Design of Digital Workplace Stress-Reduction Intervention Systems: Effects of Intervention Type and Timing

Esther Howe*
Microsoft Research
Redmond, WA, USA
University of California, Berkeley
Berkeley, CA, USA
howe@berkeley.edu

Jina Suh*
Microsoft Research
Redmond, WA, USA
jinsuh@microsoft.com

Mehrab Bin Morshed
Microsoft Research
Redmond, WA, USA
Georgia Institute of Technology
Atlanta, GA, USA
mehrab.morshed@gmail.com

Daniel McDuff
Microsoft Research
Redmond, WA, USA
damcduff@microsoft.com

Kael Rowan
Microsoft Research
Redmond, WA, USA
kael.rowan@microsoft.com

Javier Hernandez
Microsoft Research
Redmond, WA, USA
javierh@microsoft.com

Marah Ihab Abdin
Microsoft Research
Redmond, WA, USA
marah.abdin@microsoft.com

Gonzalo Ramos
Microsoft Research
Redmond, WA, USA
goramos@microsoft.com

Tracy Tran
Microsoft Research
Redmond, WA, USA
tracy.tran@microsoft.com

Mary Czerwinski
Microsoft Research
Redmond, WA, USA
marycz@microsoft.com

ABSTRACT

Workplace stress-reduction interventions have produced mixed results due to engagement and adherence barriers. Leveraging technology to integrate such interventions into the workday may address these barriers and help mitigate the mental, physical, and monetary effects of workplace stress. To inform the design of a workplace stress-reduction intervention system, we conducted a four-week longitudinal study with 86 participants, examining the effects of intervention type and timing on usage, stress reduction impact, and user preferences. We compared three intervention types and two delivery timing conditions: *Pre-scheduled* (PS) by users and *Just-in-time* (JIT) prompted by the system-identified user stress-levels. We found JIT participants completed significantly more interventions than PS participants, but post-intervention and study-long stress reduction was not significantly different between conditions. Participants rated low-effort interventions highest, but high-effort interventions reduced the most stress. Participants felt JIT provided accountability but desired partial agency over timing. We present type and timing implications.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; *Empirical studies in ubiquitous and mobile computing*; *Field studies*; • **Applied computing** → Health informatics.

*Both authors contributed equally to this research.

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9157-3/22/04.
https://doi.org/10.1145/3491102.3502027

KEYWORDS

Workplace stress, digital micro-interventions, just-in-time, stress reduction, psychotherapy

ACM Reference Format:
Esther Howe, Jina Suh, Mehrab Bin Morshed, Daniel McDuff, Kael Rowan, Javier Hernandez, Marah Ihab Abdin, Gonzalo Ramos, Tracy Tran, and Mary 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 (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA*. ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3491102.3502027

1 INTRODUCTION

Stress is prevalent and costly: Over 83% of Americans suffer from work-related stress [43] and over half experience stress throughout most of the day [48]. Work-related stress increases the risk of mental and physical health disorders, decreases productivity due to absenteeism and burnout, decreases overall job satisfaction and increases rates of stress-related accidents and employee medical, legal, and insurance costs [5, 8, 12, 63]. Workplace stress can also spillover into life outside of work, disrupting the overall well-being of workers [20]. Workplace stress intervention strategies such as organizational changes, individual stress management skills training, and therapeutic counseling have been recommended and evaluated as important components of long-term stress reduction [13]. Individual-based stress management interventions (e.g., cognitive-behavioral skills training, meditation, exercise, etc.) have been shown effective on psychological, physiological, and organizational outcome measures [50].

However, integrating individual-based stress management interventions into the workday and actively engaging with them can be challenging in workplace cultures with psychologically unsafe climates [15] or where there are high task demands but taking personal time for stress management is not supported [22]. Furthermore, while employees may learn about stress management strategies via in-person stress-reduction workshops or individual therapy, shifting from *understanding* evidence-based strategies to *using* these strategies when they are most needed (i.e., in moments of high stress) can be quite difficult psychologically, even in healthy workplace contexts that value employee well-being [21]. Smartphone-based mental health apps have gained popularity because they are always available (unlike a therapist), and can be used without drawing attention to the user or disrupting a work environment. However, most available mental health apps are not evidence based [24, 44], and those that have been rigorously studied – although shown to be effective in improving relevant outcomes [29, 32] – also have high attrition and low adherence [18, 31]. Therefore, promoting active engagement in interventions over time is an important step to improving user engagement and long-term outcomes for technology-delivered interventions [42, 45].

Recent efforts to improve user engagement over time for technology-delivered interventions have been focused on improving usability, delivery timing and content selection factors. Bite-sized, "digital micro-interventions" have been a promising content design approach to lowering the barrier to entry and to reducing the effort needed to engage with the content [1]. Because such interventions can be accessible via computers at employee workstations (i.e., do not require context switching), they may be easily integrated into the workday. To further improve engagement, adaptive or personalized micro-intervention delivery systems can incorporate individual preferences and contexts to choose the appropriate intervention content and delivery timing [14, 28, 45, 54]. For example, just-in-time, adaptive interventions (JITAI), powered by ubiquitous sensing technologies, promise to deliver the right interventions at the right opportunistic or vulnerable times to minimize disruptions and optimize efficiency [42, 52]. Just as therapists select the most appropriate intervention for a given moment, users could benefit from the personalization promised by JITAI systems [42]. Although JITAI have the potential to become more intelligent over time and improve their predictive ability to identify opportune moments for intervening [52] and the ideal content, most conceptualizations of this approach do not include opportunities for individuals to *proactively* engage in and exercise control over their stress management processes, which has been shown to improve long-term psychological outcomes in individual psychotherapy [11, 25] and to lead to health behavior change [58]. Accordingly, our research examines the role of system- and user-controlled intervention timing and content selection in promoting engagement and improving stress-reduction impact.

In this paper, we present a four-week, between-subjects study with 86 information workers. Our aim was to understand the impact of digital micro-intervention *delivery timing* and *content* on usage patterns and stress reduction throughout the workday to inform the design of effective and engaging workplace stress reduction intervention systems. Leveraging a desktop application to facilitate passive data collection and a Teams chatbot for intervention delivery, our study compared three categories of intervention content and two delivery timing conditions: *Pre-scheduled* (PS) by users and *Just-in-time* (JIT) according to passively-sensed and user-reported stress-levels.

We found our interventions to significantly reduce momentary stress. Although we found no significant difference between JIT and PS conditions in study-long or momentary stress reduction, participants preferred automated "nudges" over scheduling their own interventions, while simultaneously desiring control over their schedule with system assistance for intelligent planning. While our users rated the shorter, "easier" interventions as more enjoyable, we found that the longer, more difficult to perform interventions were in fact significantly more effective. Our findings suggest that both system-initiated intervention delivery and user-initiated intervention scheduling are promising directions for integrating stress-reduction interventions at work. In fact, users may benefit from a combination of the two, wherein system-initiated interventions offer ease and increase overall usage and user-initiated, pre-scheduled interventions promote a sense of control and could lead to healthy behavior change. In both cases, our findings suggest users benefit from having access to a healthy balance between easy-to-do and effective interventions. Based on these findings, we present opportunities to guide the design of personalized JITAI and planned intervention systems to reduce stress and enhance well-being in the workplace.

2 RELATED WORK

2.1 Workplace Stress Interventions

Intervention strategies for workplace stress are commonly grouped into three categories: primary, secondary, and tertiary [13, 49, 50, 55]. Primary strategies refer to action taken to directly change or eliminate stressors; these strategies often involve organizational-level changes (i.e., a culture shift), which might be difficult to achieve or even study because such changes are costly and may cause high-profile disruptions to the organization [13]. Secondary strategies are the most common; they target the individual experiencing stress and aim to detect and reduce their stress to prevent the development of chronic, stress-related mental and physical health issues. Individual talk-therapy, workshops teaching stress reduction and time-management, as well as mindfulness or meditation applications like HeadSpace are all popular examples of secondary strategies. Tertiary prevention concerns treatment for and recovery from stress-related mental and physical health issues through counseling or supportive services such as Employee Assistance Programs (EAPs). Some argue that addressing the source of stress through organizational changes, rather than assisting individuals, is necessary for long-term beneficial effects [13], and ample evidence points to the benefit of individual-focused interventions which are easier to implement and study [50]. Others suggest that such distinctions are not an important area for focus because individuals trained with stress management skills can bring about organization-level changes [7]. Furthermore, it has been shown that people perceive individual-level stress management techniques to be more effective than organization-level interventions [27]. While we acknowledge the importance of directly modifying or eliminating stressors, our study focuses on workplace stress interventions that empower individuals to manage their own stress responses (i.e., secondary strategies).

2.2 Technology-Delivered Mental Health Interventions

In the past decade, mental health and well-being innovators, researchers and therapists have explored the potential for new technology to overcome key barriers to mental health care access and to improve overall mental health and well-being outcomes for all. In particular, smartphone-based products have garnered attention for their ability to offer direct, real-time support to individuals trying to change their thoughts, feelings, and behaviors – changes which are typically discussed in therapy but for which individuals are usually responsible for making on their own, in between sessions. Such products are designed to capitalize on near-continuous access to users to create an opportunity for individuals to have support whenever they want – for example, while stressed at work. Given the potential for significant impact, many innovators and researchers have created and studied the efficacy of smartphone-delivered mental health interventions via randomized clinical trials [29, 32] and a myriad of consumer-oriented mental health applications exist to provide a range of support online, such as self-guided meditation or symptom management (e.g., Headspace[1], Calm[2], or Noom[3]), peer-support (e.g., Talklife[4], Supportiv[5]), or counseling (e.g., Talkspace[6], Sanvello[7]). However, despite the large number of academic and industry efforts in this area, significant room for improvement exists, specifically regarding user adherence in academic efforts, empirical testing in industry efforts, and intervention tailoring to person and context across both [44].

Many of the most promising technology-delivered interventions are narrow in scope and short in length (e.g., 1-minute meditation). These "digital micro-interventions" leverage technology affordances to provide individual components of traditional psychotherapy focused on managing proximal symptoms (e.g., relaxation for stress) in the hopes of achieving broad, distal objectives (e.g., overcoming depression) [1]. At best, systems delivering these interventions can take advantage of the usage and interaction data for personalization, increasingly recommending activities that are likely to be effective [53], used [28], preferred [45], or performed at the right time [42, 51, 56]. The fullest extension of such personalization, termed *Just-in-time adaptive interventions* (JITAI) have been introduced to deliver personalized, contextualized, and adaptable interventions incorporating the dynamic human behavior data captured through ubiquitous sensing technologies [42, 52, 57]. Prior studies focused on one iteration of JITAI, wherein micro-intervention delivery timing is improved based on ecological momentary assessment (EMA) or other passively sensed data [52, 56]. However, no prior work has examined how JITAI could be integrated into the everyday workflow over time or explored the differences between JITAI and a manual, planned approach that provides individuals full control over timing of the interventions. While the JITAI approach holds great potential for improving adherence and engagement, efforts to implement and test JITAI systems *in context* are still in their infancy and, as such, their automated features have not been tested against manual options. Our study aims to examine different ways of engaging with digital micro-interventions over time, specifically looking at the timing and the content of interventions situated in work contexts. To understand the effectiveness of the different engagement timing and intervention content on stress management for both the short- and long-term, we measured user stress levels at multiple time points: (1) using a single item question regarding stress levels just before and after each intervention use, as well as five times per day spaced throughout the day, and (2) in a weekly survey which included a full, validated self-report stress scale [35].

2.3 Personalizing Technology-Delivered Interventions

As individual needs and behaviors change over time, how they want to engage with these technology-delivered stress interventions (i.e., whether a technology-initiated intervention or human-initiated intervention is appropriate or preferred) may depend largely on that individual's context, history, and characteristics. One's vulnerability to stressors and perception of stress intensity is influenced by individual differences that tend to be relatively stable over time (such as personality, demographics (e.g., age, gender), and past experiences from developmentally-sensitive ages), as well as variables which may fluctuate more dramatically over time (such as cognitive appraisal abilities, coping strategies, and available social support [3, 13, 30, 59, 61]. The effectiveness and user engagement of stress interventions can also be influenced by individual characteristics as well as situational factors surrounding the user's current stress level [47], current receptivity to interventions [42], and other mediators of change (e.g., acceptance of undesirable thoughts and feelings [4]). Incorporating individual preferences has been shown to improve the engagement and outcomes of stress interventions [14, 28, 45, 54]. Therefore, our study investigates user preferences for stress intervention types and timing to improve intervention engagement and stress reduction impact.

3 METHOD

The goal of this work was to identify design opportunities for systems that integrate digital micro-interventions into every day work contexts. Our research questions were:

RQ1. How does intervention timing impact intervention usage, stress reduction, and user preference?

RQ2. How do different types of interventions impact intervention usage, stress reduction, and user preferences?

RQ3. What aspects of the intervention timing and content do participants find most useful or needed?

To examine the impact of different delivery timings (RQ1), we conducted an experimental study that directly compared two conditions: *Pre-scheduled* (PS) by users and *Just-in-time* (JIT) according to user stress-levels. We designed and developed an intervention system with a chatbot that facilitated different delivery timing of stress reduction interventions to information workers (Section 3.1). In the PS condition, the chatbot helped participants browse through the catalog of interventions and schedule them

[1] https://www.headspace.com/
[2] https://www.calm.com/
[3] https://www.noom.com/
[4] https://www.talklife.com/
[5] https://www.supportiv.com/
[6] https://www.talkspace.com/
[7] https://www.sanvello.com/

in their calendars; in the JIT condition, the chatbot nudged participants to perform interventions when it detected high stress levels. To examine the impact of different intervention types (RQ2), we adapted evidence-based psychotherapy interventions into digital micro-interventions and categorized them into three types according to their function and required user effort (Section 3.2). Participants selected from the three intervention types throughout the study. Finally, to solicit user feedback from in-vivo usage of the system (RQ3), we deployed this intervention system to 86 information workers and conducted a longitudinal study with four weeks of observation that compared two delivery timings and three intervention content types (Section 3.3).

3.1 System Design

We designed and developed an intervention system that composed of (1) a *stress score* component, (2) a *JIT* component, and (3) a *chatbot* component. Our *stress score* component computes the user's current level of stress based on passively sensed data. Then a *JIT* component leverages the stress scores and the user's self-reported stress levels to determine when to nudge users to perform stress-reduction interventions. Finally, a *chatbot* component proactively sends messages to the user and facilitates the delivery of ecological momentary assessments (EMAs), surveys, and intervention content. We designed our system to capture salient signals in people's work context where they are more likely to be in front of their computers, but offered flexibility through the chatbot so that they could utilize their computers or mobile devices for engaging in EMAs, surveys, or interventions as needed.

3.1.1 Stress Score Component. The *stress score* component is responsible for inferring user's stress level based on passively-sensed information. Sano et al. [52] has highlighted that computer usage, calendar and email usage, intervention history, activity and heart rate variability are together useful features for predicting intervention timing. Our *stress score* component uses similar features but optimizes for detecting moments of high stress, such that interventions can be delivered at moments of need (i.e., reduce momentary high stress).

We capture contextual and behavioral information about people through custom logging software that runs on their primary work computer. From this logged information that includes computer activity (e.g., window switching, keyboard usage), behavioral and physiological signals (e.g., facial expression, breathing rate), we compute a stress score. Specifically, the logging software has three main capabilities. The first is to capture email, calendar and application data from the users' desktop applications usage (all participants used Microsoft Outlook as their primary email and calendar software). The second is to use their webcam to capture their position and facial actions while they are in front of their workstation. The third is to use their webcam to capture their heart rate and breathing rate using a non-contact measurement technique [34].

We designed the stress score to capture five components that previous work has identified as sources of stress. They are defined as follows:

- **Email** (w_1): The volume of emails received in a given day has been linked to higher stress in information workers [38, 39]. The *email* component (w_1) at X hours into the day was computed as the number of emails received until that time of day / 2400.
- **Calendar** (w_2): The lack of breaks and number of appointments in a work day (e.g., meetings) is a known stressor for information workers [37]. The *calendar* component (w_2) of the score was computed as the total number of appointments in a given day / 15.
- **Time** (w_3): People are also more likely to experience negative emotional states (such as stress) later in the day, in general [41]. The percentage into the day (w_3), was the *time* component of the stress score.
- **Facial** (w_4): Previous work has identified that facial expressions during information work can capture changes in affect [40, 41]. To describe facial behaivor we use the Facial Action Coding System (FACS) [17], the most commonly used and descriptive taxonomy. The *facial* component (w_4) was computed as corrugator (AU04) and lip depressor (AU15) activity minus zygomatic major (AU12). Corrugator (i.e., brow furrowing) and lip depressor (i.e., frowning) are typically linked to negative experience (e.g., confusion, frustration, displeasure), where are zygomatic major (i.e., smiling) is typically linked to positive experience (e.g., joy, pleasure) [26].
- **Physiological** (w_5): Changes in heart rate are also associated with increased stress during computer based work [23, 60]. The *physiological* component of the stress score (w_5) was computed as the current heart rate (in beats/min) divided by 100 beats/min.

We normalize each component of the stress score to create a value between 0 and 1. If any of these numbers was greater than 1, it was rounded to 1. We combined and normalized these components to compute the stress score as:

$$S = \frac{w_1 + w_2 + w_3 + w_4 + w_5}{5} \tag{1}$$

This stress score was stored in the database for retrieval by the *JIT* component.

Our stress score includes aspects of work demand (i.e., email volume, calendar volume) and available resource (i.e., time into the day) as well as behavioral and/or physiological changes in reaction to stress (i.e., facial expression, heart rate). These components were designed based on insights from prior work and crafted to create a simple and explainable continuous estimate of how likely an individual was to be experiencing stress. Although a more elaborate "machine-learned" stress score could be used in the future, for the purposes of our study, we found that this was a practical estimate of stress. We ran a retrospective analysis of the correlation between our stress scores and participants' self-reported ratings on their momentary stress levels via EMAs (1=Not at all stressed; 3=Moderately stressed; 5=Extremely stressed), we found a significant positive correlation between the two (N=1318, Pearson r=0.2, p<0.01). We note that the stress score was not always available at the time participants reported their momentary stress levels because the participant may have disabled the sensing software temporarily or responded to the EMAs when they are not at their

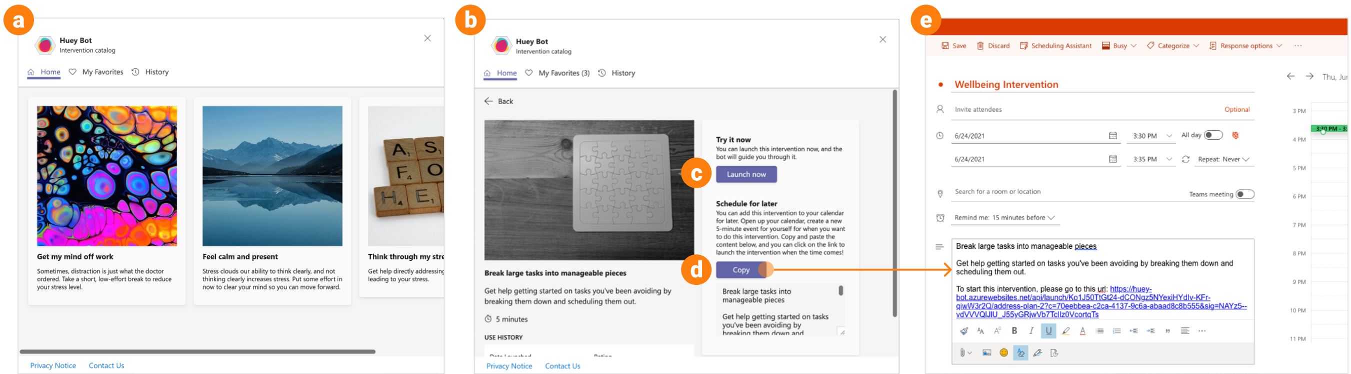


Figure 1: Users can browse the intervention catalog to choose from different intervention types (a) and navigate to an intervention to view its details (b). Users can choose to launch the intervention at that moment (c) or copy relevant information about the intervention, including a link to launch the intervention, to perform the intervention at a later time (d). Users can paste this information into a calendar event to schedule it at a desired future time (e).

desk. Therefore, we rely on both stress scores and self-reported stress levels for the *JIT* component, which we discuss next.

3.1.2 JIT Component. Our *JIT* component is responsible for determining if the system should nudge the user to perform a stress-reduction intervention at that time. We leverage the computed stress score and also apply heuristics that incorporate users' self-reported stress levels to maximize potential efficacy of the stress interventions [56]. Self-reported stress levels are obtained from EMAs or at the end of an intervention usage.

Based on the data from the first week of using the system during the four-week observation period, we compute each user's average for computed stress scores and self-reported stress levels. We then use these averages as individual baselines (week one of four) and as thresholds for delineating high-stress from low-stress during the subsequent weeks (weeks two to four). Our logic for JIT interventions also checks historical intervention usage and system nudges to ensure that the system does not aggressively prompt for engagement. Our system will send a nudge to engage in a stress intervention if and only if:

- Computed stress score is greater than or equal to the user's baseline (or 0.5) in the past 5 minutes *or* self-reported stress level is greater than or equal to the user's baseline (or Moderately Stressed) in the past 30 minutes,
- It is during the weekday and the user's working hours,
- There is no scheduled stress interventions during the remainder of the day,
- The user has not completed an intervention in the past hour,
- There has not been a system-initiated nudge in the past two hours, and
- There has not been four or more nudges so far that day.

These nudges are sent as messages from the chatbot, which we describe next. Our system is configurable such that the *JIT* component can be active for a subset of users.

3.1.3 Chatbot Component. We use the Microsoft Teams chatbot[8] as a platform to maximize success in delivering and interacting with EMAs, surveys, and intervention content because all of our users regularly used Teams for work-related communication and had Teams clients readily available on desktops and mobile phones. We designed our chatbot, named Huey, to proactively initiate conversations with the users, where it would remind users to complete EMAs or surveys or to help them engage in stress-reducing interventions. Most prompts are presented as Adaptive Cards[9], either with predefined response options (e.g., scales for stress levels) or a button to open a task module dialog that hosted web-based contents (e.g., videos, surveys). We implemented Huey using Microsoft's Bot Framework[10].

Huey provides seamlessly integrated experiences for browsing the intervention catalog or consuming the stress interventions through Teams task modules (i.e., embedded web controls) so that the users can achieve all tasks within the Teams app. Users can browse the intervention catalog to drill into different intervention types (Figure 1a), navigate to an intervention they like (Figure 1b), and launch the intervention within the same dialog flow (Figure 1c). If the users want to perform the intervention at a later time, they can copy intervention metadata, which includes a link to launch the intervention at any time (Figure 1d). This information can be easily pasted into a calendar event and scheduled at a time that works better for them (Figure 1e). When Huey nudges users to perform an intervention, users can choose to perform the intervention at that moment or to postpone it to a later time that day (Figure 2a). When the users opt to perform interventions at that moment, they can choose an intervention type they are interested in performing (Figure 2b). Then Huey selects a random intervention from that category that is used the least frequently. Or users can opt to postpone the intervention and select a later time for Huey to check back in (Figure 2c).

Just before the users perform their intervention, Huey asks them to rate their momentary stress using a 5-level stress rating (1=Not at all stressed; 3=Moderately stressed; 5=Extremely stressed; Figure 3a). After performing the intervention, Huey asks users to

[8]https://docs.microsoft.com/en-us/microsoftteams/platform/bots/what-are-bots

[9]https://adaptivecards.io/

[10]https://dev.botframework.com/

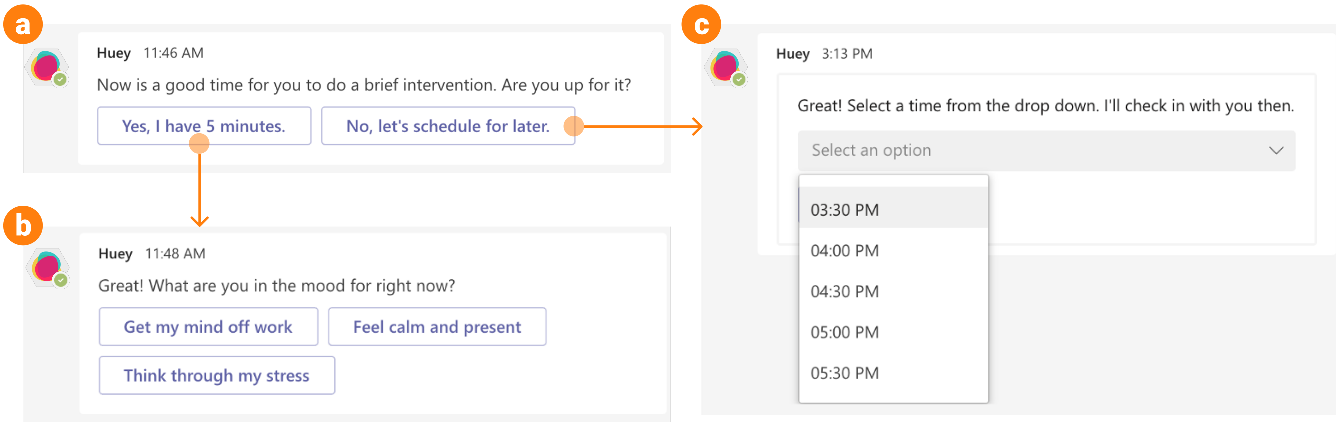


Figure 2: When the system determines that an intervention is needed at the moment, Huey sends a nudge to users to perform an intervention (a). When the users opt to perform interventions at that moment, they can choose an intervention type they are interested in performing (b). Or they can opt to postpone the intervention and select a later time for Huey to check back in (c).

reflect on how the intervention went and to rate the intervention (1=Very poor; 3=Acceptable; 5=Very good; Figure 3d,e). Finally, Huey concludes the intervention by asking users to rate their momentary stress and comments on the changes in stress levels from before performing the intervention (Figure 3f,g). Figure 3 shows an example of this intervention consumption flow.

Huey supports three different intervention modalities: (1) A video-based intervention provides a brief description of the content, followed by a task module dialog that played the video (Figure 4a); (2) A single-turn text prompt intervention provides a brief instruction for the users to engage in an activity, followed by a prompt to answer a reflective question (Figure 4b); (3) A conversation-based intervention provides a dialog that guides users through a series of prompts (Figure 4c).

3.2 Intervention Design

We designed interventions based on components of Cognitive Behavioral Therapy (CBT) and Dialectical Behavioral Therapy (DBT), two empirically supported front-line psychotherapy modalities that are used to flexibly treat a wide range of mental health and well-being concerns [9, 36]). We created all interventions to take under five minutes, and they were comprised of either a short video, a single-turn text prompt, or a brief therapeutic conversation with Huey. We then categorized the interventions depending on the function served for users as well as approximate effort required.

- **Get my mind off work** (Low effort): 'Get my mind off work' interventions are translations of the DBT *pleasant activities schedule*, a list of positive activities that help individuals regulate their emotions by becoming engaged with something that elicits positive feelings [33]. Examples include watching a short video of penguins playing in Antarctica or listening to a favorite song. These interventions are simple and likely similar to activities many individuals do naturally throughout the day when attempting to take a break at work. However, they steer clear of activities which may feel pleasant in the moment but research shows may lead to more distress in the long-term, such as scrolling on social media or eating a high quantity of sugary food. These interventions are intended to capture users' attention with low levels of user effort and investment, and are deployed using the single-turn text prompt and video formats.
- **Feel calm and present** (Medium effort): 'Feel calm and present' interventions draw inspiration from the mindfulness practices taught in CBT and DBT, which help individuals re-focus on the present moment in order to gain perspective and increase control over their thoughts, feelings, and behaviors. Examples include using the five senses to notice and name components of the immediate environment or writing a self-affirming statement ten times with one's non-dominant hand – an activity which takes a substantial amount of focus and tends to bring individuals in states of stress into the present moment. These interventions typically require moderate levels of user effort but have significant empirical support as stand-alone interventions capable of decreasing stress [33]. They are deployed using the single-turn text prompt and video formats.
- **Think through my stress** (High effort): 'Think through my stress' interventions help users directly address and problem solve stress-inducing components of their lives using strategies from CBT and DBT, such as cognitive re-framing, pros and cons lists, and reaching out to a friend or co-worker for help with emotional processing or getting productive [62]. These interventions require the highest amount of user effort as they necessitate direct engagement with stressful content. They are delivered via single-turn text prompts and therapeutic conversations with Huey.

3.3 Study Design

We conducted a four-week, between-subjects user study, where our participants engaged with our system through the Huey chatbot, which delivered stress-reducing micro-intervention content and facilitated different study requirements and protocols.

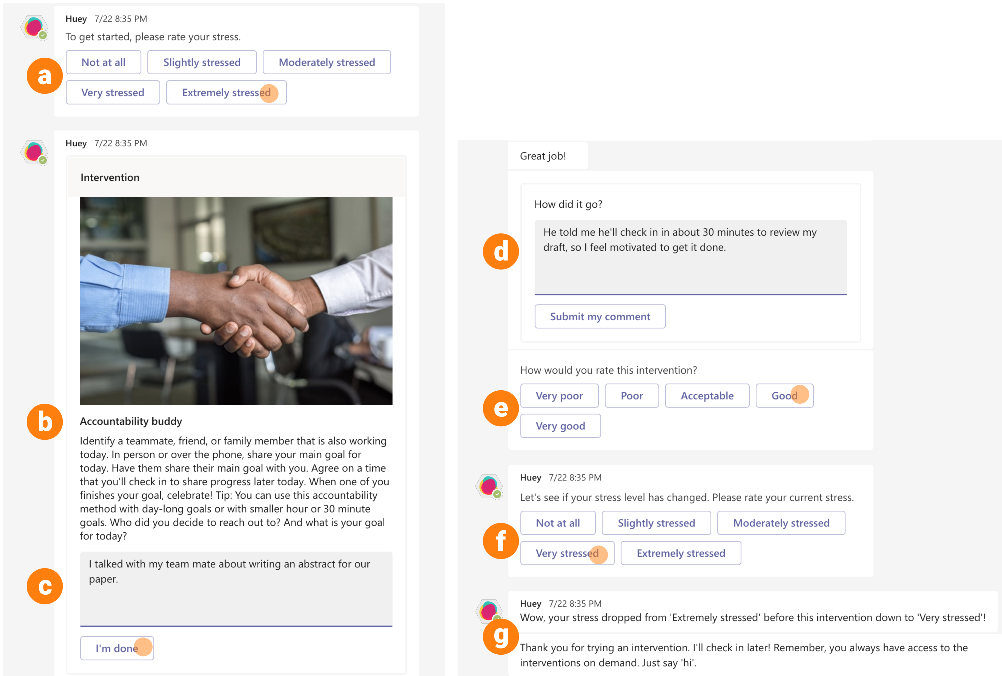


Figure 3: An example dialog flow of how Huey facilitates intervention consumption. Huey first asks users to rate their current stress level (a) before presenting the intervention content to the users (b). The intervention may ask the users to respond to a prompt (c). After completing the intervention, Huey asks users to reflect on the intervention (d) and provide a rating (e). Finally, Huey asks users to rate their stress level (f) and comments on the changes in stress levels (g).

3.3.1 Participants. We recruited information workers from a large technology organization by sending email advertisements about the study to a randomly sampled set of employees from the organization's employee database. Interested participants completed a brief screener survey about their demographics (e.g., age, gender, role) and work set up (e.g., primary device specification and OS, web camera availability). Eligible participants, whose primary device specification met our sensing software requirements, were asked to install and run the study software for 30 minutes to confirm system compatibility. We then enrolled participants on a first-come, first-served basis. In total, we enrolled 87 participants into the study. Participants were randomly assigned to one of the two conditions while maintaining equal gender distribution between the conditions, as prior work shows women report higher overall workplace stress than men [22]. One participant dropped out and another participant switched conditions during week one of the four-week observation period, both due to unforeseen technical issues.

Of the final set of 86 participants who successfully completed the study, 65.1% identified as male and 32.6% as female. 38.4% were in the age range of 36-45 years old, 23.3% in 26-35 and 23.3% in 46-55 year old ranges. 54.7% worked in Engineering/Development roles, 22.1% in Sales and Marketing roles, 8.1% in Operations and Services roles, 5.8% in Business Development and Strategy roles, and 4.7% in Administrative Assistant or Human Resources roles. 86.2% of the participants worked remotely from home.

Huey also supports special commands to allow accessing study-related instructions (via "help") or on-demand interventions (via "hi") at any time. Messaging "hi" to Huey would initiate dialogs for browsing the intervention catalog and performing interventions on demand. The timing of reminders for EMAs and surveys as well as which dialog flows are available are configurable per user.

We were unable to capture webcam-based signals from 16 of 86 participants (eight in each condition) due to unforeseen performance issues with the sensing software that interfered with their daily work. However, our *stress score* component is robust to missing data such that this was not an issue for the study.

3.3.2 Procedure. The study procedure included one week of on-boarding, four weeks of observation of intervention usage and engagement, and one week of off-boarding (Figure 5).

During the preparation week, participants were asked to install the required sensing software and the chatbot and to complete an intake survey. The intake survey asked for the participant's local time zone and their typical start and end times of the work day, which were used to configure the system's interaction with the participants. The intake survey included the Depression, Anxiety, and Stress Scale 21 (DASS-21), a 21-item self-report questionnaire designed to assess for clinical levels of depression, anxiety, and stress [35], and the Emotional Regulation Questionnaire (ERQ), a 10-item scale designed to measure respondents' tendency to regulate their emotions through cognitive reappraisal and expressive suppression [19]. We asked participants to report their current

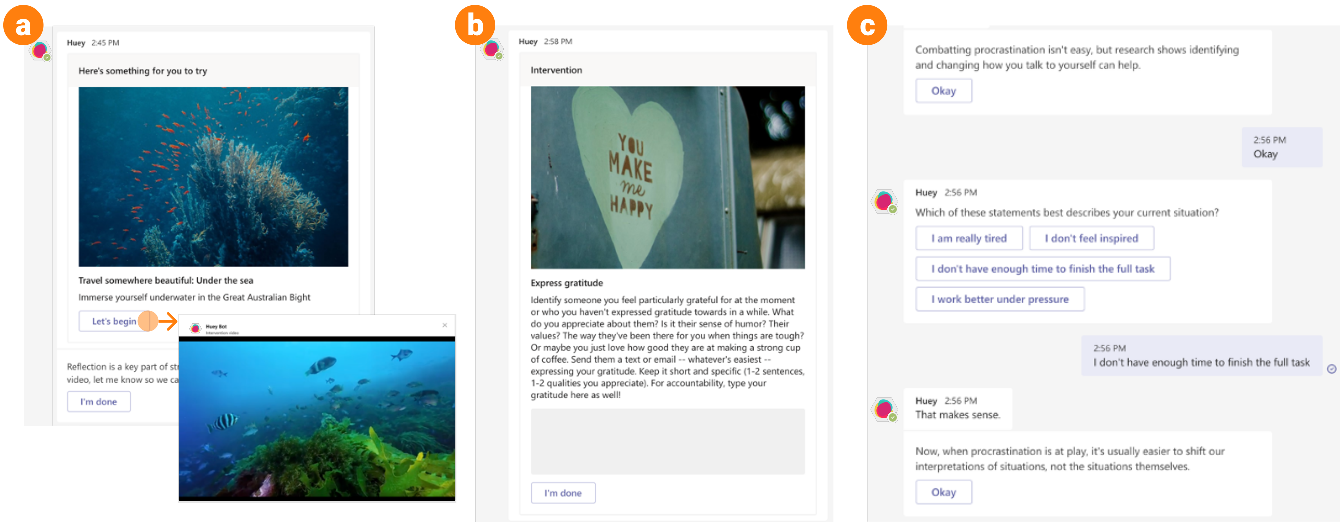


Figure 4: Huey supports three intervention modalities: (a) A video-based intervention provides a brief description of the content, followed by a task module dialog that played the video; (b) A single-turn text prompt intervention provides a brief instruction for the users to engage in an activity, with an open-ended prompt to answer a reflective question; (c) A conversation-based intervention provided a dialog that guides engage through a series of prompts.

stage of behavior change to reduce stress at work from four stages of behavior change (Stage 1: Pre-contemplation, Stage 2: Contemplation, Stage 3: Taking action, Stage 4: Maintenance) adapted from the Transtheoretical Model [46] The intake survey also included questions for personality style, stressful life events, emotional resilience, and self-care style.

During the four-week observation period, participants were asked to interact with Huey to engage in stress-reducing interventions, where we configured the intervention system to enable features specific to their assigned conditions:

- **Pre-scheduled Engagement (PS)**: In this condition, participants were asked to plan their interventions in advance. On Fridays prior to each study week, Huey asked participants to browse the catalog of stress-reducing interventions, choose specific interventions they would like to try, and schedule at least one intervention for the upcoming week into their work calendar (Figure 1). From the intervention catalog, participants could copy intervention-specific information into their calendars with a link to launch the intervention. These participants leveraged the calendar's built-in reminder functionality for the interventions. On Mondays of each study week, Huey reminded participants to review their scheduled interventions and to adjust them appropriately. When the scheduled time arrived, participants clicked on the link in the calendar event to engage with Huey to carry out the intervention. Participants in this condition could also access the intervention catalog on demand, where they could launch a selected intervention at that moment or copy intervention-specific information into their calendar.
- **Just-in-time Engagement (JIT)**: In this condition, participants were asked to engage with an intervention based on our system's JIT component (Section 3.1.2). When our JIT component determined that a stress-reducing intervention is needed and appropriate for the participant, Huey sent a message to the participant, presenting an option to perform the intervention at that moment or to postpone it to a later time that day (Figure 2). When the participants opted to perform interventions at that moment, Huey asked them to choose between one of three intervention categories we described in Section 3.2. Once a category was selected, Huey chose a random intervention from that category that was used the least frequently. Participants then engaged with Huey to carry out the intervention. As in the PS condition, participants in the JIT condition could also access interventions on demand, where they could perform the intervention at that moment or schedule it to a later time that day.

Based on each participant's reported work hours, Huey asked participants to complete five EMAs per day during the weekday, roughly equally spaced apart to be on the hour or half-hour (e.g., at 9 AM, 11 AM, 12:30 PM, 2:30 PM, and 4 PM if work hours span 9 AM to 5PM), and to complete two optional EMAs during the weekends (e.g., at 11 AM and 3 PM). Each EMA consisted of two parts. The first part was required for the study and asked participants to rate their stress level during the past 30 minutes using the same 5-level stress rating. The second part included questions about work demands, available resources, arousal, valence, food consumption, and social interactions. EMA questions can be found in the Supplementary Information. Participants were also asked to complete morning surveys 15 minutes before the start of each work day, evening surveys 15 minutes before the end of each work day, and weekly surveys during the afternoon on Fridays. The morning survey included questions from the Census Sleep Diary [10], and the evening survey captured food and beverage intake throughout

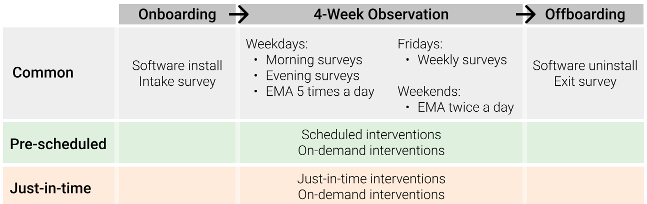


Figure 5: Our study procedure included on-boarding with software install and intake survey, four weeks of observation of intervention usage, surveys, and EMAs, and off-boarding with software uninstall and exit survey.

the day. On Fridays of the first three weeks, participants were asked to complete weekly surveys which included questions from the DASS-21 and the stages of behavior change.

After the four-week observation period, participants were asked to complete an exit survey that included scales for DASS-21, stressful life events, emotional resilience, and the stages of behavior change. The exit survey presented 8 questions probing the usability of the assigned conditions including ease of use, satisfaction, and frustration. The exit survey also included condition-specific and open-ended questions probing their preferences for engaging with the interventions, appropriate timing of interventions, and how participants compared accessing interventions on-demand, to scheduling interventions in advance, or to being nudged to do an intervention by a system. It also solicited comments about the intervention content, what factors motivated them to perform the interventions, and any perceived helpfulness or impact of the interventions on stress reduction. These questions were the identical for both conditions. Exit survey questions can be found in the Supplementary Information.

Each participant was compensated with a $400 Amazon gift card for their participation and data. Our study was approved by the Microsoft Research Institutional Review Board (IRB).

3.4 Data Processing & Analysis

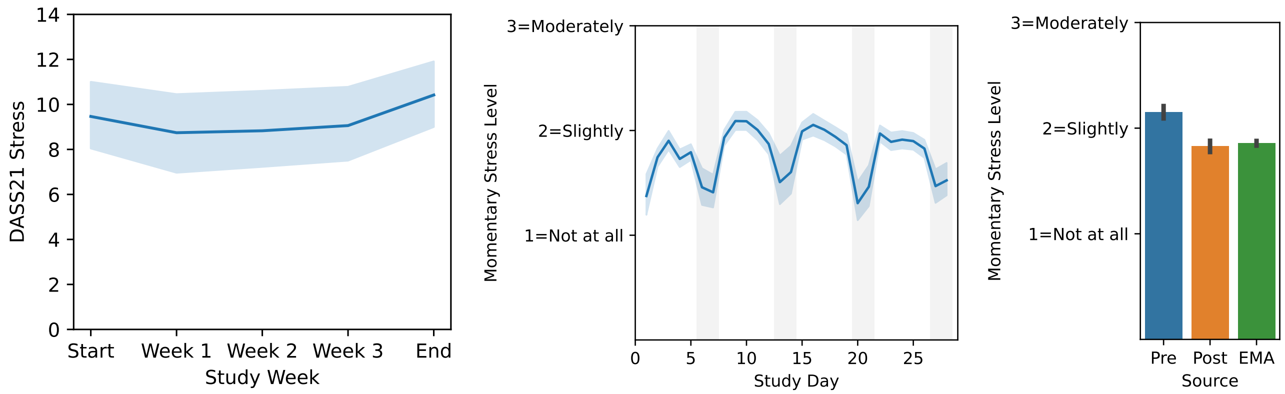
We combined data from the system usage logs and survey responses to understand engagement patterns, intervention usage, and outcomes. We leveraged the system usage logs for our analysis of intervention usage. Each intervention use instance was associated with one of three intervention categories (Section 3.2), whether or not it was used on-demand, timestamps of when it was started and completed, stress levels immediately before and after the intervention use, user ratings, and any free-form comments the participants wanted to provide. We extracted 1651 unique intervention attempts during our study. 28.9% (477/1651) of those were started but never completed. 96.5% (1133/1174) of the completed interventions were followed by user ratings, and 92.4% (1085/1174) of the completed interventions had both pre-intervention and post-intervention stress levels. We collected 6685 stress levels from EMAs, 1174 from pre-intervention use, and 1085 from post-intervention use, for a total of 8944 stress levels. We had both intake and exit DASS-21 measures for each of the 86 participants and 217 DASS-21 measures from weekly surveys, for a total of 389 DASS-21 measures. We computed *momentary stress reduction* by subtracting the pre-intervention stress levels from the post-intervention stress levels and *study-long stress reduction* by subtracting the DASS-21 stress sub-scale responses from the intake surveys from that of the exit surveys, where positive values indicate higher stress reduction. We aggregated the intervention usage and stress level data for each participant for analysis. We mapped the participants’ reported stages of behavior change to numerical values based on their reported stage (Stage 1: Pre-contemplation, Stage 2: Contemplation, Stage 3: Taking action, Stage 4: Maintenance) and examined the *change in the stages of behavior change* between the study start and end. Additional data we collected from surveys around depression/anxiety, personality, life events, resilience, sleep, food and beverage intake, etc. and from the sensing software are out of the scope of analysis for this paper due to lack of time for the analysis.

For comparing the means of the two conditions (JIT vs. Pre-scheduled), we used the Welch Two Sample t-test. Wherever applicable, we used Benjamini-Hochberg procedure [2] on the t-test results to correct for multiple comparisons. For comparing differences within participants, we used paired t-tests. We used one-way analysis of variance (ANOVA) tests to examine differences in outcome variables (e.g., stress reduction) with multi-level factors (e.g., intervention categories). When we found significant results, we then investigated pairwise differences, employing Tukey’s HSD procedure to correct for the increased risk of Type I error due to unplanned comparisons. We used linear mixed-effects models to investigate the relationship between per-participant characteristics and outcome variables, again investigating any pairwise differences using Tukey’s HSD procedure to adjust for repeated testing. When we included gender as a variable, we included the subset of participants (N=84) who identified themselves as male or female, due to the small sample size of other gender identity categories (N=2). For correlation analyses, we used Pearson’s correlation. We used Python and R for processing the data and for statistical analyses.

Two researchers qualitatively coded the open-ended survey responses using inductive thematic analysis [6]. We identified several topics of interest (e.g., the timing and frequency of engagement with the bot, motivating factors, preferences for engagement and interventions, desired functionalities), categorized participant responses into themes within each topic, and quantified their occurrence.

4 FINDINGS

We first describe the temporal trajectory of participants’ self-reported stress throughout the study to contextualize the overall impact of the study (Section 4.1). Then we organize our findings according to our research questions. We first address RQ1, presenting the impact of the two engagement timing conditions on the overall intervention usage, momentary and study-long stress reduction, and user ratings (Section 4.2). We also include the impact of on-demand intervention usage. We then address RQ2 and present the impact of the three intervention types on usage, stress reduction, and ratings (Section 4.3). Finally, we address RQ3 and summarize participants’ feedback on overall system usability, engagement timing, and interventions (Section 4.4).



(a) Weekly DASS21 stress level

(b) Daily momentary stress level

(c) Momentary stress level across sources

Figure 6: Self-reported stress levels throughout the study. (a) shows average DASS-21 stress subscale from intake, weekly, and exit surveys. DASS-21 stress scores of 0-14 indicates 'Normal' stress range. (b) shows average momentary stress levels per day from EMAs and pre-/post-intervention uses. Vertical grey bands denote weekends to highlight daily patterns on momentary stress levels. (c) shows average momentary stress levels across all pre-intervention uses (Pre), post-intervention uses (Post), and EMAs. All error bands and bars indicate 95% confidence intervals.

4.1 Stress over Time

Study-long stress. Overall, the present sample did not experience extreme levels of stress or statistically significant changes in stress levels from study start to study finish. Average participant stress at study start was within a non-clinical range (i.e., within normal limits; $\bar{x}$=9.47, σ=6.96 at study start) and remained within this range throughout the study with little variation (Figure 6a), as measured by the stress sub-scale of DASS-21 [35]. Stress levels at study start and stress levels at study end ($\bar{x}$=10.42, σ=7.23) indicated no statistically significant study-long change in stress levels.

Short-term stress. Momentary stress levels, as assessed via EMAs and pre-/post-intervention stress levels, were relatively stable throughout the course of the study as well. Participants reported an average of 104 momentary stress levels over the course of the four-week observation ($\bar{x}$=4.19, σ=2.39 per day per participant). Average momentary stress level was 1.87 (σ=0.91), between '1=Not at all stressed' and '2=Slightly stressed,' with minimal variation over the course of the study. We also saw lower stress levels during the weekends (Figure 6b). Average EMA stress level was generally below pre-intervention and above post-intervention stress levels (Figure 6c).

Pre-/post-intervention stress. Pre-intervention ($\bar{x}$=2.16, σ=0.95) and post-intervention ($\bar{x}$=1.82, σ=0.87) stress levels indicate that interventions resulted in a statistically significant momentary decrease in stress levels ($t(1084)$=18.113, $p \ll 0.001$).

4.2 RQ1: Engagement Timing Impact

Quantity of interventions used. Participants completed an average of 13.65 interventions over the four-week study ($\bar{x}$=10.39, min=2, max=55). We found a statistically significant difference in number of interventions completed between the two conditions; JIT participants completed significantly more interventions than PS participants (19.74 vs. 7.56 per participant, $t(63.633)$=-6.696, $p \ll 0.001$), but this is likely due to the study design, wherein JIT participants were prompted throughout the day.

Stress reduction. Despite differences in completed usage of interventions, we found no statistically significant difference in both momentary and study-long stress reduction between the two conditions. We also found no correlation between the total number of completed interventions and study-long stress reduction (Pearson r=-0.06).

User ratings. Participants had a generally positive reaction to the interventions, giving them an average rating of 3.65, between '3=Acceptable' and '4=Good' (σ=0.98). We found that JIT participants rated interventions significantly lower than PS participants by about 0.256 points ($\chi^2(1)$=5.962, $p<0.05$).

Behavior change stage. At the beginning of the study, 48.8% of participants were in 'Stage 3: Taking action' stage of behavior change, with 32.6% in 'Stage 2: Contemplation', 15.1% in 'Stage 4: Maintenance', and 3.5% in 'Stage 1: Pre-contemplation' stages. Controlling for behavior change stage at study start, we found statistically significant difference in advancement through the behavior change stages between conditions: PS participants reported significantly more advancement through the behavior change stages compared to JIT participants ($F(1)$=6.834, $p<0.05$) and no statistically significant interaction effect between the intake stage and condition.

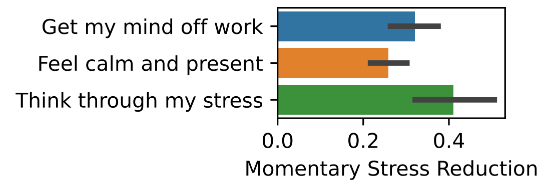
On-demand usage. Although users could access on-demand interventions in both conditions, PS participants completed statistically significantly more on-demand interventions compared to JIT participants (2.63 vs. 0.02; $t(42.125)$=7.552, $p \ll 0.001$). On average, PS participants completed interventions on-demand 38.2% of the time (σ=0.33). 46.5% of PS participants completed on-demand 50% of the time or more. For PS participants, interventions used on demand reduced statistically significantly more stress than those used at pre-scheduled times ($\chi^2(1)$=10.587, p<0.01) by about 0.23 points. Again, this could have been an artifact of the study design, as PS participants might have wanted interventions at different times than for which they had scheduled them. According to their pre-intervention stress levels, we found that PS participants used on-demand interventions when they were slightly more stressed than at pre-scheduled times (2.23 vs 2.06), but the effects were not statistically significant ($t(229.81)$=-1.862, p=0.064). We found no statistically significant difference in subjective ratings between interventions used on-demand versus those used at pre-scheduled times.

4.3 RQ2: Intervention Type Impact

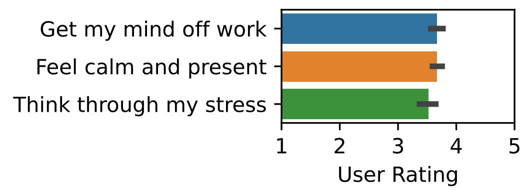
In both PS and JIT conditions, participants were able to choose from the three intervention types whenever completing an intervention.

Quantity of interventions used by type. On average, participants selected 'Get my mind off work' interventions 36.8% (σ=0.254) of the time and completed 71.6% of those selected. They selected 'Feel calm and present' 45.5% (σ=0.251) of the time and completed 72.7% of those selected, and selected 'Think through my stress' 17.7% (σ=0.201) of the time and completed 100% of those selected. JIT participants completed statistically significantly more 'Feel calm and present' interventions ($t(82.631)$=-2.978, p<0.01) and statistically significantly fewer 'Get my mind off work' interventions ($t(80.364)$=-2.073, p<0.05) compared to PS participants. There were no statistically significant usage differences between JIT and PS for the 'Think through my stress' intervention type. We modeled the impact of baseline DASS-21 stress, emotion regulation strategies, behavior change stage, age, and gender on the completion rate per intervention type and found the baseline DASS-21 stress to have a statistically significant effect on the completion rate of 'Feel calm and present' interventions ($F(1)$=5.630, p<0.05). We found no other statistically significant effects.

Stress reduction by type. When we examined the impact of the completion rate per intervention type on study-long stress reduction, we found that a higher rate of completed uses of 'Get my mind off work' among all completed uses had a statistically significant improvement on study-long stress reduction ($F(1,83)$=6.055, p<0.05). Of the 1085 completed intervention uses with pre- and post-intervention stress levels, 'Think through my stress' reduced momentary stress by 0.41 points on average (σ=0.61), 'Get my mind off work' reduced momentary stress by 0.32 points on average (σ=0.55), and 'Feel calm and present' reduced momentary stress by 0.26 points on average (σ=0.53; Figure 7a). Intervention type had a statistically significant effect on momentary stress reduction ($\chi^2(1)$=9.77, p<0.01). Pairwise comparisons of intervention type revealed that 'Get my mind off work' interventions reduced momentary stress more than 'Feel calm and present' interventions, and 'Think through my stress' interventions reduced momentary stress more than 'Feel calm and present'–both to a statistically significant extent. There was no statistically significant difference in momentary stress reduction between 'Think through my stress' and 'Get my mind off work.' We found similar results when controlling for condition (JIT/PS),



(a) Momentary stress level across intervention types



(b) User rating across intervention types

Figure 7: Average momentary stress reduction (a) and user rating (b) across three intervention types. Positive stress reduction indicates that stress is reduced after intervention use on a 5-point scale (1=Not at all stressed; 3=Moderately stressed; 5=Extremely stressed). Higher user rating on a 5-point scale (1=Very poor; 3=Acceptable; 5=Very good) indicates that the user liked the intervention after intervention use. The error bars indicate 95% confidence intervals across all intervention uses.

baseline stress, emotion regulation style, behavior change stage, gender, and age.

User ratings by type. On average, participants' subjective ratings of the interventions were 3.67, between '3=Acceptable' and '4=Good,' for 'Get my mind off work' (σ=1.02), 3.67 for 'Feel calm and present' (σ=0.96), and 3.52 for 'Think through my stress' (σ=0.96; Figure 7b). Intervention category had a statistically significant effect on user rating ($\chi^2(1)$=7.44, p<0.05), such that 'Get your mind off stress' interventions were rated significantly higher than 'Think through my stress' interventions.

4.4 RQ3: User Feedback on Intervention Timing and Types

Pre-scheduled participant feedback. PS participants (N=43) used a variety of factors for determining when they would choose to place interventions on their calendars. Some participants chose beginning of the day or end of the day (N=24), and several chose to space them throughout the week (N=7). Many looked for free spots on their calendar (N=13), after several back-to-back meetings when they knew they would be stressed, or afternoons when they knew they would be tired.

PS participants liked that having interventions on the calendar held them accountable (N=13): "I didn't forget because it was on the calendar"; "it calmed me seeing it was there." There was a subgroup (N=13) that especially liked to plan interventions out or make them recur, while others specifically mentioned their ease of use (N=5), that they could use them on demand if they needed to (N=3), and that they liked having breaks in the calendar to learn something new (N=6). Although some users complained that the chatbot did not schedule interventions automatically for them based on their availability, others noted that free times were not necessarily indicative of stressful moments, meaning a chatbot might not pick the best time based on a simple free time algorithm.

While 11 out of 43 PS participants told us that they liked pre-scheduled and on-demand interventions equally, 30 participants stated a strong preference for on-demand interventions because they were easy to access and perform when they were stressed (in the moment) and because it was too hard to predict when they would be stressed in the future. Though on-demand interventions were easy to access, it was noted that it was hard to remember to do them. 33 participants said that they wanted automatic "nudging" by the bot based on their stress levels.

JIT participant feedback. JIT participants (N=43) thought that the JIT interventions were a good reminder to take time out of their day, especially when stressed (N=30). They thought that the interventions were convenient and helpful (N=17). In terms of improvements to the design of the system, JIT participants raised timing and frequency issues: nudges were too frequent and disruptive of focus (they thought that the system should make them easier to ignore). Although they wanted automatic detection and interventions based on stress levels and an overly crowded calendar, they suggested intelligent timing based on their availability and task context. Many JIT participants liked the agency to perform the interventions when they wanted or needed to by using the on-demand feature, which ended up being less disruptive to workflow and kept the users in control (N=18).

System feedback. On a 5-point agreement scale (1=Strongly disagree; 5=Strongly agree), participants in both conditions agreed that the intervention system they used during the study made it easier to engage in interventions compared to before the study ($\bar{x}$=4.07, σ=0.98), that the system was easy to use ($\bar{x}$=4.03, σ=1.13), that they found themselves engaging in more interventions compared to before the study ($\bar{x}$=3.88, σ=1.12), and that the system met their requirements for engaging in interventions ($\bar{x}$=3.76, σ=1.09). Participants in both conditions also agreed that, if given the opportunity, they would continue to use the system ($\bar{x}$=3.38, σ=1.20), but this had the lowest agreement scale among positive usability statements. Participants disagreed that using the system was a frustrating experience ($\bar{x}$=2.37, σ=1.18). JIT participants found that their condition-specific system to be easier to use than PS participants with statistic significance (4.28 vs. 3.79, p<0.05). We found the differences between conditions to be not statistically significant in all other usability ratings. Condition-specific agreement ratings are illustrated in Figure 8.

Intervention type feedback. Generally, participants had the most positive reaction to 'Feel calm and present' and 'Get my mind off work' interventions. Across all participants, 30 participants felt that 'Feel calm and present' interventions were most helpful and 26 felt that 'Get my mind off work' interventions were most helpful in immediate stress reduction; only 4 participants felt that 'Think through my stress' interventions were helpful in immediate stress reduction. 16 and 12 participants felt that 'Feel calm and present' and 'Get my mind off work' interventions had the biggest impact on long-term stress reduction, respectively, and only 9 participants felt that 'Think through my stress' interventions had the biggest impact on long-term stress reduction.

Some participant reactions to specific intervention content were polarized, according to participant open-response feedback: one participant perceived 'Think through my stress' to be "most helpful in channeling [their] energy in a new direction" while another thought that interventions that required introspection, such as those in the 'Think through my stress' category, to be least helpful because "then [they were] stressed about what [they were] writing."; one participants perceived watching videos of nature or an interesting place were "good for getting mind off current activities of the day that were the contributors to stress" while another thought "the ones which required that [they] sit at [their] computer to watch a video" to be least helpful.

Participants indicated an interest in accessing a wide variety of interventions, expressing an overall preference for interventions that varied in their physical environment (e.g., being physically away from the desk vs. doing interventions at the desk), in the level of focus on stress (e.g., think about stress vs. take mind off stress), in the social interactions (e.g., involved other people vs. alone), in familiarity (e.g., surprising and new vs. known and expected), and in effectiveness (e.g., interventions that I benefited from). Participants overwhelmingly wanted intervention content that was simple, easy to do, and required low effort or burden.

5 DISCUSSION

In the four-week, between-subjects study presented herein, we examined the impact of digital micro-intervention *delivery timing* and *content type* on usage patterns and stress reduction throughout the workday for N=86 information workers. Through testing two delivery timing conditions (PS and JIT) and three categories of intervention content type, we showed that digital micro-interventions were effective at reducing momentary stress (stress change from pre- to post-intervention), regardless of intervention engagement timing and content type. Although delivery timing did not have a statistically significant impact on momentary or study-long stress reduction, we did observe noteworthy differences between user perceptions of delivery timing conditions: JIT intervention engagement was perceived as easy to do and motivating, while PS intervention engagement was perceived as nicely customizable to user work schedules. We also found that PS intervention engagement was associated with statistically greater advancement through the stages of behavior change (i.e., advancement towards long-term integration of stress-reduction behaviors into every day life). Lastly, we found that while low-effort, positive distraction interventions were perceived to be more helpful, high-effort, problem-solving interventions were indeed more effective. Understood alongside our qualitative findings, which suggested user preferences for engagement timing and content type are versatile and wide ranging both between people and within individuals over time, we propose stress reduction intervention systems should support both PS and JIT intervention use, and offer a wide variety of content type.

Figure 8: Average agreement on eight statements about the system usability across *Pre-scheduled* (PS) condition in blue and *Just-in-time* (JIT) condition in orange. Agreement was measured using a 5-point scale (1=Strongly disagree; 5=Strongly agree). The error bars indicate 95% confidence intervals across participants. * indicates that there is a statistically significant difference between the two conditions.

5.1 Design Considerations and Research Needs for Workplace Intervention Systems

Extrapolating from our quantitative and qualitative findings, we offer key design suggestions and opportunities for workplace stress-reduction intervention systems.

5.1.1 Integrate digital micro-interventions into the workplace. The present study revealed that using digital micro-interventions throughout the workday reduces momentary stress, and that this effect is present across variation in delivery timing (JIT, PS and on-demand) and content type (low- to high-effort). Digital micro-intervention systems will continue to be optimized for the greatest stress-reduction impact, but even without such refinement, simply having easy access to such interventions may empower employees to reduce their stress levels within a matter of minutes. Further, digital micro-interventions were shown to reduce workplace stress when used as stand-alone individual-level interventions, i.e., as secondary strategies aimed at targeting the individual to reduce stress, without employing primary strategies at the organizational level aimed at eliminating the stressors themselves [13, 49, 50, 55]. Based on these finding, we propose that workplaces provide digital micro-intervention access to their employees as a first-line secondary strategy for reducing employee stress when employers cannot engage with primary strategies or as short-term harm reduction when primary strategies, such as changing fundamental components of an organizational structure, take months or even years to enact. Future research should compare the effectiveness of digital micro-interventions to other secondary strategies, such as longer breaks during the workday. Future research should also investigate whether certain primary strategies, such as overall workload reduction, could multiply the stress-reducing effects of digital micro-interventions and whether digital micro-intervention access could be used as a tertiary strategy for helping employees recover from stress-related mental and physical health issues.

5.1.2 Provide a personalized balance between automation and agency. Our findings suggest that digital micro-intervention systems should offer users multiple levels of control over the timing and content of interventions, from low-control/high-automation options to high-control/low-automation. The majority of participants across both conditions preferred having intervention timing determined by the automated stress detection system (JIT) for ease of use – an opinion based either on lived experience from being assigned to the JIT condition, or on reading a description of the JIT condition after having completed the study in the PS condition. Yet participants also requested concurrent access to interventions on-demand, the ability to pre-schedule interventions at their discretion, and the ability to "snooze" the entire system. Our findings also revealed that the JIT system tested was not sufficiently intelligent for some users due to issues like receiving intervention nudges while busy. Further, participants in the PS condition reported more advancement through the stages of behavior change over the course of the study, compared to those in the JIT condition, with the majority of participants who advanced shifting from Stage 2: Contemplation to Stage 3: Taking action. In other words, our results suggest that, despite JIT being the preferred condition, participating in the PS condition may have shifted users' self-perceptions towards being individuals capable of taking action, while participating in the JIT condition did not change users' self-perceptions. Overall, despite user preferences for JIT interventions and promises of intelligent adaptability and personalization of JITAI systems [42, 52, 57], there were benefits to user-initiated on-demand and pre-scheduled options, especially while JIT system metrics are undergoing refinement. Future research should systematically test various ratios of system automation versus user control and seek to establish whether (a) user-initiated intervention engagement promotes greater advancement through behavior change stages than future iterations of JIT systems with more sophisticated timing algorithms, and (b) which type of intervention engagement – user-versus system-initiated – is the best match for each stage of behavior change [46].

Our study also revealed a parallel user interest in system-selected content. Specifically, participants wanted to be provided with the "right" intervention for the given moment, i.e., an intervention they would *like* and that would address their momentary needs. Participants also indicated an interest in accessing a wide variety of interventions, suggesting that novelty in-and-of-itself may be an important component of user engagement and, secondarily, intervention impact. Systems delivering digital micro-interventions should have the ability to intelligently select interventions for users depending on their momentary needs, including the need for novelty. Future research should test the frequency with which new content should be introduced and implement a content-renewal system at an optimal frequency. These features will likely lead to more sustained user engagement and stress reduction over time.

5.1.3 Promote self-experimentation on intervention content that compares effectiveness and effort. Although incorporating individual preferences has been shown to improve the engagement and outcomes of stress interventions [14, 28, 45, 54], our findings suggest that users may not always be aware of which content helps reduce their stress the most. In our study, the highest-effort interventions we tested ('Think through my stress', i.e. problem solving) had the largest stress reducing effect over the course of the full study, across conditions and baseline characteristics. However, these interventions were selected significantly less frequently than 'Feel calm and present' interventions and were rated significantly lower than 'Get my mind off work' interventions. Future systems should offer users feedback and the opportunity to reflect on their past experiences of stress reduction and content ratings, as this may prompt different and potentially more effective choices when selecting intervention content. For example, users could benefit from a dashboard that summarizes recent trends in self-reported and passively-sensed stress levels, as well as intervention use, impact, and rating history. By exploring past behavior, users could learn about themselves, prompting them to make more informed decisions when choosing interventions in the future.

5.1.4 Solicit user feedback to adapt intervention timing and content. Just as the system providing feedback to users may facilitate their change and growth, users providing feedback to the system can help the system improve. Participant feedback in our study indicated that users are eager to provide suggestions and believe it will improve their user experience. Given the wide range of participant preferences for timing and content and overwhelming need for personalization, this hypothesis is likely correct. Offering users intervention timing and frequency that is "just right" and providing users with personalized content for each user/context pairing will require quite sophisticated system intelligence. Opportunities for the user to train the algorithms to perform optimally will also be necessary. For example, with the help of the aforementioned user dashboard that summarizes trends in stress levels and intervention behavior, users could review their recent activity to identify patterns the system may not otherwise detect. Future research should design and test dynamic assessment and integration of user preferences into sensing and intervention delivery systems.

5.2 Limitations

The present study and associated findings have some important limitations.

Our sample presented with low levels of stress at the beginning of the study. Users of all stress levels can benefit from stress reduction interventions, but interventions are likely to be most effective and beneficial for individuals and organizations when used by individuals with moderate to high stress. As such, interventions should be designed for and tested on a higher stress sample. Additionally, the relative homogeneity of the sample (all information workers, majority engineers and majority male-identifying) limits the generalizability of our findings.

Both JIT and PS participants had access to interventions on-demand, complicating comparisons between the two conditions. Further, PS participants selected their intervention content days in advance, while JIT participants selected content only a moment in advance. Future studies should be designed to clearly separate out the effects of JIT, PS, on-demand, and the duration between content selection and completion. While intervention content was inspired by evidence-based stress reduction strategies and similar to digital micro-interventions tested elsewhere, the particular content had not been tested prior to the present study. Future work should test intervention content and delivery timing separately.

The stress metric employed for the JIT condition was not refined prior to study implementation, and therefore may have prompted interventions at inopportune times. For example, although the stress metric incorporated the number of calendar events per day, it was not capable of distinguishing between work-related and personal events. As personal events could have included self-care activities with stress reducing impacts, the assumption that a greater number of calendar events per day was associated with greater stress may not have been fully accurate. Additional system limitations included: (1) PS participants were required to manually schedule interventions for themselves without the assistance of a calendar integration; and (2) eight participants had to turn off their cameras due to heavy system load, which constrained the stress metric employed for JIT participants.

5.3 Privacy and Ethics

User privacy is a major concern with any application that tracks user behavior. Privacy in the context of work related stress is even more sensitive, since, in a toxic work environment, work stress related concerns can be stigmatized [16]. Hence, privacy regarding tracking stress related data is very sensitive, and it must be well regulated within respective organizations. Note that for inferring stress, we used high level activity data from each participant (e.g., total number of emails in a given window, total number of minutes in meetings, etc.). Such data pose relatively few privacy challenges. Irrespective of the granularity of such data, strong regulations need to be established regarding this data collection. In addition, ethical decisions about when to intervene, the granularity of intervention (e.g., individual level, community level, etc.) and how such interventions align with the individual preferences for receiving interventions needs to be well thought out through user centered design and ethical review boards within our respective communities. We intend to embrace all of these challenges in our efforts to confront workday stress and all of its adverse side effects for information workers.

6 FUTURE WORK AND CONCLUSION

Reducing workplace stress is of critical importance for employees and employers alike. To gather and formulate design recommendations for the development of workplace stress reduction intervention systems, we conducted a four-week longitudinal study testing the delivery timing and content type of digital micro-interventions in a sample of information workers at a large technology company. Our findings showed that digital micro-interventions are effective at reducing short-term workplace stress, suggesting that these interventions should be integrated into workplaces now for immediate, positive impact. Further, our findings suggested that personalization of delivery timing, content type, and balance between user and system control may improve user engagement and stress reduction outcomes. Our findings reveal that people can benefit from experiences that help them understand the effects of interventions on their stress levels and, in doing so, bridge differences between perceived and objective effectiveness. Finally, as noted earlier, despite the large number of academic and industry efforts in this area, significant room for improvement exists. Our study provides one step towards bridging previous work in academia and industry, addressing issues of adherence, empirical testing and intervention tailoring. Future work should strive to further develop tools for personalization, tools for users to reflect on and guide future intervention usage, and opportunities for users to offer continual feedback for ongoing system improvement.

REFERENCES

[1] Amit Baumel, Theresa Fleming, and Stephen M Schueller. 2020. Digital micro interventions for behavioral and mental health gains: core components and conceptualization of digital micro intervention care. *Journal of medical Internet research* 22, 10 (2020), e20631.

[2] Yoav Benjamini and Yosef Hochberg. 1995. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal statistical society: series B (Methodological)* 57, 1 (1995), 289–300.

[3] George A Bonanno and Charles L Burton. 2013. Regulatory flexibility: An individual differences perspective on coping and emotion regulation. *Perspectives on Psychological Science* 8, 6 (2013), 591–612.

[4] Frank W Bond and David Bunce. 2000. Mediators of change in emotion-focused and problem-focused worksite stress management interventions. *Journal of occupational health psychology* 5, 1 (2000), 156.

[5] 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.

[6] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[7] Rob B Briner. 1997. Improving stress assessment: Toward an evidence-based approach to organizational stress interventions. *Journal of Psychosomatic Research* 43, 1 (1997), 61–71.

[8] Elizabeth Brondolo, Kahaema Byer, Peter J. Gianaros, Cindy Liu, Aric A. Prather, Kamala Thomas, and Cheryl L. Woods-Giscombé. 2017. APA Working Group on Stress and Health Disparities.(2017). *Stress and Health Disparities: Contexts, Mechanisms, and Interventions among Racial/Ethnic Minority and Low-Socioeconomic Status Populations. Retrieved from http://www.apa.org/pi/health-disparities/resources/stress-report.aspx* (2017).

[9] 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.

[10] Colleen E Carney, Daniel J Buysse, Sonia Ancoli-Israel, Jack D Edinger, Andrew D Krystal, Kenneth L Lichstein, and Charles M Morin. 2012. The consensus sleep diary: standardizing prospective sleep self-monitoring. *Sleep* 35, 2 (2012), 287–302.

[11] Rachel A Coleman and Robert A Neimeyer. 2015. Assessment of subjective client agency in psychotherapy: A review. *Journal of constructivist psychology* 28, 1 (2015), 1–23.

[12] Thomas W Colligan and Eileen M Higgins. 2006. Workplace stress: Etiology and consequences. *Journal of workplace behavioral health* 21, 2 (2006), 89–97.

[13] Cary L Cooper and Sue Cartwright. 1997. An intervention strategy for workplace stress. *Journal of psychosomatic research* 43, 1 (1997), 7–16.

[14] Gavin Doherty, David Coyle, and John Sharry. 2012. Engagement With Online Mental Health Interventions: An Exploratory Clinical Study of a Treatment for Depression. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* https://doi.org/10.1145/2207676.2208602

[15] Maureen F Dollard. 2012. *Psychosocial safety climate: A lead indicator of workplace psychological health and engagement and a precursor to intervention success.* Routledge.

[16] Daniel Eisenberg, Marilyn F Downs, Ezra Golberstein, and Kara Zivin. 2009. Stigma and help seeking for mental health among college students. *Medical Care Research and Review* 66, 5 (2009), 522–541.

[17] Rosenberg Ekman. 1997. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS).* Oxford University Press, USA.

[18] Gunther Eysenbach. 2005. The law of attrition. *Journal of medical Internet research* 7, 1 (2005), e402.

[19] James J Gross and Oliver P John. 2003. Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. *Journal of personality and social psychology* 85, 2 (2003), 348.

[20] Joseph G Grzywacz, David M Almeida, and Daniel A McDonald. 2002. Work–family spillover and daily reports of work and family stress in the adult labor force. *Family relations* 51, 1 (2002), 28–36.

[21] Sylvia Helbig and Lydia Fehm. 2004. Problems with homework in CBT: Rare exception or rather frequent? *Behavioural and cognitive psychotherapy* 32, 3 (2004), 291–301.

[22] Susana García Herrero, Miguel Ángel Mariscal Saldaña, Javier García Rodriguez, and Dale O Ritzel. 2012. Influence of task demands on occupational stress: Gender differences. *Journal of Safety Research* 43, 5-6 (2012), 365–374.

[23] Nis Hjortskov, Dag Rissén, Anne Katrine Blangsted, Nils Fallentin, Ulf Lundberg, and Karen Søgaard. 2004. The effect of mental stress on heart rate variability and blood pressure during computer work. *European journal of applied physiology* 92, 1 (2004), 84–89.

[24] Anna Huguet, Sanjay Rao, Patrick J Mcgrath, Lori Wozney, Mike Wheaton, Jill Conrod, and Sharlene Rozario. 2016. A Systematic Review of Cognitive Behavioral Therapy and Behavioral Activation Apps for Depression. *PLOS One* 11, 5 (2016), E0154248. https://doi.org/10.1371/journal.pone.0154248

[25] Ronald M Kadden and Mark D Litt. 2011. The role of self-efficacy in the treatment of substance use disorders. *Addictive behaviors* 36, 12 (2011), 1120–1126.

[26] Karim Sadik Kassam. 2010. *Assessment of emotional experience through facial expression.* Harvard University.

[27] Gail Kinman and Fiona Jones. 2005. Lay representations of workplace stress: What do people really mean when they say they are stressed? *Work & stress* 19, 2 (2005), 101–120.

[28] Artie Konrad, Victoria Bellotti, Nicole Crenshaw, Simon Tucker, Les Nelson, Honglu Du, Peter Pirolli, and Steve Whittaker. 2015. Finding the adaptive sweet spot: Balancing compliance and achievement in automated stress reduction. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* 3829–3838.

[29] Eric Kuhn, Nitya Kanuri, Julia E Hoffman, Donn W Garvert, Josef I Ruzek, and C Barr Taylor. 2017. A randomized controlled trial of a smartphone app for posttraumatic stress disorder symptoms. *Journal of consulting and clinical psychology* 85, 3 (2017), 267.

[30] Richard S Lazarus. 2020. Psychological stress in the workplace. In *Occupational stress.* CRC Press, 3–14.

[31] Jake Linardon and Matthew Fuller-Tyszkiewicz. 2020. Attrition and adherence in smartphone-delivered interventions for mental health problems: A systematic and meta-analytic review. *Journal of consulting and clinical psychology* 88, 1 (2020), 1.

[32] Oliver Lindhiem, Charles B Bennett, Dana Rosen, and Jennifer Silk. 2015. Mobile technology boosts the effectiveness of psychotherapy and behavioral interventions: a meta-analysis. *Behavior modification* 39, 6 (2015), 785–804.

[33] Marsha Linehan. 2014. *DBT Skills training manual.* Guilford Publications.

[34] Xin Liu, Josh Fromm, Shwetak Patel, and Daniel McDuff. 2020. Multi-Task Temporal Shift Attention Networks for On-Device Contactless Vitals Measurement. *NeurIPS* (2020).

[35] 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.

[36] Anita Lungu and Marsha M Linehan. 2016. Dialectical behavior therapy: A comprehensive multi-and transdiagnostic intervention. *The Oxford handbook of cognitive and behavioral therapies* (2016), 200–214.

[37] Gloria Mark, Shamsi Iqbal, and Mary Czerwinski. 2017. How blocking distractions affects workplace focus and productivity. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. 928–934.

[38] Gloria Mark, Shamsi T Iqbal, Mary Czerwinski, Paul Johns, Akane Sano, and Yuliya Lutchyn. 2016. Email duration, batching and self-interruption: Patterns of email use on productivity and stress. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. ACM, 1717–1728.

[39] Gloria Mark, Yiran Wang, and Melissa Niiya. 2014. Stress and multitasking in everyday college life: an empirical study of online activity. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 41–50.

[40] Leah M Mayo and Markus Heilig. 2019. In the face of stress: Interpreting individual differences in stress-induced facial expressions. *Neurobiology of stress* 10 (2019), 100166.

[41] Daniel McDuff, Eunice Jun, Kael Rowan, and Mary Czerwinski. 2019. Longitudinal Observational Evidence of the Impact of Emotion Regulation Strategies on Affective Expression. *IEEE Transactions on Affective Computing* (2019).

[42] Inbal Nahum-Shani, Shawna N Smith, Bonnie J Spring, Linda M Collins, Katie Witkiewitz, Ambuj Tewari, and Susan A Murphy. 2018. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine* 52, 6 (2018), 446–462.

[43] Ron Neal and Joanne Sibug. [n.d.]. Workplace Stress on the Rise With 83% of Americans Frazzled by Something at Work. https://www.globenewswire.com/news-release/2013/04/09/536945/10027728/en/Workplace-Stress-on-the-Rise-With-83-of-Americans-Frazzled-by-Something-at-Work.html (Accessed 2021-08-20).

[44] Martha Neary and Stephen M Schueller. 2018. State of the field of mental health apps. *Cognitive and Behavioral Practice* 25, 4 (2018), 531–537.

[45] Pablo Paredes, Ran Gilad-Bachrach, Mary Czerwinski, Asta Roseway, Kael Rowan, and Javier Hernandez. 2014. PopTherapy: Coping with stress through pop-culture. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare*. 109–117.

[46] James O Prochaska and Wayne F Velicer. 1997. The transtheoretical model of health behavior change. *American journal of health promotion* 12, 1 (1997), 38–48.

[47] Candace M Raio, Temidayo A Orederu, Laura Palazzolo, Ashley A Shurick, and Elizabeth A Phelps. 2013. Cognitive emotion regulation fails the stress test. *Proceedings of the National Academy of Sciences* 110, 37 (2013), 15139–15144.

[48] Julie Ray. 2019. Americans' Stress, Worry and Anger Intensified in 2018. https://news.gallup.com/poll/249098/americans-stress-worry-anger-intensified-2018.aspx (Accessed 2021-08-20). , 207–209 pages.

[49] Shirley Reynolds. 1997. Psychological well-being at work: is prevention better than cure? *Journal of psychosomatic research* 43, 1 (1997), 93–102.

[50] Katherine M Richardson and Hannah R Rothstein. 2008. Effects of occupational stress management intervention programs: a meta-analysis. *Journal of occupational health psychology* 13, 1 (2008), 69.

[51] 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. https://doi.org/10.1109/ACII.2015.7344623

[52] 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.

[53] Jessica Schroeder, Jina Suh, Chelsey Wilks, Mary Czerwinski, Sean A Munson, James Fogarty, and Tim Althoff. 2020. Data-Driven Implications for Translating Evidence-Based Psychotherapies into Technology-Delivered Interventions. In *Proceedings of the 14th EAI International Conference on Pervasive Computing Technologies for Healthcare*. 274–287.

[54] Stephen M Schueller. 2010. Preferences for positive psychology exercises. *The Journal of Positive Psychology* 5, 3 (2010), 192–203.

[55] Stuart D Sidle. 2008. Workplace stress management interventions: What works best? *Academy of Management Perspectives* 22, 3 (2008), 111–112.

[56] Joshua M Smyth and Kristin E Heron. 2016. Is providing mobile interventions "just-in-time" helpful? An experimental proof of concept study of just-in-time intervention for stress management. In *2016 IEEE Wireless Health (WH)*. IEEE, 1–7.

[57] Donna Spruijt-Metz and Wendy Nilsen. 2014. Dynamic models of behavior for just-in-time adaptive interventions. *IEEE Pervasive Computing* 13, 3 (2014), 13–17.

[58] Victor J Strecher, Brenda McEvoy DeVellis, Marshall H Becker, and Irwin M Rosenstock. 1986. The role of self-efficacy in achieving health behavior change. *Health education quarterly* 13, 1 (1986), 73–92.

[59] Sujit Sur and Eddy S Ng. 2014. Extending theory on job stress: The interaction between the "other 3" and "big 5" personality traits on job stress. *Human Resource Development Review* 13, 1 (2014), 79–101.

[60] Tanja GM Vrijkotte, Lorenz JP Van Doornen, and Eco JC De Geus. 2000. Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability. *Hypertension* 35, 4 (2000), 880–886.

[61] Sarah B Watson, Yong Wah Goh, and Sukanlaya Sawang. 2011. Gender influences on the work-related stress-coping process. *Journal of Individual Differences* (2011).

[62] Amy Wenzel. 2017. Basic strategies of cognitive behavioral therapy. *Psychiatric Clinics of North America* (2017).

[63] Mina Westman and Dalia Etzion. 2001. The impact of vacation and job stress on burnout and absenteeism. *Psychology & Health* 16, 5 (2001), 595–606.