede, The Netherlands
v.evers@utwente.nl

ABSTRACT

We present a comparative analysis of motivational messages designed with a theory-driven approach. A previous study [4] involved crowdsourcing to design and evaluate motivational text messages for physical activity, and showed that these peer-designed text messages aligned to behavior change strategies from theory. However, the messages were predominantly rated as motivating in the later stages of behavior change, not in the earlier stages, including those strategies intended for the earlier stages. We speculated that the peers that designed the messages aligned to the strategies did not have sufficient expertise to motivate people in earlier stages. Therefore, we replicated the study with experts. We found that for two of the strategies expert-designed messages were found more motivating in the earliest stage, while for several of the strategies peer-designed messages were rated more motivating for later stages. We conclude that when using these strategies in behavior change technology, expert-designed messages could be more motivating in the earliest stage, while peer-designed messages could be more motivating in the later stages.

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: User Interfaces - Theory and methods; J.4 Social and Behavioral Sciences: Psychology

PervasiveHealth '17, May 23–26, 2017, Barcelona, Spain

ACM ISBN 978-1-4503-6363-1/17/05…$15.00
https://doi.org/10.1145/3154862.3154875

Author Keywords

motivational messages; exercise adherence; health behavior change; Transtheoretical Model; processes of change; stages of change; pervasive health technology;

INTRODUCTION

Increasing someone's motivation to change their behavior is difficult, but recent developments in persuasive technology and systems for health behavior change have had promising results. The application areas are diverse, with systems that assist people to prevent obesity [7], assistive technology that supports healthy lifestyles [9], and technology-based interventions that promote physical activity [13]. A successful approach to support long-term behavior change with technology is to apply strategies from behavior change theories and models [6, 17]. One effective model for behavior change is the Transtheoretical Model (TTM) [20]. According to the TTM, behavioral change consists of five *stages of change* (Table 1). These stages demarcate people's willingness to change their behavior, ranging from long-term inactive (i.e., in the Precontemplation stage) to long-term active (i.e., in the Maintenance stage). The TTM also provides ten strategies to move through the stages of behavior change: the *processes of change* (Table 2). Each of the processes is associated differently with moving through the stages [15]. For example, Social Liberation (i.e., making someone aware of the opportunities in the environment to perform the desired behavior) will be more effective for an inactive person than for an already active person. Strategies to change someone's behavior should be adapted to that specific behavior and to the stage of behavior change that the person is in [22]. Stage-based interventions can be more effective than neutral interventions [14]. However, in what form should the strategies or interventions be communicated through technology? Using text-based messages (which we will refer to as 'text messages' in this paper), tailored to the stages of change has proven effective in various contexts, for

example physical activity [18]. However, it is essential to ground how these messages are designed or acquired. But there is no standardized method to translate behavior change strategies into practically applicable text messages. To overcome this limitation, research dedicated to discerning how to translate theories and models of behavior change (like the TTM) into practically applicable interventions (like text messages) and evaluating the effect of those interventions on the user is increasing.

For example, Redfern et al. [19] designed (based on behavior change techniques) 137 text messages tailored to the user's name. Patrick et al. [19] developed 3000 SMS and MMS messages for tailoring to the user's preferences on timing and frequency of the messages. In Kaptein et al. [8], two researchers designed 42 text messages for six strategies tailored to the user's susceptibility to persuasion. For most studies, it is the authors or other experts who designed the messages. However, peer-designed text messages can be more engaging and more relevant to the user in comparison to expert-designed text messages, as is shown by Coley et al. [2]. In a similar vein, in our previous work [4] we used crowdsourcing to collect peer-designed motivational text messages, and we aligned these messages to behavior change strategies (the ten processes of change) through the use of coding and a codebook. We evaluated how motivating these peer-designed strategies (the processes of change) were perceived across the self-assessed stages of change of the participants. We found that peer-designed text message strategies (aligned to the processes of change) were predominantly rated as motivating by people in the later stages of change, in which people are arguably easier to motivate because they are already changing their behavior. However, none of the text message strategies were found really motivating in the earlier stages, in which people are arguably hardest to motivate, because they are not yet willing to change their behavior. In this paper, we again focus on the validated motivational strategies (processes of change) from the TTM to ground the design of our motivational text messages for behavior change technology. We hypothesize that peers from our previous work did not have the expertise needed to motivate people across all the five stage of change. Therefore, we investigate whether text messages designed by experts will be perceived as more motivating in the earlier (harder) stages as well as in the later (relatively easier) stages. We evaluate how people rate the expert-designed motivational messages representing the ten processes, and compare this to the data from our previous study involving peers. We expect expert-designed messages to be perceived as more motivating than peer-designed messages. We hypothesize that:

H1: Expert-designed messages representing the processes of change will be perceived as more motivating than peer-designed messages representing the processes of change, across all five stages of change.

METHOD

We followed the approach of our previous work [4] aimed at eliciting motivational messages. We designed two surveys: one (online and paper-based) elicitation survey and one evaluation survey.

Stage of change
Precontemplation (PC): The individual is not willing to change in the foreseeable future (measured as the next 6 months). Individuals in this stage are mostly uninformed or demoralized.
Contemplation (C): The individual is willing to change in the next 6 months. Individuals in this stage are aware of some pros of behavior change, but are still more inclined to value the cons.
Preparation (P): The individual is willing to change in the foreseeable future (measured as the next month) and has already taken some small steps towards change (in the past year). Individuals in this stage usually have some plan on how to tackle this inactiveness.
Action (A): The individual has changed, but not longer than 6 months. Individuals in this stage have 'changed', but have not reached the duration which exemplifies real behavior change.
Maintenance (M): The individual has changed, longer than 6 months. Individuals in this stage have changed and are working not to relapse.

Table 1. The stages of change and a description, based on [20].

Processes of change
Consciousness Raising (CR): The individual seeks increased knowledge about the causes, consequences and cures for their problem behavior.
Dramatic Relief (DR): The individual's emotions about the problem behavior and possible solutions are evoked.
Environmental Reevaluation (ER): The impact that the individual's problem behavior has on their environment is reevaluated.
Social Liberation (SOL): Attempts are made to increase alternatives for the individual's former problem behavior.
Self-reevaluation (SR): Cognitions and emotions regarding the individual with respect to their problem behavior are reevaluated.
Self-liberation (SEL): The individual has the belief that he can change and commits to it by choosing a course of action.
Helping Relationships (HR): The individual seeks trust and open discussion about the problem behavior as well as support for the healthy behavior change.
Counterconditioning (CC): The individual substitutes positive behaviors for the individual's problem behavior.
Reinforcement Management (RM): Changes made by the individual are rewarded when in a positive direction or punished when in a negative direction.
Stimulus Control (SC): Stimuli that may cue a lapse back to the problem behavior are avoided and prompts for more healthier alternatives are inserted.

Table 2. The processes of change consisting of Experiential (first five) and Behavioral processes (last five) with a short description, based on [15].

Elicitation survey

In the online and paper-based elicitation survey, participants were asked to design motivational messages for different scenarios. We used five scenarios, each one describing a fictional person who was in one of the five stages of change (see Table 3 for an example of one of the five scenarios and for an example of designed messages). We reached out to fitness coaches, behavioral coaches and researchers in health psychology to design a maximum of six motivational messages for each of three randomly selected scenarios. In total, 377 messages were designed by 25 experts (average of 5 per scenario) of Dutch (16) and German (9) nationality (age between 21 and 62 years, M = 33.6, SD = 11.0; 17 female). To see in what way the content of these expert-designed messages reflected

Stage of change scenario and three expert-designed text messages
Precontemplation: "Consider a middle-aged person, with a steady personal life and solid friend foundation. This person lacks regular exercise in his/her daily life and is unwilling to consider starting with this, at least not within the next 6 months."
Consciousness Raising: "Regular exercise will make you less vulnerable for diseases."
Dramatic relief: "The longer you push exercise aside, the longer you are denying yourself a better quality of life."
Social Liberation: "Look around in your neighbourhood if you can find any facilities to do some sports."

Table 3. Example of a scenario used in the elicitation survey and examples of experts' responses coded as processes of change.

the processes of change, we translated the processes to distinct coding categories (identical to our previous approach, for more details about the coding see [4]). Two coders iteratively coded the messages and reached a Cohen's kappa of 0.62, which was deemed a substantial agreement [12]. To represent the categories in the follow-up online evaluation survey, we selected five representative messages for each of the ten categories (50 messages in total).

Evaluation survey

In the online evaluation survey, participants rated the expert-designed text messages on how motivating they perceived the messages to be. Respondents were recruited on Amazon Mechanical Turk[1] (AMT) and were redirected via a link to the survey hosted on SurveyMonkey[2] (compensation 1.5 US dollars). In total we had 341 respondents (age between 20 and 75 years, M = 38.8, SD = 12.2; 171 female). To ensure consistency and a high quality of responses, the AMT requirements were: respondents should have completed >1000 tasks on AMT, with >98% of these tasks approved and respondents had to be located in the United States. In the survey, respondents were presented a total of fifty expert-designed motivational messages, five representative messages for each of the ten processes of change categories. We asked the respondents to rate each message according to how motivating they found the message to be for themselves in their own situation. Each message was rated on a one-item scale from 1 (*"Very demotivating"*) to 5 (*"Very motivating"*) with 3 as neutral (*"Neither demotivating nor motivating"*). The order of the messages was randomized for each of the respondents. Afterwards, respondents' stage of change was measured with the validated 1-item stage of change measure for exercise[3] [16]. The Cronbach's alphas (Table 6) for the ten categories show that the five messages for each of the categories were internally reliable.

ELICITATION SURVEY RESULTS

We report how the 377 messages that experts designed were divided between the higher-order experiential and behavioral processes and between the ten different processes in Table 4. In total, experts came up with 147 experiential messages (39%)

[1] https://requester.mturk.com/

[2] https://www.surveymonkey.com/

[3] web.uri.edu/cprc/exercise-stages-of-change-short-form

Processes	peers (n = 2886)		experts (n = 377)			std. res.
of change	n (obs)	%	n (exp)	n (obs)	%	
Exp.	800	27.7	104.5	147	39.0	**4.16**
CR	138	4.8	18.1	32	8.5	**3.27**
DR	59	2.0	7.6	7	1.9	-0.22
ER	79	2.7	10.2	5	1.3	-1.63
SOL	11	0.4	1.5	14	3.7	**10.21**
SR	513	17.8	67.2	89	23.6	**2.66**
Beh.	2086	72.3	272.5	230	61.0	**-2.57**
SEL	939	32.5	122.8	79	21.0	**-3.95**
HR	238	8.2	31.0	38	10.1	1.26
CC	289	10.0	37.8	32	8.5	-0.94
RM	512	17.7	66.9	63	16.7	-0.48
SC	108	3.7	14.0	18	4.8	1.07

Table 4. Observed frequency (n (obs)) and percentages are reported for peer and expert messages, and also expected frequency (n (exp)) for expert message categories. Reported standardized residuals represent the difference between the observed and the expected expert frequency. Numbers in bold are $p < .05$.

Categories/Stage scenarios		PC-S	C-S	P-S	A-S	M-S	Total
Experiential	Count	53	29	52	5	8	147
	Expected	25.7	23.8	33.9	37.8	25.7	147
	Std. residual	**5.4**[3]	1.1	**3.1**[2]	**-5.3**[3]	**-3.5**[3]	
Behavioral	Count	13	32	35	92	58	230
	Expected	40.3	37.2	53.1	59.2	40.3	230
	Std. residual	**-4.3**[3]	-0.9	**-2.5**[1]	**4.3**[3]	**2.8**[2]	
Total	Count	66	61	87	97	66	377

Table 5. The distribution of all the codes over the two higher-order process categories (experiential and behavioral) and five stages of change scenarios. [1] $p < .05$, [2] $p < .01$, [3] $p < .001$

and 230 behavioral messages (61%). A Chi-square test (see Table 5) was carried out to see if the scenarios (i.e., the stages) the experts had to design messages for had an effect on the counts within the higher-order processes. The results show that there is a significant association between the stages and the higher-order processes ($\chi^2(4) = 131.733, p < .001$). The values of the standardized residuals are used to further interpret the results.

A positive value indicates there were more messages of that higher-order process than expected and a negative value points to less messages than expected ($p < .05$ for a z-value higher than 1.96 or lower than -1.96 [5]). The residuals show that for the Experiential processes in the Precontemplation ($z = 5.4$) and Preparation stage ($z = 3.1$), there is a significant overrepresentation of the processes, and in the Action ($z = -5.3$) and Maintenance ($z = -3.5$) stage there is a significant underrepresentation of the processes. For the Behavioral processes this is reversed, in the Precontemplation ($z = -4.3$) and Preparation stage ($z = -2.5$), there is a significant underrepresentation of the processes, and in the Action ($z = 4.3$) and Maintenance stage ($z = 2.8$) there is a significant overrepresentation of the processes. This means that there are more Experiential messages in the earlier stages and fewer in the later stages than the Chi-square model predicts, and there are fewer Behavioral

	peers			experts		
Coded category	*M*	*SD*	α	*M*	*SD*	α
Consciousness Raising	3.75	0.66	.76	3.67	0.61	.73
Dramatic Relief	3.02	0.94	.82	2.67	0.90	.78
Environmental Reevaluation	3.13	0.97	.88	3.13	0.69	.75
Social Liberation	3.15	0.69	.73	3.13	0.67	.82
Self-reevaluation	3.67	0.69	.73	3.68	0.66	.75
Self-liberation	3.74	0.59	.72	3.75	0.60	.69
Helping Relationships	3.69	0.67	.74	3.45	0.70	.72
Counterconditioning	3.51	0.61	.68	3.57	0.61	.68
Reinforcement Management	3.86	0.70	.85	3.96	0.71	.83
Stimulus Control	3.21	0.65	.73	3.17	0.66	.70

Table 6. Averages (M), standard deviations (SD), and Cronbach's alpha's (α) for all the evaluated motivational peer and expert-designed text message categories. Messages were rated on as scale from 1 ("Very demotivating") to 5 ("Very motivating") with a 3 as neutral ("Neither demotivating nor motivating"). (N = 350)

messages in the earlier stages and more in the later stages than the Chi-square model predicts. Unfortunately, there were not enough expert-designed messages to also investigate the lower-order (the ten processes) distribution over the stages.

To evaluate how the distribution of 377 expert messages over the processes was proportionately different from the distribution of the 2886 peer messages over the processes from our previous work [4], we performed a Chi-square goodness-of-fit test. The results show that there is a significant difference ($\chi^2(9) = 143.026, p < .001$) between the proportional distribution of the expert messages and the proportional distribution of the peer messages (Table 4). Again, the values of the standardized residuals are used to further interpret the results of the Chi-square test. As can be seen in Table 4, experts provide proportionately more messages than peers for Experiential processes (CR, SOL, SR), which are related to the earlier stages, and significantly less for Behavioral processes (SEL), which are related to later stages of change.

EVALUATION SURVEY RESULTS

The self-assessed stages of change of the 341 participants were as follows: 22 rated themselves in Precontemplation, 53 in Contemplation, 71 in Preparation, 72 in Action and 123 in Maintenance.

To investigate whether, across the stages, categories of expert-designed messages aligned to the processes were perceived as more motivating than the categories of peer-designed messages aligned to the processes, we combined our dataset of rated expert-designed messages ($N = 341$) with the dataset of rated peer-designed messages from our previous work [4] ($N = 350$). In Table 6 descriptive statistics are reported.

To test whether there were differences in how motivating people perceived the messages between the designer of the messages (expert or peer) for the text message categories across the stages of change of the participant, we ran a linear mixed-effects model analysis in R [21] with the lme4 package [1] and to output significant differences, the lmerTest package [11]. To accommodate for hierarchical completeness, our model

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
object	4	90782.34	90816.14	-45387.17	90774.34			
..1	13	90716.82	90826.67	-45345.41	90690.82	83.53	9	**0.0000**
..2	17	90703.82	90847.47	-45334.91	90669.82	21.00	4	**0.0003**
..3	18	90702.86	90854.96	-45333.43	90666.86	2.97	1	0.0850
..4	54	90685.12	91141.43	-45288.56	90577.12	89.74	36	**0.0000**
..5	63	90514.41	91046.77	-45194.20	90388.41	188.71	9	**0.0000**
..6	67	90512.61	91078.77	-45189.31	90378.61	9.80	4	**0.0440**
..7	103	90527.83	91398.20	-45160.92	90321.83	56.78	36	**0.0151**

Table 7. An analysis of variance comparison of incrementally built models. Starting from the baseline model (object), where value is predicted from the intercept and random effects of respondent and messages, adding category (1), stage (2), designer (3), category:stage (4), category:designer (5), stage:designer (6), and category:stage:designer (7).

included the motivational rating of the text messages as the outcome, and as fixed effects: the categories of the text messages (processes of change), designer of the message (expert or peer), the stages of change of the participant, the interaction between the categories and designer of the message, the interaction between the categories and the stage of the participant, the interaction between the designer of the message and the stage of the participant, and the three-way interaction between category, designer, and stage. The participants and messages were included as random effects. An analysis of variance comparison of the incrementally built models is reported in Table 7 (built with the xtable package [3]).

We report on a summary of the model (see Table 8 and 9, and in the Appendix, Table 11, 12, 13, and 14) with the Consciousness Raising (CR) category as the reference level (benchmark to which to compare the scores of the other categories relatively). To make the intercept more interpretable, we recoded the stages from 0 to 4 (e.g., stage 0 = Precontemplation, stage 4 = Maintenance). This means that the intercept score is now the score on the reference level while all other factors are zero, meaning designer of the message 0 (expert) and stage 0 (Precontemplation).

Results of the model

Looking at the main effects (see Table 9, row: Designer (peer)), we found that the expert or peer version of the messages does not make a significant difference in the rating of the CR category (intercept reference level) for stage 0 (Precontemplation). We are interested in all these main effects between the designer of the messages and all the categories across each of the stages. The summary of the model, however, reports the significance of these relations only *in comparison to* the reference level category CR (see Appendix Table 11, 12, 13, and 14). Through the interaction estimates, the summary of the model does report on the relationship between categories and the designer of the message and the stages, because the estimate for the designer of the message, in relation to a selected stage and a selected category other than the reference category is equal to the estimate of the designer or the message for the reference category plus the estimate for the interaction between the designer and the selected category plus the estimate for the designer and the selected stage plus the estimate for the designer, the selected category and the selected stage. However, this does not incorporate the significance of the relation

Random effects	variance	*SD*
RespondentID (Intercept)	0.226	0.475
Text Message (Intercept)	0.030	0.173
Residual	0.755	0.869

Table 8. The variance and standard deviation of the random effects of the model: respondents and text messages. Number of observations: 34550, number of respondents: 691 (350 + 341), number of text messages: 50.

between the designer of the messages and a category for that stage, only the significance of those relations in comparison to the reference level category. To see if there were significant main effects for the designer of the message on the other nine categories across the different stages, we ran the same model with changed reference levels (see Table 10) for the categories and for the stages. Note that the cell in the first row and first column of Table 10 reports the identical estimate to the estimate reported for the designer of the message in Precontemplation stage (0 level) in Table 9, where we also report the standard error.

A statistically significant difference was found between the designer of the message (expert or peer) and the categories within four out of five stages. In Table 10, the estimate scores are reported for the designer of the message and the categories between the stages (changing reference levels). The p-values show that the expert-designed messages of the categories ER and RM are rated significantly more motivating than the peer-designed messages of ER and RM by people in the Precontemplation stage. The expert-designed messages of RM are also rated significantly more motivating than the peer-designed messages of RM by people in the Preparation stage. The peer-designed messages of DR and HR are rated significantly more motivating than the expert-designed messages of DR and HR by people in the Preparation, Action and Maintenance stages. Also, the peer-designed messages of SOL and SC are rated significantly more motivating than the expert-designed messages of SOL and SC by people in the Action stage, and for SC also for people in the Maintenance stage.

Fixed effects	Estimate	Std. Error	*t*-value
(Intercept)	**3.436**[3]	0.152	22.622
DR	**-1.027**[3]	0.160	-6.415
ER	-0.255	0.160	-1.590
SOL	-0.245	0.160	-1.533
SR	-0.082	0.160	-0.511
SEL	0.164	0.160	1.022
HR	-0.164	0.160	-1.022
CC	-0.055	0.160	-0.341
RM	**0.455**[2]	0.160	2.839
SC	**-0.345**[1]	0.160	-2.157
Contemplation	0.183	0.156	1.173
Preparation	0.271	0.150	1.808
Action	**0.316**[1]	0.149	2.117
Maintenance	0.227	0.142	1.598
Designer (peer)	-0.166	0.198	-0.837

Table 9. The estimates and standard error of the fixed effects of the model: process-categories, stage of change of the participant, and designer of the message. Significant effects reported for $^1p < 0.05$, $^2p < 0.01$, $^3p < 0.001$**.**

Peer-Expert	Stages of change				
est. diff.	PC	C	P	A	M
CR	-0.166	-0.060	0.133	0.160	0.140
DR	0.132	0.095	**0.396**	**0.645**	**0.363**
ER	**-0.582**	-0.024	-0.036	0.207	0.097
SOL	-0.156	-0.098	-0.025	**0.292**	0.038
SR	-0.119	-0.139	-0.160	0.094	0.149
SEL	-0.282	-0.163	-0.070	0.050	0.152
HR	0.080	0.145	**0.217**	**0.254**	**0.320**
CC	-0.241	-0.179	-0.112	-0.022	0.085
RM	**-0.550**	-0.143	**-0.210**	-0.043	0.025
SC	-0.314	-0.124	0.012	**0.237**	**0.179**

Table 10. Estimates of peer- versus expert-designed message categories across alll the stages. Reported scores are estimates of the peer process ratings in relation to the expert process ratings means. All negative numbers represent a higher estimate for experts in that cell, all positive numbers represent a higher estimate for peers in that cell. Numbers and stages in bold are $p < .05$**.** ($N = Exp/Peer; PC : N = 17/22, C : N = 73/53, P = 94/71, A = 46/72, M = 120/123$)

DISCUSSION

The results of our two surveys provide two contributions: i) the design of behavior change strategies as text messages (elicitation survey) and ii) the comparative analysis of how motivating expert versus peer-designed text messages are perceived across the stages of changes (evaluation survey comparison).

Similar to the approach of our previous research [4], our selection of coded messages representing the process categories had substantial reliability. This shows that these messages were a good fit for the processes of change they presented. We investigated whether experts designed messages that fit the stage-scenarios (following the theoretical distribution) they were given in the elicitation survey. Looking at higher-order processes, the Experiential process messages were more prevalent in the earlier stages while Behavioral process messages were more prevalent in the later stages, as would be expected following the theoretical distribution. This strengthens our belief that designing messages for behavior change theory this way, both by experts or peers, is a valid approach.

The distribution of expert messages over the processes was proportionately different from the distribution of the peer messages over the processes. Experts provide proportionately more messages than peers for Experiential processes (CR, SOL, SR), which were related to the earlier stages, and significantly less for Behavioral processes (SEL), which were related to later stages of change. This is in line with work from Kristan and Suffoletto [10] who found that 82% of their expert-designed messages were informational or strategy facilitating messages, which loosely correspond to our categories of Consciousness Raising and Social Liberation (both of which have significantly more messages in our results compared to the peer results of our previous work [4]). Note that the result of Social Liberation needs to be interpreted with caution because

the expected number of messages for this category was lower than five.

Our results indicate only partial acceptance for H1. We hypothesized that expert-designed messages representing the strategies would be perceived as more motivating than peer-designed messages representing the strategies, across the five stages of change. Our results show that, *expert-designed* messages were perceived more motivating by people who are yet unwilling to change their exercise behavior, but unexpectedly, *peer-designed* messages were perceived more motivating by people who are either planning on changing, are already taking steps to change, or are maintaining their exercise behavior. Interestingly, Table 10 shows that, while messages from only three expert process categories are perceived as significantly more motivating than the peer categories, two of these process categories are in the first stage (Precontemplation), arguably the most difficult stage to motivate people in. On the other hand, messages from nine peer process categories are perceived to be more motivating than expert categories. This is in line with results from Coley et al. [2], where they found peer messages to be more motivating, and where it could be argued that most of their participants were not in the Precontemplation stage (the exact distribution is not reported), since they participated in a behavior change program (therefore not strictly fitting the definition of Precontemplation: *not* willing to change their behavior).

The fifty messages for experts and the fifty messages for peers were compared between all the stages, but we did not find any differences for any of the categories across *all* the five stages. This indicates that the stage that people are in seems to play a role in how the expert- or peer-designed motivational messages of the same strategies are perceived.

Limitations of the current work

Our setup has some limitations, because we elicited the expert messages from Dutch and German experts, and tested them on Americans, while in our previous work [4], we elicited the peer messages from Americans and tested them on Americans. We mitigated this by not using all the messages for the evaluation, only a small, selected subset, and for consistency the same trained coder (coder 1) selected both sets of messages. At this time, we do not know if these text messages will also be motivating when used in applications or other formats (e.g., a virtual agent) or modalities (e.g., audio) or other populations (e.g., non-AMT Americans). Moreover, we do not know yet if messages that are perceived motivating will actually result in behavior change. Nevertheless, we think that the comparison of expert-designed versus peer-designed text message categories across the stages of change might have a real impact. Motivational text messages have been shown to work before in real contexts, and tailoring to stage of change has been shown to be good way to further increase effectiveness. Therefore, we expect our findings to generalize to other contexts. Future work will need to test our theory-driven design approach in various contexts. Our next study will test the effectiveness of the expert-designed and peer-designed motivational messages *in-the-wild*, in an smartphone application promoting physical activity.

CONCLUSION

In this paper, we described two surveys, one where we asked experts to come up with motivational text messages for a scenario, and a follow-up one where we asked people to rate fifty of these expert-designed text messages, aligned to the ten processes of change (behavior change strategies), on how motivating they are. We compared these results to our previous study [4] where people rated peer-designed text messages aligned to the ten processes of change. Through this comparison, we evaluated whether there was a difference in how motivating expert- or peer-designed text messages aligned to the ten processes of change were rated. We examined these differences across the five stages of change of the participants. Our findings are relevant for researchers who want to design motivational text messages with a theory driven-approach and our findings contribute to the field with two clear recommendations for the design of motivational strategies (text messages) for behavior change systems:

(i) People perceive messages tailored to relevant user characteristics as more motivating. As our findings show, for the perception of how motivating the different behavior change strategies are, the user's stage of change plays a role. In terms of system design, this means that the system should accommodate users to, not only fill in their regular demographic information, but also more behavior specific information, like their stage of change. Strategies should be selected that fit the stage of change of the user.

(ii) People perceive differences in how intervention strategies are designed. When you are identifying which strategies work on users in different stages of change, consider that the designer of that message for that strategy has an effect on how motivating the message itself is perceived. In terms of a system that uses text messages to motivate users, this means that the system should be able to select multiple sources of the same strategy, adjusting to the situation of the user. Specifically, expert-designed messages should be used to get people exercising, peer-designed messages should be used to keep people exercising.

REFERENCES

1. Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67, 1 (2015), 1–48. DOI:http://dx.doi.org/10.18637/jss.v067.i01
2. Heather L Coley, Rajani S Sadasivam, Jessica H Williams, Julie E Volkman, Yu-Mei Schoenberger, Connie L Kohler, Heather Sobko, Midge N Ray, Jeroan J Allison, Daniel E Ford, Gregg H Gilbert, and Thomas K Houston. 2013. Crowdsourced peer- versus expert-written smoking-cessation messages. *American journal of preventive medicine* 45, 5 (Nov. 2013), 543–50. DOI: http://dx.doi.org/10.1016/j.amepre.2013.07.004
3. David B. Dahl. 2016. *xtable: Export Tables to LaTeX or HTML*. https://CRAN.R-project.org/package=xtable R package version 1.8-2.
4. Roelof A. J. de Vries, Khiet P. Truong, Sigrid Kwint, Constance H C Drossaert, and Vanessa Evers. 2016.

Crowd-Designed Motivation: Motivational Messages for Exercise Adherence Based on Behavior Change Theory. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 297–308. DOI: http://dx.doi.org/10.1145/2858036.2858229

5. Andy Field. 2013. *Discovering statistics using IBM SPSS statistics*. Sage.
6. Eric B. Hekler, Predrag Klasnja, Jon E. Froehlich, and Matthew P. Buman. 2013. Mind the Theoretical Gap: Interpreting, Using, and Developing Behavioral Theory in HCI Research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 3307–3316. DOI: http://dx.doi.org/10.1145/2470654.2466452
7. Setia Hermawati and Glyn Lawson. 2014. Managing obesity through mobile phone applications: a state-of-the-art review from a user-centred design perspective. *Personal and Ubiquitous Computing* (2014), 1–21. DOI:http://dx.doi.org/10.1007/s00779-014-0757-4
8. Maurits Kaptein, Boris De Ruyter, Panos Markopoulos, and Emile Aarts. 2012. Adaptive persuasive systems: a study of tailored persuasive text messages to reduce snacking. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 2, 2 (2012), 10.
9. Catriona M Kennedy, John Powell, Thomas H Payne, John Ainsworth, Alan Boyd, and Iain Buchan. 2012. Active assistance technology for health-related behavior change: an interdisciplinary review. *Journal of medical Internet research* 14, 3 (2012). DOI: http://dx.doi.org/10.2196/jmir.1893
10. Jeffrey Kristan and Brian Suffoletto. 2015. Using online crowdsourcing to understand young adult attitudes toward expert-authored messages aimed at reducing hazardous alcohol consumption and to collect peer-authored messages. *Translational behavioral medicine* 5, 1 (2015), 45–52. DOI: http://dx.doi.org/10.1007/s13142-014-0298-4
11. Alexandra Kuznetsova, Per Bruun Brockhoff, and Rune Haubo Bojesen Christensen. 2016. *lmerTest: Tests in Linear Mixed Effects Models*. https://CRAN.R-project.org/package=lmerTest R package version 2.0-32.
12. J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics* (1977), 159–174. DOI: http://dx.doi.org/10.2307/2529310
13. Patrick WC Lau, Erica Y Lau, Del P Wong, and Lynda Ransdell. 2011. A systematic review of information and communication technology–based interventions for promoting physical activity behavior change in children and adolescents. *Journal of medical Internet research* 13, 3 (2011). DOI:http://dx.doi.org/10.2196/jmir.1533
14. Bess H Marcus, Beth C Bock, Bernardine M Pinto, Leigh Ann H Forsyth, Mary B Roberts, and Regina M Traficante. 1998. Efficacy of an individualized, motivationally-tailored physical activity intervention. *Annals of Behavioral Medicine* 20, 3 (1998), 174–180.
15. Bess H Marcus, Joseph S Rossi, Vanessa C Selby, Raymond S Niaura, and David B Abrams. 1992a. The stages and processes of exercise adoption and maintenance in a worksite sample. *Health Psychology* 11, 6 (1992), 386. DOI: http://dx.doi.org/10.1037/0278-6133.11.6.386
16. Bess H Marcus, Vanessa C Selby, Raymond S Niaura, and Joseph S Rossi. 1992b. Self-efficacy and the stages of exercise behavior change. *Research quarterly for exercise and sport* 63, 1 (1992), 60–66. DOI: http://dx.doi.org/10.1080/02701367.1992.10607557
17. Susan Michie, Marie Johnston, Jill Francis, Wendy Hardeman, and Martin Eccles. 2008. From Theory to Intervention: Mapping Theoretically Derived Behavioural Determinants to Behaviour Change Techniques. *Applied Psychology* 57, 4 (2008), 660–680. DOI: http://dx.doi.org/10.1111/j.1464-0597.2008.00341.x
18. Adity U Mutsuddi and Kay Connelly. 2012. Text messages for encouraging physical activity Are they effective after the novelty effect wears off?. In *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2012 6th International Conference on*. IEEE, 33–40. DOI:http://dx.doi.org/10.4108/icst.pervasivehealth.2012.248715
19. Kevin Patrick, Fred Raab, Marc A Adams, Lindsay Dillon, Marian Zabinski, Cheryl L Rock, William G Griswold, and Gregory J Norman. 2009. A text message–based intervention for weight loss: randomized controlled trial. *Journal of medical Internet research* 11, 1 (2009).
20. James O Prochaska and Carlo C DiClemente. 1983. Stages and processes of self-change of smoking: toward an integrative model of change. *Journal of consulting and clinical psychology* 51, 3 (1983), 390–395. DOI: http://dx.doi.org/10.1037/0022-006X.51.3.390
21. R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
22. Craig S Rosen. 2000. Is the sequencing of change processes by stage consistent across health problems? A meta-analysis. *Health Psychology* 19, 6 (2000), 593. DOI: http://dx.doi.org/10.1037/0278-6133.19.6.593

APPENDIX

Fixed effects	Estimate	Std. Error	t-value
(Intercept)	**3.436**[3]	0.152	22.622
Contemplation:Designer (peer)	0.106	0.227	0.466
Preparation:Designer (peer)	0.299	0.220	1.357
Action:Designer (peer)	0.326	0.230	1.420
Maintenance:Designer (peer)	0.306	0.213	1.434

Table 11. The estimates and standard error of the two way interaction effects between: designer of the message and stage of the participant. Significant effects reported for [1] $p < 0.05$, [2] $p < 0.01$, [3] $p < 0.001$.

Fixed effects	Estimate	Std. Error	t-value
(Intercept)	**3.436**[3]	0.152	22.622
DR:Designer (peer)	0.298	0.177	1.678
ER:Designer (peer)	**-0.416**[1]	0.177	-2.344
SOL:Designer (peer)	0.010	0.177	0.057
SR:Designer (peer)	0.047	0.177	0.262
SEL:Designer (peer)	-0.117	0.177	-0.657
HR:Designer (peer)	0.246	0.177	1.386
CC:Designer (peer)	-0.075	0.177	-0.422
RM:Designer (peer)	**-0.384**[1]	0.177	-2.163
SC:Designer (peer)	-0.149	0.177	-0.838

Table 12. The estimates and standard error of the two way interaction effects between: process-categories and designer of the message. Significant effects reported for [1] $p < 0.05$, [2] $p < 0.01$, [3] $p < 0.001$.

Fixed effects	Estimate	Std. Error	t-value
(Intercept)	**3.436**[3]	0.152	22.622
DR:Contemplation	0.073	0.139	0.521
ER:Contemplation	**-0.296**[1]	0.139	-2.126
SOL:Contemplation	**-0.317**[1]	0.139	-2.273
SR:Contemplation	0.131	0.139	0.939
SEL:Contemplation	0.021	0.139	0.153
HR:Contemplation	0.005	0.139	0.037
CC:Contemplation	-0.040	0.139	-0.285
RM:Contemplation	-0.092	0.139	-0.662
SC:Contemplation	-0.130	0.139	-0.933
DR:Preparation	-0.009	0.134	-0.070
ER:Preparation	**-0.374**[2]	0.134	-2.786
SOL:Preparation	**-0.340**[1]	0.134	-2.539
SR:Preparation	0.161	0.134	1.198
SEL:Preparation	-0.071	0.134	-0.527
HR:Preparation	-0.039	0.134	-0.292
CC:Preparation	-0.072	0.134	-0.538
RM:Preparation	-0.105	0.134	-0.785
SC:Preparation	-0.209	0.134	-1.562
DR:Action	-0.084	0.134	-0.626
ER:Action	**-0.309**[1]	0.134	-2.311
SOL:Action	**-0.430**[2]	0.134	-3.208
SR:Action	0.135	0.134	1.005
SEL:Action	-0.075	0.134	-0.558
HR:Action	-0.056	0.134	-0.417
CC:Action	-0.015	0.134	-0.111
RM:Action	-0.168	0.134	-1.258
SC:Action	-0.180	0.134	-1.341
DR:Maintenance	0.096	0.127	0.751
ER:Maintenance	**-0.279**[1]	0.127	-2.191
SOL:Maintenance	-0.228	0.127	-1.790
SR:Maintenance	0.038	0.127	0.298
SEL:Maintenance	-0.157	0.127	-1.235
HR:Maintenance	-0.092	0.127	-0.720
CC:Maintenance	-0.066	0.127	-0.517
RM:Maintenance	-0.248	0.127	1.950
SC:Maintenance	-0.159	0.127	-1.247

Table 13. The estimates and standard error of the two way interaction effects between: process-categories and stage of the participant. Significant effects reported for [1] $p < 0.05$, [2] $p < 0.01$, [3] $p < 0.001$.

Fixed effects	Estimate	Std. Error	t-value
(Intercept)	**3.436**[3]	0.152	22.622
DR:Contemplation:Designer (peer)	-0.143	0.203	-0.704
ER:Contemplation:Designer (peer)	**0.452**[1]	0.203310	2.223
SOL:Contemplation:Designer (peer)	-0.048	0.203	-0.236
SR:Contemplation:Designer (peer)	-0.126	0.203	-0.618
SEL:Contemplation:Designer (peer)	0.014	0.203	0.068
HR:Contemplation:Designer (peer)	-0.041	0.203	-0.201
CC:Contemplation:Designer (peer)	-0.044	0.203	-0.219
RM:Contemplation:Designer (peer)	0.301	0.203	1.481
SC:Contemplation:Designer (peer)	0.084	0.203	0.415
DR:Preparation:Designer (peer)	-0.036	0.197	-0.181
ER:Preparation:Designer (peer)	0.246	0.197	1.248
SOL:Preparation:Designer (peer)	-0.169	0.197	-0.856
SR:Preparation:Designer (peer)	-0.340	0.197	-1.724
SEL:Preparation:Designer (peer)	-0.087	0.197	-0.441
HR:Preparation:Designer (peer)	-0.162	0.197	-0.822
CC:Preparation:Designer (peer)	-0.171	0.197	-0.865
RM:Preparation:Designer (peer)	0.041	0.197	0.208
SC:Preparation:Designer (peer)	0.027	0.197	0.137
DR:Action:Designer (peer)	0.187	0.206	0.910
ER:Action:Designer (peer)	**0.463**[1]	0.206	2.250
SOL:Action:Designer (peer)	0.121	0.206	0.590
SR:Action:Designer (peer)	-0.112	0.206	-0.547
SEL:Action:Designer (peer)	0.006	0.206	0.029
HR:Action:Designer (peer)	-0.153	0.206	-0.742
CC:Action:Designer (peer)	-0.108	0.206	-0.525
RM:Action:Designer (peer)	0.180	0.206	0.878
SC:Action:Designer (peer)	0.226	0.206	1.099
DR:Maintenance:Designer (peer)	-0.074	0.191	-0.390
ER:Maintenance:Designer (peer)	0.373	0.191	1.952
SOL:Maintenance:Designer (peer)	-0.112	0.191	-0.586
SR:Maintenance:Designer (peer)	-0.038	0.191	-0.197
SEL:Maintenance:Designer (peer)	0.128	0.191	0.672
HR:Maintenance:Designer (peer)	-0.066	0.191	-0.344
CC:Maintenance:Designer (peer)	0.020	0.191	0.106
RM:Maintenance:Designer (peer)	0.269	0.191	1.409
SC:Maintenance:Designer (peer)	0.188	0.191	0.983

Table 14. The estimates and standard error of the three way interaction effects between: process-categories, designer of the message, and stage of the participant. Significant effects reported for [1] $p < 0.05$, [2] $p < 0.01$, [3] $p < 0.001$.