Finding the Adaptive Sweet Spot: Balancing Compliance and Achievement in Automated Stress Reduction

Artie Konrad[1,2], Victoria Bellotti[2,1], Nicole Crenshaw[3,2], Simon Tucker[2], Les Nelson[2], Honglu Du[2], Peter Pirolli[2], and Steve Whittaker[1]

[1]UC Santa Cruz
1156 High Street, Santa Cruz, CA
{akonrad, swhittak}@ucsc.edu

[2]Palo Alto Research Center
3333 Coyote Hill Rd.,
Palo Alto, CA
{bellotti, stucker, lnelson, hdu, pirolli}@parc.com

[3]UC Irvine
Irvine, CA
crenshan@uci.edu

ABSTRACT

Automated coaching systems offer a convenient, cost-effective way to reduce stress, which can be a serious health issue. However, one concern with such systems is *compliance*; users fail to achieve daily stress reduction goals because goals are too easy or too difficult. To address this, we built DStress (Design for Stress), a theoretically grounded system that sets adaptive goals in three coaching dimensions: Exercise, Meditation and Accessibility. DStress modifies goal-difficulty based on the individual's immediately previous performance. In a 28-day deployment with 65 users, DStress reduced scores on one direct measure of stress almost in half, significantly more than two other non-adaptive coaching strategies. However, on a second direct stress measure, no improvement was found. There were also no improvements on other indirect stress measures. Analysis of 2842 user-generated reports suggests our findings were the result of DStress balancing compliance against the degree of challenge of the goals it would set.

Author Keywords

Stress reduction; mental health; compliance; behavior change.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

The majority of healthcare coaching systems target physical health, but mental well-being is increasingly seen as critical [16,27,32]. Stress is an unavoidable fact of life. Mild stress can be adaptive, but chronic stress profoundly affects individual and societal health [43]. Chronic stress increases mortality, morbidity, immunosuppression, anxiety, depression, and absenteeism [9,34 ,38]. Job stress alone is estimated to cost the U.S. economy more than $300 billion a year [43]. We describe the design and evaluation of an adaptive, personalized healthcare system that addresses stress.

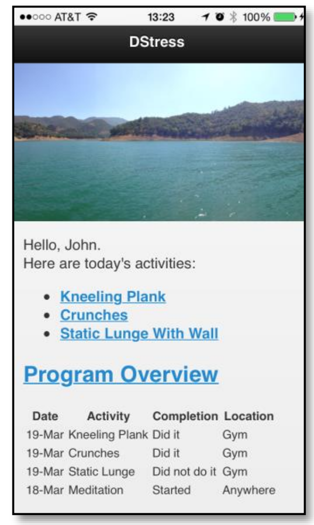

Figure 1. The DStress system. Program overview and activities for the day.

Two successful approaches to stress reduction are *physical exercise* and *mindfulness meditation*. These are effective when implemented under expert supervision, but they require costly personalization to ensure compliance [4,15]. This has prompted calls for new approaches to provide *simple*, *accessible*, *scalable*, and *sustainable* regimens that meet user needs [8]. Computational methods show significant promise as a class of solutions. In a 2000 survey, 91% of respondents said they wanted access to computational interventions. These are cheap, highly convenient and users need not worry about perceived social stigma of interacting with professionals [18].

But while automated mental healthcare systems are affordable and private, they have limitations. One common problem is *compliance,* where users fail to persist with proposed health-enhancing behaviors. A review of 46 computerized interventions for anxiety and depression, found the median completion rate was only 56% [47]. Recent work on mental and physical health indicates that compliance failure can result from lack of personalization. [16] improved compliance with an online mental health intervention by emphasizing personalization among other strategies. [13] investigated goal-setting strategies for fitness, but found it hard to design effective programs from

CHI 2015, April 18–23, 2015, Seoul, Republic of Korea.

ACM 978-1-4503-3145-6/15/04…$15.00.
http://dx.doi.org/10.1145/2702123.2702512

national recommendations. Participants stated that recommended activities often did not fit personal goals or expertise. This is a major limitation of many commercial health programs (e.g. [1,22]) and national recommendations (e.g. [40,45]).

However, designing effective personalized systems is complex. Prior research on behavior change shows an apparent paradox. On the one hand, people need to be given challenging activities [29]. On the other, if activities are *too* difficult, this demotivates as goals are seen as unattainable [5].

How then might we design stress reduction systems that avoid this apparent paradox? Building on successful approaches to physical well-being [6,28] we designed an *adaptive* system, DStress (see Figure 1), which *modifies goal difficulty depending on participants' immediately previous performance*. Incorporating behavior change theory [29] participants are given successively more challenging stress reducing activities. If they succeed in these activities, this should motivate. But participants who fail on an activity are immediately given an easier goal. Succeeding on this easier goal should re-establish self-confidence and thus motivation, increasing the probability that they will persist with the system [5]. This adaptive mechanism aims to optimize system compliance while still maintaining a challenging level of difficulty so that users experience the greatest reductions in their stress.

We address compliance in two other ways. A second critical system design feature is that we support stress reduction via both Exercise and Meditation dimensions. By providing *multiple strategies* for success, we counter a different critique of current physical healthcare systems; that they are overly simplistic [19,39]. A final more pragmatic design dimension is Accessibility. Users have to be provided with goals they can achieve in *multiple contexts*. Suggesting a fitness goal that requires users to always attend a gym may restrict their ability and willingness to attempt that exercise. Thus the Accessibility dimension complements the Exercise dimension by adjusting the location of workouts, if the user is having trouble making it to the gym.

We conducted a one-month field trial deployment comparing our adaptive stress coaching system, DStress, with non-adaptive versions of the system, to address the following questions:

- *Stress*: Does DStress actually reduce stress?
- *Adaptivity*: Does adaptivity reduce stress more than non-adaptive approaches?
- *Compliance versus Achievement*: We also explore underlying mechanisms for achieving stress reduction. Is simple compliance sufficient or must users also carry out tasks that present a challenge?

We show that an adaptive approach can lead to greater stress reduction. However, compliance alone is insufficient, with participants only receiving well-being benefits when they successfully tackled activities that were optimally challenging.

RELATED WORK

Behavior Change Theory

There are many behavior change theories. Some focus on the capacity to control behavior, e.g. the *Theory of Planned Behavior* [2], and Bandura's *Self-Efficacy Theory* [5]. The *Transtheoretical Model of Behavior Change* [41] considers behavior change as a series of stages. *Goal-Setting Theory* describes different types of goals and predicts which are most motivating [29]. Goffman's *Presentation of Self in Everyday Life* has been applied to coaching systems that exploit social motivations (e.g. [14]). We focus on Goal-Setting Theory and Self-Efficacy Theory because they address our concerns of adapting goal difficulty to optimize challenge and maintain self-confidence, as we shall discuss.

Goal-Setting Theory: Difficult is Better

Goal-Setting Theory (GST) proposes a linear positive relationship between goal difficulty and the effort put into achieving that goal [35]. Also, specific goals elicit more effort than abstract goals, e.g. "do 20 push-ups," as opposed to, "do your best" [29]. In Goal-Setting Theory, goals represent a discrepancy between a person's current state and the state they seek to achieve. According to the theory, achieving a goal confers satisfaction by closing that gap and more challenging goals are more motivating because they require more effort to achieve than easier goals.

Self-Efficacy Theory: Easier is Better

Self-efficacy is the belief in one's ability to succeed in a given situation [5]. If people observe themselves failing at a desired goal, their confidence in their ability to succeed in that situation drops. Past achievement is critical to influencing behavior change, so Self-Efficacy Theory holds that *successful completion* of planned goals is more important than the goal itself [7]. So less challenging goals are therefore more desirable because, by effortlessly realizing these goals, participants feel confident that they can achieve future goals and therefore remain compliant.

Technology to Support Health

The past few years have seen the emergence of many healthcare-coaching technologies (see [26] for a review). The vast majority of these systems are designed to address physical health, e.g. providing social persuasion to motivate exercise [44]. There has been less research on systems built to support mental health and stress, although new research is emerging [23]. For example, [36,37] suggest "just in time" coping techniques when the user exhibits elevated stress and negative mood. [31] designed a wearable biofeedback device to help monitor and self-regulate stress. [30] built a smartphone application that detects stress from changes in speech production. Many of these mental health systems are therapy-based (invoking Cognitive Behavioral

Therapy or Positive Psychology) [16,27,32]. There has also been interest in mindfulness-based design for mental health [46]. However, with the exception of [16], [27] and [32] very few mental health systems have been designed to address underlying issues with compliance.

Coaching Strategies

Some systems employ *predetermined* coaching strategies that profile the user along different dimensions to set goals that remain fixed regardless of user achievement. For instance, Chick Clique, Houston, and Ubifit set personalized goals but these are not continually reevaluated in line with user progress [12,13 ,44]. Another example is the Mobile Lifestyle Coach where users earn "lifestyle points" for exercising and eating healthily. Users attempt to earn the same predetermined number of lifestyle points each day [17].

While predetermined strategies are simple to implement, they may be problematic from a compliance perspective. If a user surpasses a predetermined easy goal, they may lose motivation (following Goal-Setting Theory), or if goals are too difficult users might lose self-efficacy (following Self-Efficacy Theory). Adaptive systems attempt to avoid these scenarios by providing goals that dynamically adjust to users' progress, encouraging them to persist with the coaching system.

Adaptive systems are more complex to implement, however. Some rely on 'embedded' professionals to evaluate progress and set new personalized goals. For example, some systems log health information that is continually reviewed by a healthcare practitioner who provides feedback and adjusts goals [25,48]. Automated approaches that don't rely on professionals are less common but becoming popular. For example in online learning, adaptive and intelligent web-based educational systems (AIWBES) tailor lessons to the development of the student. Systems like ELM-ART and ActiveMath, adapt the content and sequence of the curriculum to the individual [10]. In adaptive coaching for health, systems like Fish'n'Steps, SMS Coach, and "Laura" assess disposition and user performance to tailor appropriate goals [6,19 ,28]. Goals are adjusted depending on current progress.

While these systems provide adaptivity, one user critique is that they often provide limited strategies for success. Users of SMS Coach, for instance, frequently reported that they wanted multiple variables to be considered to help change their behavior, as well as more diverse feedback [19]. Participants in Obermayer's smoking cessation study desired messaging that was more tailored to their individual challenges [39]. So, like a real coach, an automated, adaptive system should be flexible along multiple dimensions to meet user needs.

Accessibility has been shown to be important in past medical research. For example, accessibility of soap stations encourages handwashing compliance among nurses [24]. Adaptive systems also need to be accessible to provide coaching that fits a user's lifestyle [14,47]. For instance, if it were more convenient to join friends on a bike ride than to go for a walk, Fish'n'Steps users would not get any credit for their exercise [28]. Thus, adaptive systems must be flexible enough to provide goals that can be achieved in different contexts.

Finally, recent review articles argue that the majority of automated health systems are not based on theory [11,42]. In some systems that do claim theoretical motivations, the specifics of how theory informs their design are never discussed (e.g. [20]). Systems that are informed by theory can address the bottlenecks in behavior change, resulting in more effective interventions [13,42].

Finding the Adaptive Sweet Spot

We designed DStress to address these limitations and adaptively select goals derived from behavior change theory. By balancing Goal-Setting Theory and Self-Efficacy Theory we aim to design a program that is the best of both worlds. Our objective is to maximize compliance by balancing goals that are not too difficult (maintaining self-efficacy) and not too easy (maintaining motivation). DStress does this by adapting difficulty based on the user's past performance. These daily adjustments occur in three different coaching dimensions: *Exercise*, *Meditation*, and *Accessibility*. Furthermore, each of these dimensions is *complementary* and *reinforcing*. Compliance with meditation provides greater relaxation to recover from physical exercise. The Accessibility dimension adjusts convenience of exercise location to induce greater compliance with the Exercise dimension. We seek to evaluate our multidimensional approach as well as the influence of our adaptive mechanisms on compliance, achievement, and ultimately stress.

METHOD

We developed and evaluated three versions of a coaching system; an adaptive DStress, a non-adaptive *Easy* version, and a non-adaptive *Difficult* version. The Easy version was intended to maximize one aspect of compliance by presenting straightforwardly achievable activities. The Difficult version was meant to motivate another aspect of compliance with challenging activities. The DStress version was designed to balance ease and challenge, and our overall goal was to determine which of the three systems most reduced stress.

Participants

We recruited 77 participants through Facebook and an internal database of participants. We also used a snowball-recruiting scheme where participants could recommend others. They were paid $15 for starting the study, and another $35 for completing. People who recruited friends that completed the study received an additional $10 per person up to a maximum of $50 dollars. Participants were first screened to confirm they were physically capable of performing the exercises in this study using the Physical

Activity Readiness Questionnaire. Four participants were removed who did not pass this screening. Eight dropped out because of unexpected surgery, international travel, busyness at work, or because they ceased communication with the research team. This left 65 (42 female) aged 19-59 (*M*=31.12, *SD*=9.29), of which 45 used iPhones, 16 used Androids, and 1 person used another type of smartphone. Participants were randomly assigned to a group and each group was assigned an Easy, Difficult, or DStress version of the system. Participants were blind to which group they were in and were not informed that there were different groups. There were 24 participants in the Easy (16 female, age *M*=31.75, *SD*=9.26), 22 in the Difficult (13 female, age *M*=31.68, *SD*=8.94) and 19 in the DStress group (13 female, age *M*=29.68, *SD*=10.04).

The DStress System

DStress is a web-based coaching system (see Figure 1) that runs in a browser, making it accessible from any smartphone. The system provides information about a participant's stress-related activities program (exercise, meditation), details about how to execute activities, methods for logging progress with activities, reminders about activities and logging, and a record of past activities. To manipulate accessibility we also provided participants with information about *where* to carry out their activity.

Users log-in to access their personalized program of activities for the day. There are three types of days: exercise, meditation and rest. On *exercise* days, exercises are presented as clickable links to detailed instructions on how to perform each one and how many repetitions and sets to do. They are presented with a recommended *location* (either the gym or a convenient place, such as at home). On exercise days, users receive one upper body exercise, one lower body exercise, and one circuit training exercise. If the day is a *meditation* day, DStress presents a link to instructions and how many minutes to meditate that day. And on *rest* days, users do no exercise or meditation.

Users can click a link on the home screen to see their entire *program overview*. This also displays all previous activities and which ones were completed. Users can also see the day's scheduled activity to help with planning. Users can browse activities further in the future which are titled as "Exercise", "Meditation" or "Rest" days without details, since the specifics are determined adaptively based on user performance.

After completing activities, users click a link to *report* how they did. On exercise days, DStress asks whether users made it to the gym. And for each exercise users report whether they completed it successfully ("*Complete*"), started but didn't finish ("*Start*"), or didn't start the exercise ("*Fail*"). The same responses are collected for meditation days allowing users to report progress against the day's goal (number of meditation minutes).

Each morning at 9am users receive an automatic email reminder to complete the day's activities with a link to login to DStress. And at 8pm, if users have not yet reported their progress for the day, they are emailed a reminder to log their activities.

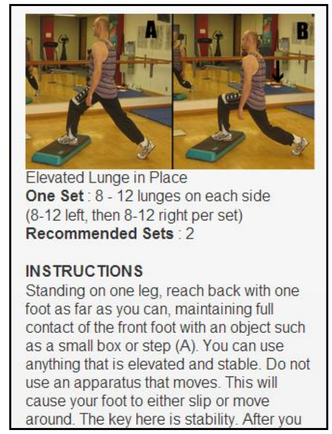

Figure 2. DStress stress reduction exercise. Elevated lunge instructions.

Procedure

We wanted to assess changes in stress level as a result of using DStress. Participants therefore took an online *pre-test* survey to measure baseline stress along with demographics. They then worked with the system for 28 days, after which they took a *post-test* survey with the same scales. Comparing the pre and post surveys allowed us to measure stress reduction. Participants took the surveys a day before their first activity and a day after their last activity. We emphasized the importance of reporting daily progress, regardless of whether participants completed activities, to ensure we had sufficient data to analyze. Next we describe the three coaching dimensions (Exercise, Meditation and Accessibility), which adapted to the individual for the DStress group, followed by the predetermined schedules for the other two groups.

Exercise

We worked with three certified trainers to develop 46 exercises in total. Example exercises include: Wall Pushups, Standing Knee Lifts, Squats, and Burpees. Figure 2 shows the Elevated Lunge exercise with instructions. Each exercise was independently rated for difficulty from 1 to 10 by at least 2 of the trainers (most were rated by all three and trainers were generally consistent in their evaluations). The difficulty ratings for each exercise were averaged and rounded to the nearest whole number to establish clear difficulty levels for scheduling purposes. This yielded exercises ranging in difficulty from 1 to 8.

When Participants logged-in, a link explained: how to warm up with light cardio and stretching; perform exercises with safety and good form; and cool down afterwards. The trainers also iteratively developed detailed instructions for each exercise, which we pre-tested on colleagues with very little exercise experience. Each description prescribed the number of repetitions to be completed (typically 8-12) and the number of times to complete the exercise (typically 2 sets) to standardize the exercises (see Figure 2). Some exercises required equipment usually found in a gym, but we developed equipment-free versions of these for home deployments. Exercises were always scheduled for Mondays, Wednesdays, and Fridays.

Meditation

Meditation days were designed to both reduce stress and allow recuperation to maximize exercise benefits. Meditation days fell in between exercise days (on Tuesdays, Thursdays, and Saturdays) with Sundays being a rest day with nothing scheduled. On Meditation days, participants practiced mindfulness meditation for varying lengths of time. A member of our research team has previous experience as a meditation teacher. He developed simple standard instructions, centered around observing one's thoughts with moment-to-moment awareness, labeling them as "thoughts" without judging them, and attending to breathing when the mind wanders. These instructions were piloted to ensure that participants were clear about what they should do.

Accessibility

We expected that the burden of travelling to the gym would be an important component affecting compliance [24]. So the adaptive version of DStress adjusted the proposed location for each workout based on past performance. To accomplish this, for each exercise that required equipment, a non-equipment version was created that worked the same muscle groups, was rated at the same level of difficulty, and that could be completed wherever was convenient. This gave us both more Accessible and less Accessible versions when creating exercise schedules. Location was not specified for Meditation, since any calm place is acceptable.

We now describe the monthly schedule for the three groups. Two were predetermined (Easy, Difficult) and one was adaptive (DStress).

Easy Condition

The Easy group received exercises that were rated as easy, but slightly increased in difficulty throughout the study. Participants received exercises with difficulty ratings of 1 for the first half of the study, and then those rated 2 for the last half. Their meditation days followed a similar pattern; they were asked to meditate for 5 minutes for the first half of the study, then 10 minutes for the last half. Thus, both exercise and meditation schedules had a slight increase across the study so that activities were easy but still had some variety (to be more realistic as a stress reduction program). Lastly, so that the Accessibility dimension would also be easy in this condition, participants were told that all exercises could be completed anywhere convenient.

Difficult Condition

The Difficult group received exercises that increased in difficulty to be challenging to participants. Since there were 12 exercise days, and 8 difficulty levels rated by the trainers, we chose the difficulty to increase across the 12 days as follows:

Level: 1, 2, 2, 3, 4, 4, 5, 6, 6, 7, 8, 8.

Meditation days increased similarly in 5-minute increments:

Minutes: 5, 10, 10, 15, 20, 20, 20, 25, 30, 30, 35, 40, 40.

Thus, both exercise and meditation increased steeply across the study. As in the Easy condition, both were non-adaptive, i.e. they did not depend on user performance. So that the Accessibility dimension would also be more demanding in this condition, participants were told to complete all exercise activities at the gym.

DStress Condition

The DStress group received exactly the same exercise, meditation, and accessibility schedule as the Difficult group, except that the system adapted to their performance. For example, if an individual successfully completed all three exercises that day, then they would advance to the next level along the same trajectory as the Difficult group. Thus, if a DStress user completed all exercises, did them all at the gym, and completed all meditations, they would progress through the levels at exactly the same rate as the Difficult group. The difference was in how DStress adaptively handled non-completions.

If a user did not complete an exercise or meditation level, they were given an easier level. If they completed this lower level, advanced again to the higher level, and were unable to complete this level a second consecutive time, the system locked onto the lower level indefinitely. This was because we didn't want users continually cycling back and forth to a level that was too difficult. The system therefore attempted to preserve self-efficacy and only presented exercise or meditation at the previous successful level. However, at any point, if the user reported that they did *more* exercise or meditation than was suggested, then the system advanced once again to the higher level to be more motivating. Meditating or exercising beyond what was suggested was taken as an indication that participants had mastered the lower level and were ready to advance.

If a user did not complete a specific exercise, then the next exercise day the system reduced the difficulty of all three exercises in the activity following guidance from our certified trainers. The rationale is that if users struggle with exercises in one area of their body, instead of increasing difficulty in other areas (risking potentially more failure), the system should step back altogether to offer a better chance of strengthening the weak link. For incomplete meditations, the system also stepped down meditation minutes the next meditation day.

For accessibility, our goal for DStress was to provide an adaptation for travel convenience. Initially DStress suggested that participants perform the exercises at the gym. If, however, a DStress user reported not getting to the gym, then the next exercise day the system adapted by providing a workout that could be completed wherever was convenient. If this was completed, as with exercise and meditation difficulty, the system then ramped back up and next proposed a gym workout. A second failure to get to the gym locked the system in non-gym mode until the user reported going to the gym. In this way, DStress adapted

Accessibility based on user performance, to balance self-efficacy and motivation.

Although DStress users generated reports reliably, if they failed to report progress on any day (16% of occasions), there was no information to drive adaptation, so the three dimensions remained unchanged. In contrast the other two groups progressed through the schedule regardless of their reported actions. All three dimensions (Exercise, Meditation, Accessibility) were treated independently so if a user stepped up or down in one dimension, this would not affect the schedules of other dimensions.

Data Collected

Progress reports: We collected daily progress reports from participants, allowing us to assess compliance and difficulty level achieved for each of the three dimensions (Exercise, Meditation and Accessibility).

Stress Measures: Because there is no universal measure for stress, we include DASS, PSS, and CHIPS scales which assess both direct and indirect aspects of stress across different time intervals. *Direct* measures probe participants' perceived stress, whereas *indirect* measures assess related manifestations such as physical symptoms, depression or anxiety. All stress measures were administered at both pre-test and post-test.

- *Perceived Stress Scale (PSS):* 10-item scale assessing perceived stress levels over the past month that provides a *direct* stress measure.
- *Depression, Anxiety, Stress Scale (DASS):* 21-item short version of the original 42-item scale, chosen to reduce survey burden on participants. Measures negative emotional states of depression, anxiety and stress in three independent 7-item sub-scales which are not typically combined. Participants assess emotional states over the past week. The Stress sub-scale *directly* measures stress. The Anxiety and Depression sub-scales are *indirect* measures because they are comorbid with stress [3].
- *Cohen-Hoberman Inventory of Physical Symptoms (CHIPS):* 33-item scale measuring concerns about physical symptoms over the past 2 weeks. Included as an *indirect* measure because stress is often manifested in such symptoms [33].

To exclude possible confounds, at pretest we measured body mass and exercise frequency, and at posttest we measured perceptions of system usability.

- *Body Mass Index (BMI):* Calculated from height and weight and is a measure of a person's body shape (a rough proxy for body fat percentage).
- *Godin Leisure-Time Exercise Questionnaire (GLTEQ):* 4-item scale measuring the frequency of physical activity during one's leisure time.
- *System Usability Scale (SUS):* 10-item scale designed to check whether perceived system usability was interfering with the ability to perform and record daily exercises. We wanted a brief survey, so we selected just two items (questions 1 and 8): "I think I would like to use this system frequently" and "I found the system very cumbersome to use."

These scales all demonstrate good psychometric properties such as discriminant and convergent validity and test-retest reliability in multiple populations.

STUDY HYPOTHESES

We had four research hypotheses (RHs) concerning compliance, difficulty, and stress reduction.

RH1- Superior Outcome with Adaptivity

The DStress group should show the greatest improvements on our direct and indirect stress measures due to higher compliance rates.

RH2- Exercise and Meditation Compliance

DStress' adaptive multi-dimensional approach should induce more exercise and meditation compliance than the other non-adaptive groups.

RH3- Compliance Benefits for Accessibility

Compliance should be greater for exercise activities overall if the location can be adapted to be convenient instead of the gym default. Adapting the location occurs when the user is having difficulty getting to the gym, and should increase self-efficacy, and consequently, compliance.

RH4 - Differences in Difficulty

The Difficult group should complete more difficult levels than DStress, and DStress more than Easy in the Exercise and Meditation dimensions. DStress adapts to failure by presenting easier exercises, reducing average completed difficulty level.

FINDINGS

First we analyze potential confounds to our results. We then present pre- and post-test survey results to determine effects on stress measures, followed by analysis of user-generated reports to examine compliance.

Manipulation Check

We confirmed there were no initial group differences at baseline that could confound our results. There were no baseline group differences in *Age* (p=.73), *Gender* (p=.80), *Body Mass Index* (BMI: p=.71), *Exercise Frequency* (GLTEQ: p=.65), *Perceived Stress* (PSS: p=.82), *Depression (DASS)* (p=.63), *Anxiety (DASS)* (p=.92), *Stress (DASS)* (p=.17), or *Physical Health Symptoms* (CHIPS: p=.32). At post-test there were no group differences in the *System Usability* questions. All three groups were above neutral in their desire to use the system frequently, and below neutral in finding the system cumbersome (except the Easy group which was slightly above neutral).

	Pre Stress	Post Stress	Difference
Easy	17.33	15.08	-2.25
Difficult	17.64	15.00	-2.64
DStress	21.79	12.32	-9.47***

Table 1. DStress reduces stress more than other conditions for DASS. Mean DASS Stress scores and difference scores before and after working with the three system versions. Highest possible score is 42, and lower scores indicate less stress (*p<.001)**

Survey Data

Stress Measures

To compare effects of the three system versions on our stress measures, we ran a MANOVA with one between-subjects factor: Condition (Easy, Difficult, DStress) and one within-subjects factor: Time (pre-test, post-test). The dependent variables were the following self-report scales: Perceived Stress Scale (PSS), Depression, Anxiety, and Stress (separate subscales of DASS), and Physical Symptoms (CHIPS). Using Pillai's trace, the main effect for pre-test vs. post-test was significant, V=.33, F(5,58)=5.63, p<.001, η_p^2=.33, showing that overall the intervention improved the combined stress measures, for the three groups combined. More interestingly, we also found a significant interaction effect of Time by Condition, V=.29, F(10, 118)=1.96, p=.04, η_p^2=.14, with a medium to large effect size. Univariate tests of this interaction indicated that this was caused by the Stress subscale of the DASS (F(2, 62)=4.37, p=.02). A simple main effects analysis showed that the DStress group reduced DASS Stress (p<.001), while the Easy group (p=.24) and the Difficult group (p=.11) did not. Table 1 shows the pre-test and post-test means of DASS Stress for each group. However, we did not find evidence that the DStress group improved more on *all* stress measures. Aside from DASS Stress, there were no group interactions for the other scales (PSS: F(2, 62)=.36, p=.70, CHIPS: F(2, 62)=1.47, p=.24, DASS Depression: F(2, 62)=1.92, p=.16, DASS Anxiety: F(2, 62)=.72, p=.49).

Progress Report Data

To understand why one direct measure of stress was reduced by DStress but not the Easy and Difficult groups, we analyzed user progress reports to examine compliance. Users reported on their progress 2842 times (84%) out of 3380 possible reports. There were no significant differences between the groups in number of reports generated.

Difficult Group Less Compliant for Exercise and Meditation

Compliance is assessed by program adherence, so both completing and starting an exercise were considered compliant. A person is no less compliant simply because they are not strong enough to finish an exercise. We therefore defined compliance as percentage of *Starts* and *Completes* from the total number of user generated reports (omitting the no-reports which provide no information).

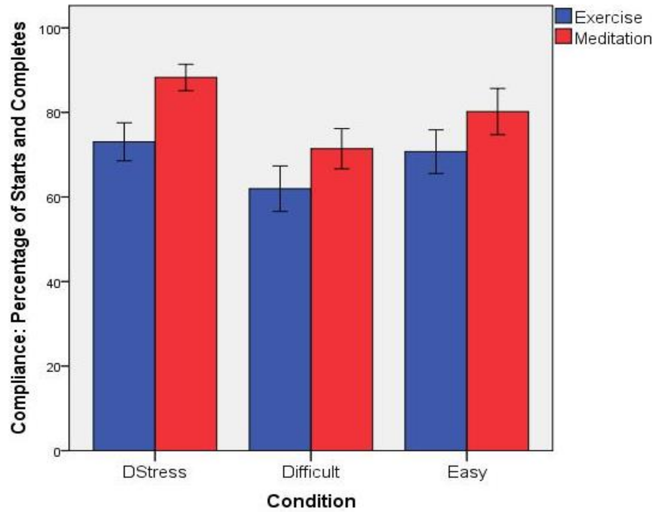

Figure 3. Compliance is greater for DStress and Easy conditions for exercise and meditation. Compliance is represented by percentage of Starts and Completes. Error bars are 1 standard error in all figures.

We carried out a one-way MANOVA to compare the effects of Condition (Easy, Difficult, DStress) on both Exercise and Meditation compliance. Using Pillai's trace, the main effect for Condition was significant, V=.17, F(4, 124)=2.96, p=.02, η_p^2=.09, showing group differences in overall compliance. Univariate ANOVAs revealed group differences in both Exercise, F(2, 62)=4.66, p=.01, η_p^2=.13 and Meditation dimensions, F(2, 62)=5.34, p=.01, η_p^2=.15. Figure 3 shows the compliance percentages of Exercise and Meditation for each group. Using Tukey's HSD, post-hoc tests revealed that for Exercise, the DStress group was significantly more compliant than the Difficult group (p=.02) but not the Easy group (p=.82). The Easy group was trending toward more compliance than the Difficult group (p=.052). Likewise in the Meditation dimension, the DStress group was significantly more compliant than the Difficult (p=.01) but not the Easy group (p=.63), and the Easy was trending toward more compliance than the Difficult group (p=.054). So, for both Exercise and Meditation, the DStress and Easy groups had comparable levels of compliance, while the Difficult group was less compliant than both of them.

Accessibility: DStress More Compliant at Home than at Gym

For each user-generated exercise progress report, we collected data on whether the system proposed gym or non-gym (wherever convenient) exercises. For the DStress group, the system proposed 444 gym exercises (M=23.37, SD=8.58), and 240 non-gym exercises (M=12.63, SD=8.58). Also, we wanted to test our prediction that adapting the exercise location to the individual would improve compliance for the DStress group. We assessed compliance within the DStress group only. We excluded Difficult and Easy groups because for them accessibility was confounded with difficulty levels.

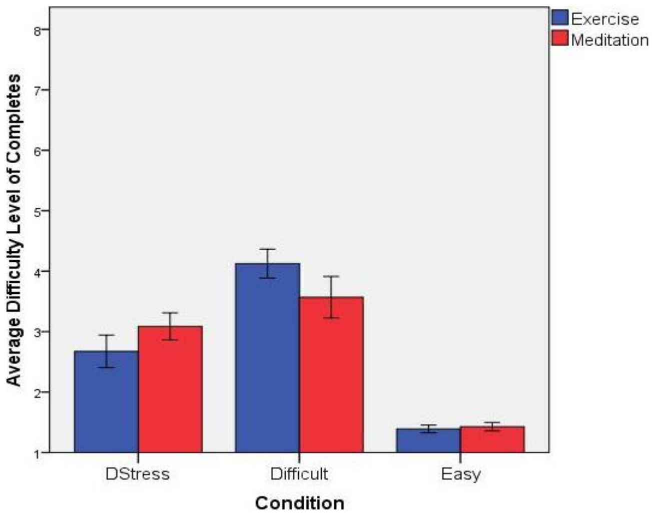

Figure 4. Achievement is greater for Difficult condition. Average difficulty level for exercise and meditation completions by condition. Length of meditation (in minutes) was converted to levels from 1 to 8 for depiction purposes.

We first compared compliance of proposed gym exercises with compliance of proposed non-gym exercises. DStress only proposes a non-gym workout if the user reports exercising at a non-gym location for the prior workout. Thus the system is adapting when proposing non-gym workouts versus the default gym workouts. A paired *t*-test of compliance percentages for proposed non-gym vs. proposed gym exercises showed differences, $t(16)=3.04$, $p=.01$, $d=.74$ with greater compliance for non-gym proposed exercises ($M=.84$, $SD=.20$) than gym exercises ($M=.69$, $SD=.19$). Higher compliance might represent an advantage of the system learning when a person has entered a busy moment in time (and is unable to make it to the gym). Also, proposing a non-gym workout after previously presenting gym workouts might increase user self-efficacy as the location becomes more convenient. This encouragement might also contribute to compliance. However, these results do not distinguish whether the increased compliance rates for non-gym exercises were generated from the adaptivity of DStress, or whether non-gym exercises in general encourage more compliance than gym exercises due to added convenience.

The Difficult Group Succeeded with Harder Tasks

We examined the effects of difficulty, exploring what difficulty levels participants had completed across the different conditions. For Exercise, a one-way between subjects ANOVA comparing the effects of Condition (Easy, Difficult, DStress) on Level (exercise difficulty) found significant differences in average exercise level completed, $F(2, 62)=40.95$, $p<.001$, $\eta^2=.57$. Post-hoc tests revealed that the Difficult group completed higher levels than DStress ($p<.001$) and Easy groups ($p<.001$). Also DStress completed more difficult exercises than the Easy group ($p<.001$). These differences support our predictions.

A one-way between subjects ANOVA also revealed group differences in Meditation level completed, $F(2, 62)=19.74$, $p<.001$, $\eta^2=.39$. Post-hoc tests show that the Easy group meditated for fewer minutes than both DStress ($p<.001$) and Difficult groups ($p<.001$). However, contrary to our prediction, there were no differences in the Meditation levels completed between DStress and Difficult groups ($p=.43$). It may be that the meditation schedule for the Difficult group became more challenging to participants quicker than the exercise schedule, resulting in lower meditation completions. Alternatively, a difficult exercise goal might be inherently more motivating than a difficult meditation goal that, in a sense, requires "non-doing". Figure 4 shows mean difficulty levels of Exercise and Meditation completed by group. Minutes of meditation was converted to difficulty level (1 to 8) for depiction purposes.

DISCUSSION

Our results are highly encouraging adding to our emerging knowledge about automated coaching systems for well-being, and to our understanding of compliance. A large-scale month-long intervention showed that a system that combines adaptivity on multiple dimensions (Exercise, Meditation, Accessibility) reduces scores on one direct measure of stress by almost 50%. We were also successful in exploring theory led design, teasing apart opposing predictions from behavioral theory, informing the ongoing debate about the role of theory in healthcare systems [11,13,14,20,42]. The study also extends our empirical understanding of how different designs encourage compliance, as well as the importance of achievement on system efficacy.

Easy tasks and adaptivity led to better compliance than difficult tasks. Furthermore, compliance was greater for adaptive non-gym proposed workouts than non-adaptive default gym workouts. But we show that compliance alone does not guarantee stress reduction, as only the adaptive DStress group showed reduced DASS Stress scores. Stress reduction with DStress seemed to result from adaptively balancing compliance *and* achievement. Increasing difficulty irrespective of individual performance may reduce self-efficacy and exceed limitations in ability. In contrast, providing simple tasks may not be as effective at inducing health benefits.

Our exercise program here comprised only resistance training but to support multiple success strategies, future designs might include other forms of exercise including running, swimming, or tennis. Social activities may also be beneficial [26]. Combining different physical activities in cross training might provide benefits consistent with the multiple strategies for success observed here. Similarly, other forms of meditation such as Transcendental Meditation, walking meditation, or guided imagery might provide successful alternatives to those who find mindfulness doesn't suit them. Furthermore, if an activity is started but incomplete, the system might suggest easier activities that can be supplemented that day, to maximize benefit and rapidly support self-efficacy. While DStress provided different strategies for reducing stress (Exercise and Meditation), these activities occurred on predetermined days, whereas offering multiple options on the same day might add another layer of flexibility.

Our adaptive approach was complex to implement, as we defined appropriate sets of exercises with qualified exercise coaches. Future designs might exploit pre-existing graded exercise programs to expedite this step. Another issue with our approach is reporting costs. While reporting was designed to be lightweight, the requirement to do so is an imposition on users. Future work might explore more automated, lightweight ways to record progress. GPS might track whether a participant went to the gym and an accelerometer might sense compliance. Physiological responses might help track exercise and meditation activity levels, suggesting a need for easier or more challenging activities. We might also aim to measure stress levels automatically to allow dynamic monitoring of program efficacy, although we are aware that such measures are currently under development [30,31].

There are limitations to this work. DStress reduced DASS Stress scores while the other system versions did not. However there were also inconsistencies across our scales. Why for instance, did DStress reduce one direct measure of stress (DASS) but not the other (PSS)? One explanation is that our DASS stress finding might have been statistical error. While this result was highly significant (p<.001), it is possible that the other non-significant measures were more accurate. An alternative explanation stems from the time scale of the probe question. The PSS probes stress over the past month, while the Stress subscale of the DASS probes it over the past week only. Stress reduction may have been more pronounced in the last week of the study compared to baseline, than over the entire study compared to baseline. Administering stress measures over multiple time points during a future study (or employing physiological sensors) would assess the time scale of benefits to resolve these inconsistencies. Furthermore, PSS measures the degree of stress attributed to situations in one's life, whereas DASS Stress measures psychological tension and irritability. It's also possible that DStress was more effective at reducing appraisals of psychological rather than external sources of stress.

The indirect stress measures such as physical symptoms (CHIPS) and the Depression and Anxiety subscales of DASS, may be less sensitive to exercise and meditation than the direct measures. Alternatively, differences on these other measures may take longer to emerge than the 28-day deployment we explored here.

DStress almost halved DASS stress scores in a month-long field trial setting. But compliance becomes a greater issue over longer periods of time [21]. Longer-term studies with adaptive systems should explore whether DStress promotes longer lasting changes of behavior and stress reduction. In addition, our study provided multiple success strategies, so further research must determine whether exercise, meditation or a combination of both was most effective for stress reduction.

We also need to examine further issues around compliance. Goal-setting theory predicts that the DStress group would be more compliant than the Easy group because they were doing more difficult, motivating exercises while maintaining self-efficacy. We did not find this. It may be that in this context, self-efficacy overshadows the motivations of more challenging exercises. We were also unable to determine whether the increased compliance rates for non-gym exercises in DStress can be attributed to adaptivity or only to the convenience of the location. And since our adaptive components were not manipulated, exercise compliance was simultaneously influenced by both accessibility and task difficulty. Future research may help to tease apart difficulty from accessibility, as well as location convenience from adaptability.

Our results suggest important practical and theoretical directions for computational systems to promote well-being. We designed an automated health system to address multiple facets of compliance, by providing personalization that is driven by theory. By achieving balance between self-efficacy and goal-setting theories, DStress provided an effective coaching strategy for reduced stress.

ACKNOWLEDGMENTS

We wish to thank the certified coaches who helped us generate the exercise content. And thanks too to our participants who were daring enough to try something new. This research was partially supported by NSF grant IIS-1321102.

REFERENCES

1. 4 Minute Abs. http://www.4minuteabs.com/
2. Ajzen, I. From intentions to actions: A theory of planned behavior. *Springer Berlin Heidelberg* (1985), 11-39.
3. Antony, M. M., Bieling, P. J., et al. Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales. *Psychol Assess*, *10*,2, (1998). 176-181.
4. Astin, J. A. Stress reduction through mindfulness meditation. *Psychother Psychosom 66*,2 (1997), 97-106.
5. Bandura, A. Self-efficacy: toward a unifying theory of behavioral change. *Psychol Rev*, *84*,2 (1977), 191-215.
6. Bickmore, T. W., Caruso, L., & Clough-Gorr, K. Acceptance and usability of a relational agent interface by urban older adults. *Ext Abstracts CHI'05,* ACM Press (2005), 1212-1215.
7. Bodenheimer, T., Lorig, K., et al. Patient self-management of chronic disease in primary care. *JAMA, 288,*19 (2002), 2469-2475.
8. Bower, J. L., & Christensen, C. M. Disruptive technologies: catching the wave. *Harvard Bus Rev,* (1995), Jan-Feb, 43-53.

9. Brenner, M. H. Mortality and the national economy. *The Lancet 314,*8142 (1979), 568-573.
10. Brusilovsky, P., & Peylo, C. Adaptive and intelligent web-based educational systems. *IJAIED 13*,2 (2003), 159-172.
11. Cole-Lewis, H., & Kershaw, T. Text messaging as a tool for behavior change in disease prevention and management. *Epidemiol Rev*, *32*,1 (2010), 56-69.
12. Consolvo, S., Everitt, K., et al. Design requirements for technologies that encourage physical activity. *Proc CHI'06,* ACM Press (2006), 457-466.
13. Consolvo, S., Klasnja, P., et al. Goal-setting considerations for persuasive technologies that encourage physical activity. *Persuasive'09*, ACM Press (2009), 1-8.
14. Consolvo, S., McDonald, D. W., & Landay, J. A. Theory-driven design strategies for technologies that support behavior change in everyday life. *Proc CHI'09,* ACM Press (2009), 405-414.
15. Cramer, S. R., Nieman, D. C., & Lee, J. W. The effects of moderate exercise training on psychological well-being and mood state in women. *J Psychosom Res 35*,4 (1991), 437-449.
16. Doherty, G., Coyle, D., & Sharry, J. Engagement with online mental health interventions: an exploratory clinical study of a treatment for depression. *Proc CHI'12,* ACM Press (2012), 1421-1430
17. Gasser, R., Brodbeck, D., et al. Persuasiveness of a mobile lifestyle coaching application using social facilitation. *Persuasive'06,* Springer (2006), 27-38.
18. Graham, C., Franses, A., et al. Psychotherapy by computer. *Psychiatr Bull*, *24*,9 (2000), 331-332.
19. Haug, S., Meyer, C., et al. Continuous individual support of smoking cessation using text messaging: a pilot experimental study. *Nicotine Tob Res 11,8* (2009), 915-923.
20. Hurling, R., Catt, M., et al. Using internet and mobile phone technology to deliver an automated physical activity program. *J Med Internet Res, 9*,2 (2007), e7.
21. Ice, R. Long-term compliance. *Phys Ther 65,*12 (1985), 1832-1839.
22. Insanity. http://www.beachbody.com/
23. Isaacs, E., Konrad, A., et al. Echoes from the past. *Proc CHI'13,* ACM Press (2013), 1071-1080.
24. Kaplan, L. M., & McGuckin, M. Increasing handwashing compliance with more accessible sinks. *Infect Control 7,*8 (1986), 408-410.
25. Kim, S. I., & Kim, H. S. Effectiveness of mobile and internet intervention in patients with obese type 2 diabetes. *Int J Med Inform 77,*6 (2008), 399-404.
26. Klasnja, P., & Pratt, W. Healthcare in the pocket. *J Biomed Inform 45,*1 (2012), 184-198.
27. Lederman, R., Wadley, G., et al. Moderated online social therapy. *TOCHI 21*,1 (2014), 1-26.
28. Lin, J. J., Mamykina, L., et al. Fish'n'Steps: Encouraging physical activity with an interactive computer game. *UbiComp'06*, ACM Press (2006), 261-278.
29. Locke, E. A., & Latham, G. P. Building a practically useful theory of goal setting and task motivation. *Am Psychol 57,*9 (2002), 705-717.
30. Lu, H., Frauendorfer, D., et al. StressSense: Detecting stress in unconstrained acoustic environments using smartphones. *UbiComp'12,* ACM Press (2012), 351-360.
31. MacLean, D., Roseway, A. & Czerwinski, M. MoodWings: A Wearable Biofeedback Device for Real-Time Stress Intervention. *PETRA'13*, ACM Press (2013), 66-73.
32. Matthews, M., & Doherty, G. In the mood: engaging teenagers in psychotherapy using mobile phones. *Proc CHI'11,* ACM Press (2011), 2947-2956.
33. McFarlane, A. C., Atchison, M., et al. (1994). Physical symptoms in post-traumatic stress disorder. *J Psychosom Res*, *38*,*7* (1994), 715-726.
34. McKee, G. H., Markham, S. E., & Scott, K. D. (1992). Job stress and employee withdrawal from work. In J. C. Quick (Eds.), *Stress & Well-Being At Work*. (pp. 153-163). Washington, D.C.: APA
35. Mento, A. J., Steel, R. P., & Karren, R. J. A meta-analytic study of the effects of goal setting on task performance: 1966–1984. *Organ Behav Hum Decis Process 39,*1 (1987), 52-83.
36. Morris, M., & Guilak, F. Mobile heart health: project highlight. *Pervasive Comput, 8,*2 (2009), 57-61.
37. Morris, M. E., Kathawala, O., et al. Mobile therapy: case study evaluations of a cell phone application for emotional self-awareness. *J Med Internet Res 12*,2 (2010), e10.
38. O'Leary, A. Stress, emotion, and human immune function. *Psychol Bull 108,*3 (1990), 363-382.
39. Obermayer, J. L., Riley, W. T., et al. College smoking-cessation using cell phone text messaging. *J Am Coll Health 53,*2 (2004), 71-78.
40. President's Council on Fitness, Sports & Nutrition. http://www.fitness.gov/about-pcfsn/our-history/
41. Prochaska, J. O., & DiClemente, C. C. The transtheoretical approach. *Handbook of Psychotherapy Integration, 2* (2005),147-171.
42. Riley, W. T., Rivera, D. E., et al. Health behavior models in the age of mobile interventions: are our theories up to the task? *Transl Behav Med 1*,1 (2011), 53-71.
43. Rosch, P. J. The quandary of job stress compensation. *Health and Stress, 3* (2001), 1-4.
44. Toscos, T., Faber, A., et al. Chick clique. Ext Abstracts *CHI'06,* ACM Press (2006), 1873- 1878.
45. US Surgeon General. http://www.surgeongeneral.gov/
46. Vidyarthi, J., & Riecke, B. E. Interactively mediating experiences of mindfulness meditation. *Int J Hum Comput Stud 72,* 8-9 (2014), 674-688.
47. Waller, R., & Gilbody, S. Barriers to the uptake of computerized cognitive behavioural therapy. *Psychol.Med 39,5* (2009), 705-712.
48. Walters, D. L., Sarela, A., et al. A mobile phone-based care model for outpatient cardiac rehabilitation. *BMC Cardiovasc Disord 10,*1 (2010), 1-8.