Exam Eustress: Designing Brief Online Interventions for Helping Students Identify Positive Aspects of Stress

Mohi Reza
University of Toronto
Toronto, Canada
mohireza@cs.toronto.edu

Angela Zavaleta Bernuy
University of Toronto
Toronto, Canada
angelazb@cs.toronto.edu

Emmy Liu
Carnegie Mellon University
Pittsburgh, USA
mengyan3@andrew.cmu.edu

Tong Li
University of Toronto
Toronto, Canada
tongli@cs.toronto.edu

Zhongyuan Liang
University of Toronto
Toronto, Canada
zhongyuan.liang@mail.utoronto.ca

Calista Barber
University of Toronto
Toronto, Canada
calista.barber@mail.utoronto.ca

Joseph Jay Williams
University of Toronto
Toronto, Canada
williams@cs.toronto.edu

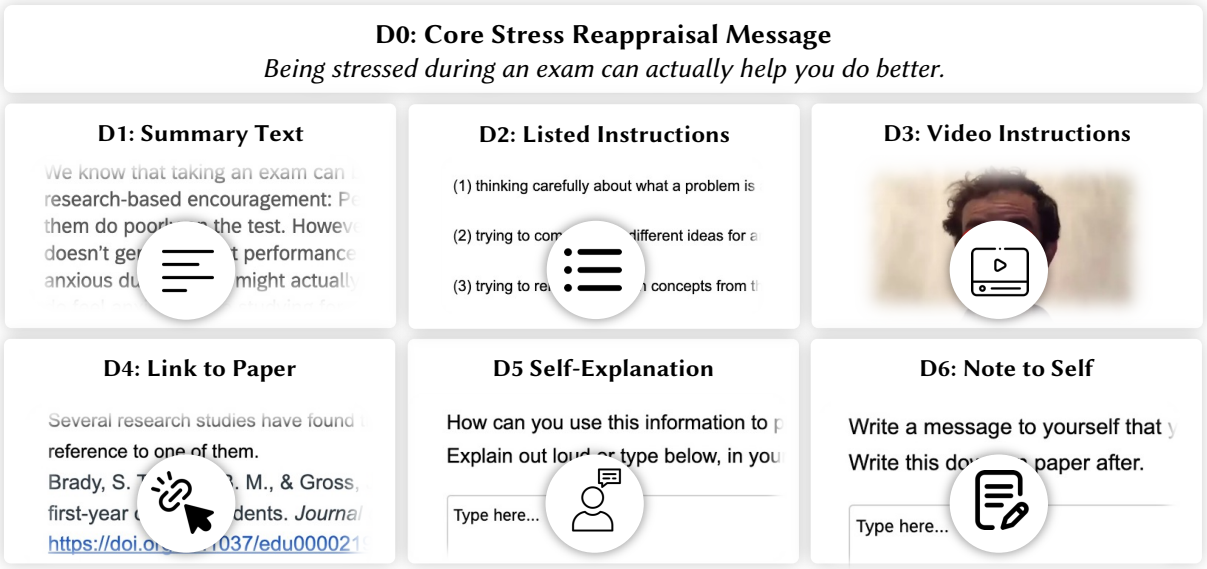

Figure 1: Our design space for exam eustress interventions consists of a core stress reappraisal message (D0) reinforced by 6 design factors: (i) D1 offers explanatory context for reappraisal in paragraph form, (ii) D2 gives explicit suggestions for what to do during exams, why stress could help, and how to use this information, (iii) D3 includes a talking-head video from an instructor explaining the explicit suggestions (iv) D4 provides a citation and link to a research paper, (v) D5 prompts students to self-explain the concept by typing or voice, (vi) D6 prompts students to write a note that they could revisit before the exam.

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04…$15.00
https://doi.org/10.1145/3544548.3581368

ABSTRACT

Stress reappraisal interventions try to shift students' negative perceptions towards *eustress*, stress that can be beneficial, and help them perform better. However, it is less clear how to present them to users as online interventions that are brief, voluntary, and scale well in real-world contexts. We explore the design of online *exam eustress* interventions by generating six design factors (D1-6) that reinforce a core reappraisal message (D0), and evaluate them through: (i) user interviews ($N = 20$) revealing six findings (F1-6) on the importance

of elaboration, layout, modality, and source of intervention content; (ii) a field experiment ($N = 1283$) showing a significant positive effect on exam scores ($p = 0.003$). Subgroup analyses indicate a significant effect for first-year but not for upper-year students, and no detectable gender differences. Our work offers insight into how students interact with online mindset interventions and design considerations for incorporating them into large courses.

CCS CONCEPTS

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

KEYWORDS

stress-reappraisal, eustress, exams, randomized field experiments, online interventions, user interviews

ACM Reference Format:

Mohi Reza, Angela Zavaleta Bernuy, Emmy Liu, Tong Li, Zhongyuan Liang, Calista Barber, and Joseph Jay Williams. 2023. Exam Eustress: Designing Brief Online Interventions for Helping Students Identify Positive Aspects of Stress. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3544548.3581368

1 INTRODUCTION

Stress has become increasingly prevalent in higher education, and exams are a key stressor[42]. While COVID-19 has further exacerbated the stress levels of students [44], it has also accustomed them to using digital learning tools [15], creating opportunities for instructors to leverage online interventions. These interventions can be used to help students not only reduce stress but also to find a way to use stress as a resource for handling challenging situations such as exams. Students often interpret normal physiological reactions to stress as wholly harmful. These negative perceptions about stress and its effects on test performance can lead to increased *distress*.

Introducing students to the concept of *eustress*, i.e. a positive psychological response to a stressor, has the potential to help them reappraise stress as not necessarily something to be eliminated but something to be embraced as a signal that the body is trying to help them focus and perform better. There is increasing evidence that this kind of messaging can benefit students in exam situations, such as when they are required to read some materials [27] or when they receive information right before the exam [6]. In this paper, we explore how to provide online activities that support students by helping them change how they think about stress's benefits.

For this exploration to be constructive, we must be aware of the many design constraints and considerations in developing online activities for students that communicate reappraisal messaging in an actionable way. In this paper, we explore the design space for activities that are **simple** (single webpage), **brief** (takes less than 5 minutes), **scalable** (requires no instructor feedback), and equip students with the ability to apply this information to their lives through self-reflecting on how to leverage eustress during exams. We think these particular elements are essential because they make online interventions feasible and impactful in real-world contexts, where students may have limited attention, and instructors are often overloaded with limited capacity for one-on-one interactions, despite wanting to help students.

Within this challenging context, we explored the design of multi-component online interventions by varying several complementary factors to reinforce engagement with a core stress reappraisal message – being stressed during an exam can help students do better. We varied the factors in terms of three design dimensions: (i) the *amount of information* and explanatory text (balancing the trade-off of limited attention and visual clutter against the degree of elaboration on how stress can be helpful), (ii) the *modality of presentation* (text versus video), and (iii) the potential value of *reflection prompts* for students to think aloud and type notes on how their future selves could use the information during their next exam.

We evaluated the range of design components through semi-structured interviews with 20 students and a randomized field experiment deployed to 1283 students in a programming course. This provided insights into the relative importance of different factors and their impact on different students in different contexts, such as when text versus video presentation might be effective, what kinds of additional information are compelling versus redundant, and what kinds of interface prompts for students are more or less impactful in terms of helping students retain and utilize the reappraisal messaging. Although the activity was delivered via a brief and optional online intervention that took an average of under 3 minutes, it positively impacted student performance on a test, with a significant effect on first-year students but not upper-year students. Our findings provide insight into how instructors and others can design online mindset interventions based on contextual features of the interface and student characteristics. We also offer design directions for a range of potential future work in the better digital delivery of mindset interventions. The main contributions of this work are:

(1) An exploration of the design space for online exam eustress interventions that are simple, brief, and scalable, focusing particularly on how students interact with different presentation modalities (text, video, or both), levels of elaboration (explanation, instruction, or paper citation), and reflection prompts to learn how to identify beneficial aspects of stress during exams. Our methodology can also be applied to design other scalable stress-management interventions.
(2) An evaluation of our multi-component design using semi-structured interviews with 20 students, as well as a large-scale online field experiment with over 1000 students providing evidence on the positive impact of our intervention on exam performance and insight into how students interact with various components described in figure 1.

2 RELATED WORK

2.1 Psychological Research into Practical Technology-Mediated Interventions

There is accumulating evidence that, in specific contexts, stress reappraisal strategies can be effective for managing feelings of worry about anxiety [23, 28] and enhancing performance in situations of acute stress [28], such as exams [6, 26, 29, 35]. Past research has shown examples where videos [10], emails [6], and text instructions can be better than no reappraisal. However, it

remains unclear how to combine them into brief online interventions because there has been little direct comparison between such presentation modalities. We explore the design space of different modalities, explicitly comparing video, text, and a combination of video and text. Given how important it is to preserve student attention and to empower instructors to help more students without becoming overloaded, we focus our exploration on voluntary, brief and scalable intervention combinations. While it is particularly challenging to design interface components that meet all of these three criteria, we think they are essential because interventions that meet them have the potential to translate psychological research into *practical* technology-mediated interventions that can positively impact thousands of students.

2.2 HCI Research on Prompting People to Change their Beliefs and Attitudes

We contribute to the growing body of existing work within HCI that focuses on leveraging online technology to promote positive belief change at scale by designing systems that apply various behavioural science approaches to real-world problems such as mental health treatment using online chat tools [3, 40] and conversational coaches [37], engaging users to reflect on physical activity [33], cognitive behavioural therapy (CBT) [17] based apps to reduce depression and anxiety [5, 21], and just-in-time (JIT) based interventions [22] to reduce digital workplace stress [24]. In this paper, we present a design exploration of how different content modalities can be used to communicate stress-reappraisal messaging to students in large classrooms effectively. We give qualitative insight into how students interact with different modalities to alter their stress mindset and quantitative evidence showing that our designs improve exam performance in a large-scale field experiment.

2.3 Shifting from Designing *Distress* Interventions to *Eustress* Interventions

Current HCI research in the online intervention space has focused primarily on applying *reduction-based* coping strategies to manage stress instead of trying to help users embrace positive *eustress*[34]. We believe this reflects the dominant historical orientation of past research toward negative aspects of stress. For example, previous studies [9, 11] have demonstrated that prolonged stress lowered overall academic performance [9, 11], as indicated by low exam scores and low overall grade point average. These studies suggest that test anxiety, a common source of stress, is a cause of poor academic performance as it negatively affects critical factors that affect the learning process, such as sleep and biological systems that mediate the body's responses to stress. However, building upon pioneering work from the 1980s [19], a new class of studies has hinted at the multiple potential benefits stress can have on academic performance when used methodically [2, 4, 36, 45]. Compared to most existing HCI contributions in the stress-management space, one contrasting feature of our intervention is that it focuses on reappraisal rather than stress reduction.

Our decision to focus on the positive aspects of stress is informed by previous studies on emotional regulation that have compared the use of reappraisals for stress management with other strategies such as suppression [23, 32], self-distraction [16, 30], and acceptance [20, 23, 48]. These studies suggest that reappraisal is a more effective strategy than suppression and acceptance for moderating physiological arousal and the subjective feeling of stress [23]. The reappraisal strategy can also help reduce the negative emotional experience of stress [16, 35]. Both reappraisal and self-distraction may be effective for attenuating emotional reactions [30]; however, we choose reappraisal for our intervention as we aim to measure its effectiveness in an exam setting.

2.4 Situating Design Factors in Prior Work

To generate our six design factors, we draw guidance from research on multimodal information presentation in HCI [8, 49], effective communication strategies from market research [39], instructional design [14, 31], multimedia learning [38] and reflective learning [47]. Convincing and concise reporting often requires incorporating data from multiple sources, methods, and modalities [8, 38, 39]. To leverage this approach, we layer design factors on top of each other to reinforce a core reappraisal message. Literature on instructional design and multimedia learning informs our choice to test different content layouts (e.g. paragraphs and bullet points) and mediums (e.g. text and video). Prior work has shown the potential instructional benefits of spoken words in videos [31], how bulleted lists may improve information retention [1, 25], and how mixing modalities could improve learning by offering learners parallel opportunities for information processing [50]. One of the biggest design challenges is how to help learners retain the reappraisal information and potentially change their behaviour on the upcoming exam. To address this, we turned to prior work on reflective learning via writing and voice, which has shown the success of reflection prompts in guiding future behaviour. We incorporate these insights in the final part of our intervention by designing two reflection prompts for assisting recall [41] and behaviour change [18, 47].

3 THE DESIGN SPACE FOR ONLINE EXAM EUSTRESS INTERVENTIONS

3.1 Design Constraints and Considerations

We want to use online interventions to impact students' behaviour by helping them rethink exam stress as being useful rather than detrimental. We consider the constraints posed by brief online webpage interfaces. On one hand, such interfaces are ideal for technology-mediated online intervention because students can easily access them via their computer or smartphone browsers. However, for an intervention embedded within such a setting to be effective, it needs to be (i) *voluntary*, because students don't *have* to do them, and (ii) *scalable*, so it can be sent to many students without requiring instructor intervention. These two constraints are considerably challenging to meet due to the counter-intuitive nature of the reappraisal message. One could imagine that reappraisal messaging is best delivered in person, by someone talking to the students, explaining the idea to them, sharing stories, and asking questions.

To investigate the design of components of a brief, simple, scalable, online intervention interface that can be self-administered by students, we considered content modalities and information that target three elements of users' cognition and behaviour:

(1) Information presentation that communicates the message effectively to students.
(2) Information content that engages them in deeper processing to understand the information.
(3) Prompts them to consider specific actions so they would be more likely to remember the ideas in the future.

A key guiding consideration is that students' attention in online environments may be incredibly limited and that instructors are often overloaded and unable to offer one-on-one in-person interactions with students at scale. Therefore we asked what the considerations are in deciding what information and activities merit being included to convey a reappraisal message that is effective for different students in varying contexts, which accounts for the trade-off between limited student attention and offers a thorough explanation, and results in students internalizing the message.

3.2 The Six Design Factors

In section 2.4, we discussed how our intervention design factors drew insights from prior work on multimodal information presentation in HCI [8, 49], effective communication strategies drawn from market research [39], instructional design [14, 31], multimedia learning [38] and reflective learning [47]. In this section, we revisit some of those insights and describe the specifics of our design and our rationale behind each factor.

D0: Core Stress Reappraisal Message. All students received the following core message:

> *"Being stressed during an exam can actually help you do better."*

We suspected that this brief message alone would not be sufficient in convincing students that stress can be helpful, which was later confirmed in our user interviews (see F1 in section 4.4). An effective communication strategy from market research involves the integration of multiple sources and methods into well-synthesized content [39]. We adopted this strategy and explored ways to reinforce D0 using six design factors (D1-6).

D1: Explanatory Elaboration Providing Research-based Rationale and Encouragement. We provided additional explanatory context by laying out the specific logic of reappraisal to understand how important this information was to students, and which aspects of this elaboration were useful in different contexts. Furthermore, we wanted to evaluate whether framing the stress reappraisal concept as being *research-based* would convince students, in alignment with science communication literature on conveying accurate scientific information through persuasive scientific narratives [12].

The text that could be included or not was:

> *"We know that taking an exam can be a stressful experience, and so we wanted to provide a note of research-based encouragement: People think that feeling anxious while taking a test will make them do poorly on the test. However, recent research suggests that increased levels of stress doesn't generally hurt performance on tests and can even help performance. People who feel anxious during a test might actually do better. This means that you shouldn't feel concerned if you do feel anxious while studying for or taking the upcoming exam. If you find yourself feeling anxious, simply remind yourself that your stress and its higher energy could be helping you do well."*

D2: Explicit Suggestions for What to Think During Exams, Why Stress Could Help, and How to Use the Information. While offering explanatory elaboration may help convince students, we wondered if they needed explicit suggestions on what to do. Our rationale was to evaluate whether it was helpful to guide students through directed prompts [13, 46] containing specific instructions and explanations on how to reappraise stress. We also wanted to see if this information was redundant or even potentially unhelpful as it could reduce student attention and dilute the impact.

> *"During your exam, try to remember that feeling stressed might actually help you perform better, by making you more alert, and helping you work harder.*
>
> *Try to use the feeling of stress as a cue, to put energy into: (1) thinking carefully about what a problem is asking you, (2) trying to come up with different ideas for answering questions, and (3) trying to remember which concepts from the semester are relevant.*
>
> *How might stress help you do better on an exam? Your brain is recruiting resources to make you pay attention, so that you can have more energy to work hard and think deeply.*
>
> *If you find yourself feeling stressed during the exam, remind yourself that this is normal, and not necessarily bad – it may even be helping you do better than if you weren't stressed."*

In light of prior studies showing the benefits of spoken words in instructional videos [31] and the potential to improve information retention and learning through mixing modalities [41], we wondered whether varying how the explicit suggestions were presented to learners, either as text or video, would be helpful in our context.

D3: Presentation by Instructional Video. We developed an instructional talking-head video where a faculty member researcher explained the concept in an enthusiastic manner in order to explore how students' preferences for the modality varied and if there was value in both. For simplicity of presentation, we chose to do it for just the explicit suggestions from D3.

D4: Validation by Explicit Citation of a Source Paper. Providing explicit evidence that research studies support stress reappraisal, versus simply giving the explanatory elaboration of why this could be useful.

> *"Several research studies have found that sharing messages like this can help people do better on exams – here is the reference to one of them.*
>
> *Brady, S. T., Hard, B. M., & Gross, J. J. (2018). Reappraising test anxiety increases academic performance of first-year college students. Journal of Educational Psychology, 110(3), 395–406. https://doi.org/10.1037/edu0000219"*

Finally, to tackle the considerable challenge of helping students retain and apply the intervention information during their exam, we turned to the literature on behaviour change through reflective learning [18, 47] and designed two prompts that encouraged students to consider how to apply this information moving forward.

D5: Prompt to Reflect on How to Use the Information During the Exam. Self-monitoring or generic prompts encourage students to reflect and produce more coherent ideas than directed prompts [13]. There have also been studies showing the benefits of self-explaining through voice or writing [41]. Therefore, we designed a self-monitoring prompt where students were told they could talk out loud, type, or both.

> *"How can you use this information to perform well on your exam? Explain out loud or type below, in your own words"*

Our goal was to prompt students to deepen their understanding of the information by being active instead of passive.

D6: Prompts to Type out a Message to Look at Before the Exam and Write It on Paper. To maximize information retention and chances of behaviour change during the exam, we asked students to write a message that they could revisit before the exam.

> *"Write a message to yourself that you can look at right before the exam, as a reminder of how to use this information. Write this down on paper after."*

We wanted to help students visualize what they could say to themselves before the exam to further increase the chances they remember the reappraisal message. It was less important whether the students actually wrote it down.

These factors explored a complex set of design components, but suggest many directions for other components that could be explored in future research. We discuss some of these directions in section 6.3. Each factor was presented in the same order as their names, i.e. D0 to D6. We chose this order because it illustrates the design principles from prior work as a logical structure, i.e. having the core reappraisal message at the very beginning, augmented by some explanatory text specifying the underlying logic (or not), some concrete steps for using the knowledge (or not), a video on those steps (or not), a paper citation for added credibility (or not), and finally, the two reflection prompts (or not). To evaluate our design, we conducted two studies, as discussed in sections 4 and 5.

4 STUDY 1: USER PERSPECTIVES ON INTERVENTION COMPONENTS

The first study was an exploration of user perspectives to characterize the impact and trade-offs of including particular components of the design. We conducted interviews with students where we showed all the different components of the six design factors of our online intervention interface. We asked users to reflect on each component and share what they thought or felt as they read it, to better understand the impact and trade-offs of including a particular component of the design.

4.1 Participants

Our participants consisted of 20 students (14 women and 6 men) aged 18 and above, from a large and diverse introductory programming course. We recruited them through a call for participation via email to students who previously expressed interest in research activities. To gather diverse perspectives, we recruited students from different years and disciplines. In terms of their year in the program, ten participants (P2-6, P8-12) were starting their second year, six were going into their 3rd year (P1, 7, 13, 14, 17, 19), three were in their 4th year (P8, 16, 18), and one just graduated (P15). Participants were studying a range of subjects including Accounting, Actuarial Science, Biology, Computer Science, Economics, Mathematics, Physics, Physiology, and Statistics. Their diverse academic backgrounds sparked rich, wide-ranging conversations about exam stress that were not tied to viewpoints from a specific discipline or year in the program.

4.2 Procedure

Participants attended the interview using online video conferencing software. They completed a consent form and gave us permission to record the session before starting. Each interview lasted between 45 and 60 minutes, and participants were compensated 15$/hr for their time. The interviewer started with a brief conversation about the participants' past experience with exam stress to understand any pre-existing notions they may have about whether stress is good, bad or both. After this conversation, the participant was asked to share their screen, and to go through every component (D0-6)–first individually on separate webpages, so we could understand how they felt about each component in isolation, and then all on one webpage, so we could understand how they compared components with each other. When going through the components individually, we counterbalanced the order in which participants saw the text (D2) and the video (D3) because we were interested in understanding subjective user impressions immediately after seeing each factor. As the participant went through the activity, the interviewer asked them to think aloud and share any thoughts and feelings as they arose. A silent notetaker was also present during the interviews to write down observations.

4.3 Analysis

Our qualitative data consisted of interview transcripts that were generated from the recordings, observation notes, and video recordings. We coded and analyzed this data using reflexive thematic analysis [7] through an inductive lens, drawing from the rich theory on stress mindset and reappraisal as a pre-existing code.

4.4 Findings

Our findings (F1-F6) indicate a clear need for reinforcing the core reappraisal message presented in D0 with the six design factors (D1-6). We order them considering their prevalence in our data and our judgement on their importance.

F1: A Short Stress-reappraisal Message on Its Own Is Not Sufficient Enough to Convince Users.

> *"D0 is just a small sentence, I think. It is just conveying a conclusion but we don't know how this conclusion came about. I don't think that it is a good choice to convey this idea on its own." – P2*

Based on comments from several participants (P1-4, 6, 8, 10-12), we found that D0 alone is not sufficient for convincing users that stress can be beneficial to them because user agreement with the brief stress reappraisal message in D0 largely depended on past experience with stress. P1 and P2 noted how D0 was too short and did not offer an explanation of how it reached the conclusion that

stress can be helpful during exams. P3 and P4 mentioned that they mostly agreed with D0 because of their past positive experience with stress. F1 is further exemplified through comments from those who did not initially agree with D0, such as P8, 10-12. After expressing disagreement, they described how past exams where they experienced stress did not go well.

We also found that users could potentially misinterpret the message in D0 when presented alone because of how brief and counter-intuitive it was. After reading the message, P6 said that they found it a "bit weird because everyone would experience stress during the exam" and "with the way D0 is framed, it's like we should try to eliminate stress entirely". The latter half of the comment clearly shows that the user took away the exact opposite message to the one we were trying to convey, i.e. not to eliminate stress but rather to embrace it as being helpful. Furthermore, users also appeared to have missed the point about timing. The stress reappraisal message referred to stress *during* the exam. However, among those who said they agreed with the message, such as P3, further discussion on why they agreed with it revealed that they were thinking of stress before the exam and how that stress helped them prepare better.

F2: Users Find Content That Is Described as Being Research-based More Convincing.

> *"I think stress and anxiety are different, but having read D1...it will be more convincing...because there is research showing that this idea, this conclusion about stress is not something that people just imagined, but did research on to come to the conclusion." – P2*

A noteworthy element of the wording in D1 that stood out to participants was that we mentioned how the suggestions contained in the message were *research-based*.

As exemplified by P2's comment above, participants felt convinced by D1 because we mentioned that it was research-based. P2 felt convinced by D1 because it "summarizes research about stress during exams". Similarly, P1 said "knowing that this is research-based…I don't think I would feel like it would be a bad thing". They even mentioned, "I think I would be grateful for a little bit of stress when writing my exams". P8 mentioned that their thoughts on anxiety changed after seeing D1 "because research suggests that feeling anxious during your exam could lead to better results actually. So I am kind of questioning my actual personal experience". P10 found D1 to be a "definitely reassuring text" and P12 concurred, saying the "whole paragraph was definitely comforting, especially since it's research-based encouragement".

F3: Users Are Unlikely to Click on a Citation Link but They Value the Added Credibility.

> *"I would not click on links to research articles but maybe news articles, for example, Globe and Mail..." – P1*

Participants expressed mixed reactions to being presented with a link to a research article (D4). Most of our interviewees were in their second year of undergraduate studies, and likely had limited exposure to reading research articles. While some students such as P4 observed that "it's an interesting study and it's really recent from 2018" and were able to summarize the key takeaways after quickly skimming through the abstract, several participants such as P1, 2 and 11 appeared confused by the complexity of the writing. P1 noted that they would prefer to read news articles: "I think I would not click on links to research articles but maybe news articles, for example, Globe and Mail…as long as they have a citation at the bottom that links to something credible". The last part of their comment indicates that even if they may not necessarily want to read a research article, they value having a citation link.

F4: Structuring Content as Listed Instructions May Help with Recall.

> *"I will remember D2 because you have (1), (2), (3)...and with numbers, it's easy to remember. So this would be much more useful for me during the exam." – P5*

D2 had some notable structural differences compared to D1, such as being separated into multiple paragraphs (despite being roughly the same overall length) and having concrete instructions listed as numbered points. We found that participants such as P5, 6 and 12, noticed these structural differences and commented on them without being explicitly prompted. P6 said "I do like how it's put into these strategic points" and shared that it's harder to remember things when you are stressed. P5 mentioned that having ideas as a numbered list may help them remember the points during exams.

F5: Stressed or Lonely Users May Find Comfort in a Short Talking-head Instructional Video.

> *"...when you are feeling extremely stressed, you don't even have the mind to read a paragraph." – P8*

We wanted to see whether the presentation modality (i.e. video vs text) made a meaningful difference with regard to how participants perceived the message. In D3, a faculty member read the instructions from D2 in a talking-head video. P8 observed that "when you're feeling extremely stressed, you don't even have the mind to read a paragraph", and that "you can always finish a video that's only 40 seconds", suggesting that short videos may be easier to consume compared to text when students are experiencing high levels of stress. They also commented on how if the message is "from your professor or examiner, you will definitely watch a comforting message like this" leading to "a better effect compared to any text-formed message". Adding to P8's observation, P10 noted that the video modality felt better because "having someone explain it to you rather than just reading it makes a big difference as you feel like you're being guided through the stress".

In addition to being stressed, users may also prefer a talking-head video over text due to other factors related to mental health such as feeling lonely. For instance, P5 mentioned that "if I have been studying for several nights alone, and there's nobody talking to me, then I would prefer to watch a video like this. But if I'm reviewing for the exam with my friends and I don't feel lonely, I would prefer the text form". However, some students who are not particularly stressed about exams in general, such as P1 and P12, may not find talking-head videos appealing. P1 noted that having instructions in text form was easier to follow because the video had more "distracting" elements such as "his voice, his accent, and gestures", and P12 shared that reading the text allowed them to "go through it at my own pace", which felt much harder to do with the video as it required pausing and rewinding.

F6: Writing May Make Non-native English Users More Conscious of Grammar and Word Choice.

> *"When I'm typing, I'll pay more attention to getting the right word, having the right grammar, and finding a better word to explain." – P3*

When discussing the two reflection prompts in D5 and D6, we found that several participants such as P2, P3, P5, and P7 preferred speaking out loud over typing because they felt more comfortable sharing rough ideas without worrying about grammar and word choice. P2, who was an international student whose native language is not English, noted that "when I type, it is kind of everywhere…I will also choose to correct my grammar and will have to look up certain words". In contrast, when describing speaking, they said "I will concentrate more on new ideas…I will just speak them…if I am writing them down, I may ignore the new ideas because I will (only) write down the things that I'm sure of". Similarly, P5 mentioned that "when I'm typing, I would think in my first language, Chinese, but I may have trouble translating my thoughts into written English. But if I'm speaking, I won't have time to translate things, so I would just choose some other easy way to say what I am thinking". Opting for less complex vocabulary may be tied to worries about being judged when writing, as hinted by P7's comment: "I feel like you wouldn't be criticized that much for talking incorrect [sic] compared to writing poorly". Elaborating on the latter half of P2's comment about only writing ideas that they are sure of, P6 noted that when talking out loud, they are more likely to share their "inner voice" which has "much more content" that gets lost when typing.

Students who felt more comfortable with writing, such as P8, P9 and P11, did not feel like they would share different ideas when speaking vs typing and described some advantages of the written form. For instance, P8 noted that they would focus on the main ideas, leaving word choice and grammar checks for the end. They also mentioned that "writing something down and actually considering your word choice would be a better way to remember everything". P11 preferred typing because they could "change" their words and "be more organized".

5 STUDY 2: LARGE-SCALE RANDOMIZED FIELD EXPERIMENT

To further evaluate our design, we conducted a large-scale randomized field experiment where we looked at the impact of the intervention on exam performance, and whether there were any differences in terms of how it affected students of different genders and year of study. The results are summarized in figure 2.

5.1 Intervention Deployment

We embedded the stress intervention into an online course activity (as an optional section) that was distributed to students during the fourth week of class, about 10 days before an upcoming exam. This activity was one of five graded activities that we asked students to complete during the term, and students were given 2% of their final marks after their completion. The other activities were unrelated to stress reappraisal and contained multiple parts, meaning that our intervention had to compete for students' limited time and attention. The inclusion of those other activities may add more noise to the analysis of the stress intervention effect; however, we intentionally randomized those activities and parts separately to minimize the potential for inducing a fake effect as much as possible.

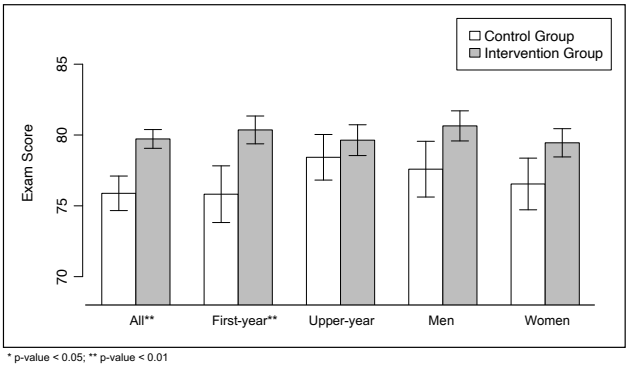

Figure 2: The differences in average exam scores between the control group and the intervention group at several levels. From left to right, we do the comparison within all data, first-year students, upper-year students, men, and women. The error bars show the standard error of each group.

5.1.1 Participants. The intervention was deployed to 1283 undergraduate students enrolled in an Introduction to Programming course at a large research-intensive post-secondary institution in Canada in the Spring 2020 semester. 59.8% of the sample were first-year students, 22.8% second-year students, 9.3% third-year students, 5.3% fourth-year students, and 2.8% students in their fifth year or higher. They came from various disciplines including the physical sciences, natural sciences, social sciences, life sciences, humanities, and commerce. 54.6% of students took the course to fulfill program requirements, while 20.0% took the course as an elective. An additional 25.5% of students took the course to fulfill general education requirements or for other reasons. The gender identity of students in the class was 46.5% male and 51.4% female; 0.2% of students specified another gender, while the remaining declined to answer. Student demographic data was collected from a voluntary survey administered mid-semester. Note that all questions were voluntary, so a different number of students may have answered each demographic question.

5.1.2 Randomization. Students were randomly assigned to either a control group, which did not receive the intervention or a treatment group, which received the intervention. Those who were in the treatment group saw the core stress-reappraisal message and a random combination of one or more variants of additional content illustrated in Figure 1. The randomization was automatic and dual-anonymous. Neither the instructors nor the students had access to the assignment policy. We chose a 2:1 split between treatment and control due to ethical concerns and felt that it was not necessary for us to exclude half the students from the potential benefits of our intervention, given our large sample size. Balancing treatment and control is usually preferred because they result in a higher power. However, given that our sample size was sufficiently large, this was not a concern for us.

5.1.3 Data Cleaning. The intervention was delivered to 1284 participants in total. 1014 of them clicked the intervention link and 931 participants completed it. 92 out of those 931 participants either dropped the class or did not finish the midterm, leaving us with 839 participants. Of those 839, 175 finished the intervention late, i.e., after the midterm, and so we had to exclude them from our study because they did not follow our experimental protocol. We allowed late students to access the activity after the end date of our experiment due to ethical considerations. We wanted to give students the opportunity to receive a grade for completing the activity even if their data was not useful to us for our research purposes. This gave us our final count of 664 participants in the cleaned dataset.

5.1.4 Quantitative Analysis. To test whether the treatment group outperformed the control group, we used one-sided independent samples t-tests, with midterm scores as our dependent variable. We also constructed linear regression models to see if adding one or more sub-treatments affected the intervention outcome.

5.2 Field Experiment Results

Our quantitative analysis results are summarized in figure 2.

5.3 Effects on Exam Scores

We first compared the treatment and control groups by conducting a one-sided independent samples t-test. As shown in table 1 and figure 2 ('All, control group' vs. 'All, intervention group'), the stress-reappraisal intervention improved students' midterm scores significantly by 3.8 % (p-value = 0.003, Cohen's d=0.25).

5.4 Result Validation

To verify that we did not randomly assign more higher-performing students to the intervention group, thereby inducing a fake effect, we used a bootstrap approach to estimate the probability of observing a similar event by chance. More specifically, from the 664 students in our cleaned dataset, we randomly placed 203 participants into one group (so that the number matches what we actually have in the control group) and put the remaining participants in another group. Then we calculated the difference between the average of these two groups and repeated this process 10,000 times to calculate the chance that the absolute value of such a difference is greater than what we observed in our true experiment, which was 3.84%. We found that the chance of this happening is less than 0.3%.

5.5 Subgroup Differences based on Gender Identity and Year of Study

We also analyzed the differences in treatment effects based on gender identity and year of study. As shown in table 1 and figure 2, the stress-reappraisal intervention had a significant effect on first-year students (p-value = 0.02), while no significant effect was observed for upper-year students (p-value = 0.27). Moreover, the effect of the stress reappraisal intervention across different gender identities is quite similar: the average improvement among men was 3.05. Among women, it was 2.91. On the other hand, the effect of our intervention on first-year students was 4.54, which is considerably higher than that for upper-year students, where we observed an increase of only 1.21. These sub-group results, however, are mostly not statistically significant as we split the data into smaller subsets and more than 100 participants chose not to inform us of their gender identity or year of study.

We placed our data in a linear regression context to further examine the significance level of the intervention effect in first-year versus upper-year groups. The output variable for the linear regression model is the midterm score, and the input variables are 'received intervention', 'upper-year', and 'upper-year and received intervention'. The results are not significant, and hence we do not have enough evidence to claim that the effect of the intervention is significantly different in the two groups.

5.6 Relative Effects of different variants

To assess the relative effects of the six different design factors, we conducted a linear regression where the midterm score is the output variable and the input variables are the six design factors: D1, D2, D3, D4, D5, and D6. These factors indicate whether the participant received a specific type of intervention. The regression was conducted only on the data of the intervention group, and the estimates measure the *add-on effect* of the six design factors, and not the independent effect of each factor as we cannot separate them from our main intervention by design. For example, the interpretation of the estimate of D1, which is -0.03, is: given that a student entered the intervention group and received our main stress intervention, the average *additional* influence on their midterm score if they also received the design factor D1 is -0.03. As shown in table 3, we found that including the design factor D4 can significantly reduce the effects of stress reappraisal intervention (p-value = 0.015), while all other design factors show no significant individual effects within the treatment group.

To check whether the number of interventions in a certain treatment influences the treatment effect, we analyze the *add-on effect* of including more design factors to the effect of the main intervention. The linear regression result is shown in table 4, where the p-value is 0.334. We also conducted an ANOVA test of the midterm scores among participants in the intervention group that received D0 through D6. In this ANOVA test, we set the number of design factors received as a categorical variable, so as to see if a certain number of sub-interventions have a statistically significant difference among other possibilities. The p-value of the test was 0.43, which is not significant. To ensure our analysis was robust, we repeated the regression and the ANOVA analysis on data that does not count D4 as a design factor or only on students who were in the intervention group but did not receive D4. We removed D4 from our analysis because it had a statistically significant and negative effect. However, in any version of that analysis, such as with D4 included, the result is not significant.

To conclude, we found our main intervention had a positive and statistically significant effect on students' test performance, and that we should be cautious about including potentially counterproductive design factors such as the reference link in the intervention.

6 DISCUSSION

In the following section, we begin by summarizing the key findings of the two studies and underscore their relevance to intervention designers. Then, we elucidate design implications tied to elaboration,

Table 1: Summary of sample sizes, group means, Cohen's d values and p-values from t-tests conducted between different groups. Note that there are 113 students who did not indicate their genders and 107 students who did not indicate their school year.

Group	Sample size	Group mean	Effect Size	Cohen's d	p-value
All, control group	203	75.88	3.84	0.252	0.003**
All, intervention group	461	79.72			
First-year students, control group	94	75.82	4.54	0.283	0.022*
First-year students, intervention group	212	80.36			
Upper-year students, control group	80	78.43	1.21	0.085	0.27
Upper-year students, intervention group	171	79.64			
Students who identify as men, control group	77	77.59	3.05	0.208	0.09
Students who identify as men, intervention group	155	80.64			
Students who identify as women, control group	93	76.54	2.91	0.184	0.08
Students who identify as women, intervention group	226	79.45			

* p-value <0.05; ** p-value <0.01

Table 2: Results of the linear regression model: midterm score ~Intervention + upper-year + upper-year and intervention. The value of "Intervention" is 1 if the participant was in the treatment group, and 0 otherwise; the value of "upper-year" is 1 if the participant is an upper-year student, and 0 otherwise; the value of "upper-year and intervention" is 1 if the participant is a first-year student and was in the treatment group, and 0 otherwise.

	Estimate	Standard Error	p-value
(Intercept)	75.82	1.58	<0.0001
Intervention	4.54	1.89	0.017*
upper-year	2.61	2.32	0.26
upper-year and intervention	-3.33	2.80	0.24

* p-value <0.05

modality, prompts, timing, and target audience. Finally, we describe opportunities for future research and outline the limitations of our work.

6.1 Key Findings

Online *eustress interventions* that help students embrace the positive aspects of stress to focus and perform better on exams have great potential, but it is less clear how designers and researchers can make specific decisions about the digital delivery of such information to students. To address this issue, we explored six design factors (D1-D6) that embody components for online interventions that are *brief*, *voluntary*, and *scalable*, and work by reinforcing a core reappraisal message (D0) through a layered approached. We systematically evaluated our designs through two studies:

In Study 1 (described in section 4), we thematically analyzed in-depth interviews with 20 participants to derive six findings (F1-F6) that underscore the need for reinforcing the core message using various sources and methods. We found that a short stress-reappraisal message, while powerful, isn't sufficient for convincing users (F1) and that users valued *research-based* guidance and encouragement (F2). Users also shared that while they were unlikely to click on a paper citation link, they often considered the *mere presence* of the citation as an indication of increased credibility (F3) for the stress reappraisal idea. We also saw how content structured as listed instructions was perceived favourably by participants because they thought listed information would be easier to recall during the exam (F4). We observed noteworthy subjective differences in users' perspectives on the same information being presented as text or video. Participants expressed a preference for instructional videos in situations where they were stressed or lonely because they valued the comfort and guidance afforded by a talking-head video from their instructor (F5). Finally, we identified some differences in how users express their ideas when writing or talking out loud, such as focusing more on grammar and word choice when typing, and expressing more ideas when speaking (F6).

In Study 2 (described in section 5), we found that our design had a significant positive effect on exam scores (p = 0.003, d = 0.252) in a large programming class. Our subgroup analysis indicated a significant effect for first-year students but not for upper-year. We did not detect significant gender differences. In F3 from Study 1, participants expressed that they found the presence of explicit paper citations to be more credible. In contrast, in the analysis of the relative effects of our design factors in section 5.6, we found that including the paper citation can significantly *reduce* the effects of stress reappraisal intervention (p-value = 0.015), while all other design factors show no significant individual effects within the treatment group. Therefore, we caution instructional designers when considering whether to include similar design components which involve external links. We suspect that participants who clicked the link may have become distracted. Furthermore, the perception of increased credibility does not automatically translate into increased internationalization of the reappraisal message. The significant and positive impact on exam performance was especially surprising and remarkable because participants spent on average, *only three minutes* on the intervention. We validated our results using a bootstrap approach described in section 5.4.

6.2 Design Implications

In this section, we synthesize our findings into five design implications that are tied to the value of additional elaboration, presentation modality, reflection prompts, timing, and target audience.

Table 3: Results from the linear regression model where the midterm score is the dependent variable and D1 to D6 are predictor variables. The value of 'Di', where i = 1, 2, 3, 4, 5, and 6, is 1 if the participant is in the intervention group and received the design factor i as one of the interventions. We only look at data in the intervention group and all the estimates are measuring the add-on effect of the six design factors upon the main effect. Each participant can receive multiple design factors.

	Estimate	Sample Size	Standard Error	p-value
(Intercept)	81.05	664	1.79	<0.0001
D1: Summary Text	-0.03	233	1.34	0.98
D2: Listed Instructions	-0.47	229	1.33	0.72
D3: Video Instructions	-0.88	226	1.33	0.51
D4: Link to Paper	-3.27	230	1.33	0.015*
D5: Self-Explanation	1.68	241	1.34	0.21
D6: Note to Self	0.21	231	1.32	0.87

* p-value <0.05

Table 4: Results from the linear regression model where the midterm score is the dependent variable and the number of interventions is the predictor variable. We only use data from the intervention group. The number of interventions is the total number of design factors received by the participant.

	Estimate	Standard Error	p-value
(Intercept)	81.32	1.78	<0.0001
Number of interventions	-0.53	0.55	0.334

6.2.1 ***Additional Elaboration: Providing More Information versus Preserving Focused Attention.*** Firstly, a key question that designers should consider is the trade-off between adding more information to ensure students understand and apply the concept (explanatory context in D1, concrete suggestions in D2), and the limits on attention and the potential visual clutter, given that students spent an average of only 3 minutes on this intervention. Our findings from the first study on exploring user perspectives showed the value of having additional elaboration as many students found it useful to have the research-based explanatory elaboration from D1, and some who were more skeptical appreciated seeing the citation from D4, even if they may not necessarily click to read a long research paper. Students also felt that having actionable instructions as a list in D2 would make it easier for them to remember how to apply the intervention during the exam. Therefore, instructional designers wishing to incorporate similar mindset interventions that are counter-intuitive can take a multi-pronged and layered approach to convince their students of the idea by making ample use of research-based encouragement in their content.

6.2.2 ***Presentation Modality: Delivering the Same Text as a Simple Video.*** Secondly, a simple video reading the same exact text from D3 is valued by students because they can *see* a person explaining why they should believe in the message and how to apply it. Designers might consider including such simple conversational videos as students shared that they felt guided through the stress reappraisal information when presented in video form, especially when feeling stressed (F5). This latter point about feeling stressed is important because, in situations where participants may not be predisposed to consuming content in a certain medium, intervention designers can help learners by offering the same information in a different form. This can also have the added benefit of improved accessibility.

6.2.3 ***Reflection Prompts: Internalizing and Applying Eustress to Everyday Life.*** Thirdly, one insight was that the reflection prompts guide students to think through how to apply the eustress intervention to everyday life, and that is certainly possible. A reflective learning approach helps in addressing a key challenge for instructional designers, which is to help learners retain and apply intervention content after finishing the activity. Therefore, we encourage intervention designers to consider how to augment their activities with reflective exercises, especially near the end of their interventions.

6.2.4 ***Intervention Timing: Delivering Interventions when Learners are Most Receptive.*** Moreover, during the interviews, certain students mentioned that they were less likely to go through such activities closer to the exam because, around that time, they would rather focus on exam content as they are worried about finishing the syllabus. Such comments signify the need for intervention designers to consider the tradeoff between delivering the information closer to the exam so they remember the intervention content, and the need for students to focus on other things that compete for their limited time and attention.

6.2.5 ***Target Audience: Eustress interventions may be more helpful to first-year students.*** Finally, in the analysis of our field experiment results in Study 2, the intervention had an overall positive significant effect on exam performance. However, when we conducted the subgroup analysis for first-year vs upper-year students, we only observed a significant effect for the first-year group. This finding is in alignment with prior work on stress reappraisal [6]. Certain comments from participants hinted at some plausible reasons as to why we observed this difference: students in their first year may be more stressed as they just transitioned from high school to university education, which can be challenging for many. Furthermore, first-year students may also be more receptive to adopting new exam strategies whereas students in more senior years may have already formed strong opinions on how to best tackle exam stress. Intervention designers can therefore benefit

from analyzing subgroup differences when evaluating their interventions, such as any difference between first-year and upper-year students.

6.3 Future Work

The content of our intervention as fully described in section 3.2, did not contain any references to course-specific materials, and as such can be easily generalized to other courses, and thus, this work can inform several directions for future research in the expansive design space for online eustress interventions. Firstly, from our exploration of the value of additional elaboration, we saw that users like articles or activities that are framed as being "evidence-backed" but may be less likely to explore the papers linked in the citations. We also found in the field experiment that the citation link had a significant and negative add-on impact in our analysis of the relative effects of each design factor. Future work can explore the specific contexts in which such an observation holds or does not hold, and whether certain student populations are more or less inclined to accept a research-based framing. Secondly, when varying presentation modalities, we saw the promise of offering the same content in a different format. Our video consisted of one individual delivering a particular message, and it is hard to know what properties would generalize to other settings. Future work can explore what dynamics affect how students react to reappraisal messaging in the video form, such as how students respond to pop-up messages on eustress when presented as in-video prompts [43] for them to do something during or after they watch the video, as well as other kinds of visualizations of the reappraisal message, such as the presence of other agents in the video including past students instead of an instructor. Thirdly, we saw how adding reflection prompts at the end of mindset interventions may help learners retain the message. Future work could examine if asking students to explain by recording a voice message helps them internalize the reappraisal information. Students could also be asked to record a voice or a video message to *themselves* that could then be sent by email or text at a particular time. These examples illustrate how our work can help suggest future directions for designs that explore how we might send these messages at the right moment, in the right format in a way that engages people in receiving reappraisal messaging to change their mindset and behaviour through technology-mediated online interventions. Finally, a more detailed exploration of how such online interventions can be integrated within different contexts such as in-person vs remote learning can further inform the design of future eustress interventions.

6.4 Limitations

Replications of our approach can be further strengthened by including pretest and post-test performance measures, i.e., having student test scores from an initial test compared to performance on a second test with the intervention in between. Furthermore, if there are multiple opportunities for testing performed at different time points after the intervention is deployed, those tests can help inform whether there is a sustained impact of such interventions or if it is a single impact that needs boosting. In our case, the final exam was cancelled due to COVID-19 when we deployed our intervention. As a result, we only had the midterm scores as a post-test measure.

It is possible, though not likely, that we randomly assigned higher-performing students to the intervention group. However, to check for this, we used a bootstrap approach and found that the chance of getting a similar or more significant event by randomly splitting students into two groups is less than 0.3%. It is also possible that higher-performing students have less stress and therefore perform better on exams, and stress may not be equally distributed across all performing levels of students.

Our intervention was accessible through a web page in a field setting. This setup maximized the ecological validity of our results. At the same time, it led to complex implementation issues that we tried to minimize, such as accounting for page refreshes by setting up our randomizer such that if participants' reloaded the activity, they would remain assigned to the same experimental group (i.e., either control or intervention) through browser caching. However, a small percentage of participants (1.7%) loaded and completed the activity from more than one device (e.g., by doing it from their phone after seeing the announcement and then doing it again on their laptop). These participants were inadvertently labelled as being in both the control group and the intervention group by the randomizer. However, because they saw the intervention, they were of course no longer part of the control. We factored this into our analysis by treating those participants as being in the intervention group. We reassessed whether our decision to leave this small percentage of participants in the intervention group induced an artificial effect by rerunning the analyses with those participants excluded. The results remained similar to our current analysis, e.g., the main effect remained significant ($p = 0.018$).

7 CONCLUSION

Our work presents an approach for designing effective interventions for shifting student attitudes of exam *stress* towards *eustress*. We explored the design space for online stress reappraisal interventions that were *brief*, *voluntary*, and *scalable*. We conducted two studies to explore design dimensions that varied in the levels of elaboration, presentation modality, and the use of reflection prompts. We instantiated these dimensions using six design factors (D1-6) that reinforce a core reappraisal message (D0) and evaluated these factors through both semi-structured interviews with 20 students, and a randomized field experiment deployed to over 1200 students in a real-world programming course. The interviews yielded practical insights into the relative importance and impact of factors on different students in different contexts, such as when text versus video presentation might be effective, what kinds of additional information are compelling versus burdensome, and what kind of interface prompts for students are more or less impactful in helping them retain and utilize the reappraisal messaging. These dimensions can be used to generate an expansive space of design variables that can be explored in future work: from the modality of videos that incorporate various visualizations, reflective activities such as students' recording voice and video messages for their future selves, to adaptive delivery systems that target message timing to the right moment and mental state. Although students only spent

an average of 3 minutes on the intervention, our field experiment revealed that combinations of the six design factors were powerful enough to boost the class average exam score from 76% to 80%, which is equivalent to an increase from a **B** to an **A-** in the grading scale used by the course where we deployed our study.[1]

ACKNOWLEDGMENTS

We are grateful to the students who participated in this work. We would also like to thank Andrew Peterson at the University of Toronto, Gerry Chan at Dalhousie University, and René Kizilcec at Cornell University for their input on the early drafts of this manuscript. This work was partially supported by grants from the Office of Naval Research (N00014-18-1-2755, N00014-21-1-2576), the Natural Sciences and Engineering Research Council of Canada (RGPIN-2019-06968), and the National Science Foundation (2209819).

REFERENCES

[1] [n. d.]. Bulleted lists and memory - an experiment. https://www.ab-lab.org/bulleted-lists.html
[2] Sami Abdo Radman Al-Dubai, Redhwan Ahmed Al-Naggar, Mustafa Ahmed Alshagga, and Krishna Gopal Rampal. 2011. Stress and coping strategies of students in a medical faculty in Malaysia. *The Malaysian journal of medical sciences: MJMS* 18, 3 (2011), 57.
[3] Azy Barak and John M Grohol. 2011. Current and future trends in internet-supported mental health interventions. *Journal of Technology in Human Services* 29, 3 (2011), 155–196.
[4] Miranda L Beltzer, Matthew K Nock, Brett J Peters, and Jeremy P Jamieson. 2014. Rethinking butterflies: The affective, physiological, and performance effects of reappraising arousal during social evaluation. *Emotion* 14, 4 (2014), 761.
[5] Dionne Bowie-DaBreo, Corina Sas, Heather Iles-Smith, and Sandra Sünram-Lea. 2022. User Perspectives and Ethical Experiences of Apps for Depression: A Qualitative Analysis of User Reviews. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 21, 24 pages. https://doi.org/10.1145/3491102.3517498
[6] Shannon T Brady, Bridgette Martin Hard, and James J Gross. 2018. Reappraising test anxiety increases academic performance of first-year college students. *Journal of Educational Psychology* 110, 3 (2018), 395.
[7] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health* 11, 4 (2019), 589–597.
[8] Yujia Cao, Mariët Theune, and Anton Nijholt. 2009. Towards Cognitive-Aware Multimodal Presentation: The Modality Effects in High-Load HCI. In *Engineering Psychology and Cognitive Ergonomics*, Don Harris (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 3–12.
[9] Mark S Chapell, Z Benjamin Blanding, Michael E Silverstein, Masami Takahashi, Brian Newman, Aaron Gubi, and Nicole McCann. 2005. Test anxiety and academic performance in undergraduate and graduate students. *Journal of educational Psychology* 97, 2 (2005), 268.
[10] Alia J Crum, Peter Salovey, and Shawn Achor. 2013. Rethinking stress: the role of mindsets in determining the stress response. *Journal of personality and social psychology* 104, 4 (2013), 716.
[11] Ralph E Culler and Charles J Holahan. 1980. Test anxiety and academic performance: The effects of study-related behaviors. *Journal of educational psychology* 72, 1 (1980), 16.
[12] Michael D Jones and Deserai Anderson Crow. 2017. How can we use the 'science of stories' to produce persuasive scientific stories? *Palgrave Communications* 3, 1 (2017), 1–9.
[13] Elizabeth A Davis. 2003. Prompting middle school science students for productive reflection: Generic and directed prompts. *The Journal of the Learning Sciences* 12, 1 (2003), 91–142.
[14] Björn B de Koning, Vincent Hoogerheide, and Jean-Michel Boucheix. 2018. Developments and trends in learning with instructional video. *Computers in Human Behavior* 89, 1 (2018), 395–398.
[15] Shivangi Dhawan. 2020. Online learning: A panacea in the time of COVID-19 crisis. *Journal of educational technology systems* 49, 1 (2020), 5–22.
[16] Liliane Efinger, Simon Thuillard, and ES Dan-Glauser. 2019. Distraction and reappraisal efficiency on immediate negative emotional responses: role of trait anxiety. *Anxiety, Stress, & Coping* 32, 4 (2019), 412–427.
[17] Kristina Fenn and Majella Byrne. 2013. The key principles of cognitive behavioural therapy. *InnovAiT* 6, 9 (2013), 579–585.
[18] Angela Fessl, Oliver Blunk, Michael Prilla, and Viktoria Pammer. 2017. The known universe of reflection guidance: a literature review. *International journal of technology enhanced learning* 9, 2-3 (2017), 103–125.
[19] Susan Folkman. 1984. *Stress: appraisal and coping*. Springer, 11 West 42nd Street.
[20] Philippe R Goldin, Craig A Moodie, and James J Gross. 2019. Acceptance versus reappraisal: Behavioral, autonomic, and neural effects. *Cognitive, Affective, & Behavioral Neuroscience* 19, 4 (2019), 927–944.
[21] Aleesha Hamid, Rabiah Arshad, and Suleman Shahid. 2022. What Are You Thinking?: Using CBT and Storytelling to Improve Mental Health Among College Students. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 441, 16 pages. https://doi.org/10.1145/3491102.3517603
[22] Wendy Hardeman, Julie Houghton, Kathleen Lane, Andy Jones, and Felix Naughton. 2019. A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. *International Journal of Behavioral Nutrition and Physical Activity* 16, 1 (2019), 1–21.
[23] Stefan G Hofmann, Sanna Heering, Alice T Sawyer, and Anu Asnaani. 2009. How to handle anxiety: The effects of reappraisal, acceptance, and suppression strategies on anxious arousal. *Behaviour research and therapy* 47, 5 (2009), 389–394.
[24] Esther Howe, Jina Suh, Mehrab Bin Morshed, Daniel McDuff, Kael Rowan, Javier Hernandez, Marah Ihab Abdin, Gonzalo Ramos, Tracy Tran, and Mary P Czerwinski. 2022. Design of Digital Workplace Stress-Reduction Intervention Systems: Effects of Intervention Type and Timing. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 327, 16 pages. https://doi.org/10.1145/3491102.3502027
[25] Nabil Issa, Mary Schuller, Susan Santacaterina, Michael Shapiro, Edward Wang, Richard E Mayer, and Debra A DaRosa. 2011. Applying multimedia design principles enhances learning in medical education. *Medical education* 45, 8 (2011), 818–826.
[26] Jeremy P Jamieson, Alia J Crum, J Parker Goyer, Marisa E Marotta, and Modupe Akinola. 2018. Optimizing stress responses with reappraisal and mindset interventions: an integrated model. *Anxiety, Stress, & Coping* 31, 3 (2018), 245–261.
[27] Jeremy P Jamieson, Wendy Berry Mendes, Erin Blackstock, and Toni Schmader. 2010. Turning the knots in your stomach into bows: Reappraising arousal improves performance on the GRE. *Journal of experimental social psychology* 46, 1 (2010), 208–212.
[28] Jeremy P Jamieson, Wendy Berry Mendes, and Matthew K Nock. 2013. Improving acute stress responses: The power of reappraisal. *Current Directions in Psychological Science* 22, 1 (2013), 51–56.
[29] Jeremy P Jamieson, Brett J Peters, Emily J Greenwood, and Aaron J Altose. 2016. Reappraising stress arousal improves performance and reduces evaluation anxiety in classroom exam situations. *Social Psychological and Personality Science* 7, 6 (2016), 579–587.
[30] Raffael Kalisch, Katja Wiech, Katrin Herrmann, and Raymond J Dolan. 2006. Neural correlates of self-distraction from anxiety and a process model of cognitive emotion regulation. *Journal of cognitive neuroscience* 18, 8 (2006), 1266–1276.
[31] Slava Kalyuga. 2012. Instructional benefits of spoken words: A review of cognitive load factors. *Educational Research Review* 7, 2 (2012), 145–159.
[32] Nicholas J Kelley, James E Glazer, Narun Pornpattananangkul, and Robin Nusslock. 2019. Reappraisal and suppression emotion-regulation tendencies differentially predict reward-responsivity and psychological well-being. *Biological psychology* 140 (2019), 35–47.
[33] Rafal Kocielnik, Lillian Xiao, Daniel Avrahami, and Gary Hsieh. 2018. Reflection Companion: A Conversational System for Engaging Users in Reflection on Physical Activity. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 2, Article 70 (jul 2018), 26 pages. https://doi.org/10.1145/3214273
[34] Mark Le Fevre, Gregory S Kolt, and Jonathan Matheny. 2006. Eustress, distress and their interpretation in primary and secondary occupational stress management interventions: which way first? *Journal of Managerial Psychology* 21, 6 (2006), 547–565.
[35] Jenny JW Liu, Natalie Ein, Julia Gervasio, and Kristin Vickers. 2019. The efficacy of stress reappraisal interventions on stress responsivity: A meta-analysis and systematic review of existing evidence. *PLoS One* 14, 2 (2019), e0212854.
[36] Jenny Jing Wen Liu, Maureen Reed, and Kristin Vickers. 2019. Reframing the individual stress response: Balancing our knowledge of stress to improve responsivity to stressors. *Stress and Health* 35, 5 (2019), 607–616.
[37] Jean-Claude Martin, Christine Lescanff, Sophie Rosset, Marilyn Walker, and Steve Whittaker. 2018. How to Personalize Conversational Coaches for Stress Management?. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable*

[1]Note: grading systems vary by province and institution in Canada, and may be different from US schools. The grading scale relevant to this study is used by certain institutions of higher education in Ontario.

Computers (Singapore, Singapore) *(UbiComp '18)*. Association for Computing Machinery, New York, NY, USA, 718–721. https://doi.org/10.1145/3267305.3267698

[38] Richard E Mayer. 2005. Introduction to multimedia learning. *The Cambridge handbook of multimedia learning* 2, 1 (2005), 24.

[39] Kathryn McAlindon, Jennifer Watling Neal, Zachary P Neal, Kristen J Mills, and Jennifer Lawlor. 2019. The bond framework: a practical application of visual communication design and marketing to advance evaluation reporting. *American Journal of Evaluation* 40, 2 (2019), 291–305.

[40] Kathleen O'Leary, Stephen M. Schueller, Jacob O. Wobbrock, and Wanda Pratt. 2018. "Suddenly, We Got to Become Therapists for Each Other": Designing Peer Support Chats for Mental Health. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3173905

[41] Tenaha O'Reilly, Sonya Symons, and Heather MacLatchy-Gaudet. 1998. A comparison of self-explanation and elaborative interrogation. *Contemporary Educational Psychology* 23, 4 (1998), 434–445.

[42] David Robotham and Claire Julian. 2006. Stress and the higher education student: a critical review of the literature. *Journal of further and higher education* 30, 02 (2006), 107–117.

[43] Hyungyu Shin, Eun-Young Ko, Joseph Jay Williams, and Juho Kim. 2018. Understanding the effect of in-video prompting on learners and instructors. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1–12.

[44] Changwon Son, Sudeep Hegde, Alec Smith, Xiaomei Wang, Farzan Sasangohar, et al. 2020. Effects of COVID-19 on college students' mental health in the United States: Interview survey study. *Journal of medical internet research* 22, 9 (2020), e21279.

[45] Joachim Stoeber and Dirk P Janssen. 2011. Perfectionism and coping with daily failures: Positive reframing helps achieve satisfaction at the end of the day. *Anxiety, Stress & Coping* 24, 5 (2011), 477–497.

[46] John Sweller, Paul A Kirschner, and Richard E Clark. 2007. Why minimally guided teaching techniques do not work: A reply to commentaries. *Educational psychologist* 42, 2 (2007), 115–121.

[47] Karran Thorpe. 2004. Reflective learning journals: From concept to practice. *Reflective practice* 5, 3 (2004), 327–343.

[48] Allison S Troy, Amanda J Shallcross, Anna Brunner, Rachel Friedman, and Markera C Jones. 2018. Cognitive reappraisal and acceptance: Effects on emotion, physiology, and perceived cognitive costs. *Emotion* 18, 1 (2018), 58.

[49] Edward R Tufte. 1990. Envisioning Information Graphics Press. *Cheshire, Connecticut* 6410 (1990), 1–35.

[50] Jesse S Zolna. 2007. The effects of study time and presentation modality on learning. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. SAGE Publications Sage CA: Los Angeles, CA, SAGE Publications Sage CA, Los Angeles, CA, 512–515.