Development and Evaluation of a Digital Single-Session Reappraisal Intervention to Shift Negative Self-Focused Social Media Interpretations

Ilana E. Ladis

Miami, Florida

B.A., Wesleyan University, 2017

M.A., University of Virginia, 2021

A Dissertation presented to the Graduate Faculty of the University of Virginia in Candidacy for the Degree of Doctor of Philosophy

Department of Psychology

University of Virginia

June 11, 2024

Committee Members:

Bethany A. Teachman, Ph.D. (Chair)

Joseph P. Allen, Ph.D.

Kara P. Wiseman, MPH, Ph.D.

Adrienne Wood, Ph.D.

Acknowledgements

Thank you to my advisor, Bethany, for allowing me to color outside of the lines during graduate school and being on board with the uncharted paths down which I (and my research) ventured. I am grateful for the ways that you model integrity, humility, and personal responsibility – as a scientist and as a person. Wherever I end up in the future, both professionally and personally, I know I'll be better for it because I have you as a role model.

Many thanks to my other wonderful committee members, Joe, Kara, and Adrienne, for lending your time and expertise to my project and for your helpful feedback and suggestions.

Thank you to my friends in the psychology department, my PACT Lab family, and my cohort for making the years spent in graduate school so enjoyable. The countless trivia nights, coffee shop writing sessions, and backyard hangs are some of my favorite memories of my time here.

Thank you to my family and friends outside of UVA for your genuine curiosity in my work and for being there for me as I developed intellectually, professionally, and personally over the last five years. Thank you, also, for the numerous trips you have made to Charlottesville despite it being extremely hard to access from Miami and Rhode Island.

Lastly, I want to thank my loving husband, Dylan Abrams, for moving to Charlottesville for me (twice), for picking up dinner regularly and with genuine enthusiasm, for always listening, and for your unwavering support of my personal and professional growth. This dissertation would not have been possible without you!

Table of Contents

Abstract

Social media is fraught with ambiguity and opportunities for making skewed interpretations. For example, seeing a post of friends hanging out without you can trigger thoughts ranging from benign ("*They seem to be having fun*") to negative ("*They're happy I'm not there*"). Negative interpretation bias, or the tendency to assign negative meanings in response to ambiguous information, is associated with mental health problems like depression and anxiety, while positive interpretation bias, or the tendency to make benign interpretations in response to ambiguity, may be protective. Interpretation bias is a modifiable treatment target in the context of both in-person therapy and digital interventions that target rigid thinking styles. Reappraisal, a cognitive skill that teaches people to respond to automatic thoughts more flexibly, may hold promise for shifting negative self-focused social media interpretations (i.e., "online" interpretation bias, which includes interpretations such as not having enough likes or followers). Through a multiphase design process with college student stakeholders, we developed a digital, single-session, self-administered reappraisal intervention to shift online interpretation bias as well as other secondary outcomes (e.g., "offline" interpretation bias; fear of missing out, known as FOMO). Feedback interviews with 20 potential end users were completed iteratively during intervention development, and the prototype was then pilot tested by 17 additional participants. The final intervention, "PRISM," (**P**rogram for **R**ethinking Negative **I**nterpretations on **S**ocial **M**edia) includes psychoeducation related to healthy social media use, training in reappraisal, and interactive practice. A randomized-controlled trial with $N = 162$ college students who reported negative self-evaluations tied to social media was then conducted to compare PRISM to a control condition that also focused on social media but did not include reappraisal. Effects on online interpretation bias and secondary outcomes were assessed immediately post-intervention and at

two-week follow-up. As hypothesized, PRISM led to greater reductions in online negative interpretation bias and increases in online positive interpretation bias than the control condition, both post-intervention and at two-week follow-up. Mirroring results for online interpretation bias, PRISM also led to greater (and, generally stronger) reductions in *offline* negative interpretation bias and increases in *offline* positive interpretation bias than the control condition, both post-intervention and at two-week follow-up. Though there were some significant time-by-condition effects for other secondary outcomes, they were generally weaker and unstable to sensitivity analyses. Results show that a digital, light-touch reappraisal intervention can shift online and offline interpretation biases in college students with problematic social media use.

General Introduction

Social media is an omnipresent force that shapes the ways we think, feel, and interact. Social media is also controversial, with some correlational research linking social media use with depression (Lin et al., 2016), anxiety (Keles et al., 2020), poor sleep quality (Woods & Scott, 2016), and suicide attempts (Sedgwick et al., 2019). Other work has recognized positive effects of social media, including its ability to foster inspiration, enjoyment, and social connectedness (Masciantonio et al., 2021; Meier et al., 2020; Pittman & Reich, 2016; Valkenburg et al., 2021). These studies challenge the narrative of social media as exclusively "bad," and reinforce that social media's impact depends on *how* it is used. This is in part because individuals navigate their digital worlds like they navigate their physical worlds: with their own lens that guides how they interpret their interactions and experiences. For individuals who struggle with psychological distress related to social media, this dissertation tested whether a novel digital, self-administered, single-session intervention can shift interpretation bias (and specifically, reduce negative and increase positive self-focused social media interpretations) in response to ambiguous information when engaging with social media.

The cognitive-behavioral perspective recognizes negative interpretation bias as critical in the etiology and maintenance of many psychological conditions, including social anxiety (Amin et al., 1998; Chen et al., 2020) and depression (Everaert et al., 2017; Hindash & Amir, 2012). If someone thinks that everyone is judging them at a party, they might not go to the next party, which eventually may have downstream consequences for mood and mental health. Consequently, negative interpretation bias (and the closely related concept of "negative automatic thinking") is often a treatment target in both face-to-face (Furlong & Oei, 2002; Muris et al., 2009) and digital interventions (Beard et al., 2021; Ji et al., 2021), and has been shown to

decrease in tandem with anxiety and depression (Furlong & Oei, 2002; Ji et al., 2021; Muris et al., 2009). Diminished positive interpretation bias is also implicated in depression and anxiety disorders (Jopling et al., 2020a) and is a frequent intervention target (Ji et al., 2021; Murphy et al., 2007). While interpretation biases can occur across a range of domains, given the inherently social nature of social media and our interest in common forms of biased thinking tied to social media (such as fear of missing out or "FOMO", tendency toward social comparison, fear of negative evaluation from others), the present study focuses on interpretation biases related to social situations.

There is a surprising paucity of research on negative and positive interpretation bias in relation to the social media environment, despite social media being rife with ambiguous content. If someone gets very few likes on their post, a negative interpretation bias might lead them to make negative self-focused interpretations, such as thinking others are judging them, which may lead them to feel anxious or depressed. Despite the high potential for interpretation biases on social media, only one study to our knowledge has examined the prevalence of interpretation biases in online settings (Miers et al., 2020). This study, conducted among 18-25 year olds, assessed "online" interpretation bias based on interpretations that participants made about hypothetical social scenarios occurring in online/social media contexts (e.g., how one interprets having few followers on social media: "*Nobody thinks my post is funny and that's why I only have one follower*") vs. "offline" interpretation bias based on social scenarios occurring in offline social situations (e.g., how one interprets someone looking at them at a party: "*They're gossiping about me*"). This investigation found that online negative interpretation bias had a moderate, positive correlation with offline negative interpretation bias (r = .53). Notably, online interpretation bias was associated with self-reported experiences of online (but not offline) peer

victimization (e.g., being sent aggressive messages) and vice versa for offline interpretation bias. This suggests that, though they are correlated, online and offline interpretation bias may manifest in different, context-specific ways. Indeed, there are features specific to online settings that add to ambiguity (e.g., not being able to see someone else's reaction in real-time) and subject users to stressful cues (e.g., awareness of others' whereabouts through features like Snapchat and Instagram Stories; Steele et al., 2020), which likely increase the potential for negative interpretations to be made. Given the extraordinarily high rate of social interaction that occurs on social media, particularly among young people, and potential for harm when negative interpretations are made, easily accessible interventions to foster healthy interpretations in the social media environment would be beneficial.

Using Reappraisal to Shift Social Media Interpretations

Reappraisal, a therapeutic technique that teaches patients to systematically question and re-evaluate their own thoughts, may hold promise for reducing negative interpretation bias and increasing positive interpretation bias in both online and offline contexts. In traditional, multi-session cognitive-behavioral therapy (CBT), reappraisal is typically taught over a few therapy sessions and then practiced in an ongoing way as problematic or unhelpful thoughts arise. Reappraisal is considered a highly effective "key ingredient" in CBT for anxiety, depression, and several other conditions (Clark, 2013). However, such treatments are often inaccessible due to barriers such as cost, stigma, and lack of available providers (Haugen et al., 2017; Marsh & Wilcoxon, 2015). Also, for some people, a single therapeutic encounter is sufficient to address their concerns (Bloom, 2001; Simon et al., 2012), and the majority who attend a first therapy session do not return for another (Talmon, 1990). There have thus been extensive research efforts centered on developing single-session mental health interventions that aim to maximize the

impact of one therapeutic encounter. These interventions are typically developed by distilling down essential components of existing interventions that are thought to drive change. The literature on single-session interventions (SSIs) has been promising, with a meta-analysis of 50 studies of in-person, single-session interventions (which analyzed both therapist-administered and self-administered interventions) for youth yielding, on average, a small-to-medium effect (g = 0.32; Schleider & Weisz, 2017). Many of these programs promote thinking styles associated with positive youth development (e.g., growth mindset interventions). In terms of reappraisal specifically, one five-minute reappraisal intervention embedded within Tumblr reduced self-reported hopelessness in teenagers immediately post-intervention (Dobias et al., 2022). For young adults, cognitively-oriented SSIs have also been linked with several positive outcomes, including increased thriving and hope, and reduced depressive symptoms and neuroticism (Bentley et al., 2018; Bernstein et al., 2021; Bu & Duan, 2019; Feldman & Dreher, 2012).

Building upon the early enthusiasm and promising results from studies of in-person SSIs, there has recently been a proliferation of research evaluating SSIs that are digital *and* self-administered. Such programs reduce access gaps by allowing anyone with internet service to complete an intervention at their own pace when it fits their schedule (Muñoz et al., 2018). Digital, self-administered SSIs – including those focused on changing thinking styles – have been rated as highly acceptable and shown efficacy in reducing depression symptoms (Osborn et al., 2020; Schleider et al., 2020; Schleider et al., 2022; Schleider & Weisz, 2018) and anxiety symptoms (Venturo-Conerly et al., 2022) among middle school-aged and high school-aged adolescents.

Single-session Interventions for Young Adults. Though most digital, self-administered SSIs have focused on middle school-aged and high school-aged adolescents (Schleider et al.,

2020), recent research suggests that digital SSIs for young adults (including college students) can also be efficacious. A dialectical behavior therapy-informed SSI reduced stress and improved self-efficacy in college students at one-week follow-up, relative to a waitlist control condition (Lee et al., 2023). Moreover, completing a crisis-oriented SSI embedded within social media was associated with immediate reductions in hopelessness among individuals aged 18 – 25, compared to the standard crisis response of referral to 988 crisis hotline (Cohen et al., 2023).

Some digital SSIs for young adults have taught reappraisal (or "cognitive flexibility") as part of a larger package of evidence-based skills (Ghosh et al., 2023; Sauer-Zavala et al., 2021; Wasil et al., 2021). A digital, cognitive-behavioral SSI designed to target fear of negative evaluation in the context of an active learning STEM university course increased confidence in class discussions compared with an active control condition (Ghosh et al., 2023). However, a digital, self-guided SSI adaptation of the Unified Protocol (which includes a module on cognitive flexibility) did not significantly impact mental health-related outcomes relative to a no-control waitlist condition in college students (Sauer-Zavala et al., 2021). It is possible that SSIs with more specific, circumscribed targets, like fear of negative evaluation in classroom settings, are more impactful than those targeting more global psychological issues like emotion dysregulation. The former intervention was also about half as long as the latter, suggesting that spending more time completing an intervention does not necessarily make it more impactful.

Other interventions for adults (of all ages) that focused exclusively on teaching reappraisal have shown positive effects. For example, reappraisal-based, writing-focused interventions (e.g., involving third-party perspective taking related to a recent interpersonal conflict) reduced problematic drinking (Rodriguez et al., 2019) and romantic conflict during the COVID-19 pandemic (Rodriguez et al., 2021) compared with credible control conditions (e.g.,

writing about one's feelings). Another digital SSI found that three separate reappraisal technique training conditions (positive reframing, self-distancing, or temporal distancing) were associated with increased general well-being and reduced emotional reactivity to a stressor two weeks later, compared with a "cope-as-usual" control condition (Ranney et al., 2017). Common elements across these trainings include some element of personalization and instruction for participants to practice reappraisal after the training ends. Taken together, these studies suggest that reappraisal can be taught in many different formats in the context of digital, self-administered SSIs. It also appears that reappraisal in the form of a bite-sized, standalone intervention (vs. as part of a larger treatment package) can be sufficient to lead to meaningful change.

Social Media-specific Interventions. The majority of existing interventions designed to reduce the negative mental health impacts of social media use have focused on social media restriction or abstinence (Herriman et al., 2024). The effects of these interventions have been inconclusive, yielding evidence for both benefits and drawbacks of cutting time spent on social media. Unsurprisingly, abstinence-focused interventions also have high relapse/non-compliance rates (Fernandez et al., 2020). These challenges highlight the need for alternatives to one-size-fits-all approaches that are sensitive to how social media is inextricably interwoven into young peoples' lives. Interventions that focus on changing the way people experience or respond to social media are likely more realistic and approachable than abstinence-focused interventions. For example, clinician-administered programs focused on educating people about social media (including increasing awareness of its negative impacts) have been associated with positive outcomes (Herriman et al., 2024). Yet, their in-person format and reliance on clinicians reduces scalability. An SSI that enhances awareness about social media and provides accessible coping

strategies other than social media abstinence could potentially reach more people aiming to use social media in a healthier way.

Only one study to our knowledge has tested whether a SSI specific to social media-related cognitions can improve psychological outcomes (Weber et al., 2022). Weber and colleague's Experiment 1 evaluated whether seeing a brief disclaimer (i.e., that Instagram users tend to only present themselves in the best light and pictures are often heavily edited to reflect a selective version of someone's life) on a hypothetical Instagram profile is associated with greater momentary positive affect, trait self-esteem, and well-being, and lower momentary negative affect and envy, relative to a control condition that had no disclaimer. Experiment 2 tested whether a brief, single-session training on either the fundamental attribution error or on growth mindsets (< 30 minutes) is associated with greater internal control beliefs (i.e., the extent to which one feels in charge of ones' own fate) assessed immediately post-intervention, compared with a non-active control condition. No intervention effects were observed for either experiment, and some effects were in the opposite direction of hypotheses (e.g., the growth mindset intervention was positively associated with envy and negatively associated with self-esteem). However, this study included adults of all ages, even though younger adults are more likely than older generations to use social media more intensely and experience negative psychological outcomes related to social media (Hayes et al., 2015; Kuss & Griffiths, 2017). Participants were also not recruited on the basis of self-reported difficulties with social media use. It is thus plausible that the null effects observed were due to the sample consisting of relatively healthy adult social media users who were not in need of such an intervention. The study also only assessed immediate outcomes, though the effects of reappraisal interventions may be more observable over time and with repeated practice (e.g., even over the course of a few weeks;

Denny & Ochsner, 2014). For example, a CBT-based app for body image-related stressors on Instagram used daily over the course of two weeks was effective in improving body image and resilience at 1-month follow-up (Aboody et al., 2020).

The Present Study

The goal of this dissertation was to develop and test whether a digital, single-session, self-administered reappraisal intervention could effectively reduce negative interpretation bias and increase positive interpretation bias in college students who tend to make negative social interpretations (e.g., related to social comparison or social rejection) on social media. Social comparison and social rejection were the focus of the present intervention as such experiences are widespread on social media and often central to its harmful effects (McComb et al., 2023). (Though other issues such as "doomscrolling" and exposure to disordered eating content on social media are also highly problematic, it was necessary to limit the intervention's scope given the brief, single-session format).

We chose to develop and test this intervention for college students given the high prevalence of mental health problems among this population (Auerbach et al., 2018), which can be exacerbated by using technology and social media in unhealthy ways (Haddad et al., 2021). College students experience unique stressors compared with early adolescents or older adults that may affect their social media use (e.g., social media can help and hinder the social transition to college; Brown, 2016; Woodley & Meredith, 2012). Recruiting college students enabled us to design the intervention with this population's specific needs in mind. Additionally, (multisession) web-based interventions have shown efficacy in reducing anxiety and depression in college students, suggesting that such interventions are a promising avenue for reaching this high-need population (Davies et al., 2014; Lattie et al., 2019). A meta-analysis of (multisession) web-based

interventions for depression, anxiety, and stress in college students showed that intervention effects were reliable, but only when the interventions were compared to a non-active control condition (vs. to an active control or alternate intervention; Davies et al., 2014). Most interventions had also not involved college students/potential end-users in the process of designing the intervention (Davies et al., 2014). The present study built on this earlier work by: 1) focusing specifically on social media, 2) involving potential college student end-users in the design process from the first stage of development, and 3) using a control condition that closely matched the intervention on activity and social media focus but did not teach reappraisal. The first phase consisted of a user-centered design, iterative development and pilot-testing phase with college students who struggle with negative interpretation bias on social media, to increase the likelihood that the intervention is well-designed to meet the needs of the target population. The second phase involved intervention deployment in a randomized controlled trial and comparison of outcomes across a reappraisal condition and a social media control condition which matched the reappraisal condition in terms of priming social media content and evaluation, time, delivery model, and included many similar task demands, but did not involve reappraisal.

As the primary outcomes, we examined whether the intervention is associated with reductions in online negative interpretation bias and increases in online positive interpretation bias. As a secondary goal, we examined whether the intervention is associated with reductions in offline interpretation bias and increases in offline positive interpretation bias. We also examined whether the intervention leads to reductions in FOMO, depression and anxiety symptom severity, fear of negative evaluation, social comparison orientation (i.e., the tendency to compare oneself to others on social media), and time spent using social media. These outcomes were chosen given their theoretical links with problematic social media use and socially-oriented interpretation bias

(Gupta & Sharma, 2021; Jopling et al., 2020b; Vogel et al., 2015) and potential responsiveness to a reappraisal-focused intervention based on prior research (Alutaybi et al., 2020; Auyeung et al., 2020; Dogaheh et al., 2011; Dryman & Heimberg, 2018; Morello et al., 2023; Smits et al., 2012). We also examined acceptability (i.e., perceptions of the intervention's usability, enjoyability, comprehensibility, and perceived relevance) and perceived effectiveness (i.e., perceptions of the intervention as helpful for using social media in a healthier way).

Phase 1: Iterative Development and Pilot Testing

The intervention was iteratively developed and fine-tuned using feedback from target end-users (i.e., college students who struggle with negative interpretation bias on social media). In Phase 1A, semi-structured individual interviews were conducted to clarify how participants engage with social media, their experiences of interpretation bias across social media platforms, and their opinions about what an intervention focused on shifting interpretation bias on social media should entail. This feedback was used to develop a prototype intervention prior to Phase 1B. For Phase 1B, a new set of participants provided feedback on the prototype intervention via synchronous co-design sessions and written ratings.

Phase 1A Method

Participants

Consistent with the literature on participatory design of digital mental health interventions for young adults (Gulliver et al., 2015; Hetrick et al., 2018; Kornfield et al., 2022; Peck et al., 2020), we recruited 20 participants for Phase 1A through UVA's participant pool in the spring of 2023. To be eligible, participants must have been ≥ 18 years old; had a personal social media account; reported using social media (on one or more platforms) several times a

day[1]; responded with either "very bothered" or "somewhat bothered" to the question, "*How bothered are you by seeing content that triggers your thoughts and feelings related to social comparison or social rejection on social media platforms such as Instagram, Snapchat, TikTok, Facebook, etc.?*"[2]; had access to Internet and a computer or smartphone; and scored ≥ 9.42 (*range* = 4-20) on the four-item "online only" subscale of the Adolescents' Interpretation and Belief Questionnaire (AIBQ 2.0; Miers et al., 2020). The AIBQ 2.0 assesses negative, positive, and neutral interpretation bias in both online and offline social scenarios. The cutoff score of 9.42 is equal to 0.5 standard deviations above the mean total score from the online only subscale in the original study, which was conducted in a convenience sample of university students (M = 7.88, SD = 3.08; Miers et al., 2020). Because the AIBQ 2.0 is a relatively new measure, there are few published studies that report norms for different samples. The score of 9.42 was chosen to help ensure that participants would experience negative interpretation bias at a sufficiently greater level than the convenience sample from the initial study. See pp. 45 - 46 for more information about the AIBQ 2.0. See Appendix A for the eligibility screener.

Procedure

All study procedures were approved by the University of Virginia (UVA) Institutional Review Board prior to recruitment. Participants who appeared eligible based on their response to the screener were contacted by the researcher to schedule interviews. Semi-structured individual interviews were conducted via Zoom in an individual format. The decision to use an individual

[1] This decision was made to establish a reasonable floor for participants' social media use frequency given that more than 70% of 18-29-year-old Snapchat and Instagram users report daily use of these platforms, and more than 50% report using these platforms multiple times per day (Auxier & Anderson, 2021).

[2] Note, the following context was provided to help ensure that participants were reporting on the kinds of concerns targeted in the intervention: "*Social media can bring up negative feelings across many different kinds of situations. For example, some content on social media can trigger people to be self-critical about their body image or feel hopeless about politics/current events. For this research study, we are specifically interested in situations where social media content triggers thoughts and feelings related to social comparison or social rejection. Some examples include thoughts that others are more successful than you, judging you, or don't want to include you*."

(vs. group) format was made given that the focus of the interviews was on participants' personal experiences and history, which we believed participants would be more willing to discuss in a one-on-one context. At the beginning of the meeting, participants were emailed an electronic consent form on Qualtrics that outlined the study procedures, potential risks and benefits, and compensation. Participants had the opportunity to ask the principal investigator questions prior to initiating the interview. After signing the electronic consent form, participants were asked to complete a demographics and mental health history questionnaire in Qualtrics.

All sessions were audio recorded, and a research assistant was present to assist with notetaking. Participants were briefly given background on the study's purpose and encouraged to share a range of their experiences with and perceptions about social media. Participants were asked about several facets of their social media use, including the platforms they use, frequency of use, content they encounter on social media, specific situations on social media that elicit negative interpretations, the extent to which they were content with or desired to change their social media use, and what they would find helpful or unhelpful in an intervention designed to help them interpret content on social media in a more neutral or positive way. Participants were debriefed and compensated with 1 hour of research credit for completing the interview. A general inductive approach was then used to organize feedback and identify key themes emerging from the interviews (Thomas, 2006). This approach is a straightforward, time-efficient, and cost-effective method of analyzing qualitative interview data.

Phase 1A Results

Participant Characteristics

See Table 1 for full demographic characteristics for Phase 1A participants ($N = 20$). With respect to sex assigned at birth, 17 participants (85%) self-identified as female and three (15%)

self-identified as male. With respect to gender identity, 19 participants (95%) self-identified as cisgender and 1 (5%) self-identified as nonbinary. With respect to race, four participants (20%) self-identified as Asian, four (20%) self-identified as multiracial, and 12 (60%) self-identified as White. With respect to ethnicity, four participants (20%) self-identified as Hispanic/Latinx and sixteen (80%) self-identified as Non-Hispanic/Latinx. With respect to sexual orientation, 18 participants (90%) self-identified as heterosexual/straight, one (5%) self-identified as bisexual, and one (5%) self-identified as queer. When asked to report annual household income, 11 participants (55%) reported > $150,000, four (20%) reported $75,000 - $149,999, one (5%) reported < $25,000, and four (20%) did not know their annual household income. When asked to identify the mental health condition(s) with which they were currently struggling from a check-all-that-apply list, four (20%) endorsed at least one mental health condition and two (10%) preferred not to respond. Among participants who endorsed at least one mental health condition, the mean number endorsed was 2.5 (*SD* = 2.4; *range* = 1 - 6). Among all participants, the mean number endorsed was 0.5 (*SD* = 1.4; *range* = 0 - 6). The mental health conditions endorsed by participants were generalized anxiety disorder (*n* = 3), depression (*n* = 2), posttraumatic stress disorder (*n* = 1), bipolar disorder (*n* = 1), eating disorder (*n* = 1), and adjustment disorder (*n* = 1).

When asked about current mental health treatment use from a check-all-that-apply list, five participants (25%) reported being in therapy, three (15%) reported using prescription medications, two (10%) reported seeking social support, one (5%) reported using self-guided help (e.g., self-help book, blogs, online intervention, mobile applications), none reported using over-the-counter medications/supplements, and none reported using other forms of help. All forms of treatment were rated as either “Helpful” or “Somewhat Helpful.”

Summary of feedback

Topic 1: Social media use patterns

Current social media use

Most participants reported using their combined social media accounts for several hours each day. The three platforms that participants reported using most were Instagram, TikTok, and Snapchat. Participants reported that they primarily use Instagram and Snapchat to follow and keep up with friends and other people they know, while TikTok is mainly for following people they do not know (e.g., influencers; accounts aligned with their hobbies and interests). Most participants reported using Snapchat to actively communicate with close friends (e.g., sending messages or "Snaps" through Snapchat's direct messaging feature). Participants reported using Instagram and TikTok much more passively (e.g., scrolling through and observing others' posts, rather than generating their own content or messaging through the app).

Developmental changes in social media use

Several participants identified distinctions between their current social media use compared to earlier in adolescence. For example, P6 described how the transition to college reduced the extent to which she keeps up with social media: "*I went to a smaller high school, there were like 200 people in my graduating class… I could keep up with what everyone was doing, and I knew everyone, so I wanted to be friends with everyone. Like, it was definitely a much more tight knit situation where I like, was able to know what everyone was doing…whereas college is just like, there's so many people here, and obviously I'm not gonna like know what other people are doing at all times… i have my close group of friends that I have made since I came here and we keep up with what one another are doing, but beyond all the people I'm friends with, I don't… I can't keep up with what they're doing all the time.*"

Some developmental changes in social media use were unique to specific platforms. For instance, some participants reported using Snapchat less intensely than they used to (P13: "*I don't use [Snapchat] that often…. When I was younger, like when I first got Snapchat, you Snapchat a bunch of people… but since coming to college, I usually just use it to actually talk to people I went to high school with, or people that I'm close with… the stories on Snapchat, I don't really look at them anyway*, *so that doesn't really affect me as much*…").

Topic II: Triggers for negative self-evaluations on social media

Upward social comparison

Many scenarios that triggered negative self-evaluation on social media involved upward social comparison. Several participants reported being triggered or bothered by seeing posts of others socializing, including posts of people they know hanging out without them (P1: "*if she's friends with other people, it makes her less friends with me*") and big groups of people spending time together (P10: "*Others' lives are so easy*"). These situations were mostly occurring in the context of Instagram specifically. When asked about the emotional impact of these types of comparisons, P4 said, "*You have to like, get used to it at some point… if you have social media, like, I feel like that happens with a lot of people… you're just like, seeing what you don't have, seeing what you lack…it's kind of like a trade-off, I guess.*"

A related theme that came up several times was FOMO. For instance, P5 said, "*I tend to think that people are doing a lot with their lives… they're like, traveling, they're going to nice restaurants, they're like going out, they're with their friends all the time, they're super happy, and then I'm like, well, I'm in my dorm right now, studying, and I'm just spending time like this… everyone's having such a better time than me*."

Some participants were also bothered by social metrics such as number of followers, likes, and Instagram birthday posts from friends (P17: "*if I don't get that many likes, I'll never make it to her level"; "they didn't post on my birthday… they might not value me as much as I value them"*).

Others reported being triggered or bothered by seeing others' accomplishments on social media (P17: "*they're better than me"; "I must suck*"). This typically came up in the context of posts related to academic or career-related achievements (P10: "*Some… people post about… little job opportunities they got… like, [they] got this really cool research opportunity, and I'm just like, yeah, I feel like I'm not doing anything with my life right now… I should be really happy for [them]…but I feel like I'm working my butt off for all the things, but I'm not really getting to the point I want to be at...*"). Also, P18 said, "*If I look at some other UVA first year, and I see that they have all this experience that I don't have, I'd be like, what did I do wrong?").*

Pressure to curate

Some participants described feeling pressure to portray a sense of cohesiveness and to have a "polished" feed, especially on Instagram. P17 said, *"Instagram is where you put your best foot forward… it's kind of like, ok, let's make sure we look like we're having fun"; "There is some pressure… you have to edit a photo... so the lighting is better, or do something so you look the best you can … if I have a photo I think I'll post, I send it to… three different people so they can put different lightings on it*." She added, "*now I think there's a totally different movement where it's like, super cultivated casual, so you see like, a really beautiful picture of a flower, but it has to be like perfectly angled, and it looks amazing, and you post that in like a slideshow post, and it has to look extremely effortless, like but you know it's not effortless, you know they spent so much time picking this out*…"

P18 described holding back from posting on Instagram due to worries about how others will perceive it: "*A lot of people do this thing called a photo dump… I wish I could do that. I just want to post some really weird pictures and show that I'm silly… are people going to think this is weird because it's not traditional?... Are they going to think, why is she posting it here, or are they going to judge it?*".

In contrast to Instagram, TikTok was sometimes identified as a platform where people can be more authentic (P6: "*On TikTok, sometimes people are a little more open, like, definitely more so than Instagram… like sometimes I'll see people posting about their mental health struggles*").

Other triggers for negative self-evaluation on social media

Other situations that participants raised during the interviews were not within the scope of the intervention. These included social media's negative impacts on body image as well as lack of diverse representation of different ethnicities and body types on social media. Others mentioned stress related to social media driven by involvement with certain types of campus groups (e.g., pressures and expectations related to how members of sororities, athletic teams, etc. are supposed to engage with social media).

Additionally, several reactions were raised during the interviews that are not necessarily reflective of biased thinking (e.g., envy when seeing other people have cool experiences; anxiety about not having secured a job before graduation). Thus, in designing the intervention, effort was taken to normalize a range of emotional reactions to social media content and focus on targeting unhelpful thought processes that could make these situations even harder.

Developmental changes in the emotional impact of social media

Some participants described having gained more perspective on social media over time, which has reduced its negative impact. P9 described how pressures to portray a certain aesthetic on Instagram peaked during for her during high school: "*When I was a kid, I did not care… I would post the most random things... like in the era of early Instagram… everyone was kind of strange… I was posting the weirdest things. And then it became, really like, you have to have this aesthetic to your page and everything… I stopped buying into that after a while...it was almost like it had to be perfect…like [social media] was such an imperfect thing… and yet it was your job to make it as perfect as you could, which I think became part of why I was like, this is ridiculous, why am I putting so much energy… why do I care so much...*".

P19 reported how likes are less important to her now than they used to be: "*I think for likes… you can totally hide them if you want, or… not hide them… I think…as I've matured, I realized they don't really matter as much, but when I was younger, for sure, I think they mattered a lot more…*". When asked if she thinks that likes still matter to other people in college, she said, "*Oh yeah, one thousand percent… it's super bad.*" Additionally, P17 said, "*When I was in middle school… I'd be like, oh my god, I only get 100 likes, and they get like 400… but I mean really, these are all the most miniscule things in the world, they literally don't matter… but of course, they mean, like, everything to my thirteen-year-old self.*" On her relationship with social media more broadly, she added, "*Comparison is like, crazy… everyone compares. I compare…. and social media gives… the material for comparison, it just… spoon-feeds it to you, even when you're trying to be like, no no no, I'm chill with myself, like everything's good… I hope to think now that I'm older and wiser… I can separate from this….it doesn't mean I'm separate from FOMO or comparison by any means…I think it's really hard to shake it, but it's to a lesser degree now, because it all literally means nothing, like… I've come to realize that, basically…*"

Topic III: Social media coping strategies and goals

Current coping strategies

Participants reported using a broad range of coping strategies including distraction (e.g., sleeping; exercising; doing work), limiting time spent on social media (e.g., taking a social media break), using positive self-mantras, and seeking support from others. Participants also described trying to think differently (including using reappraisal) to cope with social media's negative impacts (P15: "*I'm a first year… and my biggest source of anxiety this year was not having... a solid friend group. I would see on Instagram all of these people with their solid friend groups, hanging out with them all the time…and then I used to feel...very bad about that, but then when I…talked to those people in real life, they told me that it's not all that it's…cut out to be. And I'm like, on social media, I perceive it so differently… I used to think certain groups were… perfect… I had that in high school but I don't have that in college right now…but I used to be very jealous of it. Then I'd talk to friends and they'd tell me, we're not all friends, or... I have problems with them, or we don't hang out as often, we're there out of convenience… on social media, we want it to look good*. I *only post when I look and feel good, and they obviously do the same.*"). P6 also said, "*I think… if I was wherever they are, doing what they were doing, I would probably want to take pictures too... knowing the context of social media, especially of people our age, and how important it is to take pictures to post… is an important context.*"

Other participants described discounting positive aspects of others' experiences or engaging in downward social comparison. P2 said, *"it's not as good of a practice, but I might just… kind of downplay the enjoyment that they're experiencing, or say… if I were there, it probably wouldn't be that great…*". P4 said: "*it's not…the nicest thing to do, because I'll be like,*

oh, I think I'm prettier than her, or I'm definitely smarter… it kind of… pacifies… the negative thinking for a while… because it's like, well, I might not have this, but I have this…".

Desire to use social media less passively and more actively

When asked how they wished they used social media differently, most participants described a desire to use it less frequently and/or in a more active way (e.g., generating more of their own content, instead of just scrolling passively). For instance, P11 said, "*I wish I got off of… social media a lot more… I feel like that would make my life easier, like, I would have more time and also stop comparing myself…. But I also wish I just posted more, so that I had a feed of what my life was like as well… something to look back at*." When asked what interferes with her posting more, P11 said, "*I feel like I always wait for the perfect moment, and then I never end up taking pictures… and then I'm like, I only have a few pictures, I keep waiting, and then I end up never posting… I feel like I need to wait for that perfect layout, so it just ends up being that I never post*." P2 also expressed a desire to share more photos on social media, "*partially to have more to look back on, and partially because when you take photos with other people it helps strengthen the bond between you two*."

Topic IV: Suggestions for intervention design and framing

Reminders that social media is filtered

When asked what they think would be important to include in the intervention, many suggested including reminders that people selectively share positive aspects of their lives on social media (P6: "*People are not going to post the lowest lows of their life for everyone else to see, for the most part… people are going to be posting the higher points of their life, and… that's not a fair comparison to make because obviously you have the context of your everyday life when you're comparing yourself to what you see on social media*"). P8 said, "*You have to remember*

that [on] social media, people choose and curate what they're posting, so it's a very unnatural place to see what people are doing… so it's not really a reflection of real life, I guess… the best parts are what people want you to see, and I think that's helpful to remember".

Increasing positive self-focus

Some participants thought the intervention should promote positive self-reflection and mindfulness. P12 said, "*Instead of being like, what's wrong with me... or why am I not with them, why am I not like them, be like oh, here's something good about where I am that I might not necessarily have if I was with them or if I looked like them or something*." Likewise, P8 thought it would be helpful to remind users that "*Everyone is different and they have their own goals in life … what other people choose to do doesn't mean that you have to do it*." Similarly, P10 suggested reminding people, "*You're doing great as a person… you're making as much progress as anybody else… everybody else just has a different pathway in achieving their goals and you as well also have a different pathway in achieving your goals…all that matters is what you are doing and what goals you want to achieve*". P18 also suggested including a reminder that "*Other people care way less than you think they do… they're not thinking as hard as you are about yourself*".

Others suggested increasing peoples' awareness of alternative options for managing their experience on social media, such as unfollowing or muting accounts, and advocating for breaks from social media to alleviate negative feelings.

Preference for a reminder

Participants also provided feedback on their preferences for having a reminder to use their new skills after finishing the intervention. Several people said that reminders would be helpful for continued practice and implementation of skills. However, some said that reminders –

especially frequent ones – could be annoying. There was conflicting feedback on the preferred format for such a reminder, with some stating that text message would be better and others preferring email.

Potential obstacles to engagement

Though many participants said that they thought the intervention would be helpful for other students at UVA, some raised potential obstacles that could hinder engagement. For instance, some thought that a web-based delivery format would be less preferred to an app format. Another participant believed that users would only benefit from the intervention if they were intrinsically motivated to change (P18: "*It will only be helpful… for people who want it…it also has to come from inside… if someone isn't looking to change, it won't help… so I think it will help the students who are really interested in changing their mindset*).

Intervention Development Procedure

Data collected during the interviews were organized into overarching themes and used to develop content for a prototype intervention, titled PRISM ("**P**rogram for **R**ethinking Negative **I**nterpretations on **S**ocial **M**edia"). The prototype was grounded in features of established single-session interventions as well as Phase 1A user feedback. Common situations that were cited by participants as invoking negative self-evaluations were organized into a pool of vignettes from which participants could select to increase personal relevance. Feedback on the prototype was provided by members of the Program for Anxiety, Cognition, and Treatment (PACT) lab. Additionally, a control condition was concurrently developed to match PRISM on several key features, including duration, activities, and relevance to social media. PRISM and the control condition were both developed and administered in Qualtrics.

Design Choices Based on Phase 1A Feedback

- **Psychoeducation**. The psychoeducation content throughout PRISM was designed to present a balanced view of social media, rather than describing it as unequivocally problematic. The choice to present a balanced perspective was made given that many participants also described positive impacts of social media on their lives (e.g., increasing social connections; providing entertainment). For instance, early on, participants are presented with content that says, "*Social media has a bad reputation. But research shows that social media can be good or bad. It depends on* ***how*** *people use it*." The next section says, "*Social media can promote connection and self-expression. It can also make people feel left out or compare themselves*." Participants then complete an exercise where they reflect on negative and positive aspects of their own personal social media use.
- **Social media problem domains**. Social media situations described by participants as triggering negative self-focused interpretations were identified, and then used to develop six problem domains (and accompanying vignettes) that were featured in PRISM. We aimed to compile an inclusive list of domains such that most potential users would find at least a few of the domains relatable. Accordingly, with the exception of two domains that referenced Instagram, we refrained from tying the domains to any specific social media platform. The decision to reference Instagram for those domains was made because many participants cited Instagram specifically as the platform most triggering for their negative self-focused thoughts in general, and for social comparison specifically.
- Additionally, we attempted to avoid overly specific vignettes. Though we considered including "Messing up the grid" and "Weird photo dumps" as separate domains, we ultimately chose to present "Messing up your feed's aesthetic," which is broader. We also considered including separate domains specific to LinkedIn updates as well as TikTok's

"Get ready with me" videos, but to increase generalizability, chose to have "Seeing others' successes" instead. The final list of domains was "Instagram birthday posts," "Number of likes and comments," "Big group photos," "Messing up your feed's aesthetic," "Losing followers," and "Seeing others' successes."

- **Email reminder**. The decision was made to send a brief, one-time reminder email in the two-week interim between intervention completion and the follow-up questionnaires. Of note, though some participants expressed a preference for personalized email reminders, the decision was made to send a standard, non-personalized reminder to increase feasibility and scalability.

Description of Intervention Conditions

Experimental Condition (PRISM). Participants are first told that the research team is studying how to encourage healthier social media use. Psychoeducation related to positive and negative aspects of social media is included to convey a non-biased stance. To set the stage for reappraisal, participants are asked to identify their three most-used social media platforms and write one positive/negative aspect of how each platform affects their life in an embedded text box. The intervention then discusses how social media can lead to negative thoughts and how reappraisal (termed "thinking flexibly") can help mitigate the impact of negative thoughts related to social media. Following principles of "*wise interventions*" (Walton & Wilson, 2018), participants then practice thinking flexibly using two standardized vignettes: one that is about a general (non-social-media related) social situation, and one that is specific to social media (Schleider et al., 2020). Participants are given context about the limitations of thinking flexibly, including that thinking flexibly is not a strategy to normalize bullying or harassment, and

provided with appropriate resources should they be dealing with bullying or harassment, such as information for finding in-person mental health services and supportive groups around grounds.

Following other SSIs (e.g., Schleider & Weisz, 2016; Wasil et al., 2021), participants then practice the new reappraisal skill. Participants are presented with six social media problem domains that are ripe for negative self-evaluations tied to social media, which are each linked with an accompanying vignette. Following Phase 1 feedback, participants choose three domains (from six options) to increase personal relevance. The options include: “Instagram birthday posts,” “Number of likes and comments,” “Big group photos,” “Messing up your feed’s aesthetic,” “Losing followers,” and “Seeing others’ successes.” Each option links to a new page in Qualtrics where the participant is introduced to a vignette that includes an illustrated character, a brief written scenario told from first-person perspective, and an accompanying thought representing a negative interpretation the character made in response to the scenario provided. For example, participants who choose to work on “Instagram birthday posts” are presented with a character named Taylor who says: "*It was my birthday last week. A few of my friends posted for me on Instagram, but I definitely did not get as many posts as I was expecting. There's also a lot of people who didn't post for me even though I posted on their birthdays. They don't value our friendship as much as I do*." Then, in an embedded text box, the participant is then asked to write out another way of thinking about the situation that they might suggest to Taylor.

In the final section, participants are asked to identify a time when social media led to negative thinking in their own lives, so they can further practice thinking flexibly. Participants can choose from one of the six preexisting themes from earlier in the program or write a custom situation into an embedded text box. (The decision to allow a custom write-in option alongside prepopulated options was made to strike a balance between participants’ stated preferences to

generate their own responses and recognition that writing prompts can decrease engagement with self-administered interventions; Dobias et al., 2022). Participants are then asked to write a negative thought that they might have related to the situation they chose, as well as another way of thinking about this situation given what they have learned about thinking flexibly.

Following this activity, participants are provided with a summary of take-home points and reminded to try and think flexibly when faced with negative thoughts. They are then reminded that reappraisal is *not* intended to be invalidating or suggest one should accept experiences with online discrimination, bullying, or harassment. At the end, resources are provided for individuals facing such issues. See Appendix B for examples of content from PRISM related to thinking flexibly.

Social Media Control Condition. Research on single-session interventions underscores the importance of employing relevant, credible control conditions that are closely matched to the intervention (Schleider et al., 2020). Thus, the control condition for this study includes activities that prime individuals to reflect on social media and to share advice related to social media with others, but does not include psychoeducation/reappraisal training.

Participants are first told that the research team is studying how to encourage effective social media use. They are then presented with four separate pages of tables that contrast features across popular social media platforms (Instagram, Snapchat, TikTok, Twitter, and Facebook). These include communication features (e.g., Instagram's Direct Messaging), algorithm and content discovery features (e.g., TikTok's "For You" page), privacy and visibility features (e.g., "Ghost mode" on Snapchat), and special features (e.g., "Facebook Marketplace"). On each page, participants are asked to write one feature that they like and one they do not like in an embedded text box. Finally, participants are asked to write advice to a hypothetical college student trying to

decide which social media platforms to use in an embedded text box. To reduce the likelihood that the control condition is influencing participants to reflect on how social media impacts them emotionally (which is one of the goals of PRISM), control condition participants are told to limit their feedback to technical features of the social media platforms.

Phase 1B Method

Participants

The eligibility criteria for Phase 1B were identical to those in Phase 1A. For Phase 1B, we recruited 10 participants via the participant pool in the spring of 2023 to provide feedback on the PRISM prototype. During the summer of 2023, seven additional participants were recruited from the Behavioral Research at Darden (BRAD) Lab, an interdisciplinary laboratory affiliated with UVA's Darden School of Business, to supplement feedback due to low attendance at co-design sessions (see Procedure, below). All participants recruited from BRAD lab were UVA undergraduates who met the same eligibility criteria as those recruited from the participant pool.

Procedure. All study procedures were approved by the UVA Institutional Review Board prior to recruitment. Following prior studies with participatory, user-centered design, Phase 1B involved a series of synchronous, online group co-design sessions (e.g., Hetrick et al., 2018; Kornfield et al., 2022) with participants invited to attend four sessions to provide feedback on the PRISM prototype. The group format was intended to facilitate group convergence/divergence about the intervention design and content. At the beginning of the first meeting, participants were emailed an electronic consent form on Qualtrics that outlined study procedures, potential risks and benefits, and compensation. The consent form also included a confidentiality contract requesting that participants respect the privacy of one another and not share information discussed during the co-design sessions outside of the study. Participants had the opportunity to

ask the principal investigator questions prior to initiating the interview. After signing the electronic consent form, participants were asked to complete a demographics and mental health history questionnaire in Qualtrics. During each session, the principal investigator led participants through the current version of the prototype. Participants were asked to complete and provide feedback on several tasks embedded within the prototype. Participants could raise any feedback they had related to content, presentation, and/or design. Though the first co-design session was well-attended (with 9 out of 10 participants present), there was low attendance at subsequent sessions (i.e., 0 - 2 present out of 10 participants) and ultimately only three co-design sessions (out of the planned four) took place. To supplement feedback, the prototype was also reviewed by undergraduate research assistants in the PACT lab.

Once the prototype was finalized, it was sent to all Phase 1B participants along with an acceptability questionnaire to assess feedback. Participants rated acceptability using the Program Feedback Scale (PFS; Schleider et al., 2019), a measure designed to assess perceptions of single-session interventions. It contains seven Likert questions, ranging on a five-point scale from "*really disagree*" to "*really agree*", and two open-ended questions.[3] Likert questions assess factors such as usability, enjoyability, comprehensibility, and perceived relevance, while the open-ended questions ask what people liked and what they would change. Questions were also added by the research team to assess specific features of PRISM, including the extent to which participants thought that PRISM was likely to help reduce their tendencies to make negative interpretations related to social media. PFS scores were reviewed, and feedback was incorporated into the prototype prior to deployment in Phase 2. Due to a low response rate on the feedback

[3] Some a priori modifications to the wording of the PFS were made to increase relevance (e.g., for the item, "*I think the program would be helpful to other kids my age,*" 'kids' was replaced with 'college students').

survey (i.e., only three participants out of the expected 10 completed the PFS), seven additional participants were recruited from BRAD lab to review the prototype and provide feedback.

Participants recruited from the participant pool were compensated with 1-5 hours of research credit depending on how many co-design sessions they attended and whether they completed the PFS. Participants recruited from BRAD lab were compensated with a $20 Amazon gift card for reviewing the prototype and completing the PFS.

Phase 1B Results

Participant Characteristics

See Table 2 for full demographic characteristics for Phase 1B participants ($N = 17$). With respect to sex assigned at birth, 12 participants (70.6%) self-identified as female and five (29.4%) self-identified as male. With respect to gender identity, 17 participants (100%) self-identified as cisgender. With respect to race, four participants (23.5%) self-identified as Asian, two (11.7%) self-identified as Black/African American, one (5.9%) self-identified as multiracial, and nine (52.9%) self-identified as White. With respect to ethnicity, all participants (100%) self-identified as Non-Hispanic/Latinx. With respect to sexual orientation, four participants (23.5%) self-identified as bisexual, two (11.8%) self-identified as gay/lesbian, and 11 (64.7%) self-identified as heterosexual/straight. When asked to report annual household income, ten participants (58.8%) reported > $150,000, four (23.5%) reported $75,000 - $149,999, one (5.9%) reported $50,000 - $74,999, one (5.9%) reported $25,000 - $49,999, and one (5.9%) reported < $25,000. When asked to identify the mental health condition(s) with which they were currently struggling from a check-all-that-apply list, eight (47.1%) endorsed at least one mental health condition and one (5.9%) preferred not to respond. Among participants who endorsed at least one mental health condition, the mean number endorsed was 1.8 ($SD = 1.2$; *range* = 1 - 4). Among all

participants, the mean number endorsed was 0.8 (*SD* = 1.2; *range* = 0 - 4). The mental health conditions endorsed by participants were obsessive-compulsive disorder (*n* = 3), social anxiety disorder (*n* = 3), generalized anxiety disorder (*n* = 2), depression (*n* = 2), attention-deficit/hyperactivity disorder (*n* = 1), eating disorder (*n* = 1), and specific phobia (*n* = 1).

When asked about current mental health treatment use from a check-all-that-apply list, four participants (23.5%) reported using prescription medications, three (17.6%) reported being in therapy, three (17.6%) reported seeking social support, two (11.8%) reported using self-guided help (e.g., self-help book, blogs, online intervention, mobile applications), none reported using over-the-counter medications/supplements, and none reported using other forms of help. All forms of treatment were rated as either "Helpful" or "Somewhat Helpful."

Summary of Feedback from Co-design Sessions

In terms of content, participants understood and appreciated the rationale for thinking flexibly and believed it would resonate with other college students. When asked, they said they did not think that the concept of thinking flexibly would be seen as invalidating or trivializing of peoples' issues. Participants generally liked themes and accompanying vignettes offered mid-way through the intervention and found them to be relatable and realistic. People said that the vignettes covered the majority of stressful experiences most people have with social media, and they did not think any major categories were missing. Most people said they could relate to more than one vignette and would likely choose more than one if given the option. Participants also helped to fine-tune the wording of the vignettes. For instance, for a vignette about an individual who is envious of the number of likes and comments their roommate got on a photo, participants suggested increasing the number of likes from what was written to be more realistic.

In terms of flow and level of engagement, participants consistently reported liking the interactive parts of PRISM. They enjoyed identifying their most used social media platforms and writing about how those platforms affected their lives. Additionally, participants liked being able to choose the vignettes that were most relevant to them, both when practicing thinking flexibly to respond to someone else's negative thought and when practicing thinking flexibly to respond to their own negative thought. When asked whether they would prefer to complete two vignettes or three vignettes during the intervention, all participants preferred three vignettes. When asked about the amount of writing in the intervention, participants generally thought that writing was helpful, especially at the end when practicing responding to one's own thought. However, one participant said she found it difficult to identify her own scenario to work on at the end. It was decided that for the final task, users would be able to choose from the list of pre-existing vignettes or elect to write in their own custom scenario.

In co-design sessions 1 and 2, some participants expressed that the intervention was not engaging enough to hold their attention and provided suggestions for ways to make it "flashier" and more "attention grabbing." They suggested switching the vignettes from third person to first person and adding a voiceover, such that Qualtrics would have embedded audio-recordings of each vignette being read aloud in first person. After discussion amongst the research team and with participants, the decision was ultimately made to not add a voiceover, to minimize logistical issues (e.g., an auditory component could limit *where* people may access the intervention) and deception (e.g., keeping the vignettes intentionally vague and unidentifiable so as to not imply they are about real people), and to increase generalizability (e.g., a voice can imply certain demographic characteristics of the speaker that limit how much a participant sees oneself reflected in the vignettes). To make the vignettes more engaging, the written vignettes were

switched from third person to first person tense. Additionally, in response to participants' requests for more images, custom illustrations were created by a member of the study team to represent the characters in each vignette.

Summary of PRISM Descriptive Data

When participants were asked to identify their three most-used social media platforms towards the beginning of PRISM, 10 chose Instagram, eight chose TikTok, six chose Snapchat, five chose Twitter, one chose "Other" (wrote in "Pinterest"), and none chose Facebook. When participants were asked to select the areas to work on midway through the intervention, nine chose "Number of likes and comments," six chose "Seeing others' successes," five chose "Losing followers," four chose "Instagram birthday posts," four chose "Big group photos," and none chose "Messing up your feeds' aesthetic."

Summary of Ratings on Program Feedback Scale

Feedback on the prototype was generally positive. Based on responses to the PFS, participants generally found PRISM to be acceptable, especially in terms of understanding it (M_{score} = 3.8/4) and agreeing with its message (M_{score} = 3.7/4). The lowest score was on the item assessing likeliness to recommend PRISM to a friend going through a hard time (M_{score} = 2.3/4). See Table 3 for full scale results from the PFS.

The qualitative feedback from the PFS was also generally favorable. When asked what they liked about PRISM, participants said they appreciated its relevance to college students. One participant wrote, "*I liked how it felt like the researcher understood actual concerns with social media instead of usual things you see from adults who don't use social media often.*" Another wrote, "*I think the examples weren't far off from what college kids might see and think based on their social media habits. It didn't feel inauthentic to the subject matter.*" Another participant

appreciated the balanced view of social media presented: "*I liked being asked about the positive and negative aspects of the social media apps I use most. It highlights the fact that social media isn't always bad and is realistic*." One participant appreciated PRISM's flexibility, especially regarding the ability to write about their own interpretations and experiences throughout. Finally, two participants commented that they appreciated PRISM's simplicity and directness, and two commented that they liked the visual images/graphics.

When asked what they would change about PRISM, some noted that they felt it was too brief: "*I think it was a little too short to make a meaningful impact or change my thinking. I think for it to make more of an impact I would have to practice it everyday, or every time I went on social media, because I would probably forget about PRISM after doing this questionnaire*." Another participant said: "*I think maybe the explanation and solution of flexible thinking is a bit reductive. Even if changing an outlook is necessary, it's not that easy for a lot of people or they would've already done so. Maybe flexible thinking could be one of a few different proposed mental strategies to cope with social media pressure rather than the sole one*." Given that Phase 1A participants reported using several different coping strategies to manage social media's impact, we added a slide towards the end of PRISM briefly reminding participants about other strategies (e.g., seeking support; taking a break from social media).

Other participants described a desire for greater personalization and flexibility. One participant suggested that PRISM "*Offer several preconceived flexible ways of interpreting scenarios so the user is exposed to ideas and strategies they may not have thought of*." Following this feedback, some additional examples of automatic thoughts were added into the vignettes at the point thinking flexibly is introduced. Some participants also shared thoughts that they believed would be helpful to emphasize in the intervention as advice for other students. These

were paraphrased and added into a slide towards the end of PRISM (e.g., "*Try to remember that... It's okay to have bad days and not be able to think positively, but just try your best to reframe your state of mind about social media before going on"*).

Phase 1 Overall Discussion

Phase 1 entailed the collection and synthesis of data from interviews, focus groups, and written feedback questionnaires to inform the development of PRISM. Phase 1A unearthed several commonalities and distinctions in how college students at UVA struggle and cope with negative thinking related to social media, while Phase 1B enabled us to achieve consensus on content and design.

Many insights gleaned from Phase 1A were consistent with the existing social media literature, such as the way that social media serves as a catalyst for social comparison (Meier & Johnson, 2022). Also consistent with prior research, Instagram was most consistently implicated in experiences of negative thinking and social comparison in particular (Faelens et al., 2021). This was related to Instagram's role as the place where people announce important life updates, share personal and professional successes, and show off meaningful or impressive experiences. Additionally, participants' feedback about distress related to the number of likes and comments on social media posts aligns with research showing the negative mental health impacts of such social metrics (Lee et al., 2020; Radovic et al., 2017).

Many felt conflicted about social media, recognizing it as unimportant in the grand scheme of their lives yet still feeling pressure to portray themselves well and desiring positive feedback from others. Relatedly, some also felt less impacted by social media now compared to when they were in middle school or high school. This finding underscores earlier and mid-adolescence as an especially vulnerable time period for social media use (Martínez-Ferrer et al.,

2018; McCrae et al., 2017; Shafi et al., 2019). Though our intervention was designed for college students, future adaptations for younger adolescents may be beneficial.

Many participants also shared common thoughts and emotions, including envy and FOMO, that are triggered by social media. An important distinction that became apparent during interviews was that such responses are not necessarily reflective of biased thinking and constitute reasonable reactions to certain social media content. This underscored the importance of making the intervention validating to a wide range of social media experiences.

Other topics mentioned in the interviews were not within the scope of PRISM's specific focus but are worth noting as they make up the larger tapestry of individuals' experiences with social media. Consistent with prior research, some participants reported that their body image is negatively impacted by social comparison on social media (Sidani et al., 2016). Relatedly, one participant described how the lack of representation of individuals from her ethnic background on social media negatively impacts her body image and self-image more broadly. This example highlights one way that social media can enact harm towards individuals holding marginalized identities (Sobande et al., 2020). A few participants described how involvement in Greek life or university athletics compounds social media pressures, referencing strict rules and norms about how sorority members and university athletes are supposed to portray themselves online (Comer, 2022). Though the decision was made to not focus on these areas for the intervention, efforts were made to be responsive to these issues by acknowledging them and directing people to other forms of support.

Though much of the feedback in Phase 1B was positive, several issues raised by participants at that stage reflect larger debates within the digital intervention development space. For instance, the feedback that PRISM was too short to be impactful reflects the dilemma that

researchers have between making an intervention substantive enough to be effective, but short enough to hold interest. This feedback influenced our decision to include a reminder email to increase use of reappraisal (Borghouts et al., 2021; Saleem et al., 2021). The feedback from participants that a broader range of coping strategies (besides reappraisal) should be included in PRISM also relates to the issue of defining intervention scope. Though we briefly listed some alternative coping strategies at the end of PRISM, we ultimately kept the focus on reappraisal to maximize brevity and isolate specific effects.

Additionally, some participants' desires for more and different types of scenarios and ways of responding to unhelpful thoughts reflects larger issues related to the tradeoff between enhanced personalization and cost/time efficiency (Hornstein et al., 2023). Though we hoped that participants would be able to relate to most vignettes included in PRISM, we allowed write-in scenarios for the final exercise to increase personal relevance. However, we made the decision to include writing judiciously given research has shown that engagement falls off when participants are required to write in responses (Dobias et al., 2022).

Strengths of Phase 1 include the use of multiple methods of collecting user feedback, including in-depth interviews and co-design sessions, which allow for a nuanced understanding of college students' experiences and perceptions. Additionally, though samples for Phases 1A and 1B were small and primarily composed of cisgender female participants, there was diversity with respect to sexual orientation across both phases. This is a critical perspective to include as social media can be uniquely harmful for sexual minority college students but can also serve as a channel for identity expression and social connection (Talbot et al., 2022).

Several limitations must be acknowledged as well. The sample was primarily White and non-Hispanic/Latinx, which limits generalizability, though the racial/ethnic composition across

both phases was comparable to that of the overall UVA student body. Additionally, there was low attendance at co-design sessions in Phase 1B, which required supplementing feedback from additional participants. The compensation of course credits offered to participant pool participants was likely not incentivizing enough in Phase 1B given that participants were asked to meet for at least one (and up to four) synchronous meetings. Though the research team expected that most participants would opt to join more than one co-design session, making this an explicit requirement would likely have improved attendance and participation. Finally, though use of a general inductive approach for qualitative analysis was reasonable given the project's timeline and resource constraints, it is unknown how well our results would overlap with more involved qualitative analysis methods (e.g., thematic analysis; grounded theory analysis).

Phase 2: Randomized Controlled Trial

Overview and Hypotheses

Phase 2 was a randomized controlled trial to compare the effects of PRISM with the social media control condition on outcomes related to interpretation bias, mental health status, and social media use/behaviors, both immediately post-intervention and at two-week follow-up. Assessing both immediate and two-week outcomes could provide greater insight into durability of effects while also enabling participants to practice skills over time. The two-week follow-up timeframe was chosen based on similar studies evaluating digital, single-session reappraisal interventions (Ranney et al., 2017; Rodriguez et al., 2019, 2021).

All hypotheses were preregistered through Open Science Framework prior to data analysis (see https://osf.io/nwxqe). The primary test hypothesized that PRISM would lead to greater reductions in online negative interpretation bias mean scores and increases in online positive interpretation bias mean scores relative to the social media control condition,

immediately post-intervention and at two-week follow-up. There were competing hypotheses as to whether effects would be stronger immediately post-intervention or at two-week follow-up. On the one hand, effects might be stronger when the skills are fresh in participants' minds and less likely to have attenuated over time (McDanal et al., 2022). On the other hand, effects might be stronger after participants have had more time to practice and consolidate information learned via the training (Denny & Ochsner, 2014).

As secondary outcomes, we examined whether PRISM leads to the same changes in negative and positive interpretation bias immediately post-intervention and at two-week follow-up, but for the *offline* scenario subscale. While it was plausible that PRISM would shift offline interpretation bias, it was unclear how well the social media-specific training would generalize to general social scenarios.[4] Thus, we expected to see some reduction in offline bias but a weaker effect than was anticipated for the online bias. Again, there were competing hypotheses regarding whether effects would be stronger immediately post-intervention or at two-week follow-up.

We hypothesized that PRISM would also lead to greater improvements in other secondary outcomes at two-week follow-up, including reductions in FOMO, depression symptom severity, anxiety symptom severity, fear of negative evaluation, and social media-related social comparison orientation, relative to the social media control condition. There was a competing hypothesis for time spent using social media: it is plausible that considering the negative impacts of social media on one's life, which is part of PRISM, could prompt someone to cut back on their social media use. However, PRISM does not explicitly promote decreasing time spent on social media, and some people may feel encouraged to use social media after

[4] Of note, online and offline interpretation bias items loaded onto two separate factors in the AIBQ 2.0 measure development paper (Miers et al., 2020).

completing PRISM (e.g., because it no longer makes them feel as bad). Thus, while PRISM might be associated with reductions in time spent using social media, it is also possible that PRISM will have no effect on this variable. The literature also suggests that time spent using social media is not a valuable indicator of social media's impact on social/emotional wellbeing (Coyne et al., 2020; Steinsbekk et al., 2023). Notably, however, our participants were selected based on using social media frequently and with some negative emotional impact. This suggests that reduction in time spent using social media could be beneficial and was worth examining, even though this was not a primary intervention target.

PRISM was also expected to be associated with greater acceptability and perceived effectiveness relative to the social media control condition, both immediately post-intervention and at two-week follow-up. (No changes in perceived effectiveness or acceptability were hypothesized from immediately post-intervention to two-week follow-up).

Finally, for the PRISM group (vs. the control group), we expected greater endorsement of using the three intervention-related target skills (e.g., thinking flexibly) at two-week follow-up. For the control condition group (vs. the PRISM group), we expected greater endorsement of using the two control condition-related target skills (e.g., thinking about technical aspects of social media platforms) at two-week follow-up.

Phase 2 Method

Power Analysis

Prior to data collection, a power analysis using the 'WebPower' package in R (Zhang et al., 2018) was conducted to determine the sample size needed to compare the study conditions at two timepoints. The effect size used for this analysis was based on a meta-analysis of 50 studies examining the efficacy of SSIs for youth (Schleider & Weisz, 2017) which found a small-to-

medium overall effect (g = 0.32). This effect size was used in the power analysis with α (two-sided) set at .05 and power set at .90. The power analysis revealed that an N of 105 participants would be necessary to achieve power to detect a small-to-medium effect. Of note, it is typical for digital, single-session intervention studies to lose about one-third of participants at post-intervention assessment (e.g., Dobias et al., 2021; Rodriguez et al., 2021). To account for a likely dropout rate of roughly 33%, we made the a priori decision to recruit 160 participants (80 per condition).

Participants

As in Phase 1, college students (N = 165) were recruited through UVA's participant pool. Recruitment took place during the fall 2023 semester. The eligibility criteria for Phase 2 were identical to those in Phase 1 (see Appendix A).

Procedure

All study procedures were approved by the UVA Institutional Review Board prior to recruitment. Once a participant was deemed eligible based on their screener survey, they were randomized to either PRISM or to the social media control condition with a 50/50 distribution, stratified by sex assigned at birth using the "blockrand" package in R (Snow, 2020). Participants were emailed an invitation to participate in the study with the appropriate link in Qualtrics based on their condition. The email instructed participants not to begin until they had 20-25 minutes of time, strong internet access, and privacy to help them fully engage with the study.

The first Qualtrics page presented an informed consent form that outlined study procedures, potential risks and benefits, and compensation. The survey ended if participants did not consent to the study procedures. Those who provided informed consent were presented with several questionnaires collecting demographic information and assessing interpretation bias and

secondary outcomes prior to initiating their assigned study condition. Immediately after the assigned study condition was completed, participants were re-administered the interpretation bias measure and asked to complete a feedback questionnaire.

One week after the baseline session, participants were sent a brief email from the research team reminding them to engage with the study skills (see Appendix C for the emails sent to participants in each condition). Two weeks after their baseline session, participants were sent another email asking them to complete the same battery of questionnaires from baseline plus some additional feedback items. Participants who did not respond within two days of the initial email were prompted again to complete the questionnaires, receiving up to three prompts during the week following the initial follow-up contact. Participants were compensated with 0.5 hours of research credit for completing the baseline study procedures and 0.5 hours of research credit for completing the follow-up questionnaires two weeks later.

Measures

Program Acceptability. The Program Feedback Scale (PFS; Schleider et al., 2019) is a 9-item questionnaire designed to assess users' perceptions of single-session interventions. An example item reads, "*I would recommend this program to a friend going through a hard time*." For the present study, six additional questions were added to capture study-specific constructs (e.g., "*I think the program will help me think more flexibly about what I see on social media*"). The present study's version contained 13 Likert-questions on a five-point scale with response options ranging from 0 ("Really Disagree") to 4 ("Really Agree"). Additionally, two open-ended questions assessed what participants liked and what they would change about the program. Some a priori modifications were made to item wording to increase relevance (e.g., changing "*other kids my age*" to "*other college students*"). The PFS was administered as part of the immediate

post-intervention battery and two-week follow-up battery. In the current sample, internal consistency for the PFS was excellent at both timepoints (ω = .94 at post-treatment and ω = .96 at two-week follow-up).

Perceived Effectiveness. Perceived effectiveness was assessed using the standalone researcher-authored item, "*I think that having completed this program will help me use social media…*" with possible response options ranging from 1 ("Much less effectively") to 5 ("Much more effectively"). Perceived effectiveness was assessed as part of the immediate post-intervention battery and two-week follow-up battery.

Engagement with Target Skills. As a form of sensitivity analysis, engagement with specific target skills related to PRISM and to the control condition, respectively, were assessed at two-week follow-up. The PRISM-related questions assessed the extent to which, over the two weeks since completing the study, participants: 1) noticed social media was triggering their thoughts/feelings of self-comparison/not being liked, 2) tried to think differently when social media was triggering thoughts/feelings of self-comparison/not being liked, and 3) thought about what they would say to someone else for whom social media was triggering thoughts/feelings of self-comparison/not being liked. The control condition-related questions assessed the extent to which, over the two weeks since completing the study, participants: 1) thought about how technical and/or user-design aspects of social media platforms could be improved, and 2) discussed technical and/or user-design aspects of social media platforms with people in their lives. Participants answered all five questions regardless of study group to enable group-level comparisons on skill engagement.

Online and Offline Interpretation Bias. The Adolescents' Interpretation and Belief Questionnaire (AIBQ 2.0; Miers et al., 2020) is an 8-item questionnaire that was used to assess

negative and positive interpretation bias in online and offline social scenarios. After they are presented with a scenario, participants rate the extent to which they agree with a positive and negative interpretation on a scale from 1 ("Doesn't pop up in my mind") to 5 ("Definitely pops up in my mind"). An example online social scenario reads: "*You post a photo of a tasty dish that you've made on Instagram. After an hour, one of your followers responds, 'What dish is that?' What is meant by this response?*" The positive interpretation presented for this scenario was: "*This person really likes the look of the dish and would like to cook it themself.*" The negative interpretation presented for this scenario was: "*I've clearly made a mess of the dish because it isn't recognizable.*" An example offline social scenario reads: "*You've invited a group of classmates to your birthday, but a few have not yet said if they're coming. Why haven't they said something yet*?" The positive interpretation presented for this scenario was: "*They're definitely coming; they don't need to tell me that.*" The negative interpretation presented for this scenario was: "*They don't want to come because they don't like me.*" Separate sum scores are calculated for responses to positive and negative interpretation statements, with higher scores reflecting greater positive and negative interpretation biases, respectively. Some modifications were made to the measure's wording to increase relevance for participants, such as changing the term "*fellow student*" to "*classmate*." Given that the AIBQ 2.0 is a relatively new measure, there has not been formal evaluation of its psychometrics, though the parent measure (AIBQ; Miers et al., 2008) has been used extensively (de Hullu et al., 2017; Gibb et al., 2022; Sherman & Ehrenreich-May, 2018). The AIBQ was administered as part of the pre-intervention battery, immediate post-intervention battery, and two-week follow-up battery.

In the current sample, internal consistency ranged from questionable to acceptable for online negative interpretation bias (ω = .64 at baseline; ω = .77 at post-treatment, and ω = .74 at

two-week follow-up) and questionable to good for offline negative interpretation bias (ω = .68 at baseline; ω = .88 at post-treatment, and ω = .79 at two-week follow-up). Internal consistency ranged from poor to acceptable for online positive interpretation bias (ω = .57 at baseline; ω = .73 at post-treatment, and ω = .76 at two-week follow-up) and offline positive interpretation bias (ω = .52 at baseline; ω = .72 at post-treatment, and ω = .68 at two-week follow-up). Internal consistency values for negative interpretation bias were comparable to those reported in the study in which the measure was first developed (α = .71 for online bias and α = .75 for offline bias; Miers et al. 2020); however, Miers and colleagues did not report internal consistency statistics for online/offline positive interpretation bias.

Fear of Negative Evaluation. The Brief Fear of Negative Evaluation Scale (BFNE; Leary, 1983) were used to assess fear of negative evaluation. The 8-item straightforwardly-worded version of the measure was used (Rodebaugh et al., 2004; Weeks et al., 2005) as this version is more reliable and valid than the original scale (Rodebaugh et al., 2011). Participants are asked to rate the extent to which they agree with a series of statements (e.g., "*I am afraid that other people will find fault with me*") on a scale from 1 ("Not at all characteristic of me") to 5 ("Entirely characteristic of me"). Individual item scores are summed to calculate a total score, with higher scores reflecting greater fear of negative evaluation. The BFNE has strong internal consistency and construct validity when used with college student samples (Rodebaugh et al., 2004). The BFNE was administered as part of the pre-intervention battery and two-week follow-up battery. In the current sample, internal consistency for the BFNE was excellent at both timepoints (ω = .94 at baseline and ω = .95 at two-week follow-up).

Fear of Missing Out. The Fear of Missing Out Scale (FOMO Scale; Przybylski et al., 2013) is a 10-item questionnaire assessing the extent to which individuals are fearful of missing

out on social experiences. An example item reads, "*I get anxious when I don't know what my friends are up to*." Participants rate the extent to which they agree with each item on a scale from 1 ("Not at all true of me") to 5 ("Extremely true of me"). A sum score is calculated based on participants' responses to all items, with higher scores reflecting greater FOMO. Some modifications were made to the wording of the measure to increase relevance for the sample, including changing the term "in-jokes" to "inside jokes." The FOMO Scale demonstrates good internal consistency (Przybylski et al., 2013) and convergent validity (Casale & Fioravanti, 2020). The FOMO Scale was administered as part of the pre-intervention battery and two-week follow-up battery. In the current sample, internal consistency for the FOMO Scale was excellent at both timepoints (ω = .90 at baseline and ω = .90 at two-week follow-up).

Anxiety and Depression Symptoms. The Patient Health Questionnaire for Anxiety and Depression (PHQ-4; Kroenke et al., 2009) is a four-item questionnaire that was used to assess anxiety and depression symptom severity over the past two weeks. Participants are presented with the prompt, "*Over the last two weeks, how often have you been bothered by the following problems*?" and then rate separate items assessing symptoms of anxiety (e.g., "*Feeling nervous, anxious, or on edge*") and depression (e.g., "*Feeling down, depressed, or hopeless*") on a scale of 0 ("Not at all") to 3 ("Nearly every day"). In this study, the two items assessing anxiety and the two items assessing depression were summed separately to create separate indices of anxiety and depression. The PHQ-4 has strong reliability and validity (Löwe et al., 2010; Wicke et al., 2022) and scores have been shown to be sensitive to a diagnosis of anxiety or depression in college students (Khubchandani et al., 2016). The PHQ-4 was administered as part of the pre-intervention battery and two-week follow-up battery. In the current sample, internal consistency for the PHQ-4 anxiety subscale ranged from good to excellent at both timepoints (α = .84 at

baseline and α = .90 at two-week follow-up) and was good at both timepoints for depression (α = .84 at baseline and α = .87 at two-week follow-up).

Social Comparison Orientation. The Iowa–Netherlands Comparison Orientation Measure (INCOM; Schneider & Schupp, 2011) assesses individuals' tendencies to engage in social comparison. The present study used the three-item Ability subscale from the short-version of the measure (Schneider & Schupp, 2014). The Ability subscale includes items such as, "*I often compare how I am doing socially (e.g., social skills, popularity) with other people,*" with response options ranging from 1 ("I disagree strongly") to 5 ("I agree strongly"). Prior to data collection, the decision was made to modify the wording of all items to increase specificity to social media (e.g., "*I often compare how I am doing socially (e.g., social skills, popularity) with other people [on social media]*"). The INCOM was administered as part of the pre-intervention battery and two-week follow-up battery. In the current sample, internal consistency for the INCOM was good at both timepoints (ω = .78 at baseline and ω = .74 at two-week follow-up).

Time Spent Using Social Media. Time spent using social media was assessed using the single researcher-authored item, "*In the past two weeks, approximately how much time per day, on average, have you spent using social media platforms such as Instagram, Snapchat, TikTok, Facebook, etc.? To answer, please combine the time you spend across different platforms (e.g., if you spend 3 hours on TikTok and 3 hours on Instagram, you'd list 6 hours).*" Participants could choose one option from a drop-down menu reflecting two-hour increments of time (i.e., ranging from 0-2 hours to 22-24 hours). Time spent using social media was assessed as part of the pre-intervention battery and two-week follow-up battery.

Phase 2 Analytic Plan

Analyses were conducted using R version 4.3.2 (R Core Team, 2023). Systematic patterns related to missing data and completion of two-week follow-up questionnaires were evaluated as a function of demographic characteristics using logistic regression models. (These analyses were separate from those evaluating primary and secondary study outcomes of interest). Though some demographic predictors were highly imbalanced (i.e., only one individual self-identifying with a given category), we did not collapse or merge any demographic characteristics to create larger subgroups for analysis. Though this likely affected parameter estimates and power for models with sparse subgroups, this approach enabled us to avoid erasing aspects of participants' identities by collapsing demographic features into experimenter-defined subgroups (Call et al., 2023). Additionally, descriptive characteristics related to participants' use of PRISM are reported. Linear and ordinal regression models were fit to evaluate acceptability, perceived effectiveness, and engagement with target skills. Partially standardized beta estimates are reported for models with continuous outcomes (i.e., the Program Feedback Scale total score), while unstandardized estimates are reported for models with ordinal outcomes (i.e., perceived effectiveness and engagement with target skills).

Multilevel modeling was used to evaluate primary and secondary study outcomes. Multilevel modeling confers advantages over repeated-measures ANOVAs, including less stringent assumptions about data structure and increased flexibility to handle missing data (Krull & MacKinnon, 2001; Nich & Carroll, 1997). Residuals for each model were checked for linearity and normality via visual inspection of Q-Q plots (Dedrick et al., 2009). All residuals appeared to be normally distributed. Influential cases were identified using the "influence.ME" package in R, which calculates Cook's distance values at the participant level (Te Grotenhuis & Pelzer, 2012). The cut-off for determining a set of responses as influential was .0246913, based

on the recommended formula of 4/n, where *n* refers to the number of groups within the level 2 grouping factor (162 participants; Van der Meer et al., 2010). Across models, the number of participants with influential data ranged from 1 (0.62%) to 9 (5.56%). Each model was re-analyzed with influential participants removed. The pattern of results did not change for the models evaluating online interpretation bias, offline interpretation bias, anxiety, or depression. However, results changed slightly for the models evaluating fear of negative evaluation, fear of missing out, and social comparison orientation. (Influential case detection was not possible for the model evaluating time spent using social media as an outcome as it was on an ordinal scale). For transparency, both sets of analyses are reported for all models (Aguinis et al., 2013).

Analyses were conducted with the "lme4" package in R, with interactions specified between timepoint (e.g., baseline, immediately post-intervention, and two-week follow-up) and study condition (i.e., PRISM; social media control condition) to predict each outcome. Random intercepts for participant are included in each multilevel model. Though most models had linear outcomes, a multilevel ordinal regression model was fit to evaluate change in time spent using social media. Models with significant interactions were decomposed and visualized using the "emmeans," "effects," and "ggplot2" packages in R.

Standardized beta estimates were used as effect sizes as multilevel models do not yield straightforward R^2 statistics (Lorah, 2018). Partially standardized coefficients were obtained by standardizing continuous outcome variables (note, fully standardized estimates were not possible as predictor variables were categorical). Effect sizes were categorized based on Ferguson's guidelines, where 0.2 represents a "practically significant" effect, 0.5 represents a moderate effect, and 0.8 represents a strong effect (Ferguson, 2009). Additionally, values for marginal and

conditional R^2, which respectively reflect the amount of variance in each model that can be attributed to fixed effects alone vs. fixed and random effects, are reported for each model.

Phase 2 Results

Participant Characteristics

See Table 4 for full demographic characteristics for Phase 2 participants (N = 165; n = 81 randomized to the PRISM condition and n = 84 randomized to the control condition). With respect to biological sex, 137 participants (83%) self-identified as female and 28 (17%) self-identified as male. With respect to gender identity, 160 participants (97%) self-identified as cisgender and five (3%) self-identified as nonbinary. With respect to race, one participant (.61%) self-identified as American Indian or Alaska Native, 34 (20.6%) self-identified as Asian, 10 (6.06%) self-identified as Black or African American, 19 (11.5%) self-identified as multiracial, 97 (58.8%) self-identified as White, three (1.8%) self-identified as "Other," and one (.61%) preferred not to respond. With respect to ethnicity, 16 participants (9.7%) self-identified as Hispanic/Latinx and 149 (90.3%) self-identified as Non-Hispanic/Latinx. With respect to sexual orientation, six participants (3.6%) self-identified as asexual, 28 (17%) self-identified as bisexual, six (3.6%) self-identified as gay/lesbian, 119 (72.1%) self-identified as heterosexual/straight, two (1.2%) self-identified as pansexual, one (.61%) self-identified as "Other," and seven (4.2%) preferred not to respond. When asked to report annual household income, 74 participants (44.8%) chose > $150,000, 35 (21.2%) chose $75,000 - $149,999, seven (4.2%) chose $50,000 - $74,999, five (3.0%) chose $25,000 - $49,999, four (2.4%) chose < $25,000, eight (4.8%) preferred not to respond, and 32 (19.4%) did not know their annual household income. There were no significant baseline differences between PRISM and control condition participants on interpretation biases, fear of negative evaluation, FOMO, social

comparison orientation, depression symptoms, anxiety symptoms, or time spent using social media (*p*s > .200).

Rates of reported difficulties with any (i.e., at least one) mental health condition in the sample were comparable to national averages for young adults (Substance Abuse and Mental Health Services Administration, 2022). When asked to identify the mental health condition(s) with which they were currently struggling from a check-all-that-apply list, 73 (44.2%) endorsed at least one mental health condition and 8 (4.8%) preferred not to respond. Among participants who endorsed at least one mental health condition, the mean number endorsed was 2.08 (*SD* = 1.3). Among all participants, the mean number endorsed was 0.92 (*SD* = 1.35). The mental health conditions endorsed by participants were generalized anxiety disorder (*n* = 45), depression (*n* = 34), social anxiety disorder (*n* = 16), eating disorder (*n* = 16), obsessive-compulsive disorder (*n* = 12), posttraumatic stress disorder (*n* = 9), panic disorder (*n* = 7), attention-deficit/hyperactivity disorder (*n* = 3), bipolar disorder (*n* = 3), specific phobia (*n* = 2), personality disorder (*n* = 2), and dementia or other cognitive disorder (*n* = 1). Additionally, one participant who selected the "other" category reported having trauma and disordered eating habits, while another participant who selected the "other" category reported that they likely have generalized anxiety disorder.

When asked about current mental health treatment use from a check-all-that-apply list, 35 participants (21.2%) reported being in therapy, 37 (22.4%) reported using prescription medications, 41 (24.8%) reported seeking social support, 24 (14.5%) reported using self-guided help (e.g., self-help book, blogs, online intervention, mobile applications), and 3 (1.8%) reported using over-the-counter medications/supplements. Additionally, one (0.6%) participant who selected the "other" category reported using meditation. All forms of treatment were rated as "Helpful" or "Somewhat Helpful," with a few exceptions: one participant rated prescription

medications as “Not helpful,” two participants rated self-guided help as “Not helpful,” and one participant preferred not to respond regarding the helpfulness of therapy.

When asked which social media platform they used most intensely, 64 (39.5%) chose Instagram, 57 (35.2%) chose TikTok, 28 (17.2%) chose Snapchat, eight (4.9%) chose Twitter, five (3.1%) chose “other”, and none chose Facebook. When asked about their motivation to change the way they think about social media content that triggers negative self-evaluation, six reported no motivation, 34 reported low motivation, 85 reported some motivation, 36 endorsed high motivation, and one endorsed extreme motivation.

Data Processing

Of the 165 participants who enrolled in the study, 162 (including $n = 78$ in the PRISM condition and $n = 84$ in the control condition) completed their assigned study condition and questionnaires and had sufficient data for analysis of immediate outcomes.[5] One hundred thirty-nine participants, including $n = 66$ in the PRISM condition and $n = 73$ in the control condition, completed follow-up questionnaires and had sufficient data for analysis of two-week outcomes (See CONSORT diagram; Figure 1). Completing (vs. not completing) the two-week follow-up questionnaires was not significantly associated with study condition ($z = - 0.42$, $p = .677$), age ($z = 0.60$, $p = .548$), sex assigned at birth ($z = 0.58$, $p = .563$), gender identity ($zs < 0.57$, $ps > .570$), sexual orientation ($zs < 0.00$, $ps > .996$), race ($zs < 0.00$, $ps > .995$), ethnicity ($z = -1.28$, $p = .202$), or income ($zs < 0.75$, $ps > .455$).

Missing data was minimal at both timepoints. For the first set of questionnaires, there were 31 missing responses out of 14,904 across four items and 17 participants (0.21% missing).

[5] Two participants assigned to PRISM provided demographic data but did not complete PRISM or any questionnaires, while one participant provided demographic data, completed baseline questionnaires, and started PRISM, but did not finish PRISM or complete post-intervention questionnaires.

For the second set of questionnaires, there were 37 missing responses out of 10,147 across three items and 21 participants (0.31% missing). Nearly all missing responses (93.5% and 97.3% at each time point, respectively) were on the two qualitative PFS questions asking participants to write what they liked and what they would change about the intervention. It is possible that the writing requirement and framing of these questions as optional (i.e., "*Please share as many true thoughts and feelings as you would like*") could have led some participants to skip answering them. Having (vs. not having) missing data at either study timepoint was not significantly associated with study condition (z = 1.31, p = .192), age (z = 1.80, p = .072), sex assigned at birth (z = -1.84, p = .065), gender identity (zs < 1.28, ps > .081), sexual orientation (zs < - 0.00, ps > .117), race (zs < - 0.00, ps > .994), or income (zs < 0.19, ps > .469). However, missing data was associated with ethnicity, such that self-identifying as Hispanic/Latinx was associated with greater rates of having missing data (z = 1.99, p = .047).

Summary of PRISM and Control Condition Descriptive Data

For PRISM participants, the mean time it took to complete the study in minutes – including the intervention and both sets of questionnaires – was 98.74 (*SD* = 316.63). Six people took longer than three hours to complete the study, suggesting that they were likely distracted or did not complete the study in one sitting. Excluding those participants, the mean time it took to complete the study was 27.93 minutes (*SD* = 12.37). For control condition participants, the mean time it took to complete the study in minutes – including the control condition activity and both sets of questionnaires – was 109.60 (*SD* = 642.63). Three people took longer than three hours to complete the study. Excluding those participants, the mean time it took to complete the study was 27.31 minutes (*SD* = 15.98).

When participants were asked to identify their three most-used social media platforms at the beginning of PRISM, 72 (92.3%) chose Instagram, 55 (70.5%) chose TikTok, 54 (69.2%) chose Snapchat, 25 (32.1%) chose "Other" (e.g., YouTube, Tumblr), 18 (23.1%) chose Twitter, and seven (9.0%) chose Facebook. For the three vignettes where participants could practice helping others think flexibly (see vignettes in Appendix B), 60 (76.9%) chose "Seeing others' successes," 54 (69.2%) chose "Number of likes and comments," 51 (65.4%) chose "Big group photos," 28 (35.9%) chose "Instagram birthday posts," 27 (34.6%) chose "Losing followers," and 14 (17.9%) chose "Messing up your feeds' aesthetic." When asked to choose one vignette at the end with which to practice thinking flexibly on their own, 23 (29.5%) chose "Seeing others' successes," 21 (26.9%) chose "Number of likes and comments," 14 (17.9%) chose "Big group photos," seven (9.0%) chose "Instagram birthday posts," six (34.6%) chose "Losing followers," four (17.9%) chose "Messing up your feeds' aesthetic," and three (3.8%) chose to write in their own situation.

Qualitative Feedback on PRISM

In terms of positive feedback, participants liked that PRISM was engaging, easy to navigate, and clear in its presentation of information. Mirroring feedback from Phase 1B, participants also appreciated PRISM's relatability and practical examples, its balanced and realistic approach to social media, and provision of practical tips and resources. In terms of suggestions for changes, participants suggested that PRISM provide a broader range of social media scenarios to work on, both in general and for specific domains (e.g., body image; romantic relationships). Some suggested providing more examples of ways to practice thinking flexibly. Participants also suggested enhancing PRISM's visual content (e.g., providing better images and incorporating video/mixed-media elements). Moreover, though some liked the length and a few

would have preferred PRISM to be longer, most who commented on the length said they would prefer PRISM to be shorter. See Table 17 for additional examples of qualitative feedback on PRISM, including quotes from participants.

Acceptability, Perceived Effectiveness, and Engagement with Target Skills

Results from the models comparing PRISM to the control condition on acceptability, perceived effectiveness, and engagement with target skills are reported below. See Tables 5, 6, and 7 for ratings of acceptability, perceived effectiveness, and engagement with target skills across conditions, respectively.

Program Acceptability. As hypothesized, there was a significant main effect for condition, such that program acceptability was greater among the PRISM condition than the control condition, both immediately post-intervention (β = .63, *SE* = .15, *t* = 4.17, $p < .001$) and at two-week follow-up (β = .70, *SE* = .16, *t* = 4.39, $p < .001$). For the model exploring change in program acceptability from immediately post-intervention to two-week follow-up, there was no significant time-by-condition interaction (β = .05, *SE* = .12, *t* = 0.42, $p = .675$) or main effect for time (β = -.02, *SE* = .09, *t* = -0.22, $p = .829$), though there remained a significant main effect for condition (β = .63, *SE* = .15, *t* = 4.25, $p < .001$).

Perceived Effectiveness. As hypothesized, there was a significant main effect for condition, such that perceived effectiveness ratings were greater among the PRISM condition than the control condition, both immediately post-intervention (*b* = 1.64, *SE* = .35, *t* = 4.69, $p < .001$) and at two-week follow-up (*b* = 1.62, *SE* = .39, *t* = 4.15, $p < .001$). For the model exploring change in perceived effectiveness ratings from immediately post-intervention to two-week follow-up, there was no significant time-by-condition interaction (*b* = - 0.21, *SE* = .55, *z* = - 0.38,

p = .707) or main effect for time (b = 0.53, SE = .38, z = 1.40, p = .162), though there remained a significant main effect for condition (b = 2.34, SE = .52, t = 4.53, p < .001).

Engagement with Target Skills. Contrary to hypotheses, participants in the PRISM and control conditions did not differ in terms of engagement with the three intervention-related target skills (bs < .53, ts < 1.66, ps > .098) or the two control condition-related target skills (bs < .27, ts < 0.86, ps > .392) at two-week follow-up.

Changes in Online and Offline Interpretation Bias and Other Secondary Outcomes

Results from the models comparing the effects of PRISM to the control condition on changes in primary and secondary outcomes are reported below. Effects for interpretation bias-related outcomes are reported in Tables 11 – 14 and visualized in Figures 2 – 3. Effects for other outcome variables are reported in Tables 15 - 16 and visualized in Figures 4 - 5. See Table 18 for an overall summary of Phase 2 outcomes.

Online Negative Interpretation Bias. In line with hypotheses, there was a significant time-by-condition interaction such that online negative interpretation bias decreased immediately post-intervention for the PRISM condition, but not for the control condition. For the model that additively evaluated two-week outcomes, there was a significant time-by-condition interaction, such that reductions in online negative interpretation bias remained at two-week follow-up for the PRISM condition but not for the control condition. Effects were moderate immediately post-intervention (β = - 0.54) and at two-week follow-up (β = - 0.51). In the model with influential cases removed, effects were in the same direction and of similar magnitude (β = - 0.53 and β = - 0.54, respectively). There were no main effects for time or condition across models.

Online Positive Interpretation Bias. In line with hypotheses, there was a significant time-by-condition interaction such that online positive interpretation bias increased immediately

post-intervention for the PRISM condition, but not for the control condition. For the model that additively evaluated two-week outcomes, there was a significant time-by-condition interaction, such that increases in online positive interpretation bias remained at two-week follow-up for the PRISM condition but not for the control condition. Effects were small to moderate immediately post-intervention ($\beta = 0.31$) and at two-week follow-up ($\beta = 0.36$). In the model with influential cases removed, effects were in the same direction and of similar magnitude ($\beta = 0.25$ and $\beta = 0.34$, respectively). There were no main effects for time or condition across models.

Offline Negative Interpretation Bias. In line with hypotheses and consistent with the models evaluating change in *online* negative interpretation bias, there was a significant time-by-condition interaction such that offline negative interpretation bias decreased immediately post-intervention for the PRISM condition, but not for the control condition. For the model that additively evaluated two-week outcomes, there was a significant time-by-condition interaction, such that reductions in offline negative interpretation bias remained at two-week follow-up for the PRISM condition but not for the control condition. Effects were strong immediately post-intervention ($\beta = -0.86$) and at two-week follow-up ($\beta = -0.71$). In the model with influential cases removed, effects were in the same direction and of similar magnitude ($\beta = -0.83$ and $\beta = -0.74$, respectively). There were no main effects for time or condition across models.

Offline Positive Interpretation Bias. In line with hypotheses and consistent with the models evaluating change in *online* positive interpretation bias, there was a significant time-by-condition interaction such that offline positive interpretation bias increased immediately post-intervention for the PRISM condition, but not for the control condition. For the model that additively evaluated two-week outcomes, there was a significant time-by-condition interaction, such that increases in offline positive interpretation bias remained at two-week follow-up for the

PRISM condition but not for the control condition. Effects were moderate immediately post-intervention (β = 0.49) and small to moderate at two-week follow-up (β = 0.27). In the model with influential cases removed, effects were in the same direction and of similar magnitude (β = 0.42 and β = 0.25, respectively). There were no main effects for time or condition across models.

Fear of Negative Evaluation. Contrary to hypotheses, there was no significant time-by-condition interaction for fear of negative evaluation at two-week follow-up. However, there was a main effect for time, such that fear of negative evaluation decreased at two-week follow-up for both conditions, with a small effect (β = - 0.20). In the model with influential cases removed, there was a significant time-by-condition interaction in the expected direction, such that fear of negative evaluation decreased at two-week follow-up for the PRISM condition, but not for the control condition, with a moderate effect (β = - 0.39). There were no main effects for time or condition in the model with influential cases removed.

Fear of Missing Out. In line with hypotheses, there was a significant time-by-condition interaction such that fear of missing out decreased at two-week follow-up for the PRISM condition, but not for the control condition, with a small to moderate effect (β = - 0.29). There were no main effects for time or condition. In the model with influential cases removed, there was neither a significant time-by-condition interaction nor main effects for time or condition.

Anxiety Symptoms. Contrary to hypotheses, there was no significant time-by-condition interaction for anxiety symptoms at two-week follow-up. However, there was a main effect for time, such that anxiety symptoms decreased at two-week follow-up for both conditions, with a moderate effect (β = - 0.38). In the model with influential cases removed, the main effect for time on anxiety symptoms was in the same direction and of similar magnitude (β = - 0.35).

Depression Symptoms. Contrary to hypotheses, there was no significant time-by-condition interaction for depression symptoms at two-week follow-up. There were also no main effects for time or condition. In the model with influential cases removed, there was also no significant interaction nor main effects for time or condition.

Social Comparison Orientation. Contrary to hypotheses, there was no significant time-by-condition interaction for social comparison orientation at two-week follow-up. There were also no main effects for time or condition. In the model with influential cases removed, there was a main effect for condition, such that social comparison orientation was greater among the PRISM condition than the control condition across both timepoints, with a small effect (β = 0.29).

Time Spent Using Social Media. In line with (one of the competing) hypotheses, there was a significant time-by-condition interaction such that time spent using social media decreased from baseline to two-week follow-up for the PRISM condition but not for the control condition, with a small effect (b = - 1.01, SE = .51, z = -1.98, p = .047, OR = 0.36).

Phase 2 Discussion

This study evaluated a brief, digital, single-session reappraisal intervention for college students (N = 162) who struggle with negative interpretation bias on social media. By comparing PRISM to an activity-matched social media control condition, we tested its impacts on several outcomes related to problematic social media use. PRISM was seen as highly acceptable and outperformed the control condition in shifting online interpretation bias (the primary outcome) and offline interpretation bias (a secondary outcome) in the expected directions. Though there were some significant effects of PRISM on other secondary outcomes (e.g., FOMO, time spent using social media), they were smaller and less stable than the interpretation bias outcomes.

Acceptability, Perceived Effectiveness, and Engagement with Target Skills

PRISM was viewed as more acceptable and likely to be effective than the control condition at both timepoints. This aligns with PRISM's overall high acceptability and perceived effectiveness ratings in Phase 1, as well as qualitative feedback describing PRISM's relevance, helpfulness, and alignment with participants' real experiences of using social media. On the PFS, most participants chose "agree" or "really agree" on items assessing PRISM's enjoyability, understandability, usability, and agreement with the message, as well as their perceptions of PRISM as likely to be helpful to others, and the extent to which they tried their hardest when completing PRISM. Participants also typically agreed that PRISM would help themselves and others to think more flexibly, make fewer negative interpretations about what they see on social media, and make more positive interpretations about what they see on social media.

Several PFS ratings, including perceptions of PRISM's helpfulness, increased from Phase 1B to Phase 2, suggesting the changes made to PRISM prior to launching the randomized controlled trial were beneficial. Moreover, mean scores of all items assessing beliefs that PRISM would shift interpretation bias and foster flexible thinking related to social media increased from Phase 1B to Phase 2, suggesting the intervention became more credible.

Contrary to expectations, participants who completed PRISM did not significantly differ from control condition participants on engagement with any of the condition-specific target skills. In other words, completing PRISM (versus completing the control condition) was not associated with greater reported use of reappraisal skills over the two-week interim period. This is somewhat surprising, given the expected improvements on interpretation bias outcomes were observed. There are several potential explanations as to why reported engagement with target skills did not differ across conditions. First, it is possible that PRISM was successful in reducing

negative self-focused social media cognitions, as intended. If this was the case, there would be fewer instances of negative thinking that would prompt use of reappraisal, which is already a fairly effortful emotion regulation strategy (Milyavsky et al., 2019; Troy et al., 2018). Unfortunately, we did not include a measure of frequency of negative thinking or reappraisal skills at baseline, and therefore cannot assess changes in frequency of negative thinking or reappraisal skills over the course of the study. Second, there could have been a discrepancy between changes in participants' tendencies to make negative interpretations (which went down) and participants' self-concept or identity tied to social media (e.g., how they see themselves in terms of their response to social media). Recall that participants were selected in part based on their self-report that they tend to make negative self-focused interpretations when using social media. It is plausible that participants still viewed themselves as someone who struggles with social media, even though their negative self-focused interpretations were reduced. The literature is mixed regarding the extent to which cognitive-behavioral interventions improve self-concept, and prior studies have varied considerably on sample size, eligibility, intervention target, and follow-up period (Fjermestad et al., 2022; Gamer et al., 2008; O'Callaghan & Cunningham, 2015). Thus, it is not altogether surprising in retrospect that self-report about one's typical responses to social media might not shift after a single session. Also, our two-week follow-up period was chosen somewhat arbitrarily, and it is possible that explicit changes in self-concepts would emerge over a longer period and with continued practice making less negative interpretations. It will be helpful in future work to measure skill application following the interventions more directly, so the measures are less influenced by beliefs or self-concept about skill use.

Changes in Interpretation Bias (Immediately and at Two-week Follow-up)

In line with hypotheses, PRISM led to significantly greater decreases in online and offline negative interpretation bias and increases in online and offline positive interpretation bias than the control condition. This aligns with prior research showing the efficacy of single-session reappraisal interventions for shifting psychopathology-related outcomes (Ranney al., 2017; Rodriguez et al., 2019; Rodriguez et al., 2021). Decreases in negative interpretation bias were consistently stronger than increases in positive interpretation bias. As participants were deemed eligible based on having sufficiently high negative interpretation bias scores but did not need to meet a cutoff for positive interpretation bias, there may have been more room for improvement in terms of negative interpretation bias. PRISM also focused more explicitly on coping with negative thoughts, making it distinct from other cognitive training paradigms that specifically focus on training positive interpretations (e.g., some cognitive bias modification for interpretation programs; Menne-Lothmann et al., 2014).

Contrary to the competing hypothesis that the strength of effects would be stronger (either) immediately post-intervention or at two-week follow-up, effect sizes were largely consistent at both timepoints. The one exception was for the model evaluating offline positive interpretation bias, for which there was a moderate effect immediately post-intervention, but smaller at two-week follow up. This overall pattern suggests that gains from PRISM were maintained over the two-week period, especially for negative interpretation bias.

Unexpectedly, changes in interpretation bias were typically stronger for the offline (versus online) subscales, despite PRISM's focus on social media scenarios. There are several potential explanations for this. First, the training likely generalized to "offline" social scenarios in participants' lives. Some situations covered in PRISM (e.g., seeing others' successes) easily apply to several situations occurring outside of social media. It is also possible that the AIBQ's

online interpretation bias scenarios were less relevant to participants than the offline scenarios and thus not ideally suited to capture changes in social media-related interpretation bias for the present sample. For example, some platforms referenced in the online scenarios (e.g., Facebook; Twitter) were not heavily used by participants. Though our hope was that the AIBQ would gauge common experiences that cut across platforms (e.g., feeling badly about having only one follower on Twitter would generalize to experiences on other platforms), we recognize that the online scenarios in the measure may have been less well matched to participants' particular experiences. Also, internal consistency was slightly lower for online (versus offline) negative interpretation bias, so the smaller changes may simply be tied to the use of a less reliable measure.

Changes in Other Secondary Outcomes (at Two-week Follow-up)

The models evaluating two-week changes in FOMO and time spent using social media were the only ones whose results aligned with hypotheses, both with modest effects. However, when excluding participants with highly influential responses, there was no longer a significant interaction showing improvements in FOMO associated with assignment to PRISM. Also, though outlier detection for time spent using social media was not possible due to the ordinal nature of the variable, the small effect suggests that changes in time spent using social media may not be robust. Likewise, for the model evaluating changes in fear of negative evaluation, there was no significant time-by-condition interaction when all participants were included, though the hypothesized interaction emerged once participants with influential responses were removed from analyses. Collectively, these results suggest that FOMO, time spent using social media, and fear of negative evaluation may be sensitive to a reappraisal-focused intervention, but not as robustly as was observed for the interpretation bias outcomes. Though FOMO and fear of

negative evaluation in particular share similarities with interpretation bias, and were correlated with negative interpretation bias in this study, these risk factors for unhealthy reactions to social media may be more receptive to other types of interventions (e.g., those designed to address FOMO specifically; Alutaybi et al., 2020).

There were no significant time-by-condition interaction effects (either with or without the inclusion of influential cases) for the models evaluating changes in anxiety symptoms, depression symptoms, or social comparison orientation. However, anxiety decreased significantly for *both* conditions, with a moderate effect. PHQ-4 scores indicated that participants were fairly anxious compared with other college student samples (Khubchandani et al., 2016), with one-fourth of participants endorsing generalized anxiety disorder at baseline. Thus, there may have been more room for improvement in anxiety compared with other secondary outcomes. Supplemental analyses also suggest that improvements in anxiety across conditions may reflect regression to the mean (see Appendix D). The lack of improvement in depression symptoms may be because PRISM was not designed to target depressed mood and anhedonia (which are assessed by the PHQ-4), though it is plausible that PRISM positively impacted cognitive symptoms of depression that were not assessed. There are also several potential third variables connecting social media use with anxiety and mood symptoms (e.g., loneliness; interpersonal stress; sleep dysfunction) that were not evaluated in the present study. A future iteration of PRISM could expand the intervention to include such mechanisms (e.g., by offering a wider range of vignettes) and to address symptoms of depression and anxiety more directly (e.g., by building in a focus on behavioral activation or distress tolerance). Finally, though there was not an intervention effect for social comparison orientation, it is plausible that a larger

intervention dose or more training focused on changing social comparison specifically could lead to improvements in social comparison.

Strengths and Limitations

A key strength of the present study was the deployment of an intervention created collaboratively with target end-users, which likely increased PRISM's efficacy and relevance for users. Also, using an activity-matched and time-matched control condition helped to isolate intervention-specific (reappraisal training) effects from those that could be attributed to nonspecific factors, like reflecting on positive and negative aspects of social media in general. However, while the present study evaluated a social media-specific reappraisal intervention, we do not know the extent to which a general reappraisal intervention (with no specific emphasis on social media) would impact outcomes. This would be a worthy future test given the potential cost and time savings of using an intervention that requires less tailoring and impacts a broader domain of unhelpful thinking patterns.

Survey completion rates at two-week follow-up were also considerably higher than those typical of most digital intervention studies. This may be because we recruited college students who were more conscientious than typical treatment-seeking samples. Participants were also generally very motivated. At baseline, 75% of participants said they were (either) somewhat motivated, highly motivated, or extremely motivated to change how they think about content that triggers social comparison and social rejection on social media. Moreover, there was a high rate of mental health treatment engagement in this sample, with one-fifth of participants reporting that they were currently in therapy, and many others already using self-guided help. While this likely helped with intervention credibility and engagement, it is unclear how strong engagement

would be in more naturalistic settings (e.g., outside of a research study, or with participants who are less motivated or less socialized to psychological interventions).

There are also several limitations to address. Notably, internal consistency was modest for some of the interpretation bias subscales, especially for positive interpretation bias. Thus, our findings related to positive interpretation bias should be interpreted with this limitation in mind. Knowing the internal consistency also helps to contextualize the AIBQ's associations with other measures (e.g., the PHQ-4 depression subscale was not correlated with any of the interpretation bias measures, though depression and negative interpretation bias are closely linked; Everaert et al., 2017; Orchard et al., 2016). Additionally, though the sample was reasonably diverse with respect to sexual orientation and gender diversity (with one-quarter of participants self-identifying as LGBTQ+) and participants' racial and ethnic demographics mapped on to those of the overall undergraduate population at UVA, the majority of participants in Phase 2 identified as Non-Hispanic/Latinx White. Thus, less can be inferred about the suitability of PRISM for individuals from minoritized racial and ethnic groups. Future research should center the perspectives of minoritized groups who face unique stressors on social media, such as exposure to race-related discrimination. Indeed, cultural adaptations of single-session interventions for minoritized racial and ethnic groups have already shown promising effects (Shroff et al., 2023). Moreover, the sex assigned at birth and gender distribution of the sample was highly imbalanced, as participants mostly identified as cisgender and female. Thus, it is not yet clear how effectively PRISM will work for individuals who are not cisgender and female. Notably, the overrepresentation of female-identifying participants in this study aligns with research on gender differences in social media-linked difficulties, including pronounced difficulties for women resulting from social comparison on social media (Booker et al., 2018; Haferkamp et al., 2012;

Nesi & Prinstein, 2015). Finally, given that this was a first test of PRISM, and outcomes were assessed on a two-week timeframe, it is unknown whether improvements in interpretation bias would be maintained in the longer term.

Future Directions and Conclusion

Though further testing is necessary, results suggest that PRISM could have utility as a support for students struggling with negative impacts from social media. To better serve college student populations, university counseling centers could offer PRISM as a resource and/or recommend PRISM for students who endorse mental health difficulties tied to social media. PRISM could also be offered during the university onboarding process, alongside other trainings (e.g., those related to alcohol/substance use), though it will first be necessary to test PRISM with a broader range of diverse samples.

It may also be advantageous to adapt PRISM for middle and high school-aged adolescents, given that many of the adverse social media experiences PRISM is designed to address, including social media-related self-comparison and perceived rejection, are prevalent among adolescents and negatively impact adolescent mental health (Nesi et al., 2022; Weinstein, 2018; Winstone et al., 2023). Given the success of school-based mental health and digital literacy interventions (Eyal & Te'eni-Harari, 2024; Fazel et al., 2014; Kurz et al., 2022), schools could be a natural setting in which to disseminate PRISM. Other types of settings frequented by adolescents (and their caregivers), such as doctor's offices, libraries, and extracurricular activities, could also be fruitful avenues to disseminate PRISM. Adaptations to PRISM for adolescents might include changing the vignettes offered to be more adolescent-specific (e.g., building in more of a focus on Snapchat, which was less frequently cited as a platform of concern by our college student participants; Martin et al., 2018) and increasing the emphasis on

support seeking from adults (especially for younger adolescents, who may rely more on caregivers than do older adolescents).

It is also important to consider how PRISM could be adapted and scaled up across a variety of digital formats. For instance, some researchers have shared their interventions with wider audiences via brief instructional videos on YouTube (e.g., Rizvi & Kleiman, 2023). In our qualitative feedback, some participants expressed interest in video-based/multimedia content, suggesting that these kinds of adaptations would be beneficial. Moreover, while PRISM is currently only configured for a web-based format, it could be adapted into a mobile app. Finally, some SSIs are directly embedded within social media platforms (e.g., Tumblr; Dobias et al., 2020). Though this is still a new area of research, having an embedded intervention likely increases the chance that a user will be able to access the intervention when they most need it.

There are other ways that PRISM might be adapted in the future to increase its appeal and effectiveness. One of the primary pieces of constructive feedback from participants was a desire for a greater range of social media domains on which they could practice thinking flexibly. An expanded version of PRISM could offer even more domains to be responsive to a broader range of social media issues. For instance, while PRISM does not currently address body image issues and disordered eating, those areas are a natural extension for PRISM, based on our qualitative feedback. Body image could even represent a higher-order category of subdomains, given the heterogeneity of concerns falling under this umbrella. (Of course, additional testing would be necessary given that body image concerns were not the focus of this study). Additionally, though our sample was high in the tendency to do social comparison, we did not see improvements on the social comparison outcome measure. It could be helpful to offer more scenarios in the future

that address social comparison specifically (and likewise for other outcomes with unstable effects, such as FOMO).

Finally, though the focus of PRISM was on reappraisal and not on other coping strategies, a future iteration of PRISM might promote a broader range of strategies, such as seeking support, taking social media "breaks," engaging in other healthy distractions, etc. These adaptations would provide the user with greater choice and autonomy, which would be beneficial because not every situation calls for reappraisal. Finally, given that internal consistency was subpar for some of the AIBQ's interpretation bias subscales, it would be useful to test effects with a more reliable measure of negative self-focused social media cognitions (though, to our knowledge, a new one would need to be developed).

In this randomized controlled trial, PRISM was shown to be more acceptable and efficacious in shifting online and offline interpretation bias compared with a closely matched control condition. This suggests that reappraisal can effectively reduce interpretation bias, a modifiable risk factor for problematic social media use and social interactions more broadly, in a brief, self-administered format. Though this is a first test of PRISM, results align with prior studies that have found single-session interventions to benefit a broad range of psychological difficulties.

References

Aboody, D., Siev, J., & Doron, G. (2020). Building resilience to body image triggers using brief cognitive training on a mobile application: A randomized controlled trial. *Behaviour Research and Therapy*, *134*, 103723.

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods*, *16*(2), 270–301.

Alutaybi, A., Al-Thani, D., McAlaney, J., & Ali, R. (2020). Combating fear of missing out (FoMO) on social media: The FoMO-R method. *International Journal of Environmental Research and Public Health*, *17*(17), 6128.

Amin, N., Foa, E. B., & Coles, M. E. (1998). Negative interpretation bias in social phobia. *Behaviour Research and Therapy*, *36*(10), 945–957.

Auerbach, R. P., Mortier, P., Bruffaerts, R., Alonso, J., Benjet, C., Cuijpers, P., Demyttenaere, K., Ebert, D. D., Green, J. G., & Hasking, P. (2018). WHO World Mental Health Surveys International College Student Project: Prevalence and distribution of mental disorders. *Journal of Abnormal Psychology*, *127*(7), 623.

Auxier, B., & Anderson, M. (2021). Social media use in 2021. *Pew Research Center*, *1*, 1–4.

Auyeung, K., Hawley, L. L., Grimm, K., McCabe, R., & Rowa, K. (2020). Fear of negative evaluation and rapid response to treatment during cognitive behaviour therapy for social anxiety disorder. *Cognitive Therapy and Research*, *44*(3), 526–537.

Beard, C., Beckham, E., Solomon, A., Fenley, A. R., & Pincus, D. B. (2021). A Pilot Feasibility Open Trial of an Interpretation Bias Intervention for Parents of Anxious Children. *Cognitive and Behavioral Practice*.

Bentley, K. H., Boettcher, H., Bullis, J. R., Carl, J. R., Conklin, L. R., Sauer-Zavala, S., Pierre-Louis, C., Farchione, T. J., & Barlow, D. H. (2018). Development of a single-session, transdiagnostic preventive intervention for young adults at risk for emotional disorders. *Behavior Modification*, *42*(5), 781–805.

Bernstein, E. E., LeBlanc, N. J., Bentley, K. H., Barreira, P. J., & McNally, R. J. (2021). A single-session workshop to enhance emotional awareness and emotion regulation for graduate students: A pilot study. *Cognitive and Behavioral Practice*, *28*(3), 393–409.

Bloom, B. L. (2001). Focused Single-Session Psychotherapy: A Review of the Clinical and Research Literature. *Brief Treatment & Crisis Intervention*, *1*(1).

Booker, C. L., Kelly, Y. J., & Sacker, A. (2018). Gender differences in the associations between age trends of social media interaction and well-being among 10-15 year olds in the UK. *BMC Public Health*, *18*(1), 1–12.

Borghouts, J., Eikey, E., Mark, G., De Leon, C., Schueller, S. M., Schneider, M., Stadnick, N., Zheng, K., Mukamel, D., & Sorkin, D. H. (2021). Barriers to and facilitators of user engagement with digital mental health interventions: Systematic review. *Journal of Medical Internet Research*, *23*(3), e24387.

Brown, P. G. (2016). College student development in digital spaces. *New Directions for Student Services*, *2016*(155), 59–73.

Bu, H., & Duan, W. (2019). A single-session positive cognitive intervention on first-year students' mental health: Short-term effectiveness and the mediating role of strengths knowledge. *Journal of American College Health*, *67*(6), 515–522.

Call, C. C., Eckstrand, K. L., Kasparek, S. W., Boness, C. L., Blatt, L., Jamal-Orozco, N., Novacek, D. M., Foti, D., & Scholars for Elevating Equity and Diversity (SEED). (2023).

An ethics and social-justice approach to collecting and using demographic data for psychological researchers. *Perspectives on Psychological Science*, *18*(5), 979–995.

Casale, S., & Fioravanti, G. (2020). Factor structure and psychometric properties of the Italian version of the fear of missing out scale in emerging adults and adolescents. *Addictive Behaviors*, *102*, 106179.

Chen, J., Short, M., & Kemps, E. (2020). Interpretation bias in social anxiety: A systematic review and meta-analysis. *Journal of Affective Disorders*, *276*, 1119–1130.

Clark, D. A. (2013). Cognitive restructuring. *The Wiley Handbook of Cognitive Behavioral Therapy*, 1–22.

Cohen, K., Dobias, M., Morris, R., & Schleider, J. (2023). Improving uptake of mental health crisis resources: Randomized test of a single-session intervention embedded in social media. *Journal of Behavioral and Cognitive Therapy*, *33*(1), 24–34.

Comer, L. (2022). *Forced Candids, Charisma, and Confidence: The Impact of Instagram on College Women in Sororities Self-Esteem*.

Coyne, S. M., Rogers, A. A., Zurcher, J. D., Stockdale, L., & Booth, M. (2020). Does time spent using social media impact mental health?: An eight year longitudinal study. *Computers in Human Behavior*, *104*, 106160.

Davies, E. B., Morriss, R., & Glazebrook, C. (2014). Computer-delivered and web-based interventions to improve depression, anxiety, and psychological well-being of university students: A systematic review and meta-analysis. *Journal of Medical Internet Research*, *16*(5), e3142.

de Hullu, E., Sportel, B. E., Nauta, M. H., & de Jong, P. J. (2017). Cognitive bias modification and CBT as early interventions for adolescent social and test anxiety: Two-year follow-up

of a randomized controlled trial. *Journal of Behavior Therapy and Experimental Psychiatry*, *55*, 81–89.

Dedrick, R. F., Ferron, J. M., Hess, M. R., Hogarty, K. Y., Kromrey, J. D., Lang, T. R., Niles, J. D., & Lee, R. S. (2009). Multilevel modeling: A review of methodological issues and applications. *Review of Educational Research*, *79*(1), 69–102.

Denny, B. T., & Ochsner, K. N. (2014). Behavioral effects of longitudinal training in cognitive reappraisal. *Emotion*, *14*(2), 425.

Dobias, M. L., Morris, R. R., & Schleider, J. L. (2022). Single-session interventions embedded within Tumblr: Acceptability, feasibility, and utility study. *JMIR Formative Research*, *6*(7), e39004.

Dobias, M. L., Schleider, J. L., Jans, L., & Fox, K. R. (2021). An online, single-session intervention for adolescent self-injurious thoughts and behaviors: Results from a randomized trial. *Behaviour Research and Therapy*, *147*, 103983.

Dogaheh, E. R., Mohammadkhani, P., & Dolatshahi, B. (2011). Comparison of group and individual cognitive-behavioral therapy in reducing fear of negative evaluation. *Psychological Reports*, *108*(3), 955–962.

Dryman, M. T., & Heimberg, R. G. (2018). Emotion regulation in social anxiety and depression: A systematic review of expressive suppression and cognitive reappraisal. *Clinical Psychology Review*, *65*, 17–42.

Everaert, J., Podina, I. R., & Koster, E. H. (2017). A comprehensive meta-analysis of interpretation biases in depression. *Clinical Psychology Review*, *58*, 33–48.

Eyal, K., & Te'eni-Harari, T. (2024). Systematic review: Characteristics and outcomes of in-school digital media literacy interventions, 2010-2021. *Journal of Children and Media*, *18*(1), 8–28.

Faelens, L., Hoorelbeke, K., Cambier, R., van Put, J., Van de Putte, E., De Raedt, R., & Koster, E. H. (2021). The relationship between Instagram use and indicators of mental health: A systematic review. *Computers in Human Behavior Reports*, *4*, 100121.

Fazel, M., Hoagwood, K., Stephan, S., & Ford, T. (2014). Mental health interventions in schools in high-income countries. *The Lancet Psychiatry*, *1*(5), 377–387.

Feldman, D. B., & Dreher, D. E. (2012). Can hope be changed in 90 minutes? Testing the efficacy of a single-session goal-pursuit intervention for college students. *Journal of Happiness Studies*, *13*, 745–759.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*(5), 532–538. https://doi.org/10.1037/a0015808

Fernandez, D. P., Kuss, D. J., & Griffiths, M. D. (2020). Short-term abstinence effects across potential behavioral addictions: A systematic review. *Clinical Psychology Review*, *76*, 101828.

Fjermestad, K. W., Bellika, K., Matre, C., Silverman, W. K., & Wergeland, G. J. (2022). Age and depressive symptoms change predict the 4-year self-concept trajectory for youth after anxiety treatment. *Journal of Clinical Psychology*, *78*(11), 2109–2121.

Furlong, M., & Oei, T. P. (2002). Changes to automatic thoughts and dysfunctional attitudes in group CBT for depression. *Behavioural and Cognitive Psychotherapy*, *30*(3), 351–360.

Gamer, J., Schmukle, S. C., Luka-Krausgrill, U., & Egloff, B. (2008). Examining the dynamics of the implicit and the explicit self-concept in social anxiety: Changes in the Implicit

Association Test–Anxiety and the Social Phobia Anxiety Inventory following treatment. *Journal of Personality Assessment*, *90*(5), 476–480.

Ghosh, A., Cohen, K., Jans, L., Busch, C., McDanal, R., Yang, Y., Cooper, K., & Schleider, J. L. (2023). *A digital single-session intervention (Project Engage) to address fear of negative evaluation among college students: A pilot randomized controlled trial*.

Gibb, A., Wilson, J. M., Ford, C., & Shook, N. J. (2022). Does mindfulness reduce negative interpretation bias? *Cognition and Emotion*, *36*(2), 284–299.

Gulliver, A., Bennett, K., Bennett, A., Farrer, L. M., Reynolds, J., & Griffiths, K. M. (2015). Privacy issues in the development of a virtual mental health clinic for university students: A qualitative study. *JMIR Mental Health*, *2*(1), e4294.

Gupta, M., & Sharma, A. (2021). Fear of missing out: A brief overview of origin, theoretical underpinnings and relationship with mental health. *World Journal of Clinical Cases*, *9*(19), 4881.

Haddad, J. M., Macenski, C., Mosier-Mills, A., Hibara, A., Kester, K., Schneider, M., Conrad, R. C., & Liu, C. H. (2021). The impact of social media on college mental health during the COVID-19 pandemic: A multinational review of the existing literature. *Current Psychiatry Reports*, *23*(11), 1–12.

Haferkamp, N., Eimler, S. C., Papadakis, A.-M., & Kruck, J. V. (2012). Men are from Mars, women are from Venus? Examining gender differences in self-presentation on social networking sites. *Cyberpsychology, Behavior, and Social Networking*, *15*(2), 91–98.

Haugen, P. T., McCrillis, A. M., Smid, G. E., & Nijdam, M. J. (2017). Mental health stigma and barriers to mental health care for first responders: A systematic review and meta-analysis. *Journal of Psychiatric Research*, *94*, 218–229.

Hayes, M., van Stolk-Cooke, K., & Muench, F. (2015). Understanding Facebook use and the psychological affects of use across generations. *Computers in Human Behavior*, *49*, 507–511.

Herriman, Z., Taylor, A. M., & Roberts, R. M. (2024). Interventions to reduce the negative impact of online highly visual social networking site use on mental health outcomes: A scoping review. *Psychology of Popular Media*, *13*(1), 111–139.

Hetrick, S. E., Robinson, J., Burge, E., Blandon, R., Mobilio, B., Rice, S. M., Simmons, M. B., Alvarez-Jimenez, M., Goodrich, S., & Davey, C. G. (2018). Youth codesign of a mobile phone app to facilitate self-monitoring and management of mood symptoms in young people with major depression, suicidal ideation, and self-harm. *JMIR Mental Health*, *5*(1), e9041.

Hindash, A. H., & Amir, N. (2012). Negative interpretation bias in individuals with depressive symptoms. *Cognitive Therapy and Research*, *36*(5), 502–511.

Hornstein, S., Zantvoort, K., Lueken, U., Funk, B., & Hilbert, K. (2023). Personalization strategies in digital mental health interventions: A systematic review and conceptual framework for depressive symptoms. *Frontiers in Digital Health*, *5*, 1170002.

Ji, J. L., Baee, S., Zhang, D., Calicho-Mamani, C. P., Meyer, M. J., Funk, D., Portnow, S., Barnes, L., & Teachman, B. A. (2021). Multi-session online interpretation bias training for anxiety in a community sample. *Behaviour Research and Therapy*, *142*, 103864.

Jopling, E., Wilson, J., Burke, M., Tracy, A., & LeMoult, J. (2020a). Positive interpretation bias across the psychiatric disorders. *Cognitive Biases in Health and Psychiatric Disorders*, 99–117.

Jopling, E., Wilson, J., Burke, M., Tracy, A., & LeMoult, J. (2020b). Positive interpretation bias across the psychiatric disorders. *Cognitive Biases in Health and Psychiatric Disorders*, 99–117.

Keles, B., McCrae, N., & Grealish, A. (2020). A systematic review: The influence of social media on depression, anxiety and psychological distress in adolescents. *International Journal of Adolescence and Youth*, *25*(1), 79–93.

Khubchandani, J., Brey, R., Kotecki, J., Kleinfelder, J., & Anderson, J. (2016). The psychometric properties of PHQ-4 depression and anxiety screening scale among college students. *Archives of Psychiatric Nursing*, *30*(4), 457–462.

Kornfield, R., Meyerhoff, J., Studd, H., Bhattacharjee, A., Williams, J. J., Reddy, M., & Mohr, D. C. (2022). *Meeting Users Where They Are: User-centered Design of an Automated Text Messaging Tool to Support the Mental Health of Young Adults*. 1–16.

Kroenke, K., Spitzer, R. L., Williams, J. B., & Löwe, B. (2009). An ultra-brief screening scale for anxiety and depression: The PHQ–4. *Psychosomatics*, *50*(6), 613–621.

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research*, *36*(2), 249–277.

Kurz, M., Rosendahl, J., Rodeck, J., Muehleck, J., & Berger, U. (2022). School-based interventions improve body image and media literacy in youth: A systematic review and meta-analysis. *Journal of Prevention*, *43*(1), 5–23.

Kuss, D. J., & Griffiths, M. D. (2017). Social networking sites and addiction: Ten lessons learned. *International Journal of Environmental Research and Public Health*, *14*(3), 311.

Lattie, E. G., Lipson, S. K., & Eisenberg, D. (2019). Technology and college student mental health: Challenges and opportunities. *Frontiers in Psychiatry*, *10*, 246.

Leary, M. R. (1983). A brief version of the Fear of Negative Evaluation Scale. *Personality and Social Psychology Bulletin*, *9*(3), 371–375.

Lee, C. S., Bowman, M., & Wu, J. L. (2023). Preliminary outcomes from a single-session, asynchronous, online stress and anxiety management workshop for college students. *Trends in Psychiatry and Psychotherapy*, *45*, e20210448.

Lee, H. Y., Jamieson, J. P., Reis, H. T., Beevers, C. G., Josephs, R. A., Mullarkey, M. C., O'Brien, J. M., & Yeager, D. S. (2020). Getting fewer "Likes" than others on social media elicits emotional distress among victimized adolescents. *Child Development*, *91*(6), 2141–2159.

Lin, L. Y., Sidani, J. E., Shensa, A., Radovic, A., Miller, E., Colditz, J. B., Hoffman, B. L., Giles, L. M., & Primack, B. A. (2016). Association between social media use and depression among US young adults. *Depression and Anxiety*, *33*(4), 323–331.

Lorah, J. (2018). Effect size measures for multilevel models: Definition, interpretation, and TIMSS example. *Large-Scale Assessments in Education*, *6*(1), 1–11.

Löwe, B., Wahl, I., Rose, M., Spitzer, C., Glaesmer, H., Wingenfeld, K., Schneider, A., & Brähler, E. (2010). A 4-item measure of depression and anxiety: Validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. *Journal of Affective Disorders*, *122*(1–2), 86–95.

Marsh, C. N., & Wilcoxon, S. A. (2015). Underutilization of mental health services among college students: An examination of system-related barriers. *Journal of College Student Psychotherapy*, *29*(3), 227–243.

Martin, F., Wang, C., Petty, T., Wang, W., & Wilkins, P. (2018). Middle school students' social media use. *Journal of Educational Technology & Society*, *21*(1), 213–224.

Martínez-Ferrer, B., Moreno, D., & Musitu, G. (2018). Are adolescents engaged in the problematic use of social networking sites more involved in peer aggression and victimization? *Frontiers in Psychology*, *9*, 801.

Masciantonio, A., Bourguignon, D., Bouchat, P., Balty, M., & Rimé, B. (2021). Don't put all social network sites in one basket: Facebook, Instagram, Twitter, TikTok, and their relations with well-being during the COVID-19 pandemic. *PloS One*, *16*(3), e0248384.

McComb, C. A., Vanman, E. J., & Tobin, S. J. (2023). A meta-analysis of the effects of social media exposure to upward comparison targets on self-evaluations and emotions. *Media Psychology*, 1–24.

McCrae, N., Gettings, S., & Purssell, E. (2017). Social media and depressive symptoms in childhood and adolescence: A systematic review. *Adolescent Research Review*, *2*, 315–330.

Meier, A., Gilbert, A., Börner, S., & Possler, D. (2020). Instagram inspiration: How upward comparison on social network sites can contribute to well-being. *Journal of Communication*, *70*(5), 721–743.

Meier, A., & Johnson, B. K. (2022). Social comparison and envy on social media: A critical review. *Current Opinion in Psychology*, 101302.

Menne-Lothmann, C., Viechtbauer, W., Höhn, P., Kasanova, Z., Haller, S. P., Drukker, M., van Os, J., Wichers, M., & Lau, J. Y. (2014). How to boost positive interpretations? A meta-analysis of the effectiveness of cognitive bias modification for interpretation. *PloS One*, *9*(6), e100925.

Miers, A. C., Blöte, A. W., Bögels, S. M., & Westenberg, P. M. (2008). Interpretation bias and social anxiety in adolescents. *Journal of Anxiety Disorders*, *22*(8), 1462–1471.

Miers, A. C., Sumter, S. R., Clark, D. M., & Leigh, E. (2020). Interpretation bias in online and offline social environments and associations with social anxiety, peer victimization, and avoidance behavior. *Cognitive Therapy and Research*, *44*(4), 820–833.

Milyavsky, M., Webber, D., Fernandez, J. R., Kruglanski, A. W., Goldenberg, A., Suri, G., & Gross, J. J. (2019). To reappraise or not to reappraise? Emotion regulation choice and cognitive energetics. *Emotion*, *19*(6), 964.

Morello, K., Schäfer, S. K., Kunzler, A. M., Priesterroth, L.-S., Tüscher, O., & Kubiak, T. (2023). Cognitive reappraisal in mHealth interventions to foster mental health in adults: A systematic review and meta-analysis. *Frontiers in Digital Health*, *5*, 1253390.

Muñoz, R. F., Chavira, D. A., Himle, J. A., Koerner, K., Muroff, J., Reynolds, J., Rose, R. D., Ruzek, J. I., Teachman, B. A., & Schueller, S. M. (2018). Digital apothecaries: A vision for making health care interventions accessible worldwide. *Mhealth*, *4*.

Muris, P., Mayer, B., Den Adel, M., Roos, T., & van Wamelen, J. (2009). Predictors of change following cognitive-behavioral treatment of children with anxiety problems: A preliminary investigation on negative automatic thoughts and anxiety control. *Child Psychiatry and Human Development*, *40*(1), 139–151.

Murphy, R., Hirsch, C. R., Mathews, A., Smith, K., & Clark, D. M. (2007). Facilitating a benign interpretation bias in a high socially anxious population. *Behaviour Research and Therapy*, *45*(7), 1517–1529.

Nesi, J., & Prinstein, M. J. (2015). Using social media for social comparison and feedback-seeking: Gender and popularity moderate associations with depressive symptoms. *Journal of Abnormal Child Psychology*, *43*(8), 1427–1438.

Nesi, J., Rothenberg, W. A., Bettis, A. H., Massing-Schaffer, M., Fox, K. A., Telzer, E. H., Lindquist, K. A., & Prinstein, M. J. (2022). Emotional responses to social media experiences among adolescents: Longitudinal associations with depressive symptoms. *Journal of Clinical Child & Adolescent Psychology*, *51*(6), 907–922.

Nich, C., & Carroll, K. (1997). Now you see it, now you don't: A comparison of traditional versus random-effects regression models in the analysis of longitudinal follow-up data from a clinical trial. *Journal of Consulting and Clinical Psychology*, *65*(2), 252.

O'Callaghan, P., & Cunningham, E. (2015). Can a targeted, group-based CBT intervention reduce depression and anxiety and improve self-concept in primary-age children? *Educational Psychology in Practice*, *31*(3), 314–326.

Orchard, F., Pass, L., & Reynolds, S. (2016). Associations between interpretation bias and depression in adolescents. *Cognitive Therapy and Research*, *40*, 577–583.

Osborn, T. L., Rodriguez, M., Wasil, A. R., Venturo-Conerly, K. E., Gan, J., Alemu, R. G., Roe, E., Arango G, S., Otieno, B. H., & Wasanga, C. M. (2020). Single-session digital intervention for adolescent depression, anxiety, and well-being: Outcomes of a randomized controlled trial with Kenyan adolescents. *Journal of Consulting and Clinical Psychology*, *88*(7), 657.

Peck, C. E., Lim, M. H., Purkiss, M., Foley, F., Hopkins, L., & Thomas, N. (2020). Development of a lived experience-based digital resource for a digitally-assisted peer support program for young people experiencing psychosis. *Frontiers in Psychiatry*, *11*, 635.

Pittman, M., & Reich, B. (2016). Social media and loneliness: Why an Instagram picture may be worth more than a thousand Twitter words. *Computers in Human Behavior*, *62*, 155–167.

Przybylski, A. K., Murayama, K., DeHaan, C. R., & Gladwell, V. (2013). Motivational, emotional, and behavioral correlates of fear of missing out. *Computers in Human Behavior*, *29*(4), 1841–1848.

R Core Team. (2023). *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.* (R version 4.3.2 (2023-10-31)) [Computer software]. <https://www.R-project.org/>

Radovic, A., Gmelin, T., Stein, B. D., & Miller, E. (2017). Depressed adolescents' positive and negative use of social media. *Journal of Adolescence*, *55*, 5–15.

Ranney, R. M., Bruehlman-Senecal, E., & Ayduk, O. (2017). Comparing the effects of three online cognitive reappraisal trainings on well-being. *Journal of Happiness Studies*, *18*(5), 1319–1338.

Rizvi, S. L., & Kleiman, E. M. (2023). The promise of single-session interventions to inform stepped care approaches for complex mental health problems: Commentary on Schleider et al.(2023). *International Journal of Eating Disorders*, *56*(5), 885–887.

Rodebaugh, T. L., Heimberg, R. G., Brown, P. J., Fernandez, K. C., Blanco, C., Schneier, F. R., & Liebowitz, M. R. (2011). More reasons to be straightforward: Findings and norms for two scales relevant to social anxiety. *Journal of Anxiety Disorders*, *25*(5), 623–630.

Rodebaugh, T. L., Woods, C. M., Thissen, D. M., Heimberg, R. G., Chambless, D. L., & Rapee, R. M. (2004). More information from fewer questions: The factor structure and item properties of the original and brief fear of negative evaluation scale. *Psychological Assessment*, *16*(2), 169.

Rodriguez, L. M., Dell, J. B., Lee, K. D., & Onufrak, J. (2019). Effects of a brief cognitive reappraisal intervention on reductions in alcohol consumption and related problems. *Psychology of Addictive Behaviors*, *33*(7), 637.

Rodriguez, L. M., Stewart, S. H., & Neighbors, C. (2021). Effects of a brief web-based interpersonal conflict cognitive reappraisal expressive-writing intervention on changes in romantic conflict during COVID-19 quarantine. *Couple and Family Psychology: Research and Practice*.

Saleem, M., Kühne, L., De Santis, K. K., Christianson, L., Brand, T., & Busse, H. (2021). Understanding engagement strategies in digital interventions for mental health promotion: Scoping review. *JMIR Mental Health*, *8*(12), e30000.

Sauer-Zavala, S., Tirpak, J. W., Eustis, E. H., Woods, B. K., & Russell, K. (2021). Unified protocol for the transdiagnostic prevention of emotional disorders: Evaluation of a brief, online course for college freshmen. *Behavior Therapy*, *52*(1), 64–76.

Schleider, J., Burnette, J., Widman, L., Hoyt, C., & Prinstein, M. J. (2020). Randomized trial of a single-session growth mind-set intervention for rural adolescents' internalizing and externalizing problems. *Journal of Clinical Child & Adolescent Psychology*, *49*(5), 660–672.

Schleider, J. L., Dobias, M. L., Sung, J. Y., & Mullarkey, M. C. (2020). Future directions in single-session youth mental health interventions. *Journal of Clinical Child & Adolescent Psychology*, *49*(2), 264–278.

Schleider, J. L., Mullarkey, M. C., & Chacko, A. (2020). Harnessing wise interventions to advance the potency and reach of youth mental health services. *Clinical Child and Family Psychology Review*, *23*(1), 70–101.

Schleider, J. L., Mullarkey, M. C., Fox, K. R., Dobias, M. L., Shroff, A., Hart, E. A., & Roulston, C. A. (2022). A randomized trial of online single-session interventions for adolescent depression during COVID-19. *Nature Human Behaviour*, *6*(2), 258–268.

Schleider, J. L., Mullarkey, M. C., & Weisz, J. R. (2019). Virtual reality and web-based growth mindset interventions for adolescent depression: Protocol for a three-arm randomized trial. *JMIR Research Protocols*, *8*(7), e13368.

Schleider, J. L., & Weisz, J. R. (2016). Reducing risk for anxiety and depression in adolescents: Effects of a single-session intervention teaching that personality can change. *Behaviour Research and Therapy*, *87*, 170–181.

Schleider, J. L., & Weisz, J. R. (2017). Little treatments, promising effects? Meta-analysis of single-session interventions for youth psychiatric problems. *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*(2), 107–115.

Schleider, J., & Weisz, J. (2018). A single-session growth mindset intervention for adolescent anxiety and depression: 9-month outcomes of a randomized trial. *Journal of Child Psychology and Psychiatry*, *59*(2), 160–170.

Schneider, S. M., & Schupp, J. (2014). Individual differences in social comparison and its consequences for life satisfaction: Introducing a short scale of the Iowa–Netherlands Comparison Orientation Measure. *Social Indicators Research*, *115*(2), 767–789.

Schneider, S., & Schupp, J. (2011). *The social comparison scale: Testing the validity, reliability, and applicability of the Iowa-Netherlands Comparison Orientation Measure (INCOM) on the German population*.

Sedgwick, R., Epstein, S., Dutta, R., & Ougrin, D. (2019). Social media, internet use and suicide attempts in adolescents. *Current Opinion in Psychiatry*, *32*(6), 534.

Shafi, R. M., Nakonezny, P. A., Romanowicz, M., Nandakumar, A. L., Suarez, L., & Croarkin, P. E. (2019). The differential impact of social media use on middle and high school students: A retrospective study. *Journal of Child and Adolescent Psychopharmacology*, *29*(10), 746–752.

Sherman, J. A., & Ehrenreich-May, J. (2018). Ethnicity's role in the relationship between anxiety and negative interpretation bias among clinically anxious youth: A pilot study. *Child Psychiatry & Human Development*, *49*, 396–408.

Shroff, A., Roulston, C., Fassler, J., Dierschke, N. A., Todd, J. S. P., Ríos-Herrera, Á., Plastino, K. A., & Schleider, J. L. (2023). A Digital Single-Session Intervention Platform for Youth Mental Health: Cultural Adaptation, Evaluation, and Dissemination. *JMIR Mental Health*, *10*, e43062.

Sidani, J. E., Shensa, A., Hoffman, B., Hanmer, J., & Primack, B. A. (2016). The association between social media use and eating concerns among US young adults. *Journal of the Academy of Nutrition and Dietetics*, *116*(9), 1465–1472.

Simon, G. E., Imel, Z. E., Ludman, E. J., & Steinfeld, B. J. (2012). Is dropout after a first psychotherapy visit always a bad outcome? *Psychiatric Services*, *63*(7), 705–707.

Smits, J. A., Julian, K., Rosenfield, D., & Powers, M. B. (2012). Threat reappraisal as a mediator of symptom change in cognitive-behavioral treatment of anxiety disorders: A systematic review. *Journal of Consulting and Clinical Psychology*, *80*(4), 624.

Snow, G. (2020). *blockrand: Randomization for Block Random Clinical Trials* (R package version 1.5) [Computer software]. https://CRAN.R-project.org/package=blockrand

Sobande, F., Fearfull, A., & Brownlie, D. (2020). Resisting media marginalisation: Black women's digital content and collectivity. *Consumption Markets & Culture*, *23*(5), 413–428.

Steele, R. G., Hall, J. A., & Christofferson, J. L. (2020). Conceptualizing digital stress in adolescents and young adults: Toward the development of an empirically based model. *Clinical Child and Family Psychology Review*, *23*(1), 15–26.

Steinsbekk, S., Nesi, J., & Wichstrøm, L. (2023). Social media behaviors and symptoms of anxiety and depression. A four-wave cohort study from age 10–16 years. *Computers in Human Behavior*, *147*, 107859.

Substance Abuse and Mental Health Services Administration. (2022). *Key substance use and mental health indicators in the United States: Results from the 2020 national survey on drug use and health (HHS Publication No. PEP21-07-01-003, NSDUH Series H-56). Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration.* https://www.samhsa.gov/data/report/2021-nsduh-annual-national-report

Talbot, C. V., Talbot, A., Roe, D. J., & Briggs, P. (2022). The management of LGBTQ+ identities on social media: A student perspective. *New Media & Society*, *24*(8), 1729–1750.

Talmon, M. (1990). *Single-session therapy: Maximizing the effect of the first (and often only) therapeutic encounter.* Jossey-Bass.

Te Grotenhuis, H., & Pelzer, B. (2012). *Influence. ME: tools for detecting influential data in mixed effects models*.

Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, *27*(2), 237–246.

Troy, A. S., Shallcross, A. J., Brunner, A., Friedman, R., & Jones, M. C. (2018). Cognitive reappraisal and acceptance: Effects on emotion, physiology, and perceived cognitive costs. *Emotion*, *18*(1), 58.

Valkenburg, P. M., Beyens, I., Pouwels, J. L., van Driel, I. I., & Keijsers, L. (2021). *Social media browsing and adolescent well-being: Challenging the "Passive Social Media Use Hypothesis."*

Van der Meer, T., Te Grotenhuis, M., & Pelzer, B. (2010). Influential cases in multilevel modeling: A methodological comment. *American Sociological Review*, *75*(1), 173–178.

Venturo-Conerly, K. E., Osborn, T. L., Alemu, R., Roe, E., Rodriguez, M., Gan, J., Arango, S., Wasil, A., Wasanga, C., & Weisz, J. R. (2022). Single-session interventions for adolescent anxiety and depression symptoms in Kenya: A cluster-randomized controlled trial. *Behaviour Research and Therapy*, *151*, 104040.

Vogel, E. A., Rose, J. P., Okdie, B. M., Eckles, K., & Franz, B. (2015). Who compares and despairs? The effect of social comparison orientation on social media use and its outcomes. *Personality and Individual Differences*, *86*, 249–256.

Walton, G. M., & Wilson, T. D. (2018). Wise interventions: Psychological remedies for social and personal problems. *Psychological Review*, *125*(5), 617.

Wasil, A. R., Taylor, M. E., Franzen, R. E., Steinberg, J. S., & DeRubeis, R. J. (2021). Promoting graduate student mental health during COVID-19: Acceptability, feasibility, and perceived utility of an online single-session intervention. *Frontiers in Psychology*, *12*, 1167.

Weber, S., Messingschlager, T., & Stein, J.-P. (2022). This is an insta-vention! Exploring cognitive countermeasures to reduce negative consequences of social comparisons on Instagram. *Media Psychology*, *25*(3), 411–440.

Weeks, J. W., Heimberg, R. G., Fresco, D. M., Hart, T. A., Turk, C. L., Schneier, F. R., & Liebowitz, M. R. (2005). Empirical validation and psychometric evaluation of the Brief Fear of Negative Evaluation Scale in patients with social anxiety disorder. *Psychological Assessment*, *17*(2), 179.

Weinstein, E. (2018). The social media see-saw: Positive and negative influences on adolescents' affective well-being. *New Media & Society*, *20*(10), 3597–3623.

Wicke, F. S., Krakau, L., Löwe, B., Beutel, M. E., & Brähler, E. (2022). Update of the standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. *Journal of Affective Disorders*, *312*, 310–314.

Winstone, L., Mars, B., Haworth, C. M., & Kidger, J. (2023). Types of social media use and digital stress in early adolescence. *The Journal of Early Adolescence*, *43*(3), 294–319.

Woodley, C., & Meredith, C. (2012). Supporting student transition through social media. *American Journal of Distance Education*, *26*(2), 86–95.

Woods, H. C., & Scott, H. (2016). # Sleepyteens: Social media use in adolescence is associated with poor sleep quality, anxiety, depression and low self-esteem. *Journal of Adolescence*, *51*, 41–49.

Zhang, Z., Mai, Y., Yang, M., & Zhang, M. Z. (2018). Package 'WebPower.' *Basic and Advanced Statistical Power Analysis Version*, *72*.

Tables

Table 1. Demographic Characteristics for Phase 1A Participants ($N = 20$)

Age – *M* (*SD*)	19.1 (1.15)
Sex Assigned at Birth – *n* (%)	
Female	17 (85%)
Male	3 (15%)
Intersex	0 (0%)
Gender – *n* (%)	
Female	16 (80%)
Male	3 (15%)
Transgender	0 (0%)
Nonbinary	1 (5%)
Race – *n* (%)	
American Indian or Alaska Native	0 (0%)
Asian	4 (20%)
Black or African American	0 (0%)
Multiracial	4 (20%)
Native Hawaiian or Other Pacific Islander	0 (0%)
White	12 (60%)
Ethnicity – *n* (%)	
Not Hispanic/Latinx	16 (80%)
Hispanic/Latinx	4 (20%)
Sexual Orientation – *n* (%)	
Asexual	0 (0%)
Bisexual	1 (5%)
Gay/lesbian	0 (0%)
Heterosexual/straight	18 (90%)
Pansexual	0 (0%)
Other (write-in)	1 (5%)
Annual Household Income – *n* (%)	
Below $25,000	1 (5%)
$25,000 - $49,999	0 (0%)
$50,000 - $74,999	0 (0%)
$75,000 - $149,999	4 (20%)
Above $150,000	11 (55%)
Unknown	4 (20%)

Table 2. Demographic Characteristics for Phase 1B Participants ($N = 17$)

Age – *M* (*SD*)	19.5 (1.23)
Sex Assigned at Birth – *n* (%)	
Female	12 (70.6%)
Male	5 (29.4%)
Intersex	0 (0%)
Gender – *n* (%)	
Female	12 (70.6%)
Male	5 (29.4%)
Transgender	0 (0%)
Nonbinary	0 (0%)
Race – *n* (%)	
American Indian or Alaska Native	0 (0%)
Asian	4 (23.5%)
Black or African American	2 (11.7%)
Multiracial	1 (5.9%)
Native Hawaiian or Other Pacific Islander	0 (0%)
White	9 (52.9%)
Ethnicity – *n* (%)	
Not Hispanic/Latinx	17 (100%)
Hispanic/Latinx	0 (0%)
Sexual Orientation – *n* (%)	
Asexual	0 (0%)
Bisexual	4 (23.5%)
Gay/lesbian	2 (11.8%)
Heterosexual/straight	11 (64.7%)
Pansexual	0 (0%)
Other (write-in)	0 (0%)
Annual Household Income – *n* (%)	
Below $25,000	1 (5.9%)
$25,000 - $49,999	1 (5.9%)
$50,000 - $74,999	1 (5.9%)
$75,000 - $149,999	4 (23.5%)
Above $150,000	10 (58.8%)
Unknown	0 (0%)

Table 3. Phase 1B Acceptability Ratings

Statement	M (SD)	Range
Enjoy	3.00 (.67)	2 - 4
Understood	3.80 (.42)	3 - 4
Easy to use	3.60 (.70)	2 - 4
Tried hardest	3.30 (.67)	2 - 4
Helpful	2.90 (.88)	2 - 4
Recommend to a friend	2.30 (1.57)	0 - 4
Agree with message	3.70 (.48)	3 - 4
Think more flexibly – self	2.90 (.74)	2 - 4
Fewer negative interpretations – self	2.80 (.63)	1 - 3
More positive interpretations – self	2.80 (.63)	2 - 4
Think more flexibly – others	2.90 (.74)	2 - 4
Fewer negative interpretations – others	3.00 (.67)	2 - 4
More positive interpretations – others	2.90 (.74)	2 - 4
Full scale	39.90 (4.63)	36 - 50

Table 4. Demographic Characteristics for Phase 2 Intent-to-Treat Sample ($N = 165$)

Demographic characteristic	PRISM ($n = 81$)	Control ($n = 84$)	Overall ($N = 165$)
Age – *M* (*SD*)	18.9 (.96)	18.7 (.97)	18.8 (.97)
Sex Assigned at Birth – *n* (%)			
Female	68 (84.0%)	69 (82.1%)	137 (83.0%)
Male	13 (16.0%)	15 (17.9%)	28 (17.0%)
Intersex	0 (0%)	0 (0%)	0 (0%)
Gender – *n* (%)			
Female	65 (80.2%)	68 (81.0%)	133 (80.6%)
Male	13 (16.0%)	14 (16.7%)	27 (16.4%)
Transgender	0 (0%)	0 (0%)	0 (0%)
Nonbinary	3 (3.7%)	2 (2.4%)	5 (3.0%)
Race – *n* (%)			
American Indian or Alaska Native	1 (1.2%)	0 (0%)	1 (.61%)
Asian	16 (19.8%)	18 (21.4%)	34 (20.6%)
Black or African American	6 (7.4%)	4 (4.8%)	10 (6.06%)
Multiracial	11 (13.6%)	8 (9.5%)	19 (11.5%)
Native Hawaiian or Other Pacific Islander	0 (0%)	0 (0%)	0 (0%)
White	45 (55.6%)	52 (61.9%)	97 (58.8%)
Other (write-in)	1 (1.2%)	2 (2.4%)	3 (1.8%)
Prefer not to respond	1 (1.2%)	0 (0%)	1 (.61%)
Ethnicity – *n* (%)			
Not Hispanic/Latinx	72 (88.9%)	77 (91.7%)	149 (90.3%)
Hispanic/Latinx	9 (11.1%)	7 (8.3%)	16 (9.7%)
Sexual Orientation* – *n* (%)			
Asexual	4 (4.9%)	2 (2.4%)	6 (3.6%)
Bisexual	16 (19.8%)	12 (14.3%)	28 (17.0%)
Gay/lesbian	2 (2.5%)	4 (4.8%)	6 (3.6%)
Heterosexual/straight	58 (71.6%)	61 (72.6%)	119 (72.1%)
Pansexual	1 (1.2%)	1 (1.2%)	2 (1.2%)
Other (write-in)	0 (0%)	1 (1.2%)	1 (.61%)
Prefer not to respond	2 (2.5%)	5 (6.0%)	7 (4.2%)
Annual Household Income – *n* (%)			
Below $25,000	3 (3.7%)	1 (1.2%)	4 (2.4%)
$25,000 - $49,999	3 (3.7%)	2 (2.4%)	5 (3.0%)
$50,000 - $74,999	3 (3.7%)	4 (4.8%)	7 (4.2%)
$75,000 - $149,999	13 (16.0%)	22 (26.2%)	35 (21.2%)
Above $150,000	39 (48.1%)	35 (41.7%)	74 (44.8%)
Unknown	14 (17.3%)	18 (21.4%)	32 (19.4%)
Prefer not to respond	6 (7.4%)	2 (2.4%)	8 (4.8%)

Note. *Percentages can exceed 100% as participants could select more than one response option.

Table 5. Phase 2 Acceptability Ratings

	PRISM		Control	
Statement	Post-intervention M (SD)	Follow-up M (SD)	Post-intervention M (SD)	Follow-up M (SD)
Enjoy	2.87 (0.78)	2.95 (0.73)	2.63 (0.83)	2.56 (0.87)
Understood	3.65 (0.62)	3.70 (0.55)	3.51 (0.63)	3.44 (0.73)
Easy to use	3.67 (0.68)	3.85 (0.44)	3.74 (0.44)	3.56 (0.65)
Tried hardest	3.29 (0.79)	3.39 (0.70)	3.38 (0.67)	3.25 (0.91)
Helpful	3.29 (0.88)	3.14 (0.86)	2.88 (1.02)	2.63 (1.03)
Recommend to a friend	2.99 (1.13)	2.82 (1.02)	2.25 (1.19)	2.11 (1.28)
Agree with message	3.71 (0.67)	3.74 (0.47)	2.99 (0.89)	3.11 (0.86)
Think more flexibly – self	3.03 (0.95)	3.20 (0.77)	2.50 (1.10)	2.70 (1.14)
Fewer negative interpretations – self	2.85 (0.99)	3.18 (0.84)	2.35 (1.04)	2.52 (1.08)
More positive interpretations – self	3.04 (0.93)	3.26 (0.79)	2.40 (1.12)	2.59 (0.96)
Think more flexibly – others	3.14 (0.75)	3.20 (0.81)	2.61 (1.01)	2.67 (1.00)
Fewer negative interpretations – others	3.04 (0.76)	3.15 (0.81)	2.44 (0.96)	2.59 (0.94)
More positive interpretations – others	3.08 (0.82)	3.14 (0.80)	2.50 (1.02)	2.62 (0.97)
Full scale	41.64 (7.42)	42.71 (7.04)	36.30 (8.74)	36.47 (9.37)

Note. For PRISM, $n = 78$ post intervention and $n = 66$ at follow-up. For Control Condition, $n = 84$ post intervention and $n = 73$ at follow-up.

Table 6. Perceived Effectiveness Ratings at Post-Intervention and Two-Week Follow-Up

	PRISM		Control	
Rating	Post-intervention *n* (%)	Follow-up *n* (%)	Post-intervention *n* (%)	Follow-up *n* (%)
Much more effectively	9 (11.5%)	6 (9.1%)	1 (1.2%)	5 (6.8%)
Somewhat more effectively	54 (69.2%)	53 (80.3%)	37 (44.0%)	33 (45.2%)
Neutral, neither more nor less effectively	13 (16.7%)	7 (10.6%)	45 (53.6%)	33 (45.2%)
Somewhat less effectively	1 (1.3%)	0 (0%)	1 (1.2%)	1 (1.4%)
Much less effectively	1 (1.3%)	0 (0%)	0 (0%)	1 (1.4%)

Note. For PRISM, $n = 78$ post intervention and $n = 66$ at follow-up. For Control Condition, $n = 84$ post intervention and $n = 73$ at follow-up.

Table 7. Target Engagement Skill Endorsement at Two-Week Follow-Up

Skill domain	PRISM *n* (%)	Control *n* (%)
Notice one's own negative thoughts triggered by social media (PRISM skill)		
Never	2 (3.0%)	2 (2.7%)
Rarely	14 (21.2%)	13 (19.7%)
Sometimes	30 (45.5%)	27 (37.0%)
Often	16 (24.2%)	30 (41.1%)
Always	4 (6.1%)	1 (1.4%)
Reappraisal tied to one's own negative social media thoughts (PRISM skill)		
Never	2 (3.0%)	2 (2.7%)
Rarely	8 (12.1%)	14 (19.2%)
Sometimes	30 (45.5%)	36 (39.3%)
Often	21 (31.8%)	21 (28.8%)
Always	5 (7.6%)	0 (0%)
Reappraisal tied to others' hypothetical negative social media thoughts (PRISM skill)		
Never	3 (4.5%)	7 (9.6%)
Rarely	11 (16.7%)	19 (26.0%)
Sometimes	30 (45.5%)	23 (31.5%)
Often	17 (25.8%)	22 (30.1%)
Always	5 (7.6%)	2 (2.7%)
Think about technical aspects of social media (Control Condition skill)		
Never	4 (6.1%)	3 (4.1%)
Rarely	13 (19.7%)	16 (21.9%)
Sometimes	26 (39.4%)	38 (52.0%)
Often	20 (30.3%)	11 (15.1%)
Always	3 (4.5%)	5 (6.8%)
Discuss technical aspects of social media with others (Control Condition skill)		
Never	14 (21.2%)	22 (30.1%)
Rarely	29 (43.9%)	22 (30.1%)
Sometimes	17 (25.8%)	21 (28.8%)
Often	4 (6.1%)	7 (9.6%)
Always	2 (3.0%)	1 (1.4%)

Note. For PRISM, $n = 66$. For Control Condition, $n = 73$.

Table 8. Spearman Correlations Among Target Engagement Skills at Two-Week Follow-up

	Target Engagement Skill	1.	2.	3.	4.
PRISM condition	1. Notice one's own negative thoughts	—	—	—	—
	2. Reappraisal tied to one's own thoughts	.44***	—	—	—
	3. Reappraisal tied to others' thoughts	.24**	.47***	—	—
Control condition	4. Think about technical aspects	.09	.24**	.23**	—
	5. Discuss technical aspects	.11	.11	.37***	.44***

Note. Results are bolded when significant. *$p < .05$; **$p < .01$; ***$p < .001$.

Table 9. Means and Standard Deviations for Study Outcomes Across Timepoints

	PRISM			Control		
Measure	Baseline	Post-intervention	Follow-up	Baseline	Post-intervention	Follow-up
AIBQ Online Negative	11.64 (3.14)	9.73 (3.58)	9.76 (3.26)	11.36 (3.01)	11.17 (2.81)	11.08 (3.00)
AIBQ Online Positive	14.27 (2.34)	15.33 (2.67)	15.68 (2.71)	14.21 (2.56)	14.43 (2.80)	14.84 (2.98)
AIBQ Offline Negative	14.97 (3.07)	11.91 (3.98)	12.30 (3.26)	14.39 (2.65)	14.29 (3.08)	14.11 (3.38)
AIBQ Offline Positive	10.14 (2.48)	11.78 (3.22)	11.62 (2.75)	9.87 (2.32)	10.18 (2.40)	10.99 (2.57)
BFNE	31.82 (7.10)	–	29.01 (7.59)	31.95 (4.94)	–	30.56 (5.78)
FOMO	31.24 (7.97)	–	28.30 (7.33)	31.38 (6.58)	–	30.81 (6.99)
PHQ-4 Anxiety	3.21 (1.65)	–	2.47 (1.69)	3.27 (1.81)	–	2.60 (1.89)
PHQ-4 Depression	1.72 (1.76)	–	1.29 (1.55)	1.74 (1.61)	–	1.78 (1.84)
INCOM	10.92 (2.69)	–	10.36 (2.51)	10.65 (2.26)	–	10.64 (2.02)

Note. For PRISM, $n = 78$ at baseline/post intervention and $n = 66$ at follow-up. For Control Condition, $n = 84$ at baseline/post intervention and $n = 73$ at follow-up. AIBQ = Adolescents' Interpretation and Belief Questionnaire. BFNE = Brief Fear of Negative Evaluation Scale. FOMO = Fear of Missing Out Scale. PHQ-4 = Patient Health Questionnaire for Anxiety and Depression (four-item version). INCOM = Iowa-Netherlands Comparison Orientation Measure.

Table 10. Pearson Correlations Among Study Variables at Baseline and Two-Week Follow-Up

Timepoint	Variable	1.	2.	3.	4.	5.	6.	7.	8.
Baseline	1. AIBQ Online Negative	—	—	—	—	—	—	—	—
	2. AIBQ Online Positive	- .11	—	—	—	—	—	—	—
	3. AIBQ Offline Negative	.57***	.02	—	—	—	—	—	—
	4. AIBQ Offline Positive	- .16*	.23**	-.24**	—	—	—	—	—
	5. BFNE	.28***	.07	.48***	- .26***	—	—	—	—
	6. FOMO	.33***	- .09	.34***	.07	.54***	—	—	—
	7. PHQ-4 Anxiety	.20*	- .03	.24**	- .03	.39***	.36***	—	—
	8. PHQ-4 Depression	.11	- .04	.11	- .04	.27***	.30***	.40***	—
	9. INCOM	.15*	.03	.24**	- .01	.38***	.46***	.13	- .01
Follow-up	1. AIBQ Online Negative	—	—	—	—	—	—	—	—
	2. AIBQ Online Positive	- .12	—	—	—	—	—	—	—
	3. AIBQ Offline Negative	.63***	.05	—	—	—	—	—	—
	4. AIBQ Offline Positive	- .00	.33***	- .18*	—	—	—	—	—
	5. BFNE	.28***	.14	.56***	- .21*	—	—	—	—
	6. FOMO	.39***	- .15	.40***	- .13	.54***	—	—	—
	7. PHQ-4 Anxiety	.16	- .06	.29***	- .07	.43***	.34***	—	—
	8. PHQ-4 Depression	.06	- .09	.15	- .12	.30***	.28***	.64***	—
	9. INCOM	.15	.03	.27**	.07	.32***	.44***	.14	.07

Note. $N = 162$ at baseline and $N = 139$ at follow-up.

Results are bolded when significant. *$p < .05$; **$p < .01$; ***$p < .001$.

AIBQ = Adolescents' Interpretation and Belief Questionnaire. BFNE = Brief Fear of Negative Evaluation Scale. FOMO = Fear of Missing Out Scale. PHQ-4 = Patient Health Questionnaire for Anxiety and Depression (four-item version). INCOM = Iowa-Netherlands Comparison Orientation Measure.

Table 11. Interpretation Bias Outcomes at Post-Intervention (Influential Cases Retained)

Variable Name	β (*SE*)	t	*p*	τ00 (SD)	R^2M (R^2C)
Online negative interpretation bias				.68 (.82)	.05 (.72)
Timepoint (vs. baseline)					
Post-intervention	- 0.06 (.08)	- 0.73	.468		
Condition (vs. control)					
PRISM	0.09 (.15)	0.58	.566		
Post-intervention × PRISM	**-0.54 (.12)**	**- 4.56**	**< .001**		
Online positive interpretation bias				.64 (.80)	.03 (.67)
Timepoint (vs. baseline)					
Post-intervention	0.08 (.09)	0.91	.363		
Condition (vs. control)					
PRISM	0.02 (.16)	0.13	.893		
Post-intervention × PRISM	**0.32 (.13)**	**2.51**	**.013**		
Offline negative interpretation bias				.62 (.79)	.11 (.73)
Timepoint (vs. baseline)					
Post-intervention	- 0.03 (.08)	- 0.39	.696		
Condition (vs. control)					
PRISM	0.17 (.15)	1.15	.252		
Post-intervention × PRISM	**- 0.87 (.12)**	**- 7.49**	**< .001**		
Offline positive interpretation bias				.66 (.81)	.08 (.73)
Timepoint (vs. baseline)					
Post-intervention	0.11 (.08)	1.42	.159		
Condition (vs. control)					
PRISM	0.10 (.15)	0.66	.510		
Post-intervention × PRISM	**0.49 (.12)**	**4.23**	**< .001**		

Note. $N = 162$. β = standardized beta estimate. SE = standard error. Post-intervention= immediately post-intervention.

R^2M refers to the amount of variance that can be attributed to fixed effects. R^2C refers to the amount of variance that can be attributed to random effects.

Results are bolded when significant at the $p < .05$ level.

Table 12. Interpretation Bias Outcomes at Post-Intervention (Influential Cases Removed)

Variable Name	β (*SE*)	t	*p*	τ00 (SD)	R^2M (R^2C)
Online negative interpretation bias (*n* = 159)				.64 (.80)	.06 (.76)
Timepoint (vs. baseline)					
Post-intervention	- 0.06 (.07)	- 0.82	.416		
Condition (vs. control)					
PRISM	0.02 (.15)	0.12	.906		
Post-intervention × PRISM	**-0.53 (.11)**	**- 4.97**	**< .001**		
Online positive interpretation bias (*n* = 156)				.65 (.81)	.03 (.74)
Timepoint (vs. baseline)					
Post-intervention	0.10 (.08)	1.28	.203		
Condition (vs. control)					
PRISM	- 0.00 (.15)	- 0.01	.989		
Post-intervention × PRISM	**0.30 (.11)**	**2.73**	**.007**		
Offline negative interpretation bias (*n* = 154)				.59 (.77)	.12 (.79)
Timepoint (vs. baseline)					
Post-intervention	- 0.04 (.07)	- 0.65	.520		
Condition (vs. control)					
PRISM	0.14 (.14)	0.96	.337		
Post-intervention × PRISM	**- 0.82 (.10)**	**- 8.40**	**< .001**		
Offline positive interpretation bias (*n* = 157)				.66 (.81)	.07 (.78)
Timepoint (vs. baseline)					
Post-intervention	0.09 (.07)	1.32	.190		
Condition (vs. control)					
PRISM	0.10 (.15)	0.69	.491		
Post-intervention × PRISM	**0.45 (.10)**	**4.35**	**< .001**		

Note. β = standardized beta estimate. SE = standard error. Post-intervention = immediately post-intervention.

R^2M refers to the amount of variance that can be attributed to fixed effects. R^2C refers to the amount of variance that can be attributed to random effects.

Results are bolded when significant at the $p < .05$ level.

Table 13. Interpretation Bias Outcomes at Two-Week Follow-Up (Influential Cases Retained)

Variable Name	β (*SE*)	t	*p*	τ00 (SD)	R^2M (R^2C)
Online negative interpretation bias				.67 (.82)	.05 (.72)
Timepoint (vs. baseline)					
Post-intervention	- 0.06 (.08)	- 0.72	.471		
Follow-up	- 0.06 (.09)	- 0.72	.471		
Condition (vs. control)					
PRISM	0.09 (.15)	0.58	.566		
Post-intervention × PRISM	**- 0.54 (.12)**	**- 4.53**	**< .001**		
Follow-up × PRISM	**- 0.51 (.13)**	**- 4.04**	**< .001**		
Online positive interpretation bias				.58 (.76)	.04 (.62)
Timepoint (vs. baseline)					
Post-intervention	0.08 (.10)	0.83	.406		
Follow-up	0.19 (.10)	1.85	.065		
Condition (vs. control)					
PRISM	0.02 (.15)	0.13	.896		
Post-intervention × PRISM	**0.31 (.14)**	**2.29**	**.023**		
Follow-up × PRISM	**0.36 (.14)**	**2.48**	**.014**		
Offline negative interpretation bias				.61 (.78)	.11 (.71)
Timepoint (vs. baseline)					
Post-intervention	- 0.03 (.08)	- 0.38	.707		
Follow-up	- 0.04 (.09)	- 0.49	.623		
Condition (vs. control)					
PRISM	0.17 (.15)	1.14	.257		
Post-intervention × PRISM	**- 0.86 (.12)**	**- 7.20**	**< .001**		
Follow-up × PRISM	**- 0.71 (.13)**	**- 5.55**	**< .001**		
Offline positive interpretation bias				.62 (.79)	.08 (.69)
Timepoint (vs. baseline)					
Post-intervention	0.11 (.09)	1.30	.195		
Follow-up	**0.32 (.09)**	**3.47**	**.001**		
Condition (vs. control)					
PRISM	0.10 (.15)	0.66	.513		
Post-intervention × PRISM	**0.49 (.13)**	**3.88**	**< .001**		
Follow-up × PRISM	**0.27 (.13)**	**2.00**	**.047**		

Note. $N = 139$. β = standardized beta estimate. SE = standard error. Post-intervention = immediately post-intervention. Follow-up = two-week follow-up.

R^2M refers to the amount of variance that can be attributed to fixed effects. R^2C refers to the amount of variance that can be attributed to random effects.
Results are bolded when significant at the $p < .05$ level.

Table 14. Interpretation Bias Outcomes at Two-Week Follow-Up (Influential Cases Removed)

Variable Name	β (*SE*)	t	*p*	τ00 (SD)	R^2M (R^2C)
Online negative interpretation bias (*n* = 158)				.61 (.78)	.07 (.73)
Timepoint (vs. baseline)					
Post-intervention	- 0.06 (.08)	- 0.77	.440		
Follow-up	- 0.06 (.08)	- 0.77	.441		
Condition (vs. control)					
PRISM	0.01 (.15)	0.05	.964		
Post-intervention × PRISM	**- 0.53 (.11)**	**- 4.73**	**< .001**		
Follow-up × PRISM	**- 0.54 (.12)**	**- 4.51**	**< .001**		
Online positive interpretation bias (*n* = 153)				.55 (.74)	.04 (.69)
Timepoint (vs. baseline)					
Post-intervention	0.11 (.08)	1.36	.176		
Follow-up	**0.18 (.09)**	**2.11**	**.036**		
Condition (vs. control)					
PRISM	0.02 (.15)	0.15	.883		
Post-intervention × PRISM	**0.25 (.12)**	**2.17**	**.031**		
Follow-up × PRISM	**0.34 (.12)**	**2.72**	**.007**		
Offline negative interpretation bias (*n* = 156)				.53 (.73)	.11 (.72)
Timepoint (vs. baseline)					
Post-intervention	- 0.00 (.08)	- 0.05	.963		
Follow-up	0.00 (.08)	0.01	.995		
Condition (vs. control)					
PRISM	0.18 (.14)	1.26	.208		
Post-intervention × PRISM	**- 0.83 (.11)**	**- 7.44**	**< .001**		
Follow-up × PRISM	**- 0.74 (.12)**	**- 6.21**	**< .001**		
Offline positive interpretation bias (*n* = 156)				.65 (.80)	.08 (.75)
Timepoint (vs. baseline)					
Post-intervention	0.12 (.08)	1.57	.118		
Follow-up	**0.40 (.08)**	**4.92**	**< .001**		
Condition (vs. control)					
PRISM	0.12 (.15)	0.80	.423		
Post-intervention × PRISM	**0.42 (.11)**	**3.71**	**< .001**		
Follow-up × PRISM	**0.25 (.12)**	**2.10**	**.036**		

Note. β = standardized beta estimate. SE = standard error. Post-intervention = immediately post-intervention. Follow-up = two-week follow-up.

R^2M refers to the amount of variance that can be attributed to fixed effects. R^2C refers to the amount of variance that can be attributed to random effects.

Results are bolded when significant at the $p < .05$ level.

Table 15. Other Secondary Outcomes at Two-Week Follow-Up (Influential Cases Retained)

Variable Name	β (*SE*)	t	*p*	τ00 (SD)	R^2M (R^2C)
Fear of negative evaluation				.63 (.79)	.03 (.66)
Timepoint (vs. baseline)					
Follow-up	**- 0.20 (.10)**	**- 2.10**	**.037**		
Condition (vs. control)					
PRISM	- 0.02 (.15)	- 0.14	.889		
Follow-up × PRISM	-0.24 (.14)	- 1.72	.088		
Fear of missing out				.63 (.79)	.02 (.65)
Timepoint (vs. baseline)					
Follow-up	- 0.09 (.10)	- 0.96	.341		
Condition (vs. control)					
PRISM	- 0.02 (.16)	- 0.12	.904		
Follow-up × PRISM	**- 0.29 (.14)**	**- 2.07**	**.040**		
Anxiety				.61 (.78)	.04 (.65)
Timepoint (vs. baseline)					
Follow-up	**- 0.38 (.10)**	**- 3.95**	**< .001**		
Condition (vs. control)					
PRISM	- 0.04 (.16)	- 0.25	.804		
Follow-up × PRISM	- 0.02 (.14)	- 0.17	.869		
Depression				.66 (.81)	.01 (.67)
Timepoint (vs. baseline)					
Follow-up	0.04 (.10)	0.43	.671		
Condition (vs. control)					
PRISM	- 0.01 (.16)	- 0.08	.940		
Follow-up × PRISM	- 0.27 (.14)	- 1.94	.054		
Social comparison orientation				.57 (.76)	.01 (.57)
Timepoint (vs. baseline)					
Follow-up	- 0.03 (.11)	- 0.31	.754		
Condition (vs. control)					
PRISM	0.11 (.16)	0.70	.485		
Follow-up × PRISM	- 0.19 (.16)	- 1.22	.225		

Note. $N = 139$. β = standardized beta estimate. SE = standard error. Follow-up = two-week follow-up.
R^2M refers to the amount of variance that can be attributed to fixed effects. R^2C refers to the amount of variance that can be attributed to random effects.
Results are bolded when significant at the $p < .05$ level.

Table 16. Other Secondary Outcomes at Two-Week Follow-Up (Influential Cases Removed)

Variable Name	β (*SE*)	t	*p*	τ00 (SD)	R^2M (R^2C)
Fear of negative evaluation (*n* = 156)				.53 (.73)	.04 (.70)
Timepoint (vs. baseline)					
Follow-up	- 0.09 (.08)	- 1.03	.303		
Condition (vs. control)					
PRISM	0.11 (.14)	0.79	.429		
Follow-up × PRISM	**- 0.39 (.12)**	**- 3.27**	**.001**		
Fear of missing out (*n* = 157)				.68 (.83)	.02 (.77)
Timepoint (vs. baseline)					
Follow-up	- 0.05 (.08)	- 0.71	.478		
Condition (vs. control)					
PRISM	- 0.11 (.15)	- 0.76	.450		
Follow-up × PRISM	- 0.21 (.11)	- 1.87	.064		
Anxiety (*n* = 161)				.63 (.80)	.04 (.67)
Timepoint (vs. baseline)					
Follow-up	**- 0.35 (.09)**	**- 3.67**	**< .001**		
Condition (vs. control)					
PRISM	- 0.02 (.15)	- 0.13	.897		
Follow-up × PRISM	- 0.06 (.14)	- 0.46	.643		
Depression (*n* = 153)				.70 (.84)	.01 (.79)
Timepoint (vs. baseline)					
Follow-up	0.05 (.07)	0.72	.475		
Condition (vs. control)					
PRISM	- 0.08 (.15)	- 0.54	.589		
Follow-up × PRISM	- 0.21 (.11)	- 1.96	.053		
Social comparison orientation (*n* = 153)				.43 (.66)	.02 (.58)
Timepoint (vs. baseline)					
Follow-up	- 0.02 (.09)	- 0.23	.822		
Condition (vs. control)					
PRISM	**0.29 (.14)**	**2.03**	**.044**		
Follow-up × PRISM	- 0.22 (.14)	- 1.54	.126		

Note. β = standardized beta estimate. SE = standard error. Follow-up = two-week follow-up.

R^2M refers to the amount of variance that can be attributed to fixed effects. R^2C refers to the amount of variance that can be attributed to random effects.

Results are bolded when significant at the $p < .05$ level.

Table 17. Summary of Phase 2 Qualitative Feedback on PRISM

Theme	Examples of positive feedback ("*What did you like about the program*?)	Examples of critical feedback ("*What would you change about the program*?")
Length	"I think it is the perfect length for getting the point across while not being too tedious." "I like that the program was short and to the point."	"I think that it could've been shorter" "I would make it longer, and go into more examples"
Delivery and visual format	"It was very easy to use and displayed information in an absorbable manner." "I like that each "slide" is not too packed with text and images. This makes it easy to read and not overwhelming." "I loved the artwork that went along with some of the readings." "I liked that it was interactive and we were able to think through experiences because it makes it easier to apply them in real life situations"	"I think the teaching about thought flexibility could have been in video form and that way it would have been more interesting." "The graphics/stock photos were a little boring so maybe add more pretty drawings like the ones for the giving advice section and add some color and alternative fonts to the survey" "The stock pictures really took me out of reality; if you made this a "real" program, maybe you could source pictures from real students at UVA?."
Content and personal relevance	"I liked how easy the program was to understand and included a lot of real life, common examples. It was very relatable to my everyday life, so it was easy to answer the questions related to my experiences. Some of the statements might have sounded a little obvious, but they're important ones to remember, and I definitely needed a reminder." "I liked that the program used really relatable examples from actual college students."	"The program to me seemed to be tailored to female-identifying people in regards to many of its examples. I think it could include examples or information that is more tailored to men too. I also thought there were many more topics related to social media use, like body image, that could have been talked about." "Mention comparing physical and financial insecurities that can come up from looking at social media - especially for girls comparing their looks/bodies"

	"I already knew a lot of this stuff, but it explained it in a more holistic way, which made it more relatable and easier to follow. I definitely want to practice this."	"The program should include a topic about romantic relationships. A significant source of distress for many people on social media is that they are frequently exposed to romantic relationships, even if they have little involvement in these kinds of relationships in their personal lives."
Misc.	"I liked how it did not completely bash social media, and instead made efforts to use it more mindfully" "I enjoyed that the program provided a more positive way to approach social media use, rather than preaching abstinence from all social media platforms. I think that approach, while probably beneficial for our mental health, may not be realistic." "I liked that it was more straightforward and you could practice giving advice from an outside perspective" "The scenarios where we provided advice were helpful because it allowed me to examine thoughts I've had in the past in a more objective manner" "I like how it gave you tips and specific examples. I also like how you got to choose what you struggle with and it helped you with that problem." "I liked that there was a disclaimer about harassment along with a multitude of resources listed."	"It felt as though the information provided was general information that most think of." "Some of the feedback/information was common sense and felt like stuff I knew already. I would add statistics and other forms of evidence to prove your points further." "I'd make it feel less individualistic and more focused on how social media affects college students as a student whole." "I would not include giving advice to people." "I would add in more options for things the student would like to work on regarding social media." "I think it could have… validated that sometimes there is some reality to those negative thoughts (i.e. my friend is more popular than me) and that even if they are valid it shouldn't affect self-worth. I feel like it was too focused on pointing out that most of the time those thoughts are irrational, but as someone who struggles with constantly thinking about the worst case scenario it's much more helpful for me to be told that even if it's true, I'll be okay, vs. being told to think positively."

Table 18. Summary of Phase 2 Outcomes

Benchmark Target	Post-Treatment (Outliers Included)	Two-Week Follow-Up (Outliers Included)	Post-Treatment (Outliers Excluded)	Two-Week Follow-Up (Outliers Excluded)
Acceptability				
Program Feedback Scale	√	√	N/A	N/A
Perceived Effectiveness	√	√	N/A	N/A
Target Engagement				
PRISM Skills				
Notice one's own negative thoughts	N/A	X	N/A	N/A
Reappraisal tied to one's own thoughts	N/A	X	N/A	N/A
Reappraisal tied to others' thoughts	N/A	X	N/A	N/A
Control Condition Skills				
Think about technical aspects	N/A	X	N/A	N/A
Discuss technical aspects	N/A	X	N/A	N/A
Intervention Outcomes				
Primary Intervention Outcomes				
Online Negative Bias	√	√	√	√
Online Positive Bias	√	√	√	√
Secondary Intervention Outcomes				
Offline Negative Bias	√	√	√	√
Offline Positive Bias	√	√	√	√
Fear of Negative Evaluation	N/A	X	N/A	√
Fear of Missing Out	N/A	√	N/A	X
Anxiety Symptoms	N/A	X	N/A	X
Depression Symptoms	N/A	X	N/A	X
Social Comparison Orientation	N/A	X	N/A	X
Time Spent on Social Media	N/A	√	N/A	N/A

Note. A checkmark indicates that PRISM outperformed the control condition on the benchmark target as expected. An "X" indicates that PRISM and control condition participants did not significantly differ on the benchmark target. Intervention outcomes only reflect interaction effects and not main effects. "*N/A*" indicates that the benchmark target was not assessed at that timepoint.

Figures

Figure 1. CONSORT Diagram for Phase 2

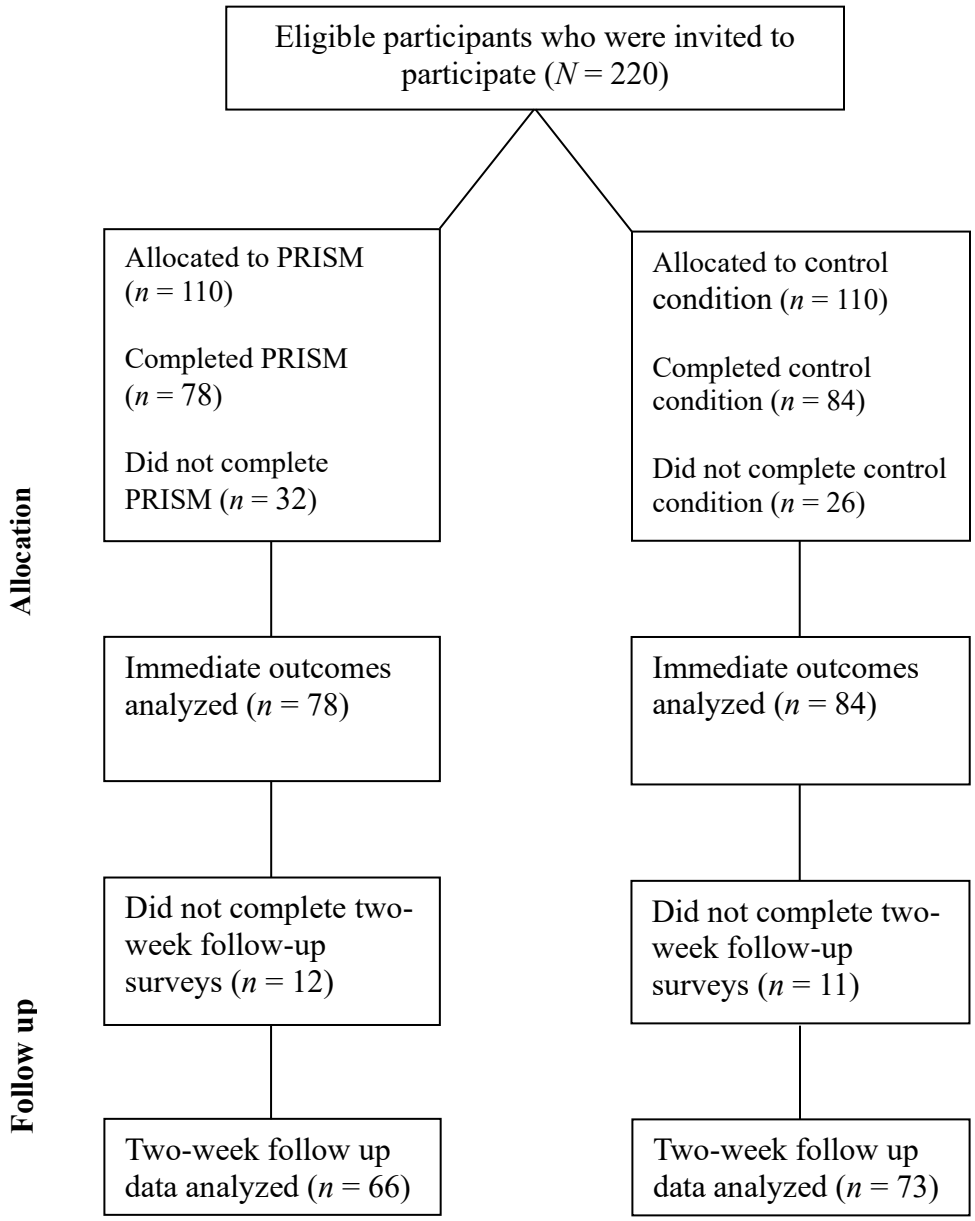
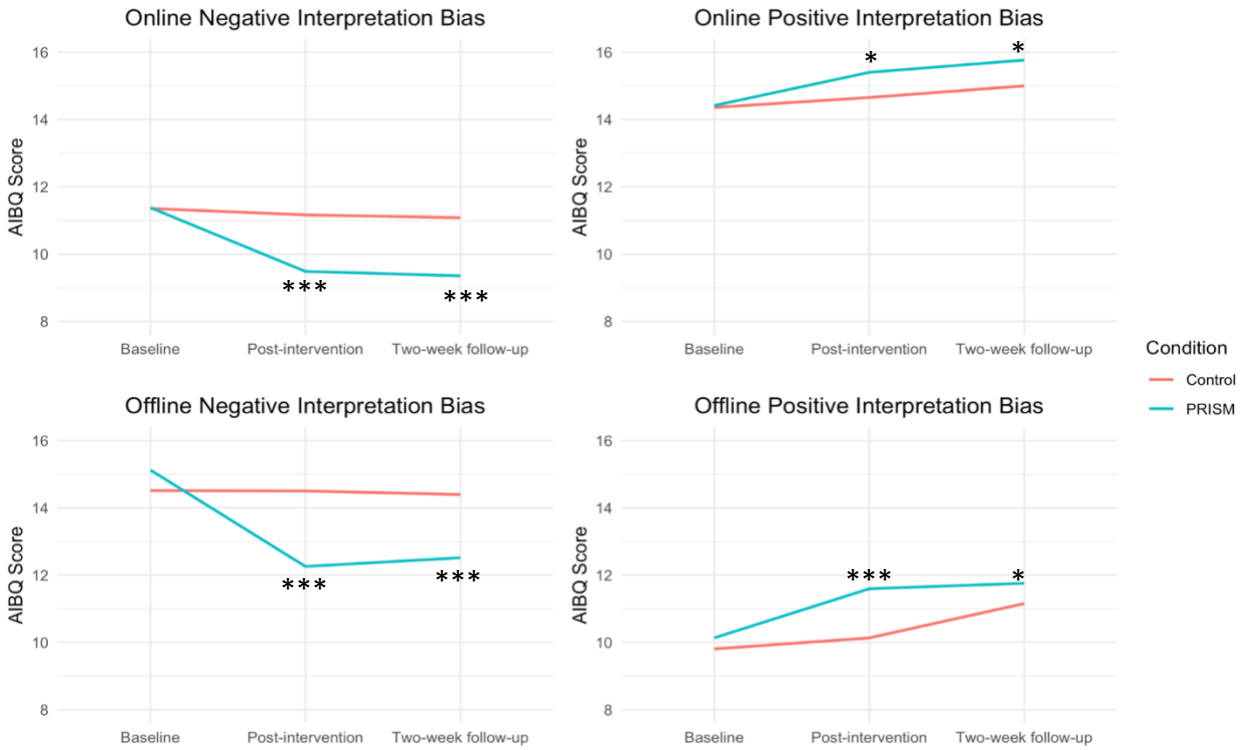


Figure 2. Interpretation Bias Outcomes (Influential Cases Retained)

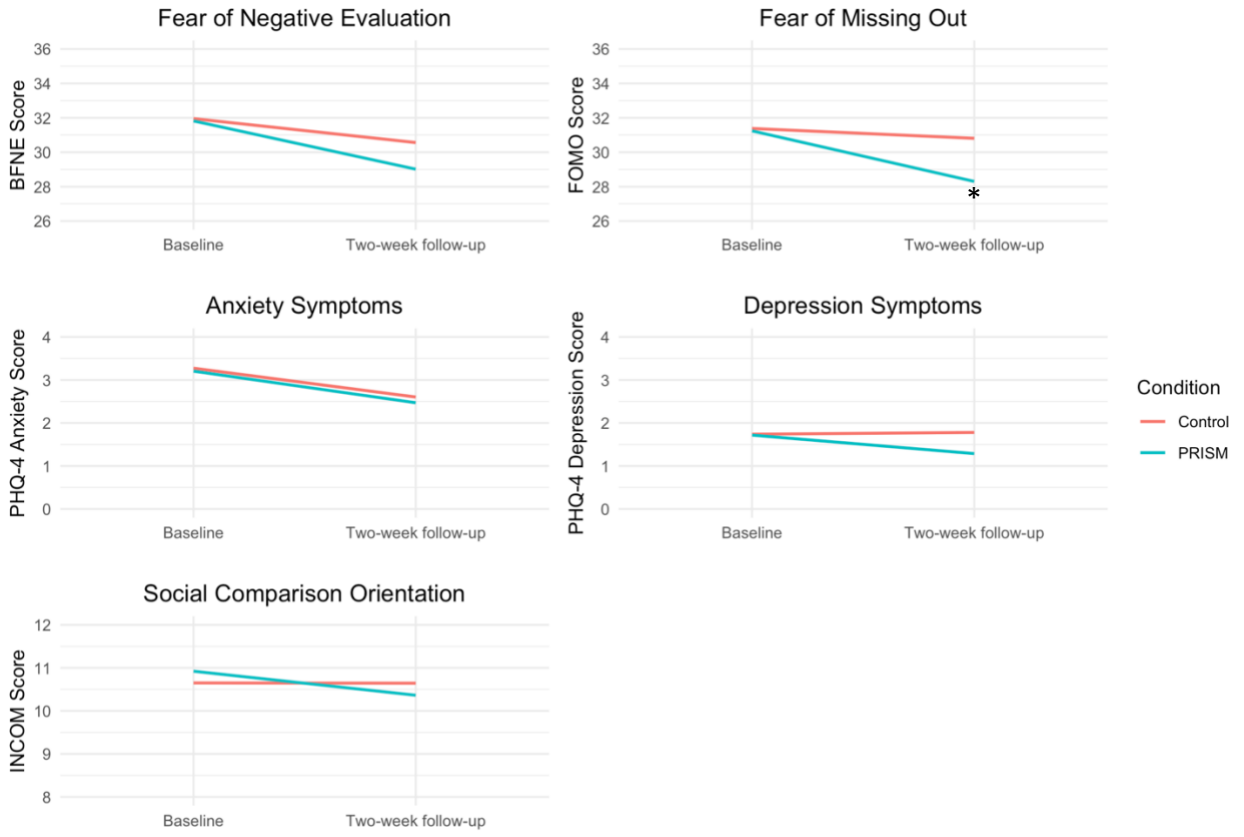
Notes. The figures above show significant effects that occurred in the context of an interaction. In the figures above, all cases are included (i.e., no outliers removed).
*$p < .05$; **$p < .01$; ***$p < .001$.

Figure 3. Interpretation Bias Outcomes (Influential Cases Removed)



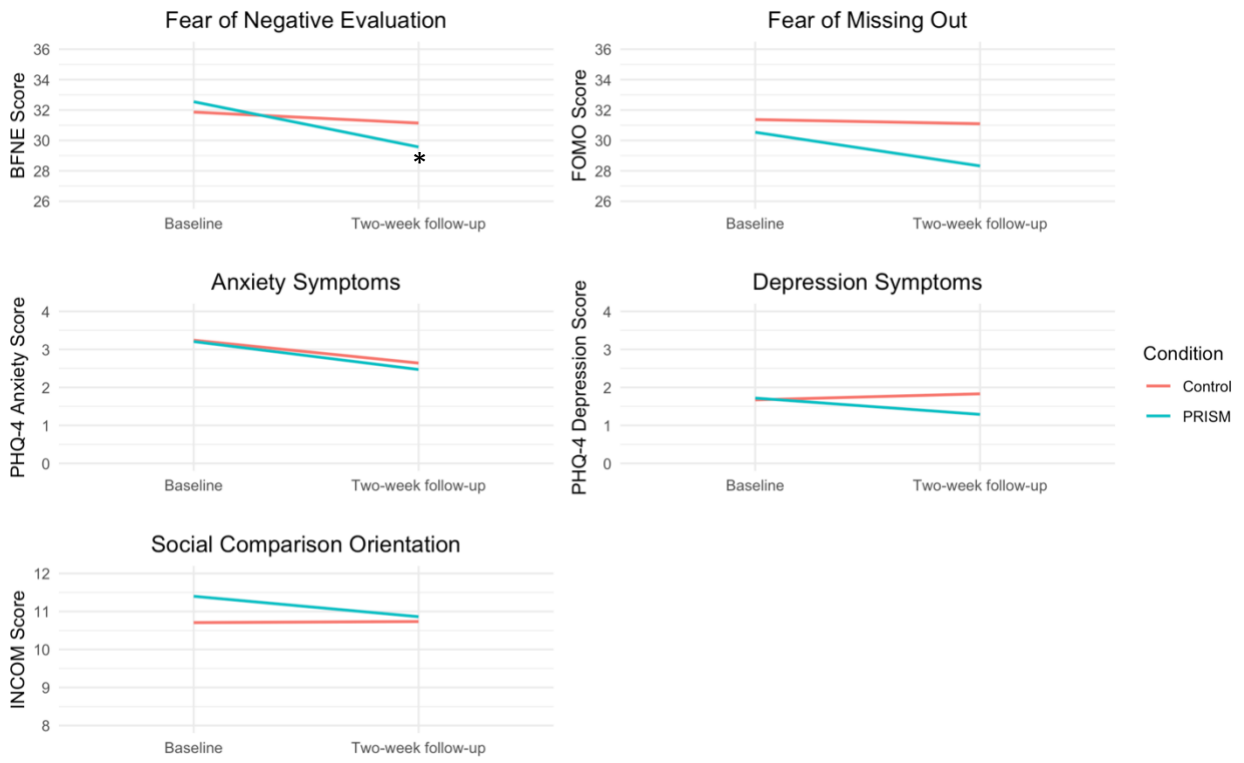
Notes. The figures above show significant effects that occurred in the context of an interaction. In the figures above, outliers were removed.
*$p < .05$; **$p < .01$; ***$p < .001$.

Figure 4. Other Secondary Outcomes (Influential Cases Retained)



Notes. The figures above show significant effects that occurred in the context of an interaction. In the figures above, all cases are included (i.e., no outliers removed).
*$p < .05$; **$p < .01$; ***$p < .001$.

Figure 5. Other Secondary Outcomes (Influential Cases Removed)



Notes. The figures above show significant effects that occurred in the context of an interaction. In the figures above, outliers were removed.
*$p < .05$; **$p < .01$; ***$p < .001$.

Appendix A. Eligibility Screening Form

1. Do you have a personal social media account?
    a. Yes
    b. No

2. Do you have regular access to the internet and to a computer or smartphone?
    a. Yes
    b. No

3. Thinking about the social media platforms you use, combined… About how often do you visit or use social media platforms such as Instagram, Snapchat, TikTok, Facebook, etc.?
    a. Several times a day
    b. About once a day
    c. A few times a week
    d. Every few weeks
    e. Less than every few weeks
    f. Not applicable/don't use social media

4. Social media can bring up negative feelings across many different kinds of situations. For example, some content on social media can trigger people to be self-critical about their body image or feel hopeless about politics/current events.

    For this research study, we are specifically interested in situations where social media content triggers thoughts and feelings related to comparing yourself to others and/or not being liked or wanted by others. Some examples include thoughts that others are more successful than you, judging you, or don't want to include you.

    How bothered are you by seeing content that triggers your thoughts and feelings related to comparing yourself to others and/or not being liked or wanted by others on social media platforms such as Instagram, Snapchat, TikTok, Facebook, etc.?
    - Very unbothered
    - Somewhat unbothered
    - Neutral
    - Somewhat bothered
    - Very bothered
    - Not applicable/don't use social media

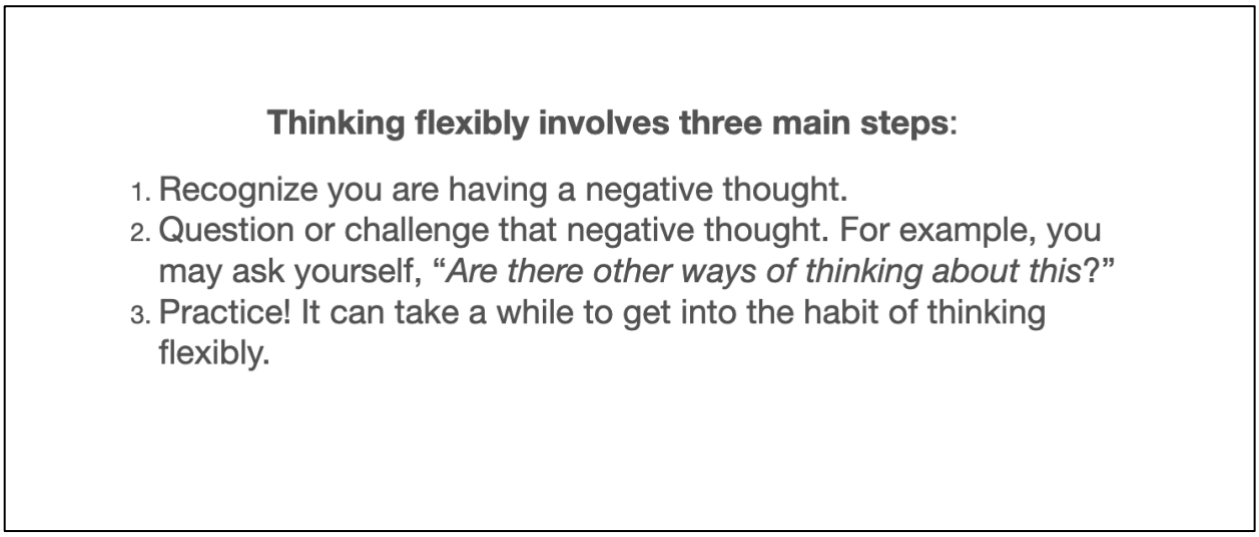
For Phase 2 only:

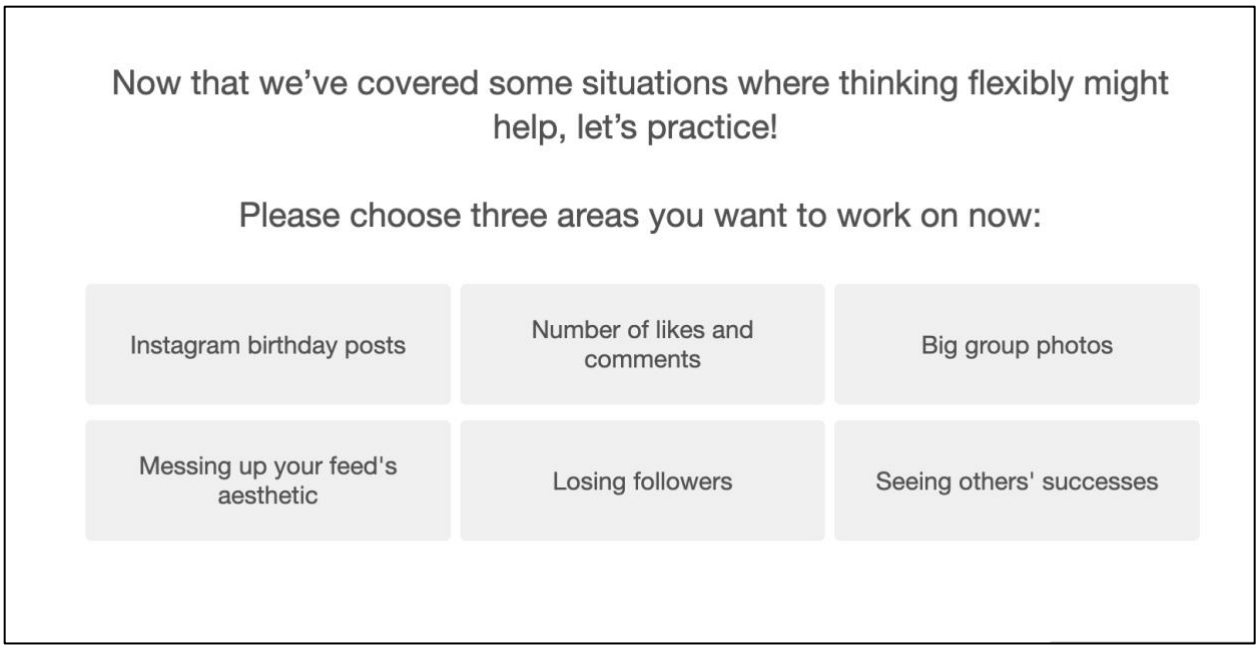
5. How much motivation do you have to change the way you think about social media content that triggers thoughts and feelings related to comparing yourself to others and/or not being liked or wanted by others?
    - No motivation
    - Low motivation

- Some motivation
- High motivation
- Extreme motivation

6. Which social media platform do you use the most intensely?
    - Instagram
    - Snapchat
    - Facebook
    - TikTok
    - Twitter
    - Another platform: _____

Appendix B. Examples of Content related to Thinking Flexibly on PRISM





LIAM

"I was scrolling through Instagram last night, and I saw a picture of my friend hanging out at a party with a bunch of people I don't know.

I must be doing something wrong because I don't have that many friends."

What's another way of thinking about the situation that you might suggest to Liam?

Before we end, we want to share some advice we heard from other college students:

Try to remember that...

- Most things that people post on social media are performative
- Pictures just capture one moment in time, and the pictures people post are ones where it appears that they are at their best
- It's okay to have bad days and not be able to think positively, but just try your best to reframe your state of mind about social media before going on

Appendix C. Reminder Emails for Both Study Conditions

Reminder email for PRISM condition

Subject line: Reminder to think flexibly

Hello,
Thank you for taking the time to participate in our social media research study! We hope the program was thought-provoking and got you reflecting on your social media use.

This is a friendly reminder to practice thinking flexibly. When faced with a negative thought, you can ask yourself, "*Is there another way of thinking about this*?" It can also help to consider what you would say to someone else who is struggling, like you practiced doing during the program.

Keep an eye out in the coming week for the final set of questionnaires from our team!

Thanks very much!
-The PRISM Study Team

Reminder email for Control condition

Subject line: Social media study reminder

Hello,
Thank you for taking the time to participate in our social media research study! We hope the program was thought-provoking and got you reflecting on your social media use.

This is a friendly reminder to continue thinking about what you like and don't like about the technical features of different social media platforms, such as their communication features and privacy settings. Considering these features can help us make informed decisions about which platforms to use ourselves and which ones to recommend to others.

Keep an eye out in the coming week for the final set of questionnaires from our team!

Thanks very much!
-The PRISM Study Team

Appendix D. Supplemental Analyses Exploring the Potential Moderating Role of Baseline Anxiety and Depression Symptoms

Supplemental analyses were conducted to explore whether baseline anxiety and depression symptoms moderated changes in those respective symptoms from baseline to two-week follow-up. There were no formal hypotheses for these analyses, but they were conducted to better understand the pattern of results observed for these symptoms, including whether results were influenced by regression to the mean. As mentioned earlier, for anxiety symptoms, there was a main effect of time but no time-by-condition interaction effect, such that participants in both PRISM and the control condition showed improvements in anxiety, and for depression symptoms, there were neither significant effects of time, condition, nor their interaction.

Anxiety symptoms. A linear model was fit to explore whether baseline anxiety symptoms, study condition, and their interaction predicted change in anxiety symptoms from baseline to two-week follow-up. There was neither a significant main effect for study condition (b = .31, SE = .51, t = 0.61, p = .546) nor a significant interaction between study condition and baseline anxiety symptoms (b = - .12, SE = .14, t = - 0.83, p = .410). However, there was a significant main effect for baseline anxiety symptoms (b = - .29, SE = .09, t = -3.22, p = .002), suggesting that participants with higher baseline anxiety scores tended to show a decrease in anxiety over time, which may indicate regression to the mean.

Depression symptoms. A linear model was fit to explore whether baseline depression symptoms, study condition, and their interaction predicted change in depression symptoms from baseline to two-week follow-up. There was neither a significant main effect for study condition (b = - .21, SE = .31, t = - 0.69, p = .495), nor a significant interaction between study condition and baseline depression symptoms (b = - .15, SE = .13, t = - 1.15, p = .253). However, there was a significant main effect for baseline depression symptoms (b = - .25, SE = .09, t = -2.73, p = .007), suggesting that participants with higher baseline depression scores tended to show a decrease in depression over time, which may indicate regression to the mean.