

Just Do Something: Comparing Self-proposed and Machine-recommended Stress Interventions among Online Workers with Home Sweet Office

Xin Tong
xin.tong@dukekunshan.edu.cn
Duke Kunshan University
Suzhou, China

Matthew Louis Mauriello
mlm@udel.edu
University of Delaware
Newark, Delaware, USA

Marco Antonio Mora-Mendoza
marcom3@alumni.stanford.edu
Stanford University
Stanford, California, USA

Nina Prabhu
ninaprabhu@alumni.stanford.edu
Stanford University
Stanford, California, USA

Jane Paik Kim
janepkim@stanford.edu
Stanford University
Stanford, California, USA

Pablo E. Paredes*
paredes@umd.edu
University of Maryland College Park
College Park, Maryland, USA

ABSTRACT

Modern stress management techniques have been shown to be effective, particularly when applied systematically and with the supervision of an instructor. However, online workers usually lack sufficient support from therapists and learning resources to self-manage their stress. To better assist these users, we implemented a browser-based application, Home Sweet Office (HSO), to administer a set of stress micro-interventions which mimic existing therapeutic techniques, including somatic, positive psychology, meta cognitive, and cognitive behavioral categories. In a four-week field study, we compared random and machine-recommended interventions to interventions that were self-proposed by participants in order to investigate effective content and recommendation methods. Our primary findings suggest that both machine-recommended and self-proposed interventions had significantly higher momentary efficacy than random selection, whereas machine-recommended interventions offer more activity diversity compared to self-proposed interventions. We conclude with reflections on these results, discuss features and mechanisms which might improve efficacy, and suggest areas for future work.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**.

KEYWORDS

Stress; Interventions; Online Workers; Machine Learning; Self-proposed; Self-management

ACM Reference Format:

Xin Tong, Matthew Louis Mauriello, Marco Antonio Mora-Mendoza, Nina Prabhu, Jane Paik Kim, and Pablo E. Paredes. 2023. Just Do Something: Comparing Self-proposed and Machine-recommended Stress Interventions

*Corresponding author.



This work is licensed under a Creative Commons Attribution International 4.0 License.

CHI '23, April 23–28, 2023, Hamburg, Germany
© 2023 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9421-5/23/04.
<https://doi.org/10.1145/3544548.3581319>

among Online Workers with Home Sweet Office. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*, April 23–28, 2023, Hamburg, Germany. ACM, New York, NY, USA, 20 pages.
<https://doi.org/10.1145/3544548.3581319>

1 INTRODUCTION

Stressors from work, family, personal health, economic status, etc. take on multiple forms and could happen recurrently. Work-related daily stressors decrease people's productivity, job satisfaction, and affect their overall well-being [2, 7, 20, 49]. As a result, self-care is increasingly recognized as an alternative for daily stress management. However, online workers usually have limited professional support and learning resources available to identify the best tools for stress self-management [2]. Designing efficacious interventions that a) accommodate to personal reactions to episodic acute stress, b) adapt to the intermittent nature of daily stressors, and c) take into consideration contextual dependencies (e.g. locations, events, preferences) is a challenging problem [25].

Rather than focusing on the full range of these tasks, our work focuses on online workers, such as Amazon Mechanical Turk (AMT) workers, since online workers suffer from a range of stressors, including deadlines, interpersonal problems, and struggles with work-life balance [1], and sit long time in front of computers. Furthermore, online workers' continuous online presence opens up the possibility to track their time spent online and to use this information to support their momentary stress management practices by recommending just-in-time micro-interventions [41], that are recommended in a subtle way that might help dissipate stress [35]. Recent research has investigated the general efficacy of stress interventions in multi-week (2-4 weeks) studies [18, 19, 40, 53], as well as the best sensing modalities [61], user's preferred intervention content, [60], and the best time to offer stress interventions [35]. Some of these studies applied artificial intelligence or machine learning (ML) algorithms to predict stress levels [35, 60] or to provide personalized interventions [35, 53]. However, whether machine-recommended interventions are more efficacious than self-proposed interventions or even a user's baseline (do-nothing) has been under-investigated.

In this work, we describe the Home Sweet Office (HSO) application prototype, a browser plugin designed to support online workers by providing access to stress-management micro-interventions. We implemented HSO as a Chrome browser extension plugin that

recommends interventions based on individual traits, personal preferences, past efficacy, and contextual information. HSO can select from over 160 micro-interventions designed based on empirically supported psychotherapy techniques (i.e., positive psychology [64, 65], mindfulness meditation [4, 13, 36, 66], cognitive behavioral therapy [12], etc.). Using HSO, we compared four groups based on their intervention selection method: (1) a control group, consisting of participants who did not install the plugin nor receive any stress interventions, (2) "HSO-Self," consisting of participants who were prompted to use their own self-proposed stress management techniques for momentary stress reduction, (3) "HSO-Random," consisting of participants who received interventions that were chosen randomly, and (4) "HSO-Bandit," consisting of participants who received interventions selected by a multi-armed bandit machine learning algorithm [43]. Using this prototype plugin, we explore how to optimally support participants towards reducing their stress levels through the following primary research questions (RQ):

- RQ1: Can browser-based micro-interventions reduce momentary stress after each intervention and multi-week stress over the course of several weeks?
- RQ2: Are machine-recommendation methods (random or machine-based) more effective than self-proposed interventions in reducing momentary stress?

A prior study compared ML-recommended vs self-selected interventions from the same set of designer-created interventions [53]. Their findings showed that ML-recommended interventions reduced more stress than self-selected ones. We extend this prior work and test self-proposed interventions vs ML-recommended interventions authored by designers, using the same nudge: the HSO browser plugin. To evaluate our system and address our research questions, we conducted two pilot studies using early iterations of the HSO browser plugin. After collecting feedback and iterating on low-fidelity prototypes of HSO, we conducted a four-week field study and evaluated the effectiveness of HSO interventions on $N = 58$ participants regarding their momentary stress reduction, using the plugin on their own devices and at their own homes. Findings from our study indicated that although there were no significant multi-week stress improvements in any of the HSO groups compared with the control group, the HSO-Bandit and HSO-Self groups both had significantly better momentary stress reduction than the HSO-Random group. While HSO-Bandit and HSO-Self were similar in effectiveness, HSO-Bandit interventions offered more rich and diverse intervention content compared to the HSO-Self group, resulting in higher engagement as measured by a longer intervention completion time, which positively correlates with stress reduction. The main contributions of this work include:

- A novel, browser-based system for providing theory-driven stress micro-interventions;
- A field comparison study of stress reduction outcomes among self-proposed, randomized, and machine-recommended micro-interventions; and finally,
- A discussion of HSO's user experiences and potential future work for researchers and application designers in the area of stress micro-interventions as well as personalized healthcare and well-being applications.

2 RELATED WORK

2.1 Workplace Stress and Stress Management

With 80% of workers in North America who report feeling stress on the job and 57% feeling stress on a daily basis, workplace stress is a widely experienced problem in the U.S. population. [8]. Workplace stress further affects workers' productivity levels and job satisfaction as well as their personal life, mental wellbeing, and physical disorders (e.g., disease and illness) [2, 7, 20, 22, 26, 49]. Moreover, healthcare expenditures are nearly 50% greater for workers who report high levels of stress [2]. Psychotherapists and psychiatrists have developed and utilized a wide variety of stress management interventions to support individuals in managing their stress including: cognitive-behavior therapy (CBT) [12], meditation and mindfulness practices [4, 14, 36], physical exercises [50], breathing techniques [10, 38, 55], emotional regulation [44, 58], and so on. Positive psychology [64, 65], for instance, is an emerging practice to help people wind down with personally targeted cues, such as asking people to express gratitude or perform compassionate acts. Positive social interactions have also been shown to improve feelings of calm and openness in social engagement [27]. CBT [12] is another effective therapy which teaches people how to recognize their sources of stress, change their negative behavioral reactions, and re-frame their thoughts. In addition to the cognitive and social techniques, somatic interventions focus on guided breathing and various physical exercises (e.g., yoga, stretches, walking, running, etc.) to promote stress reduction.

2.2 Digital Interventions for Stress

Prior work has focused on the development of promising digital interventions for stress management via mobile and web applications as well as wearable devices and biofeedback sensors [15, 33, 52, 70]. These research prototypes have integrated the aforementioned stress management techniques. For example, Sanches et al. utilized biofeedback information for detecting stress levels in a mobile device and engaged participants in self-reflection on the physiological stress reactions, which improved their stress outcomes and awareness [59]. Similarly, Morris et al. also developed the Mood Map to increase participants' self-awareness of their emotions and ways in coping with stress, which was tested in a one-month field study and found significant stress changes [51]. Heber et al. [31, 32] provided a web-based mobile app to train users' emotional regulation abilities for stress management. Paredes et al. investigated movement-based mindful interventions for commuters to reduce their stress in a car, and proposed sensation patterns on the back of the seat that could guide the mindfulness process based on user study findings [54]. In another example, Schroeder et al. implemented a web based app, PocketSkills [63], to teach Dialectical Behavior Therapy (DBT) via a conversational agent for users to manage their depression and anxiety levels. However, most of these prior work focused on a single stress intervention or a singular use case scenario.

Recent studies have begun experimenting with integrating multiple stress intervention techniques and recommending specific interventions for users (e.g., [3, 35, 53, 60, 61]). Oiva, for example, is a workplace stress management application that integrates acceptance and commitment therapy methods; early pilot work showed active use and good acceptance of the interventions and positive

effects on well-being. [3]. Paredes et al. developed PopTherapy [53], a web-based application providing a wide range of stress interventions to users based on their real-time stress levels. Their study demonstrated that participants showed higher self-awareness of stress and lower depression-related symptoms. The authors further summarized and integrated a comprehensive list of physical, psychological, and physiological stress techniques, sorting them into four intervention categories including more content than in previous studies [3], i.e., somatic, positive psychology, meta cognitive, and cognitive behavior. Sano et al. [60, 61] extended Paredes et al.'s work [53] and re-categorized their intervention types to sleep, diet, and exercise in [60] and then referred to Paredes's four intervention categories again in [61]. In recent work, Howe et al. [35] adapted cognitive behavioral therapy (CBT) and dialectical behavioral therapy (DBT) interventions into digital interventions, and categorized them into three types according to user effort, i.e., get my mind off work (low effort), feel calm and present (medium effort), think through my stress (high effort). Since the PopTherapy covered the most wide range of interventions and techniques, we adopted their interventions and categorization by adding new interventions and editing existed ones.

The effectiveness of web-based stress interventions in the work context has been examined in a body of literature that includes both multi-week field and randomized controlled trial studies with promising outcomes [6, 21, 24, 29, 30]. Contradictory findings regarding the efficacy of web-based interventions persist in the literature (e.g., [31, 32]). While [6] reported positive results in a web-based stress management intervention compared to a control group [6, 30], or conventional self-care [29], Eisen et al. found that computer-based relaxation techniques significantly reduced immediate stress, but the effect was less than the in-person group [24]. Moreover, a few studies have suggested that web-based approaches were no more effective than printed materials in reducing stress [21].

In addition, researchers have been investigating the users' decision options, (e.g., the best timing to offer/request interventions [35]), what tailoring variables (e.g., the content of interventions [60, 61]), and system recommendation rules (e.g., machine-recommended or randomly recommended [53]) to maximize the proximal and distal stress reduction outcomes. For instance, Paredes et al. [53] discovered that the machine-recommended interventions were more promising in higher stress reduction and self-awareness of stress, and lower depression symptoms than the ones participants randomly chosen. In a more recent study, Howe et al. [35] examined times to nudge and user's preferred effort for interventions. Their findings suggested no differences in pre-scheduled ones and the micro ones predicted by sensing algorithms. However, prior work all provided micro-interventions from their systems without considering users' self-proposed intervention content. Therefore, one of goals of our work is to explore how users' self-proposed interventions differ from the system prompted ones (expert-authored) and explore what intervention categories are most effective in reducing AMT workers's stress levels (RQ2).

2.3 ML for Personalized Interventions

Reinforcement Learning (RL) algorithms have been successfully applied in areas ranging from computer games [72] to health [34, 71],

and in particular have been leveraged to recommend interventions for promoting physical activity [69], and reducing stress levels [37, 60–62]. In PopTherapy [53], the authors implemented an Upper Confidence Bound (UCB) multi-armed bandit algorithm [5, 43]. The multi-armed bandit (MAB) problem describes a class of sequential decision problems in which a learner is sequentially faced with a set of available actions, chooses an action, and receives a random reward in response. At each round, the learner accumulates information about the reward compensation mechanism and learns from it, choosing the arm that is close to optimal as time elapses. The challenge of the MAB problem is that the reward that the learner has not previously chosen is unknown—therefore, the learner needs to balance exploitation and exploration, where exploitation means pulling the seemingly best arm based on current information, while exploration refers to pulling another arm to get more information. Since the MAB algorithm does not require a large initial dataset for training and can dynamically learn from newly generated data, we aimed to extend Paredes et al.'s MAB implementation [53] for personalized recommendations in our HSO system.

However, whether ML-recommended interventions recommended via a web plug-in outperforms one of that is either self-proposed or randomly selected remains under-investigated. Therefore, we aim to explore the best way to offer interventions to this population by comparing ML-recommended interventions (HSO-Bandit) random selection (HSO-Random), and user's self-proposed (HSO-Self) interventions to each other and a control group (RQ1).

3 DESIGN AND IMPLEMENTATION OF HSO

Home Sweet Office (HSO) is a prototype Chrome extension designed to provide personalized and, in the future, micro stress reduction interventions to support multi-week stress management practices for online workers. Users can access the HSO system from any device that has access to the Chrome browser. HSO records users' pre- and post-intervention stress levels and allows them to receive micro-interventions for stress management.

3.1 System User Flow

Derived from PopTherapy [53], we further explored the design of stress management micro-interventions in a web-browser context and studied the best strategies (*i.e.*, what and how) to offer these interventions to users. As shown in Figure 1 and Figure 2, HSO provides on-boarding information for initializing the interventions and a browser-based plugin interface for requesting interventions. After installing the HSO application, users first go through an on-boarding stage (Figure 1) to set their initial intervention preferences, including their nudge/prompt time period, interest in interventions that require social interactions, whether or not they want to receive audio notifications, their location, and bed time. After this on-boarding stage, users can begin requesting an intervention anytime they like (Figure 2A), or when they see the prompt, *i.e.*, the icon, blinking (see icon illustrations in Figure 1C). The nudge will stay blue when interventions are not recommended by the HSO system, then blink from blue to red when an intervention first showed up and finally stays in red if the intervention is not attended to. Once users decide to attend to an intervention, they will self-report their stress levels before (Figure 2B) and after the

intervention (Figure 2F). During the intervention, users will be recommended one intervention from the intervention pool (Figures 2C and 2D) and complete an intervention with instructions from the text descriptions or external resources (Figure 2E). Similar to PopTherapy, HSO interventions feature welcoming and friendly names, instructions, and icons to avoid theoretical and hard-to-understand therapy names as well as provides users memorable moments (Figures 2C-E).

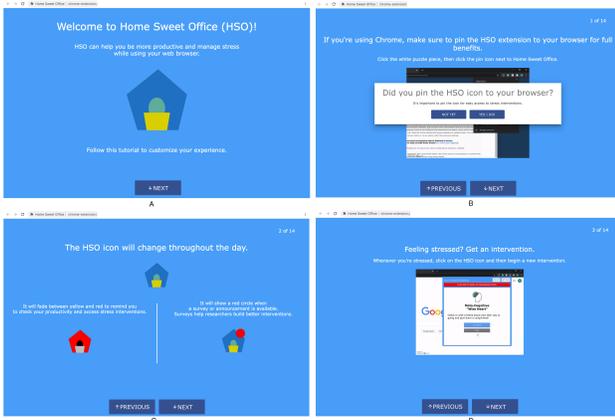


Figure 1: HSO Setup Steps. A. The starting screen introducing the application. B. Users were asked to pin the plugin to the browser so the HSO icon would appear. C. The icon will blink from yellow to blue when it recommends an intervention. And turn to red (left) if overdue after a while. D. One intervention example.

3.2 Micro-Interventions in the Pool

Adapted from the interventions in the PopTherapy prototype [53] and based on results from our two pilot studies, we designed and implemented a total of 160 micro-interventions under the same four therapy groups: Meta-Cognitive, Cognitive-Behavioral, Somatic, and Positive-Psychology. Similarly, the instructions of our micro-intervention also has two simple components (Figure 2 D): a text prompt that tells the user what to do and resources that launches the appropriate online tools to execute the micro-intervention including external websites, web applications, and social media platforms. While creating these interventions, we followed several guidelines that arose both from our iterative design process and findings from our first pilot study:

- **Short Completion Time:** Users should be able to complete the interventions within a few minutes (*i.e.*, <3).
- **Concise and Clear Description:** The descriptions of the interventions should be written in a concise and clear manner, and use short and simple sentence structures so they are easy to read and understand.
- **Simple Process and Limited Commitment:** Users should be able to complete the intervention action in simple steps, which require limited commitment, so it doesn't increase users' stress levels.

- **One Immediate Action per Interaction:** Assign one action per intervention and use minimum action verbs. The action should be something users can attend to immediately, instead of later today or at another time.
- **Provide Specific Examples:** Avoid vague and general terms in the instructions, and give proper resources or links for completing each intervention.
- **Balanced among Categories:** The initial few interventions should be balanced from each category.

Appendix Table 1 showed four therapy groups and definitions of each group, along with the number of interventions under each of the therapy groups and each technique and sample interventions. Two co-authors validated each of the interventions following the above-mentioned design guidelines and polished all of our interventions by editing, adding, or removing the current intervention content. See PopTherapy research [53] for more descriptions about the original four therapy groups and Supplementary Table 1 for more sample interventions from HSO.

3.3 System Implementation

Appendix Figure 1 demonstrates the application architecture and the server-client technical framework. The back end of the app was developed using a Node.js server and data was stored in a Google Cloud database. The multi-armed bandit recommendation algorithm was also deployed in Google Cloud. Study surveys were conducted via a system external to the HSO extension. Data aggregation and analysis took place in university servers. Users who consent to participate are provided with an on-boarding survey of their preferences for receiving interventions as soon as they download the application. At regular intervals, users are nudged via a color-changing icon to complete an intervention. Once the user has requested an intervention, they are prompted to self-report their current stress levels. After recommending the intervention, the system will prompt the user to self-report the change in their stress level. The application then keeps track of the specific intervention ID, user ID, intervention category, intervention completion time, nudge responding time, nudge states (active or not), and the user's self reported stress level before and after the intervention.

3.4 Multi-Armed Bandit Recommendations

To recommend interventions to participants in the appropriate group, we use a multi-armed bandit recommendation algorithm. In our case, the bandit must make a decision that takes advantage of commonly liked techniques (exploitation) while experimenting with new techniques that lack preference information from the user (exploration). The bandit's exploration constant, V , determines this trade-off, with a higher value biased more towards exploration. When making a decision, the bandit computes the upper confidence bound (UCB) of all techniques, which is a function of existing predictions of feedback for each technique (using a random forest model), the current input, and the exploration constant. The exploration constant is multiplied by the standard deviation between estimators in the model and added to the existing prediction, which means that techniques with more variance (because there is less feedback on the technique) will have higher UCBs with higher exploration constants. We decided to select between techniques

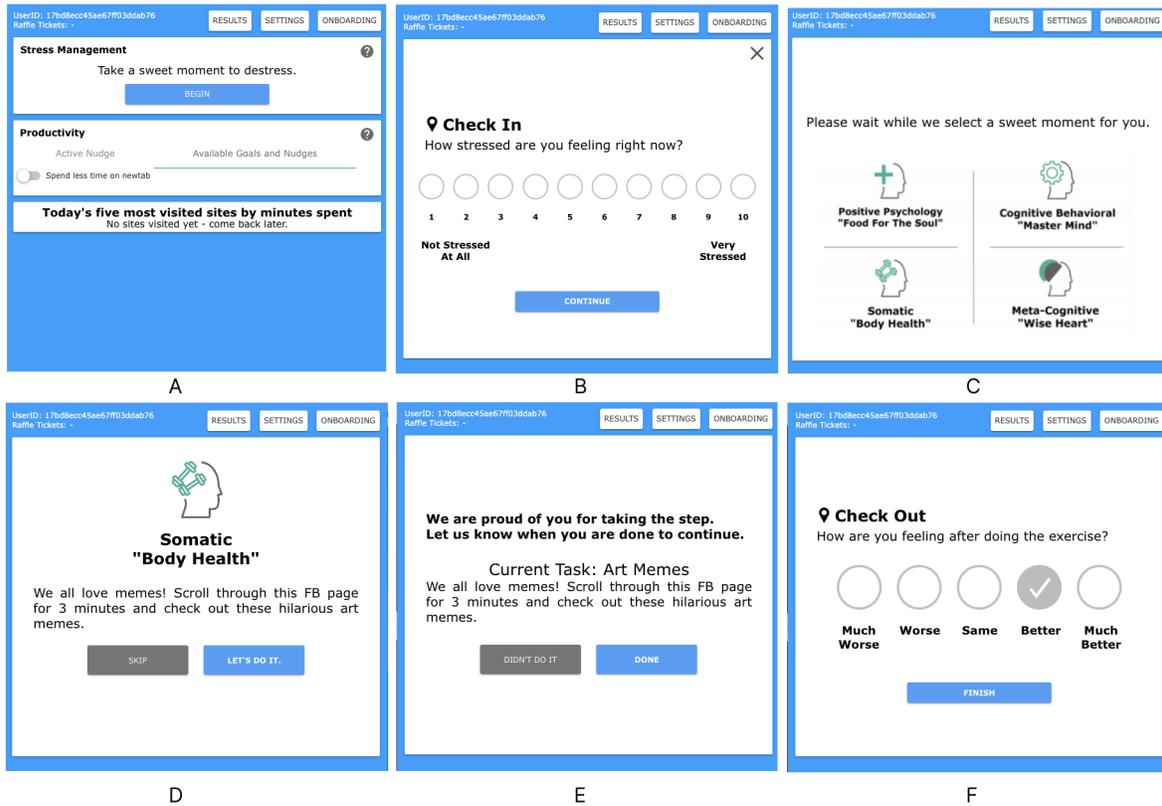


Figure 2: HSO User Flow: a user would go through steps A to F for a complete intervention from their browser drop-down screen. A. The start screen of HSO plugin. B. The self-reported pre-intervention stress-level check-in screen. C. The intervention loading screen. D. One example of the HSO-recommended interventions (the current one shows the category and title of the intervention). E. Detailed instructions of the recommended intervention (and some interventions will lead to external links and resources). F. The self-reported post-intervention stress-level check-out screen.

instead of specific interventions at the bandit level because there are 26 techniques as opposed to over 160 interventions, giving us more aggregated data per technique, whereas some interventions were never or very rarely seen by users in past trials, making them less useful to train on.

3.4.1 Inputs. The bandit takes in as input the user's current stress level (integer on 1-10 scale), what tab they are on (one-hot vector for most popular sites), how many tabs they have open, if they have the HSO app pinned (bool), and if they have HSO nudges activated (bool). Other contextual information was experimented with; however, it was dropped from our training process due to inconsistency of the availability of this information between users or subpar performance compared to the current feature set. Additionally, the user's ID and previous intervention history is not used to avoid biasing the model towards certain users.

3.4.2 Recommendations. When a user requests an intervention, the bandit algorithm uses the input it is given to recommend an

intervention technique it believes will receive the highest reward. The reward was defined as the change in stress that the respondent gave for the chosen intervention (translated to a -2 to 2 scale). For the very first decision, a random technique and intervention are chosen. After the first decision, the bandit selects the technique with the highest UCB and one intervention from that specific technique is chosen randomly to recommend to the user.

3.4.3 Training and Simulation. Bandits were trained on our existing pilot data. During training time, a technique and input combination is used if it was actually recommended to the user and therefore has a given reward. If the correct technique was chosen by the bandit, the bandit's random forest model is updated with the input, chosen technique, and reward (user feedback for that intervention). When deciding on the parameters to tune our bandit for the study, we simulated and compared three bandits with different exploration constants, as well as a purely random algorithm, for 200 time steps (decision point). 5 trials of each bandit were run, with the results shown in Figure 3. In the graph, purple is random,

and the exploration constants are $v=0.1, 0.5,$ and 1.0 for blue, red, and green, respectively. Solid lines are the median over trials and dashed lines represent the first and third quartiles. We can see that $v=0.5$ performed the best, so we used this for our main study.

3.4.4 Activation. Each user is recommended the same three interventions after downloading the HSO application to avoid drop-off as a result of early exploration, but after getting initial feedback the bandit algorithm suggests subsequent recommendations.

3.4.5 Updating. In our implementation of the multi-armed bandit algorithm, we chose to use a warm-start bandit (*i.e.*, one trained on past pilot data) for the current study as we wanted to avoid high initial drop-off from an untrained bandit’s sub-optimal recommendations. We also wanted to let the bandit learn from the participants in the study. To accomplish this goal, we updated the bandit at intervals throughout the study period using user feedback. We also wished to avoid insubstantial updates, so we manually updated the bandit with new data twice a week (*i.e.*, 8 times total throughout the study).

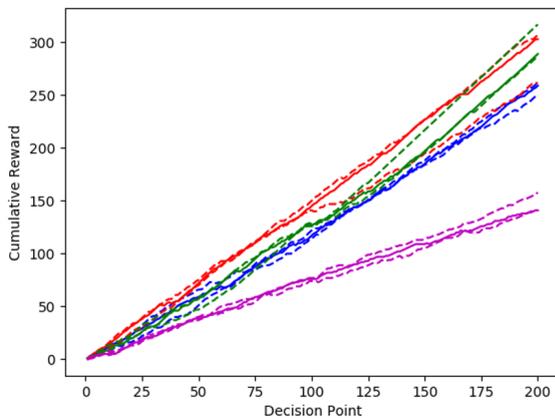


Figure 3: Results of 5 simulations over 200 timesteps for cumulative reward over 200 timesteps for random (purple) vs. bandits with exploration constants 0.1 (blue), 0.5 (red - chosen), and 1.0 (green). Dashed lines are first and third quartile, and solid lines are median of simulations.

3.5 Pilot Studies

We describe two pilot user studies and major findings that informed iterations of the HSO UI and intervention content.

3.5.1 Pilot Study 1. Fifteen participants were recruited in our first pilot study conducted April to June 2021 to better understand the usability of the HSO application, evaluate the stress interventions, and collect preliminary application log data. All of the participants were asked to install and use HSO for two-week period, during which participants received randomized interventions from our intervention pool. In addition to application logs, we collected users’ feedback about the usability of the HSO tool—in particular, the

interface design and effectiveness of the randomized interventions through pre-, weekly, and post surveys, and qualitative interviews.

A total of 64 interventions were completed by those participants with each participant completing 1 to 13 interventions in 50.4 ± 100.7 seconds (intervention completion time). Self-reported usability ratings suggested that participants generally held positive attitudes about the HSO system. Most (10/15) rated their experience as very positive, 3 rated it only slightly positive, and 2 were neutral. Similarly, from the qualitative interviews, a majority of the participants (12/15) reported that they thought HSO was useful in different ways including providing for breaks (*e.g.*, “Usefulness and productivity of good breaks” P12) and help them to acknowledge the stress they feel (*e.g.*, “Yes, it’s good to acknowledge your stress and ways to cope with it” P3). While most pilot participants did not have a preference for specific intervention categories, two reported that they preferred “Somatic” and “Positive Psychology” interventions. Approximately half of the interventions reduced participants’ stress levels. Nonetheless, no significant differences were identified in anxiety, anger, depression, and sleep survey scales before and after the study. Participants also did not complete a large amount of interventions (average of 4.27 interventions for two weeks) due to concerns about the intervention content (*e.g.*, content that required social interaction was difficult, if not impossible, to complete). Therefore, we developed our intervention design principles, described earlier, and revised all interventions in the pool to improve them. Moreover, we also used participants’ self-reported before and after stress data from the first pilot study to train the bandit recommendation algorithm used in our study.

3.5.2 Pilot Study 2. Another fifteen participants were recruited during August to October 2022. We aimed to further understand users’ responses to the ML-recommended interventions and the effectiveness of these interventions in stress management. We adopted the same mixed-method study design as described in the first pilot study and asked participants to install and keep using HSO for two weeks. But participants only received the bandit algorithm-recommended interventions, instead of random ones. The same set of surveys, browser data, and interview questions were collected.

In general, most people gave positive comments about increased awareness to their stress levels and interventions they learned to perform for stress management. Participants requested a total of 702 interventions from HSO, out of which 220 were completed (*vs.* 64 in the first pilot study). Participants completed 1 to 38 valid interventions during the two weeks (Mean \pm SD = 13.2 ± 12.6). The completion time of the stress interventions ranged from 1-10 minutes, around 3 minutes on average. Out of all interventions, most (60%) improved participants’ stress levels and less than half (38%) interventions did not have any effect on *in-situ* stress levels. Only a few (2%) made people feel worse. However, no significant differences were identified before and after the study regarding the anxiety, anger, depression, and sleep survey scales. As a result, we decided to finalize our intervention pool and user flow. We then decided to conduct a formal study with a bigger sample size to investigate how the micro-interventions from the HSO platform would engage users in stress management in the multi-week and what would be the best strategies to recommend these interventions.

4 METHODOLOGY

The goal of our study is to evaluate the effectiveness of three types of stress management intervention recommendation approaches using the HSO tool. These approaches include HSO-Self, HSO-Random, and HSO-Bandit to a non-intervention control group. Using these different conditions, we aim to investigate how to support online workers in alleviating stress via micro-interventions in the multi-week and identify the most preferred and effective intervention recommendation approach in our system for reducing stress. In this work, we recruit AMT workers with the goal of developing transferable insights which help scale our system to support and increasing number of diverse online workers.

4.1 Hypotheses and Rationale

Based on our RQs, here we introduce the following hypotheses and rationales:

H1: Participants' momentary stress levels would be reduced using HSO intervention prompts. However, their long-term stress levels might not change. We assume this because long-term stress reduction may require resolving the fundamental causes of stress and stressors [25], which HSO did not offer.

H2: The machine-recommended interventions could be more effective than the participants' self-proposed ones because the machine recommended interventions were derived from a set of expert-recommended ones and would customized to the individuals.

4.2 Study Design

Our primary research goal was to examine how users might respond to different, momentary stress management interventions. Our participants were randomly assigned to different conditions which included one for support through browser-based nudges to perform a self-proposed intervention (in the HSO-Self group) as well as conditions where users also received randomly-assigned (HSO-Random) or machine-recommended (HSO-Bandit) stress intervention content for self-management. We then study how, if at all, participants' stress would be improved after engaging with the HSO tool. More specifically, we extend prior work to test self-proposed interventions against ML-recommended interventions authored by designers using the same nudging prompt to engage participants (the browser plugin platform). As excessive prompting and surveying may increase participant stress and burden, our interventions were specifically designed to be short and focus on measuring stress relief; we did not use our *in-situ* surveys to explore the reasons for their stress. As a result, this research adopted a between-group study design to compare the three intervention approaches. The independent variable was the type of stress interventions recommendation approach that participants received. The four intervention conditions are as follows:

- Control: Participants received no interventions from HSO and they were not instructed to do anything to manage their stress levels during the study period. They only completed the weekly surveys that collected their demographic information and multi-week stress levels.
- HSO-Self: The HSO extension asks participants to employ any stress reduction management technique that they were familiar with and/or preferred. The prompt was "Take a

moment and do whatever comes to mind to reduce your stress. Write what you did below." Users could then write a brief sentence about what they did before being asked about their stress levels.

- HSO-Random: Participants received a randomized intervention from the HSO intervention pool (as introduced in section 3.3) every time they clicked on the HSO button in their browser.
- HSO-Bandit: Participants received a bandit-recommended intervention from the HSO intervention pool (as described in section 3.5) every time they click on the HSO button in their browser.

For the three HSO groups, participants were asked to install the HSO plugin in the browser and complete interventions on a daily basis while completing several standardized scales as part of a weekly survey.

The dependent variables of this study were (1) PROMIS depression, anxiety, and sleep disturbance subscales from the surveys and (2) self-reported stress levels from the browser. See below subsection 4.4 for details about the measurements and instruments of each dependent variables.

4.3 Participants

4.3.1 Recruitment. Participants were primarily recruited from AMT and Facebook online advertisements. The inclusion criteria included: (1) use of the Chrome web browser on a daily basis for work or school related tasks; (2) must be a healthy individual; (3) be aged 18 and older; (4) fluent in the English language; and (5) able to provide informed consent. Participants who were willing to join this study were first asked to fill out a screening survey containing their basic demographic information. We filtered out people who did not meet our inclusion criteria and reached out to the rest of the participants via emails with our pre-study survey and detailed study procedures and instructions. The research protocol used in this study was approved by Stanford University's Institutional Review Board (IRB).

4.3.2 Filtering and Selection. Intervention Data from the Back-End. Upon a closer examination on the intervention data, we found a technical glitch in assigning participants to different groups. Some participants in the three HSO groups only received the HSO-Self intervention in the first 2-7 days and then they were re-categorized into a specific group (either HSO-Self, HSO-Random, or HSO-Bandit).

A total of 2,258 interventions were recorded in HSO's application logs. Among all, 476 interventions were removed because of missing intervention IDs (preventing them from being matched with specific intervention content); 93 interventions were removed because of missing stress ratings; 257 interventions were removed due to missing completion times. Next, 390 more interventions were removed because some users were added to a different group for the first few days of the study before being re-assigned to a new group. Thus, we removed their initial intervention logs. We further removed 18 interventions of participants that completed less than three interventions in total (since the first three interventions were used to initialize the bandit algorithm's recommendations in the HSO-Bandit group). Finally, 1,016 completed interventions remained for analysis, completed by 58 participants in total, 31

participants from the HSO-Self group, 13 from the HSO-Random group, and 14 from the HSO-Bandit group.

Survey Data. A total of 462 participants filled out the initial pre-study survey, and 104 completed the post-study surveys (31 from the Control group, 24 from HSO-Self, 26 from HSO-Random, and 23 from HSO-Bandit). Based on the intervention data contained in our application logs, we further filtered out the participants' survey data based on the following conditions:

- No matching interventions found for the participant. This means that the participant filled out the pre- and post-surveys, but did not perform any interventions.
- Not part of the 58 participants who completed valid interventions. This refers to when participants filled out the pre- and post-surveys, but did not have any valid or completed more than three interventions during the entire four weeks.
- No valid entries for the PROMIS scales in either the pre- or the post surveys.

As a result of this filtering, there were 69 participants' survey results included, 23 in the control group, 24 in the HSO-Self group, 13 in the HSO-Random group, and 9 in the HSO-Bandit group.

Groups	Gender	Age Ranges
Control	F: 10, M: 13	18-30 (9), 30-55 (12), 55+ (2)
HSO-Self	F: 13, M: 11	18-30 (10), 30-55 (13), 55+ (1)
HSO-Random	F: 7, M: 6	18-30 (3), 30-55 (8), 55+ (2)
HSO-Bandit	F: 4, M: 5	18-30 (2), 30-55 (5), 55+ (1)

Table 1: Survey Respondents' Demographics Information

4.3.3 Demographics. Table 1 showed participants' gender and age information and Supplementary Table 2 demonstrated more demographics. Participants from both the Control and the three HSO groups were compensated with \$10 USD Amazon gift cards after completing at least the pre- and post surveys, and they also received 25 raffle tickets for completing each survey. Moreover, participants from the three HSO groups received one raffle ticket per day for four weeks if they completed at least one stress intervention per day. At the end of the study, ten raffle awards were available for ticket-holding participants, including six \$100 and two \$200 USD Amazon gift cards, and a smartphone worth approx. \$1000 USD.

4.4 Procedures

First, interested participants from AMT and other sources filled out our contact form. After participants received the pre-study survey from us via email, they were informed about the study goals and procedures, and our consent and intake procedure. In the same email, participants were also provided with detailed instructions on how to install, setup, and then use the HSO Chrome plugin for four consecutive weeks. During the study, participants were asked to try their best to complete at least one HSO stress intervention per day, with no limit to the number of stress interventions they could perform. They evaluated their stress levels before and after each HSO intervention within the plugin. At the end of Week 1, Week 2, and Week 3, participants were asked to fill out a weekly survey. Finally, after four weeks, participants were asked to fill the post-study survey. Figure 4 illustrated the detailed study procedures.

4.5 Instruments

We collected participants feedback and performances from three approaches: (1) the pre-, weekly, and post surveys (with the PROMIS scale); (2) browser intervention data; and (3) qualitative open-ended questions. In this section, we introduce data collected from each of these instruments.

4.5.1 The PROMIS Scale in the Surveys. PROMIS scale [17, 56] is short for Patient-Reported Outcomes Measurement Information System, which has been used by clinicians to evaluate people's well-being and health. Several clinical studies have validated the the psychometric properties of PROMIS sub-scales [28, 67]. We collected participants' responses of the PROMIS' depression, anxiety, and sleep subscales in the five surveys (pre-, three weekly, and post).

4.5.2 Self-Reported Stress Ratings from the HSO Browser Logs. Each time a participant requested/received an intervention, they would first have to self-report their stress level from 1 (not stressed at all) to 10 (very stressed). After completing the intervention, they would be asked to rate their recent change in stress by selecting one of five options: much worse, worse, same, better, much better. Prior work has adopted users' subjective ratings to evaluate momentary stress levels and stress delta for measuring post-test changes [53]. Similarly, we collected the stress delta ratings to (i) give participants clearer options of evaluating their stress changes without letting them thinking too much about stress after the interventions; and (ii) minimize their cognitive effort so they do not need to search their memory and compute stress levels.

4.5.3 HSO Browser Intervention Records. We also recorded each intervention's contextual information including the intervention type and ID, the intervention group (self vs. random vs. bandit), self-proposed intervention input (for HSO-Self group only), whether the intervention was completed, intervention Completion Time (CT), nudge Response Time (RT), active or inactive nudge status, and participants' open-ended qualitative feedback towards a specific intervention. Intervention CT was defined as the duration from when the participants click on the "Let's do it" button (Figure 2D), to when they click on the "Done" button (Figure 2E). Nudge RT is duration from when the HSO plugin icon is active Figure 1C to the point when participants click on the "Let's do it" button. Active or inactive nudge status refers to the state of the nudge icon (Figure 1C) at the moment when participants click on the "Let's do it" button. Our goal was to measure and evaluate participants' behaviors using HSO, including the time they devoted to each intervention and whether the nudges impacted their intervention decisions and outcomes. We then compared how these behaviors might be different across three HSO study groups.

4.5.4 Open Questions and Ratings about Usability in the Surveys. In addition, in the post-test surveys, we also added questions asking participants about their usability ratings (through a numerical rating scale) of the HSO system, their general usage of HSO interventions, their preferences about interventions and HSO app features, and their general feedback (through open-ended questions). See our Supplementary Material for all survey questions collected in pre-, weekly, and post tests.

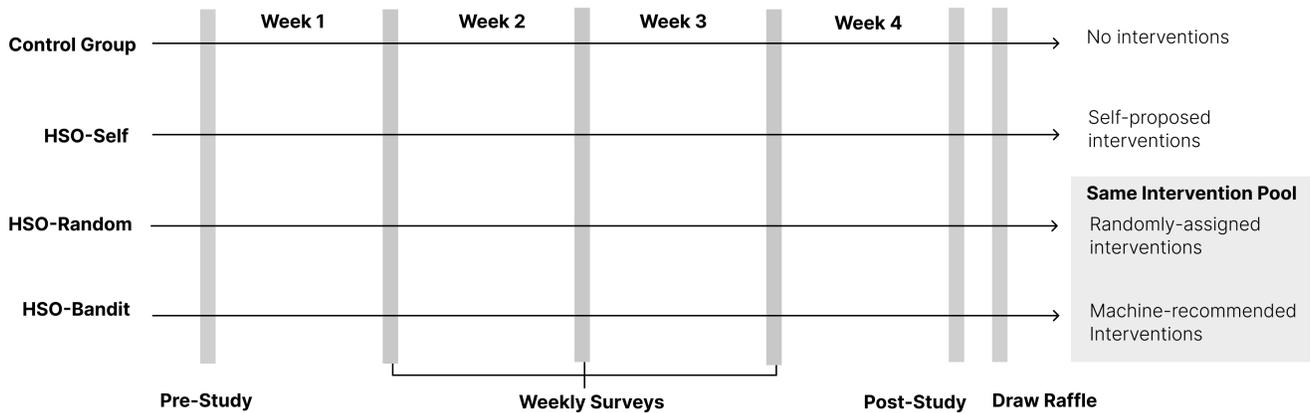


Figure 4: The study procedure diagram. Upon starting the study, participants were assigned to one of the four groups (Control, HSO-Self, HSO-Random, and HSO-Bandit) and received either different stress interventions or none at all. All participants were asked to fill out the pre-, weekly, and post-surveys during the four-week study. In the end, they both received a small fixed monetary compensation and tickets for a raffle.

4.6 Data Analysis

4.6.1 Quantitative Data Analysis. To examine the effects of the independent variable on participants stress levels (i.e., PROMIS subscales and self-reported stress levels), including the interaction effects, a Mauchly’s Test of Sphericity and repeated-measures ANOVA were performed for each dependent variable. If Mauchly’s Test of Sphericity was violated, we used a Greenhouse-Geisser correction for F and p values from ANOVA indicated by F^* and p^* . If any independent variable or combinations had statistically significant effects ($p < 0.05$), Bonferroni-corrected post-hoc tests were used to determine which pairs were significantly different. To correct for the increased risk of Type I error for significant results from pairwise differences comparison, we used linear mixed-effects models to investigate the relationship between per-participant characteristics and outcome variables and then compared pairwise differences using Tukey’s HSD tests to adjust for repeated testing.

4.6.2 Qualitative Data Analysis. Two researchers used an inductive thematic coding approach to analyze the open-ended questions from the post-test survey and from participants’ self-input interventions in the HSO-Self group. Both adhered to the following coding process: open coding to identify all concepts, axial coding to establish categories, and finally selective coding to decide on themes and meaning construction. The refinement of themes was also done through discussions with the research team. A written narrative was provided for each theme with relevant quotes from participants that support these themes.

5 RESULTS

We first report on our quantitative findings from multiple sources (i.e., surveys and *in-situ* intervention feedback). The multi-week mental well-being measures on the surveys includes the PROMIS scale data (i.e., a standardized scale that measures anxiety, depression, and sleep quality) were collected from pre-, weekly, and post-study surveys. Short-term intervention effectiveness from HSO logs

was collected every time participants self-evaluated their before and after stress levels for an intervention. We also report participants’ HSO-Self intervention content, along with other qualitative usability data, and feedback on the HSO system in general.

5.1 Quantitative Findings about Stress Changes

In this section, we report on the effectiveness of HSO interventions with respect to multi-week (PROMIS data collected from surveys) and immediate (from application logs) changes. We then compare the stress change results between the three HSO groups. Thus, with these intervention effectiveness results, we look to answer RQ1 on micro-intervention’s effectiveness for stress reduction and the best kind of recommendation strategy.

5.1.1 Anxiety, Depression, and Sleep Ratings from the PROMIS Scale.

After removing participants with incomplete data, we analyzed the rating on the PROMIS anxiety, depression, and sleep rating for all remaining participants in the control group and the three HSO groups. Figure 5 A-F showed the mean and SE values for PROMIS anxiety, depression, and sleep ratings collected in pre-, weekly, and post surveys. Two-Way 4×5 (Groups*Time) Repeated Measures ANOVA tests were carried out on the three PROMIS ratings collected in the pre-, three weekly, and post surveys from the Control (N=23), HSO-Self (N=24), HSO-Random (N=16), and HSO-Bandit (N=9) groups. We compared the four groups (i.e., Control and three HSO groups) and the five Time ratings that were collected (pre-study, week 1, week 2, week 3, and post-study). The Group was an independent factor and the Time was a within-subjects factor. However, we did not identify any significant differences of the interaction effect ($F(3, 68) = 2.06$, $p = .208$, $\eta_p^2 = .507$), or Group and Time main effect on participants’ anxiety, depression, or sleep ratings, which indicated that participants’ self-reported well-being and stress ratings did not change significantly during the four weeks. See Figure 5A-F for the mean and SE values of the self-reported anxiety, depression, and sleep PROMIS ratings

from the four groups during four weeks. While we did not find significant multi-week survey results, the Control group remains relatively flat; and anxiety, sleep and depression are trending down for the three HSO groups, as shown in Figure 5 A-C.

5.1.2 Descriptive Analysis of the Back-End Stress Data. After filtering, participants had completed 465 interventions in HSO-Self group (for each participant: Mean \pm SE = 16.62 \pm 20.13), 201 interventions in HSO-Random group (for each participant: Mean = 23.52, SE = 28.45), and 350 in the HSO-Bandit group (for each participant: Mean = 42.50, SE = 30.29). The stacked bars in Figure 6A illustrate the number of the completed interventions' category distributions and nudge activity distributions of the three HSO groups. Participants completed a majority of the interventions when being nudged by the HSO system with 84.73% interventions in HSO-Self group, 76.62% in the HSO-Random group, and 67.71% the in HSO-Bandit group, regardless of the intervention content. As shown in Figure 6A, participants in the HSO-random group completed a roughly even number of Somatic (25%), Meta-Cognitive (20%), and Cognitive-Behavior (23%) interventions, with a slightly higher amount of Positive-Psychology (32%) interventions. In the HSO-Bandit group, participants completed more Meta-Cognitive (30%) and Positive-Psychology (34%) interventions compared to Somatic (19%) and Cognitive-Behavior (17%) ones.

Figure 5G illustrated the distributions of total daily interventions completed by participants in the three HSO groups. Since the HSO-Self group included the most participants, this group showed more interventions completed by the first week. Participants in HSO-Random and HSO-Bandit groups completed a similar amount of interventions at the start, but the HSO-Random group experienced a drop in completed interventions and showed less interventions across the four weeks than the HSO-Bandit group. In general, 5G demonstrated that all three groups' total daily number of interventions dropped gradually during the four weeks. However, the HSO-Bandit group had more completed interventions in most of the days after week 1 and showed a more stable number of interventions from week 2 to 4 compared with the other two groups.

Furthermore, figures 5H and 5I showed participants' self-reported stress levels before each intervention and their self-reported improvements after each intervention. These figures did not present any tendency in participants' self-reported stress before or stress changes after the interventions. Because fewer interventions were completed in the later half of the study period, both Figure 5H and Figure 5I show a big turbulence in the later half of the figure.

5.1.3 Inferential Analysis of Stress Data from Application Logs. We further analyze participants' self-reported stress before each intervention (SB) and stress changes (SC) after each completed intervention and compared the three HSO groups differences using Independent One-Way ANOVA analysis. The between-group variable is the Group and the dependent variables were SB and SC. Figure 6C showed participants self-reported stress levels (numerical ratings from 0 to 10) before the intervention, and Figure 6D showed the distribution of stress changes in three HSO groups (-2: much worse, -1: worse, 0: no changes, 1: better, 2: much better).

The analysis revealed significant differences in SB between the three HSO groups, $F(2, 972) = 3.34, p = .036$. Tukey's HSD post-hoc test revealed that only the HSO-Bandit group (Mean \pm SE = 4.61 \pm

0.11) had a borderline higher SB compared to the HSO-Self group (Mean \pm SE = 4.22 \pm 0.12), $p = .056$. The HSO-Random group (Mean \pm SE = 4.16 \pm 0.18) was not significant different in its SB compared with the HSO-Self group ($p = .954$) nor the HSO-Bandit group ($p = .092$). Therefore, there were no significant differences of SB ratings across three HSO groups.

Independent One-Way ANOVA analysis on the stress changes (SC) found a significant difference for the Group variable on SC, $F(2, 972) = 12.73, p < .001$. From the Tukey's HSD post-hoc test, we found that both of the HSO-Bandit group (Mean \pm SE = .74 \pm .05) and the HSO-Self group (Mean \pm SE = .65 \pm 0.03) had significantly better stress improvement than the HSO-Random group (Mean \pm SE = .37 \pm .06), both of which $p < .001$. We did not observe any significant differences between the HSO-Self and HSO-Bandit groups, $p = .25$.

5.2 Nudge Response, Completion Time, & Stress

To answer RQ2, we report findings from HSO logs on nudge responding time (RT) and intervention completion time (CT) by comparing the differences among the three HSO groups.

5.2.1 Nudge RT and Intervention CT among Three HSO Groups.

Nudge Responding Time (RT). Participants in HSO-Self, HSO-Random, and HSO-Bandit groups had Mean \pm SE values of 81.70 \pm 9.63 seconds, 65.83 \pm 22.80 seconds, and 148.57 \pm 7.77 seconds nudge RT, respectively (Figure 6E). Results from an independent One-Way ANOVA analysis found significant differences in the three groups' nudge RT, $F(2, 972) = 10.28, p < .001$. Tukey's HSD post hoc test showed that HSO-Bandit group had significant more nudge RT than both HSO-Self group ($p < .001$) and HSO-Random group ($p < .001$). There was no significant differences between HSO-Random and HSO-Self groups in their nudge RT. This finding suggests that participants in the HSO-Bandit group spent more time responding to the nudge than both the HSO-Self and HSO-Random groups when an intervention had been recommended by HSO.

Intervention Completion Time (CT). Participants spent an average of 57.93 \pm 11.24 seconds in completing each intervention in HSO-Self group, 39.04 \pm 6.39 seconds in HSO-Random group, and 89.03 \pm 40.87 seconds in HSO-Bandit group (Figure 6F). However, from an independent One-Way ANOVA analysis, we did not identify any significant differences of participants' intervention CT in three groups, $F(2, 972) = .779, p = .46$. Although the statistical results did not show significant differences, we observed that participants spent more time completing the interventions in the HSO-Bandit group than HSO-Self group, and then, HSO-Random ranking last. Interestingly, the HSO-Bandit group had the largest SE value of intervention CTs compared to the HSO-Self and HSO-Random groups, as shown in Figure 6F. We theorize that participants experienced a wider range of interventions when the bandit algorithm tried to explore the pool of interventions compared with the other two groups, resulting in varying completion time spread far from the mean value. Participants were also most engaged in the HSO-Bandit group when the interventions were recommended by the ML algorithm, more engaged when they self-proposed the interventions, and least engaged when the interventions were randomly assigned from the intervention pool.

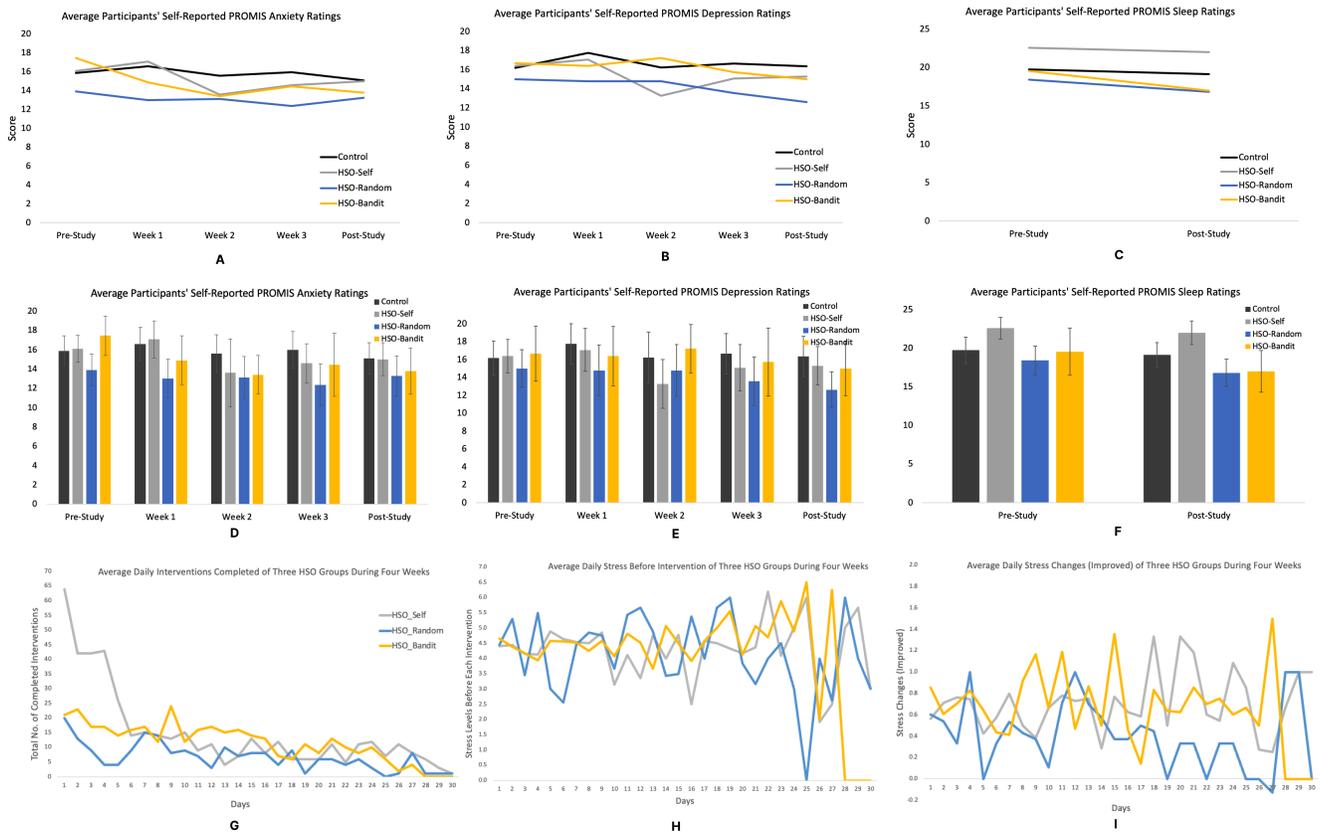


Figure 5: A-F. Mean and SE values of self-reported PROMIS anxiety, depression, and sleep ratings from pre- and post surveys. Anxiety and depression ratings also contain the three weekly results. **G.** Total number of completed interventions by the three HSO groups during four weeks. **H.** Average self-reported stress levels before each intervention of all three HSO groups during four weeks. **I.** Average self-reported stress improvements of the three HSO groups during four weeks.

5.2.2 User Requested vs. System Nudged Interventions & Stress Changes. We examined the effect of whether participants requested an intervention or responded to a system nudge on the stress changes outcomes in three HSO groups (see Figure 6A for more details about the interventions' nudge status in each group). As mentioned in the Method section, "Active" indicates that users responded to the intervention during or after the "nudge" blinks (or flashes) at them, which was designed to provide reminders at a certain time periods; while "Inactive" means that users requested the intervention when the "nudge" was not engaged by the system. Figure 6A illustrated the distributions of completed interventions and did not including incomplete ones. We run independent One-Way ANOVA tests in each of the three groups to identify differences in SC and intervention CT. We found a significant stress reduction in all three groups when the nudge was not active: HSO-Self ($F(1,452) = 5.475, p = .02$), HSO-Random ($F(1,185) = 19.329, p < .001$), HSO-Bandit ($F(1,332) = 102.434, p < .001$), which showed that user requested interventions could reduce stress significantly better than the nudged ones. As for the effect of nudge activity on intervention CT, independent One-Way ANOVA tests revealed that only HSO-Random had a significantly longer CT when the nudge was

not active, $F(1, 185) = 9.01, p = .003$. This finding suggested that participant spent longer time completing interventions when they requested the HSO randomized ones.

5.3 Intervention Content and Categories

Next, we present qualitative results of participants' self-reported intervention content and quantitative analysis results about the intervention categories in HSO-Random and HSO-Bandit groups from the back-end logs to answer RQ2.

5.3.1 Qualitative Findings: Self-Proposed Interventions. Two researchers coded participants' self-proposed interventions and mapped them back to the four categories in HSO's intervention pool. Here, we present an overview of participants self-proposed interventions and then compare them with interventions from HSO's pool.

Overview of Participants' HSO-Self Interventions. Each participant has their own patterns of proposing intervention content, which were usually limited in intervention content (range between one to five) or categories (less than two, after mapped back to HSO's four categories). A majority (80%) of the self-proposed interventions were already parts of their daily routine and were physical, such as

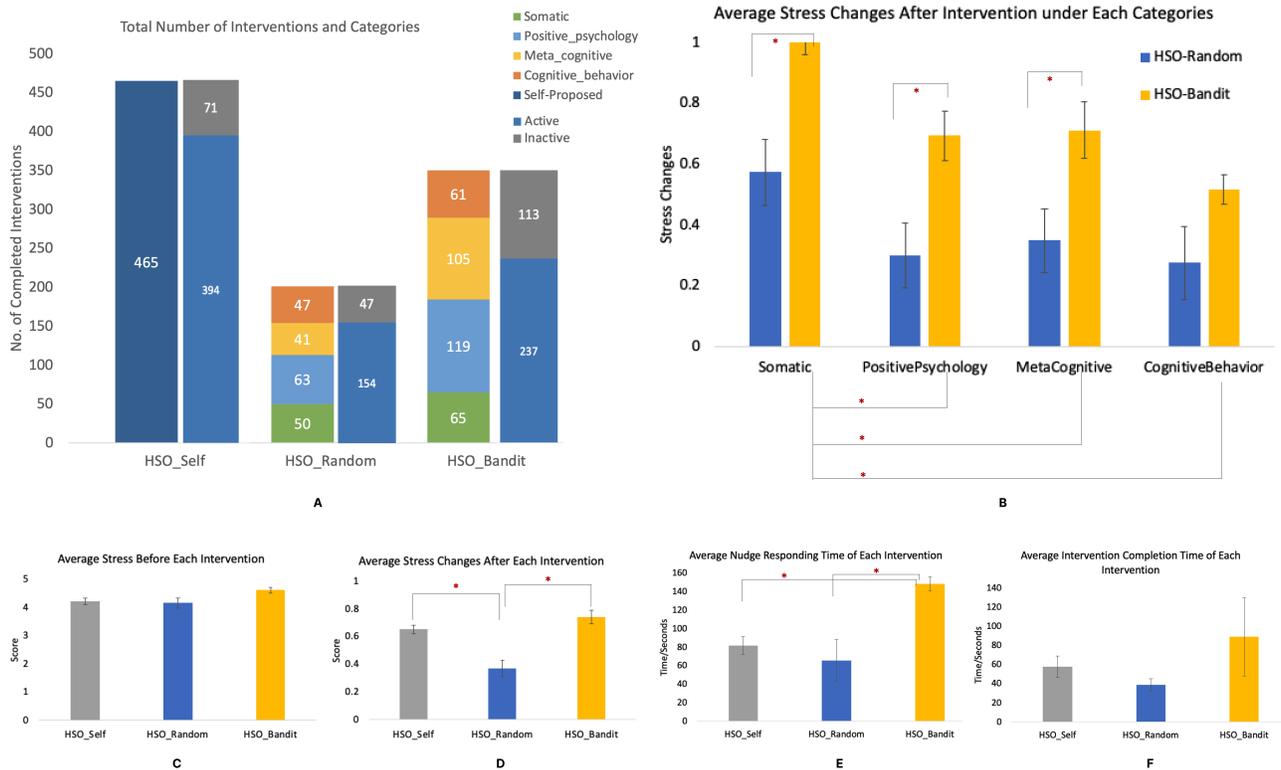


Figure 6: A. Distributions of completed intervention categories in three HSO groups. B. Mean stress changes (and Std. Error) of interventions under the four categories. C. Mean self-reported stress before interventions (SB) of three HSO groups. D. Mean self-reported stress changes after intervention (SC) of three HSO groups. E. Mean nudge responding time before an intervention of the three HSO groups. F. Mean intervention completion time of the three HSO groups. (* indicates significant differences)

walking/exercising, eating/drinking, bathing/showering, breathing, sleeping, meditating, consuming media, and so on. For instance, more than half of P31's self-proposed interventions were about controlled breathing and the rest were usually bathing or consuming media. Although more than half (59.03%) of the self-proposed interventions effectively reduced participants' stress, the rest of the interventions did not elicit any stress changes. Only a few (1.39%) interventions had negative effects and increased participants' stress levels. When participants described their self-proposed interventions, sometimes the interventions were ongoing (32%), sometimes they had completed an action (52%), and sometimes they talked about future short-term goals or plans (16%). We found that even thinking about plans in the near future helped participants reduce their stress levels.

Low Variability in the Self-Proposed Content. From the qualitative coding analysis, we found that most of the self-proposed interventions (85%) could be categorized as somatic interventions. The most adopted somatic interventions were controlled breathing, eating and drinking, and exercise. The remaining few (10%) self-proposed interventions were categorized as positive psychology (such as meditation or prayer), cognitive-behavior (e.g., socializing, distracting activities, etc.), or meta-cognitive (e.g., consuming media) types. A few participants performed combined-type (i.e.,

performing more than one intervention at time from two or more categories) or activities that did not fit with our categorization and were classified as "Other". Three participants even had several (1.39%) self-proposed mal-adaptive interventions for stress relief, such as drinking alcohol, using drugs, and smoking. In conclusion, almost all of the self-proposed interventions demonstrated low variability compared with what HSO's intervention pool offered and most were covered by HSO's intervention pool.

5.3.2 Intervention Categories and Stress Changes. As shown in Figure 6B, we demonstrated the amount of interventions completed under each categories for HSO-random and HSO-Bandit groups. Here, we further report our quantitative findings from independent One-Way ANOVA tests about the effects of intervention categories on stress changes (SC) in each of the three groups. In HSO-Random group, results showed no significant differences of SC in all four intervention categories, $F(2, 183) = 1.527, p = .209$. The mean and SE values of each category: somatic (Mean \pm SE = $.574 \pm .109$), positive psychology (Mean \pm SE = $.30 \pm .107$), meta cognitive (Mean \pm SE = $.35 \pm .105$), cognitive behavior (Mean \pm SE = $.275 \pm .119$). However, in HSO-Bandit group, our findings (Figure 6B) suggest that the four intervention types significantly affected participants stress changes, $F(3,330) = 4.166, p = .006$. More specifically, Tukey's HSD post-hoc

analysis found that somatic interventions (Mean \pm SE = 1.066 \pm .107) resulted in significantly more stress changes than almost all other three types, positive psychology (Mean \pm SE = .693 \pm .080, $p = .041$), meta cognitive (Mean \pm SE = .712 \pm .092, $p = .062$), and cognitive behavior (Mean \pm SE = .517 \pm .049, $p = .004$). This indicated that somatic interventions might have the highest effectiveness of stress reduction out of the four categories. Moreover, independent T-Tests results showed that three intervention categories in HSO-Bandit group elicited significantly more stress changes than the HSO-Random group: somatic ($t(106) = 3.177$, $p = .002$), positive psychology ($t(169) = 2.938$, $p = .004$), and meta cognitive ($t(142) = 2.224$, $p = .028$) categories, except for the cognitive behavior category ($t(96) = 2.115$, $p = .160$).

5.4 Usability, Preferences, & Challenges

Here, we further analyze findings from participants' self-reported interventions in the HSO-Self group's logs, as well as open-ended questions follow-up survey that was completed by 31 participants (response rate 31/58) from three HSO groups. Next, we summarize findings about participants' general feedback on the intervention content from different groups and their overall usability experiences of the HSO system, related design features and technical challenges.

5.4.1 Participants' general feedback on HSO's interventions. Their feedback can be summarized into three main points: (1) the intervention content was helpful because they felt a real impact on stress reduction and mental health, (2) the systems nudges for stress management provided value by reminding them to take time for self-care of their stress levels and increasing their awareness, and (3) the micro-interventions themselves were not fundamentally helpful in resolving the sources of stressors. Most participants (22/31) believed that the interventions supported them effectively in offering intervention content to manage their stress levels, for instance, P-B3 from HSO-Bandit group reported, *"It [stress levels] has changed for the better in that stress interventions can have a real impact on stress reduction and mental health."* Another HSO-Bandit participant said, *"Going into this study I was doubting the process. After doing the interventions it made me a believer. I saw changes within myself and it helped me work through changes that I probably wouldn't do on my own. This extension [HSO] is more like a forcer to think about stress..."* (P-B5). P-R5 from HSO-Random group also reported, *"I liked it progressively more as I got acclimated to it. They [the interventions] were simple, basic, yet interesting."* They also thought of HSO interventions as "guidance" or a "toolbox" that provided knowledge and ways of stress management, e.g., *"...when I interacted with (HSO), I was usually always attentive and interested because I was curious what intervention I would get, it felt like a genuinely nice daily activity to ease my mind a bit which is nice to have."* Others (N=7/31), most from the HSO-Self group, treated the nudges as a reminder, and motivator to self-evaluate their current stress levels and possibly take a break to manage their stress with a self-proposed intervention (if necessary). For example, P-S1 commented, *"[The nudge] made me think about what I was doing and slow down if needed."*

However, a few participants (2/31) from HSO-Random group considered that HSO interventions to be not helpful in supporting them to manage their stress levels because *"(HSO interventions) can*

not change things (stressors) fundamentally" (P-R2) and commented that *"they're [HSO interventions] good for short-term assistance and helping me in the moments right after a stressful event, rather than my long-term stress level since it can be hard to be affected"* (P-R4). From the participants' feedback, their stressors were mainly from hardship and difficulties in their daily life including work challenges, family conflicts, income and finance status, etc., and they thought that the interventions would not help to resolve them fundamentally. Meanwhile, although HSO interventions received positive feedback, a few participants (3/31) thought that sometimes the interventions could cause more stress for them rather than reducing their stress levels. E.g., P-R3 wrote, *"It (interventions from HSO) caused me stress. It may not for others."* and P-R1 also responded, *"Many of them didn't seem to be relevant in getting me to destress about my situation. They seemed like just another distraction."* Due to the nature of the bandit recommendation algorithms, a few (2/31) from the HSO-Bandit group felt they began to receive some similar interventions that led to frustrating or even annoying experiences, e.g., *"I would gave it a 7 out of 10 due to the fact that there were good ones (interventions). But too many repeated themselves, so I got annoyed most of the time."*

5.4.2 Promises and challenges of the HSO extension. In general, participants (28/31) reported that the HSO system and the interventions were easy to use, access, and understand, and also provided a good usability and learning experience for them, e.g., *"Overall, the HSO extension creates peace and [its] attractive to do so [request interventions]"* (P-R4) and *"I enjoyed the study and using the HSO tool. I received tangible stress reduction benefits as a result"* (P-B3). Over a third of participants (12/31) thought that they learned stress management strategies regardless of the effectiveness of the interventions, e.g., *"I loved the HSO experience. I was able learn some techniques to relieve stress, meta-cognition, etc. Everything I participated in was extremely useful and helpful"* (P-B6).

Participants also provided feedback regarding the design of the HSO interfaces, technical difficulties, and other challenges. Several mentioned the flashing icon in particular. While some thought it functioned well as a nudge or reminder, others found it being distracting and frustrating and they had to try to ignore it eventually, e.g., *"I found after 3 or 4 weeks the flashing red alert icon asking me 4 or 5 times a day to assess my mood was super annoying so eventually I just ignored the alerts all together the past 2 weeks or so"* (P-R10). Only a few participants (3/31) reported technical difficulties and challenges, for instance, failures in skipping interventions, long intervention loading times, and so on. As P-B1 responded, *"Some don't load, some take to long, the extension slowed my computer and affected other extension performance"*.

6 DISCUSSION

Here, we further discuss the implications of our study findings from the HSO system and prior literature on users' decision points, systems' decision rules, and stress reduction (proximal and distal) outcomes. We provide insights on what stress micro-intervention recommendation approaches and content work most effectively for participants, and compare the benefits and challenges of HSO micro-interventions with prior JIT intervention systems. Finally, we describe design implications for future work and study limitations.

6.1 The Proximal and Distal Stress Reduction Outcomes: Did Stress Micro-Interventions Reduce Momentary and Multi-Week Stress?

We examined the effectiveness of HSO stress micro-interventions in three ways, self-proposed (HSO-Self), randomized (HSO-Random), and AI-recommended (HSO-Bandit) in a four-week between-group study with intervention data collected from 58 HSO-participants (survey results from 43/58) and 23 in the control group. From the surveys, we did not observe any significant changes in self-reported PROMIS stress reduction data for the three HSO groups compared to our non-intervention control group. We also did not find any significant differences in stress levels before the study and stress changes after the study between any of the three HSO groups and the control group. Similarly, Howe et al.'s work also did not find any significant stress reduction in a four-week study [35]. Our study further demonstrated non-significant changes in multi-weeks by comparing the experimental groups with a non-intervention control group, which most prior studies did not include.

While other studies observed study-long changes in stress, depression [53] and dietary behaviors [60], we hypothesize such inconsistent findings may result from two possible explanations. First, each of the studies adopted a different set of pre-study stress scales and related data, e.g., PHQ9 [42] and CSQ [57] in Paredes's research [53], PANAS [68] in Sano's and Paredes's study [53, 60], DASS-21 [45] in Howe's work [35], and PROMIS [17] in our case; and each found a significant difference in stress levels and intervention efficacy. Second, long-term stress level changes may require more time to observe than allowed in our one-month study protocol. Psychologists addressed the necessity of having organizational changes to deal with the stress sources for long-term benefits rather than simply assisting individuals with *in-situ* coping [22]. On the contrary, momentary self-reported data from HSO's application logs of all three groups showed significant momentary stress reduction, validating a common result from previous research that micro-interventions work effectively for momentary stress reduction, no matter the systems that they were implemented and mediums they were recommended through (e.g., computer browsers, mobile devices) [35, 53, 60, 61]. Thus, our study provides further evidence of the proximal outcomes of micro-interventions for stress, but also points to divergent effectiveness of its distal outcomes. Future studies should consider deploying various self-reported stress scales to evaluate their trade-offs, comprehensively measure stress perceptions, and to validate micro-interventions' distal outcomes.

6.2 The System's Decision Rules: How to Recommend Micro-Interventions and What to Recommend?

Having described the outcomes of HSO's micro-interventions on stress measures, we now discuss our findings about the decision rules that a recommendation system could adopt to enhance user experience and intervention effectiveness.

6.2.1 Self-Proposed, Randomly-Assigned, or AI-Recommended? Our results suggest that micro-interventions in the HSO-Bandit and HSO-Self groups both resulted in significant momentary stress

reduction compared to HSO-Random, but we did not identify differences between HSO-Bandit and HSO-Self. Since people know themselves best, we could assume that the self-proposed intervention content would be most effective for them and HSO only functions as a "nudge" or "reminder" to improve their awareness of feeling stressed and reminds them to take action to deal with these feelings, as reported in the surveys. Meanwhile, no differences in stress changes were exhibited between the self-proposed and the AI-recommended interventions, indicating that our multi-armed bandit algorithm reduced stress at a similar level as the self-proposed interventions. Such findings align with prior studies, which identified AI-recommended interventions performed better than randomly selected ones. Building on the work of PopTherapy [53], our study further highlights that self-proposed stress interventions could be as effective as those created by experts and implemented via AI-powered recommendation systems. Another difference in the PopTherapy implementation and ours was that their interventions were served via an app on a mobile phone that users carried with them at all times. According to the authors, this required users' continuous attention and effort, leading them to be frequently reminded about their stress and resulting in higher stress and abandonment of the application. To our knowledge, HSO is the first browser-based stress management tool that addresses this concern by providing nudges and interventions in a subtle way (i.e., through an ambient alert). Finally, this work was also conducted with a larger population within a work context, demonstrating a novel application of a stress management tool for workers.

6.2.2 Self-Proposed, Somatic, Positive Psychology, Meta Cognitive, or Cognitive Behavior? In HSO, we further adapted the intervention content and updated most micro-interventions based on Paredes et al.'s original intervention design [53] and two pilot-iterations of the system where we received participant feedback. Although HSO-Bandit and HSO-Self share similar effectiveness, further analysis of the interventions' categories revealed interesting findings about how each category affected the stress reduction outcomes. Qualitative analysis of the self-reported intervention content in the HSO-Self group showed that each person has their own way of managing their stress (e.g., breathing, eating and drinking, exercising, etc.), and most people proposed a limited number of interventions (5 or less) from a subset of the content types with most falling under the somatic category, compared with HSO-Random and HSO-Bandit interventions (which selected from a diverse pool of 161 interventions across all content types). This suggests that the HSO-Self group is a reasonably fair comparison for other HSO groups as their intervention content was, while limited in diversity, similar in nature. Further, this also suggests that HSO-Bandit participants explored a more diverse set of interventions than those in HSO-Self. Further, a closer look at our back-end data showed that most of the intervention categories (except those classified as cognitive behavior) resulted in significantly more stress reduction when personalized by the AI algorithm in HSO-Bandit compared to HSO-Random. These findings indicate that AI-personalization is more optimal for a system recommending micro-interventions for stress. Within the HSO-Random and HSO-Bandit groups, somatic interventions performed significantly better than the other categories, which further validates Sano et al.'s findings –somatic

activities were preferred and reduced stress levels most [60]. Since participants' self-proposed interventions in the HSO-Self group were mostly describing somatic interventions, somatic intervention could potentially be the most effective types of interventions for momentary stress reduction.

However, our qualitative findings also indicate that participants must meet constraints when performing somatic interventions, such as needing space and/or time. From the qualitative results, we learned that participants from the HSO-Self group desired more intervention options, whereas participants in the HSO-Bandit group noted receiving certain initial interventions they did not like. Therefore, future micro-interventions could be served differently: using an AI algorithm to recommend personalized interventions, but involving human users in the recommendation loop and allowing them to integrate self-proposed interventions into the pool. That way, participants could combine their self-proposed interventions with the machine-recommended expert-authored interventions, administrated by the systems.

6.3 How Did Users' Respond to the Nudges?

In general, we found that participants spent a significantly longer time responding to an AI-recommended interventions than the self-proposed or randomized ones and once they attended to an AI-recommended intervention they spent more time completing it. Based on our qualitative findings, we theorize that participants' self-proposed interventions were easier and faster to complete compared to the system suggested intervention which would have higher novelty. Therefore, participants were possibly a bit more delayed in responding to the AI-recommended interventions, but they still received the most effective stress reductions. Howe et al. reported a similar finding that people prefer easy and quick interventions, but complicated ones work more effectively [35].

Although participants were nudged at a specific time frame, they were also allowed to request interventions when the nudge was not active. Our results also showed that interventions worked better in all three groups when participants actively requested one rather than passively receiving one when the nudge was already active. This showed that participants' decision points might affect the proximal outcome. We assumed that when participants actively requested interventions, they were consciously more aware of their stress status and more willing to manage their stress levels. Thus, such decision points might result in the best stress reduction outcome. From qualitative findings, we also hypothesized that once the nudge was triggered, for some participants, the flashing icon could visually distract participants and add another layer of stress (i.e., breaking the intention of having the nudge be a subtle signal to the user to attend to their stress). Therefore, participants experienced significantly more stress reduction when they self-requested an intervention before being nudged by HSO. We suggest that future work should iterate further on the nudging mechanism and (e.g., potentially exploring different visual designs for when an intervention is overdue). As indicated by the participants, audio effects (e.g., like the sound people hear when they receive an email) could also be tested and evaluated as a substitute alert option to avoid distraction and stress in future systems.

6.4 Users' Experiences Using HSO in-the-Wild

From the qualitative findings, most participants from all HSO groups reported that HSO was designed well and was highly useable. Generally, HSO was viewed as filling three major roles including a toolbox, a reminder, and an expert guided content provider (with some overlap between roles). For participants from the HSO-Self group, reported appreciating that HSO provided a nudge for taking breaks and dealing with their stress, which increased their awareness of their momentary stress levels. Further, HSO-Self participants' self-proposed stress interventions worked best and fell under similar categories as the interventions from the HSO's intervention pool. Surprisingly, the self-proposed interventions resulted in an equivalent level of stress reduction as the interventions authored by expert designers and recommended by AI algorithms (HSO-Bandit group). Such findings indicate the importance and potential of involving human users in designing interventions for future research. However, when user's did self-propose intervention content they occasionally (1.39%) proposed mal-adaptive and potentially harmful interventions (e.g., smoking, drinking alcohol). Therefore, if involving human (non-expert) users in content authoring, future stress management systems should also recognize and provide immediate feedback and suggestions to prohibit potentially harmful interventions (e.g., allowing for feedback while collecting users' post-intervention information). Some participants from the HSO-Random group considered HSO as a toolbox, expert guidance, or as a educational platform. For instance, from the qualitative findings, participants felt curious about what would happen and perceived receiving randomized HSO interventions to be a fun activity. Similarly, participants from the HSO-Bandit group also treated the interventions as as expert content provider guiding them to do things that, otherwise, they would not do.

As suggested in prior pilot studies and the main study, participants liked simple, fast, easy-to-execute interventions; and they disliked the ones they could not perform due to space, time, equipment limitations, and/or the inaccessibility of others to interact with. Different participants also reported varying most-liked and disliked intervention content or categories. In the HSO-Random or HSO-Bandit groups, participants might complain about being assigned to interventions they did not like when participants not given the agency to propose their own intervention content. While in the HSO-Self group, the majority of participants repeated a limited set of intervention content and categories. Similar to prior work [35, 53, 60, 61], current HSO interventions were not designed to resolve specific stressors directly but focused on understanding users' behaviors and stress changes when receiving micro-interventions. Our study examined people's stress changes and behaviors of using HSO in the wild. We did not focus on the sources of stress, i.e., the stressors, nor the possible impacts of intervention strategies and content on stress reduction caused by different stressors though Mauriello et al. [47] provides an indication that our participants were likely dealing with common everyday stressors related to their work, relationships, and health. However, we acknowledge that knowing the sources of stress may allow JIT systems like HSO to better support users with stress management by recommending more personalized and effective interventions.

Future work could further investigate this issue to help users set up a strategy to deal with specific stressors.

6.5 The Challenges in Adherence and Retention of the AMT Participants

Adherence and retention were challenges in this work, similar to prior research. In general, around 15% of the participants completed the study out of the initial population who enrolled in the three HSO groups, and around 40% completed the control group surveys. We also observe a large drop in HSO usage (*i.e.*, total number of completed interventions) across the four weeks, especially when comparing week 1 and week 4. The follow-up survey revealed that more than half of the participants either stopped using HSO interventions intentionally or simply forgot to use it regardless of their study group. We hypothesize our compensation structure and recruitment method could be the potential cause. In our work, we recruited people primarily from AMT, Facebook, and our university's mailing lists and we paid participants a fixed \$10 USD and included them in a participation-based raffle. With regard to participants' motivations to participate, studies of AMT workers continually note the importance of payment to participate in tasks [9, 39, 46]. In similar work, participants were recruited from large technology companies and paid at a fixed rate of \$200 to \$300 USD. Thus, dropout during and, particularly, after the study is expected. We received a similar number of total valid interventions from fewer participants (1061 interventions completed by 58 participants) during the same period as in Howe et al.'s study (with 1155 interventions completed by 86 participants) [35]. Future work could further investigate this issue and compare population differences in performance, stress reduction, and incentive structures.

Although our participants were mainly from AMT, they met our selection criteria of being online workers that used their browsers daily for their work or study. Our participants also covered a wide range of demographics, *e.g.*, gender, age, living areas, and educational background and these workers are frequently dealing with common everyday stressors [47]. Therefore, data about these stressors has been helpful in other contexts [48]. Prior work also indicated that AMT workers could be qualified as normal research participants [11, 16]. This further supports that our work can provide insights into how other online workers may use such tools and applications. Another concern is that AMT is, in fact, a marketplace subject to human factors, *e.g.*, monetary rewards, which affect worker performance [23]. For example, increasing the reward of a set of tasks would lead to faster results. Therefore, results from our multi-week study with the raffle compensation payment schedule can provide valuable insights, but future work could further investigate how users' motivations impact their stress outcomes by recruiting multiple types of online workers.

6.6 Limitations

Regarding study design and procedures limitations, all participants were given HSO-Self interventions due to a technical glitch in our back-end server during the first few (2-3) days of the study, rather than the specific interventions they should have received. Therefore, we re-grouped participants after the first few days and randomly assigned them to one of the three HSO experimental groups. Although

we took out the intervention data from our dataset, this change (the interventions participants received) may have impacted their interests in HSO and motivation to participate. We also included limited kinds of self-reported stress scales and related surveys in pre- and post-surveys for evaluating multi-week stress changes; future research could consider using different or comparing multiple stress scales. Moreover, we allowed HSO-Self participants to propose and execute their own preferred stress interventions (*i.e.*, whatever came to their minds at that moment). We acknowledge this study design could be a constraint and future work can further ask participants to author their own interventions with specific and comparable requirements beforehand and then adopt those interventions in the study or compare participants-authored interventions with HSO expert-authored interventions.

7 CONCLUSION

Work stress affects workers' productivity levels, job satisfaction, personal life, mental wellbeing, and physical disorders. To support online workers with access to stress interventions with empirically supported psychotherapy techniques, we developed a browser-based plugin, Home Sweet Office (HSO), which recommends interventions (expert-authored) based on individual traits, personal preferences, past efficacy, and contextual information. We further conducted a four-week field study to compare the efficacy of self-proposed and machine-recommended interventions and to understand the best intervention categories. Findings from our study indicated that although there were no significant multi-week stress improvements in any of the HSO groups compared with the control group, HSO-Self and HSO-Bandit groups both had significantly better momentary stress reduction than the HSO-random arm. While HSO-Bandit and HSO-Self were similar in effectiveness, HSO-Bandit interventions offered richer and more diverse expert-authored intervention content compared to the HSO-Self group, resulting in significantly longer intervention completion time, which positively correlates with stress reduction. Finally, we propose (1) design recommendations for future researchers and designers working to design content and delivery methods for improving efficacy and engagement with stress micro-interventions and (2) further exploring near-term/long-term effectiveness of stress micro-intervention systems.

ACKNOWLEDGMENTS

We thank the Stanford School of Medicine's Psychiatry and Behavioral Sciences' Innovation Grant for funding this research as well as the Natural Sciences and Engineering Research Council of Canada Grant (NSERC PDF-558147-2021) for supporting Xin Tong's post-doctoral research. We would like to acknowledge the contributions of Gi-Soo Kim of Ulsan National Institute of Science and Technology who provided feedback on designing and training the bandit algorithms used for this study. Moreover, we would like to thank Geza Kovacs and Michael Bernstein, who originally authored Habitlab on which HSO was built, for their valuable feedback. The contributions of Xin Tong, Matthew Louis Mauriello, and Pablo E. Paredes were made while in transition from Stanford School of Medicine to Duke Kunshan University, the University of Delaware, and the University of Maryland, respectively.

REFERENCES

- [1] 2015. *Special Report: the costs of nonpayment*. <https://blog.freelancersunion.org/2015/12/10/costs-nonpayment/>
- [2] 2020. Stress at work. (2020). <https://doi.org/10.26616/NIOSHUB99101>
- [3] Aino Ahtinen, Elina Mattila, Pasi Välikynen, Kirsikka Kaipainen, Toni Vanhala, Miikka Ermes, Essi Sairanen, Tero Myllymäki, Raimo Lappalainen, et al. 2013. Mobile mental wellness training for stress management: feasibility and design implications based on a one-month field study. *JMIR mHealth and uHealth* 1, 2 (2013), e2596.
- [4] John A Astin. 1997. Stress reduction through mindfulness meditation. *Psychotherapy and psychosomatics* 66, 2 (1997), 97–106.
- [5] Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. 2002. Finite-time analysis of the multiarmed bandit problem. *Machine learning* 47, 2 (2002), 235–256.
- [6] Douglas W Billings, Royer F Cook, April Hendrickson, and David C Dove. 2008. A web-based approach to managing stress and mood disorders in the workforce. *Journal of occupational and environmental medicine/American College of Occupational and Environmental Medicine* 50, 8 (2008), 960.
- [7] Edwin Boudreaux, Cris Mandry, and Phillip J Brantley. 1997. Stress, job satisfaction, coping, and psychological distress among emergency medical technicians. *Prehospital and Disaster Medicine* 12, 4 (1997), 9–16.
- [8] Danielle Boyd. [n. d.]. *Workplace Stress*. <https://www.stress.org/workplace-stress>
- [9] Robin Brewer and Anne Marie Piper. 2016. "Tell It Like It Really Is" A Case of Online Content Creation and Sharing Among Older Adult Bloggers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 5529–5542.
- [10] Richard P Brown, Patricia L Gerbarg, and Fred Muench. 2013. Breathing practices for treatment of psychiatric and stress-related medical conditions. *Psychiatric Clinics* 36, 1 (2013), 121–140.
- [11] Michael Buhrmester, Tracy Kwang, and Samuel D Gosling. 2016. Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data? (2016).
- [12] Andrew C Butler, Jason E Chapman, Evan M Forman, and Aaron T Beck. 2006. The empirical status of cognitive-behavioral therapy: a review of meta-analyses. *Clinical psychology review* 26, 1 (2006), 17–31.
- [13] James Carmody and Ruth A Baer. 2008. Relationships between mindfulness practice and levels of mindfulness, medical and psychological symptoms and well-being in a mindfulness-based stress reduction program. *Journal of behavioral medicine* 31, 1 (2008), 23–33.
- [14] James Carmody and Ruth A Baer. 2008. Relationships between mindfulness practice and levels of mindfulness, medical and psychological symptoms and well-being in a mindfulness-based stress reduction program. *Journal of behavioral medicine* 31, 1 (2008), 23–33.
- [15] Laura Carter, Deevakar Rogith, Amy Franklin, and Sahiti Myneni. 2019. NewCope: a theory-linked mobile application for stress education and management. *Studies in health technology and informatics* 264 (2019), 1150.
- [16] Krista Casler, Lydia Bickel, and Elizabeth Hackett. 2013. Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in human behavior* 29, 6 (2013), 2156–2160.
- [17] David Cella, William Riley, Arthur Stone, Nan Rothrock, Bryce Reeve, Susan Yount, Dagmar Amtmann, Rita Bode, Daniel Buysse, Seung Choi, et al. 2010. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *Journal of clinical epidemiology* 63, 11 (2010), 1179–1194.
- [18] Janine Clarke, Judith Proudfoot, Mary-Rose Birch, Alexis E Whitton, Gordon Parker, Vijaya Manicavasagar, Virginia Harrison, Helen Christensen, and Dusan Hadzi-Pavlovic. 2014. Effects of mental health self-efficacy on outcomes of a mobile phone and web intervention for mild-to-moderate depression, anxiety and stress: secondary analysis of a randomised controlled trial. *BMC psychiatry* 14, 1 (2014), 1–10.
- [19] Shanice Clarke, Luis G Jaimes, and Miguel A Labrador. 2017. mstress: A mobile recommender system for just-in-time interventions for stress. In *2017 14th IEEE Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 1–5.
- [20] Thomas W Colligan and Eileen M Higgins. 2006. Workplace stress: Etiology and consequences. *Journal of workplace behavioral health* 21, 2 (2006), 89–97.
- [21] Royer Cook, Douglas Billings, Rebekah Hersch, Anita Back, April Hendrickson, et al. 2007. A field test of a web-based workplace health promotion program to improve dietary practices, reduce stress, and increase physical activity: randomized controlled trial. *Journal of medical Internet research* 9, 2 (2007), e627.
- [22] Cary L Cooper and Sue Cartwright. 1997. An intervention strategy for workplace stress. *Journal of psychosomatic research* 43, 1 (1997), 7–16.
- [23] Djellel Eddine Difallah, Michele Catasta, Gianluca Demartini, Panagiotis G Ipeirotis, and Philippe Cudré-Mauroux. 2015. The dynamics of micro-task crowdsourcing: The case of amazon mturk. In *Proceedings of the 24th international conference on world wide web*. 238–247.
- [24] Katherine Pollak Eisen, George J Allen, Mary Bollash, and Linda S Pescatello. 2008. Stress management in the workplace: A comparison of a computer-based and an in-person stress-management intervention. *Computers in Human Behavior* 24, 2 (2008), 486–496.
- [25] Elissa S Epel, Alexandra D Crosswell, Stefanie E Mayer, Aric A Prather, George M Slavich, Eli Puterman, and Wendy Berry Mendes. 2018. More than a feeling: A unified view of stress measurement for population science. *Frontiers in neuroendocrinology* 49 (2018), 146–169.
- [26] Susan Folkman, Judith Tedlie Moskowitz, Elizabeth M Ozer, and Crystal L Park. 1997. Positive meaningful events and coping in the context of HIV/AIDS. In *Coping with chronic stress*. Springer, 293–314.
- [27] Barbara L Fredrickson. 2001. The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American psychologist* 56, 3 (2001), 218.
- [28] Kelsie M Full, Atul Malhotra, Katie Crist, Kevin Moran, and Jacqueline Kerr. 2019. Assessing psychometric properties of the PROMIS Sleep Disturbance Scale in older adults in independent-living and continuing care retirement communities. *Sleep Health* 5, 1 (2019), 18–22.
- [29] Paul R Grime. 2004. Computerized cognitive behavioural therapy at work: a randomized controlled trial in employees with recent stress-related absenteeism. *Occupational Medicine* 54, 5 (2004), 353–359.
- [30] Dan Hasson, Ulla Maria Anderberg, Töres Theorell, and Bengt B Arnetz. 2005. Psychophysiological effects of a web-based stress management system: a prospective, randomized controlled intervention study of IT and media workers [IS-RCTN54254861]. *BMC public health* 5, 1 (2005), 1–14.
- [31] Elena Heber, David Daniel Ebert, Dirk Lehr, Stephanie Nobis, Matthias Berking, and Heleen Riper. 2013. Efficacy and cost-effectiveness of a web-based and mobile stress-management intervention for employees: design of a randomized controlled trial. *BMC public health* 13, 1 (2013), 1–12.
- [32] Elena Heber, Dirk Lehr, David Daniel Ebert, Matthias Berking, Heleen Riper, et al. 2016. Web-based and mobile stress management intervention for employees: a randomized controlled trial. *Journal of medical Internet research* 18, 1 (2016), e5112.
- [33] Blake Anthony Hickey, Taryn Chalmers, Phillip Newton, Chin-Teng Lin, David Sibbritt, Craig S McLachlan, Roderick Clifton-Bligh, John Morley, and Sara Lal. 2021. Smart devices and wearable technologies to detect and monitor mental health conditions and stress: A systematic review. *Sensors* 21, 10 (2021), 3461.
- [34] Irit Hochberg, Guy Feraru, Mark Kozdoba, Shie Mannor, Moshe Tennenholtz, and Elad Yom-Tov. 2016. A reinforcement learning system to encourage physical activity in diabetes patients. *arXiv preprint arXiv:1605.04070* (2016).
- [35] Esther Howe, Jina Suh, Mehrab Bin Morshed, Daniel McDuff, Kael Rowan, Javier Hernandez, Marah Ihab Abdin, Gonzalo Ramos, Tracy Tran, and Mary P Czerwinski. 2022. Design of Digital Workplace Stress-Reduction Intervention Systems: Effects of Intervention Type and Timing. In *CHI Conference on Human Factors in Computing Systems*. 1–16.
- [36] Julie Anne Irving, Patricia L Dobkin, and Jeesoon Park. 2009. Cultivating mindfulness in health care professionals: A review of empirical studies of mindfulness-based stress reduction (MBSR). *Complementary therapies in clinical practice* 15, 2 (2009), 61–66.
- [37] Luis G Jaimes, Martin Llofriu, and Andrew Raji. 2015. Preventer, a selection mechanism for just-in-time preventive interventions. *IEEE Transactions on Affective Computing* 7, 3 (2015), 243–257.
- [38] Ravinder Jerath, Molly W Crawford, Vernon A Barnes, and Kyler Harden. 2015. Self-regulation of breathing as a primary treatment for anxiety. *Applied psychophysiology and biofeedback* 40, 2 (2015), 107–115.
- [39] Toni Kaplan, Susumu Saito, Kotaro Hara, and Jeffrey P Bigham. 2018. Striving to earn more: a survey of work strategies and tool use among crowd workers. In *Sixth AAAI Conference on Human Computation and Crowdsourcing*.
- [40] Predrag Klasnja, Eric B Hekler, Saul Shiffman, Audrey Boruvka, Daniel Almirall, Ambuj Tewari, and Susan A Murphy. 2015. Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. *Health Psychology* 34, S (2015), 1220.
- [41] Geza Kovacs, Zhengxuan Wu, and Michael S Bernstein. 2018. Rotating online behavior change interventions increases effectiveness but also increases attrition. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–25.
- [42] Kurt Kroenke and Robert L Spitzer. 2002. The PHQ-9: a new depression diagnostic and severity measure. , 509–515 pages.
- [43] Lihong Li, Wei Chu, John Langford, and Robert E Schapire. 2010. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*. 661–670.
- [44] Tania M Lincoln, Maike Hartmann, Ulf Köther, and Steffen Moritz. 2015. Dealing with feeling: Specific emotion regulation skills predict responses to stress in psychosis. *Psychiatry Research* 228, 2 (2015), 216–222.
- [45] Peter F Lovibond and Sydney H Lovibond. 1995. The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour research and therapy* 33, 3 (1995), 335–343.
- [46] David Martin, Benjamin V Hanrahan, Jacki O'neill, and Neha Gupta. 2014. Being a turker. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. 224–235.
- [47] Matthew Louis Mauriello, Thierry Lincoln, Grace Hon, Dorien Simon, Dan Jurafsky, and Pablo Paredes. 2021. SAD: A Stress Annotated Dataset for Recognizing

- Everyday Stressors in SMS-like Conversational Systems. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI EA '21). Association for Computing Machinery, New York, NY, USA, Article 399, 7 pages. <https://doi.org/10.1145/3411763.3451799>
- [48] Matthew Louis Mauriello, Nantanick Tantivasadakarn, Marco Antonio Mora-Mendoza, Emmanuel Thierry Lincoln, Grace Hon, Parsa Nowruzi, Dorien Simon, Luke Hansen, Nathaniel H Goenawan, Joshua Kim, Nikhil Gowda, Dan Jurafsky, and Pablo Enrique Paredes. 2021. A Suite of Mobile Conversational Agents for Daily Stress Management (Popbots): Mixed Methods Exploratory Study. *JMIR Form Res* 5, 9 (14 Sep 2021), e25294. <https://doi.org/10.2196/25294>
- [49] Bruce S McEwen and Eliot Stellar. 1993. Stress and the individual: Mechanisms leading to disease. *Archives of internal medicine* 153, 18 (1993), 2093–2101.
- [50] Kathleen Mikkelsen, Lily Stojanovska, Momir Polenakovic, Marijan Bosevski, and Vasso Apostolopoulos. 2017. Exercise and mental health. *Maturitas* 106 (2017), 48–56.
- [51] Margaret E Morris, Qusai Kathawala, Todd K Leen, Ethan E Gorenstein, Farzin Guilak, William DeLeeuw, and Michael Labhard. 2010. Mobile therapy: case study evaluations of a cell phone application for emotional self-awareness. *Journal of medical Internet research* 12, 2 (2010), e1371.
- [52] Leanne G Morrison, Adam WA Geraghty, Scott Lloyd, Natalie Goodman, Darius T Michaelides, Charlie Hargood, Mark Weal, and Lucy Yardley. 2018. Comparing usage of a web and app stress management intervention: An observational study. *Internet interventions* 12 (2018), 74–82.
- [53] Pablo Paredes, Ran Gilad-Bachrach, Mary Czerwinski, Asta Roseway, Kael Rowan, and Javier Hernandez. 2014. PopTherapy: Coping with stress through popculture. In *Proceedings of the 8th international conference on pervasive computing technologies for healthcare*. 109–117.
- [54] Pablo Enrique Paredes, Nur Al-Huda Hamdan, Dav Clark, Carrie Cai, Wendy Ju, and James A Landay. 2017. Evaluating in-car movements in the design of mindful commute interventions: exploratory study. *Journal of medical Internet research* 19, 12 (2017), e6983.
- [55] Valentina Perciavalle, Marta Blandini, Paola Fecarotta, Andrea Buscemi, Donatella Di Corrado, Luana Bertolo, Fulvia Fichera, and Marinella Coco. 2017. The role of deep breathing on stress. *Neurological Sciences* 38, 3 (2017), 451–458.
- [56] Taylor E Purvis, Brian J Neuman, Lee H Riley, and Richard L Skolasky. 2019. Comparison of PROMIS Anxiety and Depression, PHQ-8, and GAD-7 to screen for anxiety and depression among patients presenting for spine surgery. *Journal of Neurosurgery: Spine* 30, 4 (2019), 524–531.
- [57] Michael E Robinson, Joseph L Riley III, Cynthia D Myers, Ian J Sadler, Steven A Kvaal, Michael E Geisser, and Francis J Keefe. 1997. The Coping Strategies Questionnaire: a large sample, item level factor analysis. *The Clinical journal of pain* 13, 1 (1997), 43–49.
- [58] Darya Saedpanah, Shiva Salehi, and Ladan Fattah Moghaddam. 2016. The effect of emotion regulation training on occupational stress of critical care nurses. *Journal of clinical and diagnostic research: JCDR* 10, 12 (2016), VC01.
- [59] Pedro Sanches, Kristina Höök, Elsa Vaara, Claus Weymann, Markus Bylund, Pedro Ferreira, Nathalie Peira, and Marie Sjölander. 2010. Mind the body! Designing a mobile stress management application encouraging personal reflection. In *Proceedings of the 8th ACM conference on designing interactive systems*. 47–56.
- [60] Akane Sano, Paul Johns, and Mary Czerwinski. 2015. HealthAware: An advice system for stress, sleep, diet and exercise. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 546–552.
- [61] Akane Sano, Paul Johns, and Mary Czerwinski. 2017. Designing opportune stress intervention delivery timing using multi-modal data. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 346–353.
- [62] Hillol Sarker, Moushumi Sharmin, Amin Ahsan Ali, Md Mahbubur Rahman, Rummana Bari, Syed Monowar Hossain, and Santosh Kumar. 2014. Assessing the availability of users to engage in just-in-time intervention in the natural environment. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*. 909–920.
- [63] Jessica Schroeder, Chelsey Wilkes, Kael Rowan, Arturo Toledo, Ann Paradiso, Mary Czerwinski, Gloria Mark, and Marsha M Linehan. 2018. Pocket skills: A conversational mobile web app to support dialectical behavioral therapy. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–15.
- [64] Stephen M Schueller. 2010. Preferences for positive psychology exercises. *The Journal of Positive Psychology* 5, 3 (2010), 192–203.
- [65] Martin EP Seligman and Mihaly Csikszentmihalyi. 2014. Positive psychology: An introduction. In *Flow and the foundations of positive psychology*. Springer, 279–298.
- [66] Annie Shearer, Melissa Hunt, Mifta Chowdhury, and Lorena Nicol. 2016. Effects of a brief mindfulness meditation intervention on student stress and heart rate variability. *International Journal of Stress Management* 23, 2 (2016), 232.
- [67] Jeanne A Teresi, Katja Ocepek-Welikson, Marjorie Kleinman, Mildred Ramirez, and Giyeon Kim. 2016. Psychometric properties and performance of the Patient Reported Outcomes Measurement Information System®(PROMIS®) depression short forms in ethnically diverse groups. *Psychological test and assessment modeling* (2016).
- [68] David Watson, Lee Anna Clark, and Auke Tellegen. 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology* 54, 6 (1988), 1063.
- [69] Elad Yom-Tov, Guy Feraru, Mark Kozdoba, Shie Mannor, Moshe Tennenholtz, Irit Hochberg, et al. 2017. Encouraging physical activity in patients with diabetes: intervention using a reinforcement learning system. *Journal of medical Internet research* 19, 10 (2017), e7994.
- [70] Bin Yu, Mathias Funk, Jun Hu, Qi Wang, and Loe Feijs. 2018. Biofeedback for everyday stress management: A systematic review. *Frontiers in ICT* 5 (2018), 23.
- [71] Yufan Zhao, Michael R Kosorok, and Donglin Zeng. 2009. Reinforcement learning design for cancer clinical trials. *Statistics in medicine* 28, 26 (2009), 3294–3315.
- [72] Xingchen Zhou, Pei-Luen Patrick Rau, and Xueqian Liu. 2021. "Time to Take a Break" How Heavy Adult Gamers React to a Built-In Gaming Gradual Intervention System. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–30.

A HSO THERAPY GROUPS AND TECHNIQUES

Therapy Group (No. of Interventions)	Therapy Explanations	Techniques (No. of Interventions)	Sample Interventions
Meta-Cognitive (39)	Respond to ongoing experience episodes with emotions that are socially tolerable and flexible to permit spontaneous reactions or delay them as needed.	Acceptance and Commitment Therapy (9) Self-compassion (4) Self-soothing (4) Dialectic Behavioral Therapy (6) Emotion Regulation (6) Mindfulness (10)	Mindfulness: Observe how you are feeling physically, mentally, and emotionally for two minutes, without any opinions or judgement. Self-soothing: Write down a list of fun things you want to do with friends or family and schedule one of them for the upcoming weekend!
Cognitive-Behavioral (39)	Observe thoughts, their triggers and their consequences, entertain alternatives, dispute them, etc.	Cognitive Behavioral Therapy (4) Cognitive Re-framing (6) Interpersonal Skills (8) Problem Solving Therapy (9) Visualization (12)	Visualization: Visualize someone you care about, and try to feel their presence around you. Write down a note about your experience. Problem Solving: Go on Reddit and look up how others solved a problem that you are experiencing right now.
Somatic (38)	Exercises to shift physiological signs of arousal.	Relaxation (6) Laughter (10) Breathing (5) Exercise (11) Sleep (6)	Exercise: Walking outdoors can improve your mood, health, creativity, and more! Get up from your desk, put on some shoes, and take a short walk around the block. Relaxation: Call a friend and have a quick yoga session together to relax your body. Have one person screenshare the video over Zoom.
Positive-Psychology (45)	Focus on wellness and well-being, and making the positive aspects of life more salient.	Act of Kindness (12) Affirm Values (5) Best Future Self (8) Strengths (5) Thank You Letter (4) Three Good Things (10)	Affirm Values: You are an incredible person and the world is lucky to have you in it. Write down a few personal qualities you love about yourself! Best Future Self: Think about one thing you want to change about your life and write down 2 ways you can work towards it today.

Table 1: Intervention therapy groups, therapy explanations, techniques, amount of interventions, and sample interventions in each category.

B HSO SYSTEM FRAMEWORK

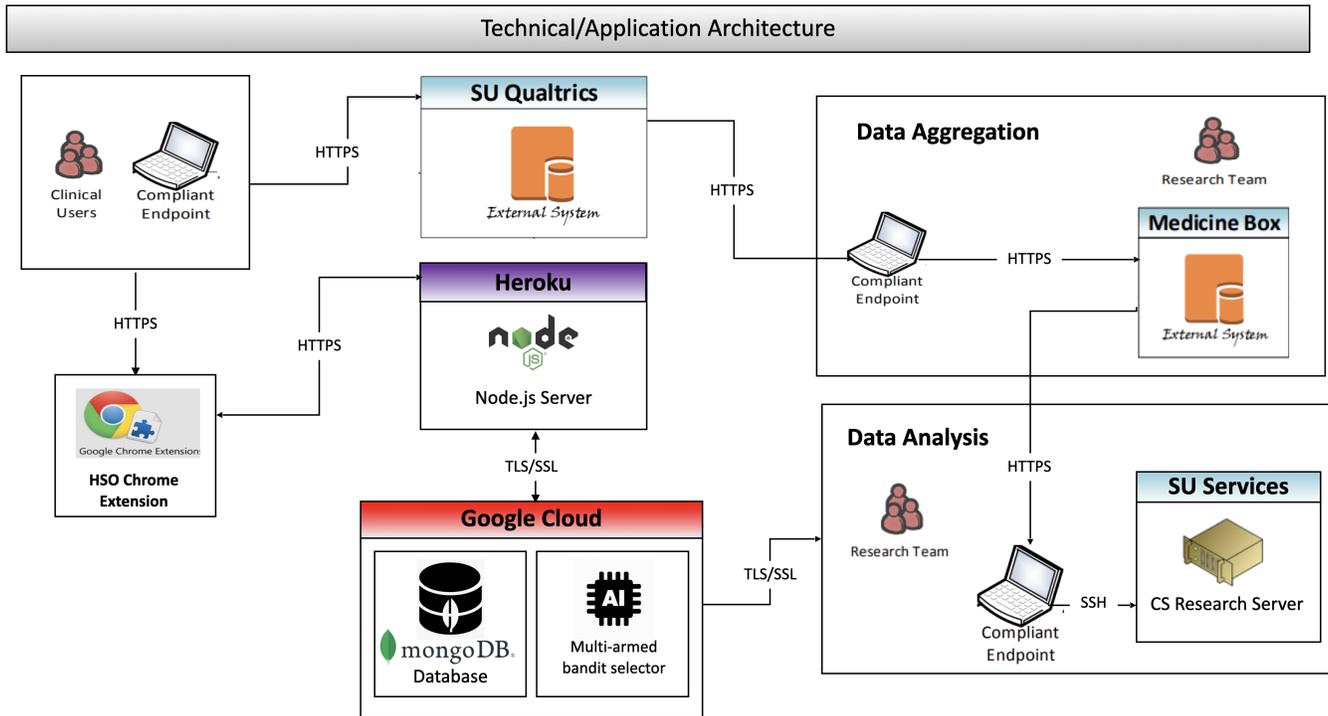


Figure 1: HSO Server-Client System Implementation.