University of Nevada, Reno

Moderators of Guided and Unguided Self-Help for Depression: The Role of Self-Regulation

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Psychology

by

Emerson M. Epstein

Dr. Anthony Papa/Dissertation Advisor

August, 2022

THE GRADUATE SCHOOL

We recommend that the dissertation
prepared under our supervision by

EMERSON M EPSTEIN

entitled

Moderators of Guided and Unguided Self-Help for Depression: The Role of Self-Regulation

be accepted in partial fulfillment of the
requirements for the degree of

DOCTORATE OF PHILOSOPHY

Anthony Papa, Ph.D.
Advisor

Amy H. Lansing, Ph.D.
Committee Member

Cynthia L. Lancaster, Ph.D.
Committee Member

Mariann Weierich, Ph.D.
Committee Member

Daniel M. Cook, Ph.D.
Graduate School Representative

David W. Zeh, Ph.D., Dean
Graduate School

August, 2022

Abstract

Self-help interventions for depression have been used to help reduce the gap in the need for treatment and the availability of resources. However, not everyone benefits from self-help and unguided self-help tends to have poorer treatment response and higher dropout than guided self-help. Nonetheless, some people still benefit from unguided self-help. Unguided self-help leans heavily on the individual to initiate and maintain change and those with self-regulation are more like to engage in change on their own. The purpose of this study was to understand the self-regulatory processes that may be used to stratify patients into guided and unguided interventions. Using theories of self-regulation, we proposed four self-regulatory processes that impact change, are known to be variable in depressive disorders, and might affect whom benefits from guided versus unguided self-help: autonomous motivation, goal specificity, response inhibition, and delay discounting. After enrolling 336 participants and included 184 in our primary analyses, we observed significant treatment effects of our two self-help groups. We did not observe a significant difference between our two experimental groups on any outcomes (e.g., treatment response, odds of completing, etc.). Significant interactions between these self-regulatory variables and our group term were seldom observed. Some self-regulatory processes did predict outcomes for our entire sample. Implications for future research are discussed.

Keywords: depression, self-help, self-regulation

Table of Contents

List of Tables

List of Figures

Moderators of Guided and Unguided Self-Help for Depression: The Role of Self-Regulation

The lifetime prevalence of a major depressive episode is estimated to be 16.9% in the United States (Andrade et al., 2003). The need for psychological services for depression out paces the supply. This is often referred to as the treatment gap, which is the difference between disorder prevalence and proportion of treated individuals. The World Health Organization has estimated that this gap is around 56% in the Americas for depression (Kohn, Saxena, Levav, & Sarceno, 2004). While stigma about seeking mental health care contributes to this gap (Corrigan, Druss, & Perlick, 2014), this problem is made even more salient when we observe that there are only 182 mental health professionals per 100,000 U.S. citizens (Robiner, 2006). These numbers have fueled efforts to develop and deliver psychological services for those with depression that do not require extensive time commitment from mental health professionals and also have the potential to obviate the stigma of going to a mental health professional's workplace for care.

Self-help interventions have been critical in minimizing this gap (Bennett et al., 2019; Bower & Gilbody, 2005; Cuijpers, Noma, Karyotaki, Cipriani, & Furukawa, 2019; Ebert et al., 2018). There are two broad categories of self-help, guided and unguided. Guided self-help entails an individual engaging in a self-help program with the aid of a therapist or coach. The role of the therapist is to troubleshoot problems and encourage engagement. Unguided self-help, on the other hand, relies solely on the individual to initiate and maintain change. Self-help interventions increase access to psychological services because they require fewer resources from therapists, are generally more affordable than traditional face-to-face therapy, and do not require going into a clinic for care (Bower & Gilbody, 2005; Ebert, et al., 2018). However, it is unclear who benefits the most from guided or unguided self-help. Given the need for and rise in popularity of self-help, it is important to understand who will likely benefit the most from these two treatment delivery methods to ensure that the right people receive the right intervention. Therefore, the aim of this study was to identify individual difference characteristics that could be used to stratify people into guided or unguided self-help programs.

Self-Help as a Treatment Delivery Method

The two major formats of self-help include bibliotherapy and online interventions. In bibliotherapy, therapeutic content is delivered in a book format with psychoeducation and skill-building worksheets embedded right in the book. Online interventions provide such content through statistic web pages or interactive web or mobile applications. Regardless of the format, when averaged across disorders, the general conclusions of self-help programs are that they tend to be more effective than non-treatment control groups, be less effective than face-to-face treatment, and have higher attrition, with guided self-help faring better than unguided self-help (Ebert et al., 2018; Riper et al., 2014; Spek et al., 2007). Examining specific psychological conditions reveals more nuanced results, however.

Adults with major depressive disorder (MDD) are one of the most widely researched populations in the self-help literature. In a meta-analysis of studies examining self-help for clinical samples of depression (Menchola, Arkowitz, & Burke, 2007), self-help was more effective than non-treatment control (Cohen's $d = 1.28$), but less effective than traditional therapist-administered treatment ($d = -0.44$). A limitation of this meta-analysis is that it aggregated numbers from guided and unguided self-help studies and did not compare these two treatment delivery methods. However, contact with a therapist (e.g., no contact at all, contact a couple times during the study, weekly contact to clarify procedures) predicted outcomes, such that more contact was related to better outcomes. This meta-analysis suggests that therapist support increases the efficacy of self-help for depression.

Other meta-analyses separating self-help conditions (i.e., guided versus unguided) corroborate the role of therapist support in self-help interventions. In a meta-analysis examining traditional face-to-face cognitive behavioral therapy (CBT) versus guided self-help (i.e., minimal contact with therapist to work through treatment procedures) for depression and anxiety, Cuijpers, Donker, van Straten, and Andersson (2010) found that both treatment delivery methods were equally effective (d = -0.02), with no difference in long-term outcomes. In a subsequent meta-analysis of treatment delivery methods for adults with depression, face-to-face CBT and guided self-help were again found equally effective (d = 0.12) and guided self-help was more effective than unguided self-help (d = 0.37; Cuijpers et al., 2019). A comparison between face-to-face and unguided self-help was not conducted. This meta-analysis also found that guided self-help and unguided self-help were more effective than waitlist control (d = 0.81 for guided group and d = 0.52 for unguided group). These studies suggest that self-help programs for depression can be effective treatment delivery methods, particularly those with guidance from a therapist. However, despite the advantage of guided self-help, some people still benefit from unguided programs.

Acceptability of Self-Help Programs

Self-help interventions appear to come with higher attrition rates and lower treatment satisfaction than typical face-to-face interventions. However, this disadvantage appears to be isolated to unguided self-help. When looking at just guided self-help, Cuijpers and colleagues (2010) found no difference in dropout rates between face-to-face and guided self-help treatments in the literature (OR = 1.14). Although Cuijpers and colleagues did not report specific dropout rates, across randomized control trials (RCTs), the average dropout rate for CBT for depression has been estimated to be 17% for face-to-face format (Cooper & Conklin, 2015) and 28% for guided self-help format (Richards & Richardson, 2012). Unguided self-help programs appear to have a higher dropout rate. In two meta-analyses, the dropout rate for unguided self-help for depression was noted to be 74% and 83% (Karyotaki et al., 2015; Richards & Richardson, 2012). Richards and Richardson (2012) found a significant difference in dropout rate between guided and unguided self-help in favor of guided self-help, (OR = 7.35). Karyotaki and colleagues (2015) found that male gender, lower education, CBT-based interventions, and comorbid anxiety symptoms increased the odds of dropping out. Increased age appeared to be a protective factor. These studies suggest that support is a critical factor in retaining people in self-help programs, but a minority of people may persist through unguided programs.

Another indicator of acceptability is treatment satisfaction. This is defined as someone's subjective evaluation of how much the person believes the treatment was helpful for him or her.

In a systematic review of treatment satisfaction with self-help for children and adolescents with depression and anxiety, Richardson, Stallard, and Velleman (2010) found that the treatment satisfaction of computerized CBT was moderate to high. With respect to the difference in satisfaction between guided and unguided programs, the extant results are suggestive. Among adults diagnosed with depression, treatment satisfaction trended in favor of guided self-help (Berger, Hämmerli, Gubser, Andersson, & Caspar, 2011). In a study examining self-help for depression, Dear and colleagues (2018) found that their guided group were more likely to endorse “Very Satisfied” in response to the question, “Overall how satisfied were you with the course?” than an unguided group. There was no difference in the likelihood of rating “Satisfied.” Overall, these studies suggest that self-help may be satisfactory for people, with some advantage to guided self-help. However, more research is needed to draw any definitive conclusions about the difference in treatment satisfaction between guided and unguided self-help. In summary, self-help programs are generally acceptable to people, with some potential advantage for guided programs.

Moderators of Self-Help

It is clear that self-help can improve symptoms and is generally acceptable to people. However, support appears to be important for boosting efficacy and acceptability. Nonetheless, some people still benefit from unguided self-help programs. What are unclear are the individual difference factors that could be used to help stratify people into guided or unguided self-help programs. Three studies have attempted to predict outcomes of self-help while utilizing a head-to-head comparison between guided and unguided self-help. Head-to-head comparisons are particular helpful because they control for extraneous variables that may otherwise account for moderation.

Berger, Hämmerli, Gubser, Andersson, & Caspar (2011)

This study included 76 Adults diagnosed with MDD or persistent depressive disorder based on the Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV; American Psychiatric Association, 1994) criteria with no active suicidality. Participants were randomized into a guided self-help, unguided self-help, or waitlist control group. The self-help program was a 10-week, CBT for depression, web-based intervention (i.e., psychoeducation, behavioral activation, cognitive modification, mindfulness/acceptance, interpersonal skills, relaxation/physical exercises/lifestyle modification, problem solving, expressive writing/forgiveness, and positive psychology interventions). The guided self-help arm included therapist guidance via email (i.e., providing feedback on usage, praising usage, problem solving barriers to use, and answering any questions). Primary outcomes were depression symptoms, general psychopathology, interpersonal problems, quality of life, treatment satisfaction, and treatment adherence (i.e., lessons completed and time spent in the program). Predictors were gender, previous treatment, medication status, or presence of co-morbidity.

Both experimental groups saw a large within-subjects effect on depression at post treatment (d = 1.24 for guided group and d = 0.80 for unguided group) and at follow-up (d = 1.26 for guided group and 0.95 for unguided group). Both experimental groups had significant post-treatment reduction in depression compared to the waitlist group (d = 1.14 for guided self-help and d = 0.66 for unguided self-help). Experimental groups did not significantly differ from one another (d = 0.30). With respect to secondary outcomes, the guided group also showed improvement in general psychopathology, interpersonal problems, and quality of life but the unguided group only showed improvement in general psychopathology. Experimental groups did not differ in these secondary outcomes. Treatment satisfaction was trending in favor of guided self-help (d = 0.54).

Lessons completed and time spent in the program did not significantly differ between experimental groups. For instance, the mean number of lessons completed was 8.5 for the guided group and 6.8 for the unguided group. All participants in the guided group completed post-treatment paperwork, while the unguided group had 22 out of 25 participants complete post-treatment paperwork. These authors attributed their relatively low attrition rate compared to other studies (e.g., Meyer et al., 2009) to their more stringent inclusion criteria and contact that they had with participants at the beginning of the study (i.e., structured diagnostic interview). No examined variables predicted outcomes. Overall, guided and unguided self-help can be effective for depression but no moderators of effectiveness were identified.

Kenter, Cuijpers, Beekman, & van Straten (2016)

This study examined 269 outpatient adults between 18-79 years old diagnosed with MDD per a clinical interview. Participants were randomized to a guided problem-solving intervention online or a bibliotherapy control group. All participants were on a waitlist for routine therapy in a community mental health center. The Internet intervention entailed modules on identifying a problem, finding solutions to the problem, selecting one solution, creating a plan to solve the problem with this solution, executing the plan, and evaluating the plan. The last session focused on long-term goals. This arm included a weekly student support component, which encouraged application use and clarified self-help content. The unguided arm included a self-help book with no student support. Outcomes included: depression symptoms, anxiety symptoms, insomnia, quality of life, dropout, and adherence (i.e., completing 4 or more of the modules). Predictors of dropout and adherence included: treatment expectancy, perceived credibility, gender, age, education, and income.

Both treatments had a medium to large within-subjects effect sizes, $d = 0.75$ for guided self-help and $d = 0.69$ for unguided self-help. There was a very small and non-significant between-subjects effect ($d = 0.07$). Between-subjects effects for secondary outcomes were not significant. They did not compare groups on dropout, adherence, and treatment satisfaction. However, non-completers were more likely to be female, younger, and less educated, and have lower income. Depressive severity did not predict dropout. Treatment expectancy and perceived credibility at baseline predicted higher adherence. However, moderation analysis was not completed.

Wade, Cassedy, McNally, et al. (2019) & Wade, Cassedy, Sklut, et al. (2019)

These two studies included the same 150 parents of adolescents hospitalized for moderate to severe traumatic brain injury from across five participating hospitals. Parents and adolescents were randomized to an unguided self-help group or a guided-self help group for depression and conduct. The unguided group included 10 online sequential sessions providing training in positive/cognitive reframing, problem solving, communication, and self-regulation/anger management with email reminders to engage in the program. The guided group also included weekly videoconference sessions for the first month and then biweekly for the next three months (i.e., review didactic content, skills, and completion of homework and problem solve around family identified goals or aims). Outcomes were depression symptoms, general psychopathology, child behavior problems, dropout rate, treatment adherence (i.e., number of lessons completed), and satisfaction. Predictors were ethnicity, education, comfort with technology, and treatment preference.

The guided self-help group saw a significant reduction in depression at the six-month follow-up (*d* = 0.53) and nine-month follow-up (*d* = 0.69), but the unguided group did not see a significant reduction in depression over time. For general psychopathology, the guided group improved from baseline to six-month follow-up (*d* = 0.53), while the unguided group improved between baseline and nine-month follow-up (*d* = 0.92). Number of sessions completed did not differ among treatment arms.

Being in the face-to-face arm, being of minority status, completing fewer sessions, and having recent child injuries were more likely to drop out prior to follow-up. Parents with less education and less comfort with technology improved more with depression in the guided self-help than other groups. No variables moderated the effect of group on general psychopathology. Furthermore, being assigned to a preferred treatment according to the adolescent came with fewer dropouts than being assigned to a non-preferred treatment. Parent and adolescent treatment preferences before treatment did not related to post-treatment satisfaction, adherence, or improvement in child behavior problems.

Summary of Findings

To date, no consistent moderators for self-help for depression have been found. Berger, Hämmerli, and colleagues (2011) found no moderation of the variables they studied (i.e., gender, previous treatment, medication status, or presence of co-morbidity). However, this study was underpowered to find a moderation effect. Twenty-six participants per group are needed to detect a large between-subjects effect size (i.e., *d* =

0.8) assuming a power of 0.8 and a two-tailed significance value of 0.05. Moderation analyses typically need two to three times what is required for a simple between-subjects analysis (Cohen, Cohen, West, & Aiken, 2003).

Kenter and colleagues (2016) found that non-completers were more likely to be female, younger, less educated, and have lower income across their groups. They did not attempt to investigate whether these variables moderated dropout rates between guided and unguided self-help. Their findings are somewhat consistent with a previous meta-analysis of unguided self-help, in which those with more education and increased age were less likely to dropout (Karyotaki et al., 2015). However, this meta-analysis also found that male, not female gender predicted dropout. Furthermore, Berger, Hämmerli, and colleagues (2011) did not find that gender predicted number of sessions completed, a related variable to dropout. Other studies of guided and unguided self-help have not shown a link between demographic variables such as age, education, and income and treatment outcomes (e.g., Berger, Caspar, et al., 2011). Demographic variables are generally inconsistent predictors of outcomes in self-help.

Kenter and colleagues (2016) also investigated baseline depression severity, treatment expectations, and perceived creditability. They found that baseline depression severity did not predict dropout, but treatment expectations and perceived creditability did. Unfortunately, moderation analysis was not performed. Whereas baseline symptoms do appear to predict outcomes across different levels of care (e.g., face-to-face CBT and pharmocotherapy for depression; Weitz et al., 2015), other studies find that baseline symptoms do not moderate outcomes in self-help with other presenting problems (e.g., Berger, Caspar, et al., 2011; Loeb, Wilson, Gilbert, & Labouvie, 2000), lending doubt to the use of baseline symptom severity as an indicator of treatment need. Similarly, treatment expectations and perceived creditability also predict outcomes across various levels of care (Beatty & Binnion, 2016; Lewin et al., 2011; Söchting, Tsai, & Ogrodniczuk, 2016). As a result, Kenter and colleague's results are not helpful for the current study. If a baseline factor either inconsistently predicts outcomes or similarly predicts them across levels of care, then it is hard to use them to stratify care.

Finally, Wade and colleagues (2019) found that those with lower education and lower comfort with technology had better symptom reduction in a guided group than in an unguided group. It is possible that those with lower education and lower comfort with technology may have a harder time utilizing self-help, particularly Internet interventions and may require additional help from a therapist or coach or may benefit more from bibliotherapy. Additional studies could be conducted to test this hypothesis. However, as noted earlier, education is mixed as to whether it moderates self-help outcomes in the literature. Barriers to utilizing self-help (e.g., comfort with technology) may be fruitful factors for future research to investigate.

The lack of significant findings consistent across research may be due to how these studies were conducted. In general, these studies lacked a theory to guide their research into potential predictors of treatment engagement and response. Given that self-help leans heavily on the person to initiate and maintain change, whether partially in guided self-help or completely in unguided self-help, then theories of self-regulation may offer specific and testable hypotheses about moderators of self-help.

Self-Regulatory Processes

There are many theories of self-regulation found in the literature: social cognitive theory (Bandura, 1977; Bandura, 2015), behavioral theory (Rachlin, 1995), self-determination theory (Deci & Ryan, 1980; 2012), control theory (Carver & Scheier, 1982; Powers, 1991; Vancouver et al., 2001), goal-setting theory (Locke, 1991; Locke & Latham, 2006), theory of planned behavior (Ajzen, 1991; 2011), and dual process models (Hofmann, Friese, & Strack, 2009; Hofmann, Schmeichel, & Baddeley, 2012; Milyavskaya, Inzlicht, Hope, & Koestner, 2015; Lally & Garnder, 2013) to name several dominant theories. These theories have a few things in common. First, to initiate self-regulation, a discrepancy between current and desired states is required (Bandura, 1977; Carver & Scheier, 1982; Deci & Ryan, 1980; Locke, 1991). Second, effortful control must be exerted on the self or environment to align actual states with desired states (Ajzen, 1991; Carver & Scheier, 1982; Deci & Ryan, 1980; Rachlin, 1995). Most of these theories explicate the motivational processes that go into self-regulation (e.g., autonomous motivation, Deci & Ryan, 2012; goal setting, Locke & Latham, 2006) or provide attitudinal constructs to account for change (e.g., behavioral intentions, Ajzen, 1991; self-efficacy, Bandura, 2015). However, these theories have not been updated to account for recent developments in dual system information processing, which entail automatic (e.g., habit formation) and controlled (e.g., executive functioning) self-regulatory processes.

The dual process model of self-regulation accounts for some of the shortcomings of past theories (Hofmann et al., 2012; Lally & Garnder, 2013). This theory suggests that self-regulation involves any goal-directed pursuit and operates on two interrelated systems: the automatic system, which entails effortless regulation requiring little or no conscious awareness (e.g., forming productive habits) and the control system, which entails effortful regulation apart of conscious awareness (e.g., controlling undesired automatic responses, making decisions consistent with long-term goals). However, these systems are not orthogonal. According to this framework, the dominant responses to be regulated are well-learned habits but the efforts to regulate them can turn into productive habits with practice (Lally & Garnder, 2013). For instance, Adriaanse, Kroese, Gillebaart, & De Ridder (2014) found that self-control negatively predicted the strength of unproductive habits, which mediated the relationship between self-control and behavior. One of the best predictors of future behavior is past behavior (Ouellette & Wood, 1998). The role of habits in self-regulation is documented in a number of contexts, including eating, exercising, sleeping, meditating, and depression (Adriaanse et al., 2014; Galla & Duckworth, 2015; Watkins & Nolen-Hoeksema, 2014). Those good at self-regulating have a history of regulating unproductive habits, which they can effortlessly call upon when required.

Habit Formation and Reformation in Depression

Unproductive automatic tendencies can make alternative actions less accessible (Lally & Garnder, 2013). The processes that maintain unproductive habits (Lally & Gardner, 2013) are found in the maintaining mechanisms of depression. MDD like other psychological disorders entail automatic and decontextualized patterns of responding that despite creating problems are maintained in a cycle (Beevers, 2005; Fisher, 2015; Hofmann, Curtiss, & McNally, 2016; Moorey, 2010; Watkins & Nolen-Hoeksema, 2014). MDD is characterized by specific cognitive distortions and behavioral avoidance patterns that maintain depression (Young, Rygh, Weinberger, & Beck, 2014). For instance, those with MSS tend to predict that they will fail at tasks, not connect with others, and not enjoy activities (Renner, Lobbestael, Peeters, Arntz, & Huibers, 2012; Riso et al., 2003). These pessimistic beliefs tend to be maintained by confirmation bias (e.g., focusing on negatives, discounting positives; Andrews, 1989; Moorey, 2010) and lead to avoiding activities commonly associated with failure, rejection, and discontent, which in turn prevent people with depression from finding evidence in the contrary to their beliefs (Andrews, 1989; Moorey, 2010). Furthermore, avoidance is maintained by intermittent negative reinforcement, as people with depression have learned to circumvent punishing environments, whether real or perceived (Ferster, 1973; Ottenbriet & Dobson, 2004). This limits the amount of contact people with depression have to positive reinforcement, thereby further perpetuating depression (Carvalho & Hopko, 2011; Moorey, 2010). Environments with minimal positive reinforcement can in turn engender passive and abstract rumination, which can limit the ability to regulate dominant response patterns (Watkins & Nolen-Hoeksema, 2014), such as the negatively biased associative processing common with depression (Beever, 2005). These responses characterize the automatic tendencies of those with MDD.

CBT is an effortful process that, with practice, can develop productive, antidepressant habits. The core clinical change processes of CBT are cognitive restructuring and behavioral activation, which directly target the cognitive distortions and behavioral avoidance found in MDD. These unproductive dominant responses decrease as a result of CBT, giving credence to these maintaining mechanisms and the utility of CBT in correcting these mechanisms (Garratt, Ingram, Rand, & Sawalani, 2007;

Lorenzo-Luaces, German, & DeRubeis, 2015; Quilty, McBride, & Bagby, 2008). Furthermore, CBT comes with a lower relapse rate compared to psychotropic treatments (Young et al., 2014), presumably because CBT teaches people the skills necessary to prevent future depressive episodes. The skills necessary to change depressive psychopathology mirror the maintaining mechanisms. For example, Jacob, Christopher, and Neuhaus (2011) developed a measure of skills developed in CBT, which include two subscales, cognitive restructuring and behavioral activation. These authors found that these skills significantly improve over the course of CBT for depression and the change in these skills predict post-treatment depression scores. CBT improves depression by targeting the factors that maintain depression.

Motivation to Change Depressive Habits

Changing these depressive habits requires various self-regulatory processes. The process that appears to initiate self-regulation across all self-regulation theories is the discrepancy between actual and desired outcomes states. Motivation is the reason for change and can often increase goal attainment (Ning & Downing, 2012). Motivation can be generated by general feelings of discontent with where a person is in his or her life and can come in the form of wanting to do something or having to do something. However, not all motivations are equal. Whereas vague goals such as “wanting to be happy” can be enough for someone to seek therapy, it may not be enough to sustain change efforts. For instance, motivations that are autonomous versus controlled and goals that are specific versus general predict effective self-regulation (Deci & Ryan, 2012; Locke & Latham, 2006; Milyavskaya et al., 2015; Sheldon & Elliot, 1999).

Autonomous motivation is about seeing the personal benefit of change (Deci & Ryan, 2012), the "want to" part of self-regulation (i.e., wanting to change). It can improve goal-attainment by reducing temptations to default to automatic tendencies (Milyavskaya et al., 2015). Those with depression have higher levels of amotivation and lower levels of autonomous motivation in early stages of change, and autonomous motivation appears to predict the maintenance of behavior change (Vancampfort et al., 2016). Furthermore, autonomous motivation early in therapy predicts change later in therapy for depression (Zuroff et al., 2007) and lack of autonomous motivation is a major reason participants give for dropping out of guided self-help for depression (Marks et al., 2003). In sum, it appears that autonomous motivation is a valuable self-regulatory asset that is impaired in those with depression.

Another aspect of motivation that is important for self-regulation is goal specificity, as specific goals provide cues to guide behavior (Brown & Latham, 2002; Seijts, Latham, Tasa, & Latham, 2004). Vague goals such as "wanting to be happy" do not provide such guidance. The Goal Specificity Task measures goal specificity by asking participants to list as many relevant, specific, and plausible goals that they can think of in a given time frame, while coders code these goals for level of specificity (Dickson & MacLeod, 2004). Using this task, those with depression were found to have fewer specific goals than healthy controls; approach and avoidance goals where both less specified (Belcher & Klangas, 2014; Dickson & MacLeod, 2004). Having vague goals reflects the abstract and overgeneral processing found in depression (Sumner, Griffith, & Mineka, 2010; Watkins & Nolen-Hoeksema, 2014) and may contribute to the difficulties that people with depression have in making changes. This suggests that goal specificity may not only be a self-regulatory deficit found in MDD, but may be an important aspect of adaptive self-regulation.

Effortful Control of Depressive Habits

Once motivation is evident, problematic dominant responses must be regulated in order to have successful outcomes, which require the ability to consciously coordinate efforts to address this goal. Although more general executive functioning abilities have been linked to self-regulation (e.g., working memory; Hofmann et al., 2012), successful self-regulation entails targeting specific dominant responses with the aid of specific executive functioning abilities. Two specific and effortful regulatory abilities with significant empirical support that discriminate those with MDD are: (1) inhibiting the attentional capture of negative self-relevant information (i.e., response inhibition; MacLeod, 1991; Mogg & Bradley, 2005), and (2) making choices that maximize long-term rewards (i.e., delay discounting; Prencipe et al., 2011; Pulcu et al., 2014). According to the dual process model of self-regulation, these regulatory abilities may start off as effortful, but with practice may require less effort to initiate (Lally & Gardner, 2013). This suggests variability in how well someone can self-regulate. Those experiencing depression with a history of exercising response inhibition and curbing delay discounting may be better suited to initiate and maintain change on their own.

Those high in response inhibition can direct attention away from prepotent information in the environment (MacLeod, 1991). The attentional capture of this type of prepotent information in the environment limits attentional engagement with other aspects of a given situation thus interfering with the ability to adaptively respond to other tasks associated with that situation (Collins & Jackson, 2015; Pessoa, Kastner, &

Ungerleider, 2002). People experiencing MDD have negative self-schemas that increase the salience of negative self-relevant information in a given situation (Young et al., 2014). This bias appears to be more pronounced with self-relevant negative information than negative information in general (Bradley, Mogg, Millar, & White, 1995; Epp, Dobson, Dozois, & Frewen, 2012; McCabe & Gotlib, 1995; Mogg & Bradley, 2005). Regulating such information suggests an effortful self-regulatory process that is deficient in those with depression. Having the ability to inhibit these automatic responses may suggest a greater ability to make changes to depressive habits.

A common measure of response inhibition among psychopathology research is the emotional Stroop task. In this task respondents are asked to ignore the semantic meaning of emotionally laden words (e.g., "failure") and name the color the word is printed (Smith & Waterman, 2003). Emotionally laden words can be tailored to specific disorders (e.g., Bauer & Cox, 1998). In a meta-analysis of the emotional Stroop effect in depression, Epp and colleagues (2012) found a between-subjects interference effect comparing clinically depressed and control participants among negative content (Hedges's $g = 0.98$) and a within-subjects effect comparing negative and neutral words among clinically depressed participants (Hedges's $g = 0.25$). This within-subjects effect was not evident in control participants. Salience of disorder-specific information and difficulties regulating attentional focus to such information, as measured by the emotional Stroop task is evident across psychological disorders (Williams, Mathews, & MacLeod, 1996). This makes response inhibition, as measured by the emotional Stroop task, a particularly important factor in the self-regulation of psychological disorders.

Delay discounting is another self-regulatory process that plays into effortful decision making related to immediate and delayed choices (Prencipe et al., 2011). Delay discounting refers to devaluing an outcome when its receipt is delayed (Odum, 2011). Those with low delay discounting tend to make choices that maximize long-term gains (Madden & Bickel, 2010). People experiencing MDD have pessimistic views of future rewards in that they perceive future rewards are less likely to occur than healthy controls (Alloy & Ahrens, 1987; Dickson, Moberly, O'Dea, & Field, 2016; Young et al., 2014). Neurological evidence corroborates this phenomenon. Those that are depressed have a diminished neurological response to anticipatory reward (Stringaris et al., 2015). Dickson and colleagues (2016) found that depressed participants reported a greater inclination to disengage from unattainable goals and a lower ability to re-engage with new goals after some goal blockage than control participants. Forgoing instant gratifications and making decisions consistent with long-term outcomes is an effortful self-regulatory ability that is deficient in those with depression. Those without this ability (i.e., high delay discounting) may find it harder to make changes on their own.

A common measure of delay discounting used in psychopathology research is the monetary choice task. In this task, respondents are asked to choose between two hypothetical monetary choices, one fixed with varying delays (e.g., $200 in one month, three months, etc.) and another that varies but is given immediately (e.g., $100 now, $75 now, etc.; Du, Green & Myerson, 2002). Discounting per the monetary choice task is evident in those with depression (Imhoff, Harris, Weiser, & Reynolds, 2014; Pulca et al., 2014; Takahashi et al., 2008). For instance, Pulcu and colleagues (2014) studied patients with current MDD, remitted MDD, and healthy controls and found that current MDD participants discounted larger (i.e., $119-$134, after adjusting for inflation) long-term rewards more than the other two groups. There was no difference between groups for small or medium long-term rewards. Pursuing immediate outcomes and discounting delayed outcomes is evident across psychological disorders (Amlung et al., 2019). This research suggests that delay discounting, as measured by the monetary choice task, may be used to predict effortful change.

Other Factors to Consider

Self-efficacy and behavioral intention are also noted in theories of self-regulation and hence merit a discussion of their potential predictive validity within a self-help context. Theories of self-regulation deviate in their treatment of self-efficacy (i.e., belief in a person's ability to achieve some outcome). Several theorists suggest that low self-efficacy can limit a person's ability for change (Bandura, 2015; Carver, 2018), while others say that high self-efficacy is a product of successful past performance (Powers, 1991; Vancouver et al., 2001). Whereas several meta-analyses have found a link between high self-efficacy and increased task performance (Gwaltney, Metrik, Kahler, & Shiffman, 2009; Multon & Brown, 1991; Stajkovic & Luthans, 1998), some have criticized this work as relying on correlations and between-subjects designs that cannot distinguish the directionality of self-efficacy and task performance (Biglan, 1987; Sitzmann & Yeo, 2013). In an illuminating meta-analysis of the within-subjects relationship between performance and expectancies, Sitzmann and Yeo (2013) found that the relationship between self-efficacy and subsequent task performance was not significant after controlling for time, an important covariate in within-subjects analyses (Singer & Willet, 2003), while the relationship between past performance and self-efficacy remained significant. This brings doubt to self-efficacy as a self-regulatory process that could moderate self-help outcomes.

Ajzen's theory of planned behavior (1991; 2011) proposes that behavioral intention (i.e., how committed someone is to something) is important for making change. However, the predictive validity of this construct is mixed in the literature (Ajzen, 2011). Two factors that moderate whether intention predicts behavior is the strength of the dominant response to be regulated (Lally & Gardner, 2013) and whether someone has the ability to inhibit dominant responses (Ajzen, 2011). That is, a person can intend to change, but with strong undesired responses and without the ability to regulate such responses, intention does not really matter. Another problem with this construct is that intention may be conflated with motivation. In a meta-analysis of studies integrating factors from the Theory of Planned Behavior and Self-Determination Theory, Hagger and Chatzisarantis (2009) found that after correcting for sampling and measurement error, motivation and intention correlated at .52, suggesting that intention may be a superfluous construct. This research suggests that intention may not be suited to moderate outcomes in self-help and may even require analyzing complicated interactions that would be difficult to interpreted (e.g., interaction between time, group, inhibition, and intention).

Current Study

Given that self-help leans heavily on the individual to initiate and maintain change, especially for unguided self-help, the current study investigated whether self-regulatory processes moderate the effect of guided and unguided self-help on our primary outcomes: depression, odds of completing the program, sessions completed, homework assignments completed, and emails sent. Theories of self-regulation offer several predictions about who may be successful in unguided self-help and who may need more support. Those with depression that have higher autonomous motivation are more likely to initiate and maintain change. People with depression also vary in the specificity of their motivation, with more explicated goals contributing to better self-regulation. Additionally, MDD is characterized by unproductive habits that require skills to combat. Having the ability to inhibit dominant responses and make future-oriented decisions makes it easier for someone to change on his or her own. Theories of and research on self-regulation suggests that autonomous motivation, goal specificity, response inhibition, and delay discounting are important processes for self-regulation, are lacking in MDD, and may moderate outcomes in guided and unguided self-help.

A couple preliminary hypotheses were tested before our primary hypotheses. First, given the prior literature, we expected to observe a similar treatment effect, in which our guided and unguided groups would demonstrate a larger reduction in depression than our waitlist group. We also expected that our two experimental groups would differ, albeit slightly, with more benefit being found in the guided group than the unguided group. Also similar to previous research, we expected that those in the guided condition would be more likely to complete the self-help program than the unguided condition. Although some have reported that sessions completed do not differ between guided and unguided groups, we predicted that we would see a different based on theoretical grounds (e.g., guided self-help is designed to bolster engagement). We are not aware of any research comparing guided and unguided self-help on assignments completed, but based on a similar rational as session completed, we expected to observe a group effect. Given that email support was built into the guided self-help condition, we expected a group effect with our emails sent outcome. All of our primary hypotheses entailed moderation of these group effects, and are as follows.

Hypothesis 1: Self-Regulation Will Moderate the Effect of Group on Symptom Reduction Over Time

Although we will explore the aforementioned auxiliary hypotheses (e.g., the guided group will have a larger reduction in depression over time than the unguided group), our primary aim was to examine if self-regulatory abilities moderate the influence of self-help on outcomes. Guided self-help is designed to increase engagement in self-help and hence bolster those that might not have the resources to engage in self-help on their own. Those with high autonomous motivation, high goal specificity, high response inhibition, and low delay discounting are more likely to initiate change on their own and are less likely to disengage from goal efforts. As a result, we hypothesized that self-regulatory abilities at baseline would moderate the effect of our treatment groups on symptom reduction over time. More specifically, we anticipated that there would be a larger group difference in depression over time among those with low self-regulation. This pattern of results also suggests that those with high self-regulation will fair better overall than those with low self-regulation. Hence, we also hypothesized that self-regulation on its own would moderate the effect of time on depression.

Hypothesis 2: Self-Regulation Will Moderate the Effect of Group on the Odds of Complete the Program

Using a similar rational as hypothesis one (i.e., more self-regulation, less goal disengagement; guided bolstering those that need it), we predicted that self-regulation would moderate the relationship between our experimental groups and the odds of completing the self-help program. We anticipated that there would be a larger group difference in the odds of completing the self-help program among those with low self-regulation. We still expected to see a group effect among those with high self-regulation, albeit smaller. This suggests a main effect of self-regulation on the odds of completing the program.

Hypothesis 3: Self-Regulation Will Moderate the Effect of Group on the Rate of Sessions and Homework Assignments Completed

Completing sessions and homework assignments is import for the effectiveness of self-help. However, it is also quite effortful. Further, guided self-help is designed to improve adherence to self-help protocols. As a result, we predicted that self-regulatory would moderate the relationship between our experimental groups and the number of sessions completed and the number of homework assignments completed. More specifically, we hypothesized that there would be a larger group difference on the rate of sessions completed and homework assignments completed among those with low self-regulation than those with high self-regulation. This too suggests a main effect of self-regulation.

Hypothesis 4: Self-Regulation Will Moderate the Effect of Group on the Amount of Emails Send to Study Personnel

Given that email correspondence was encouraged in the guided condition, we anticipated a significant group effect. We also anticipated that those with low self-regulation would need more help and hence ask for help, especially in the guided self-help condition. This suggests a significant group by self-regulation interaction, such that there would be a larger group effect among those with low self-regulation than those with high self-regulation. Given that this hypothesis is purely theoretical and no other research is available to support this claim, we were open to an alternative possibility. Just because those with low self-regulation may need more help, does not mean that they are more likely to seek help. They may not have the skills to do so. Those with high self-regulation may be better adept at utilizing their social resources to bolster their own change. As a result, we would expect a larger group difference in emails sent among those high in self-regulation. Due to these competing predictions, we considered this to be an exploratory hypothesis.

Methods

Participants

Recruitment Sources

Participants were recruited across the United States using print and online advertisements. Our participants came from five major sources. However, our major source of recruitment was Mechanical Turk (MTurk) through Human Intelligence Tasks (HITs). MTurk is a crowdsourcing platform where workers get paid from requesters to perform various tasks. Some have criticized the use of MTurk to collect data for research, such as workers misrepresenting themselves (Kan & Drummy, 2018) and practice effects from experienced workers (Robinson, Rosenzweig, Moss, & Litman, 2019). Although the quality of data obtained from MTurk is similar to other convenient samples, such as university students (Kees, Berry, Burton, & Sheehan, 2017), different convenient samples come with unique biases that should be addressed (Landers & Behrend, 2015).

We attempted to improve the quality of our data by posting our HITs with CloudResearch (formally known as TurkPrime; Litman, Robinson, & Abberbock, 2017), which is an Internet-based platform that allowed us to recruit quality participants and contact participants throughout the course of the study. First, we utilized CloudResearch's panels feature, which allowed us to only screen those who had endorsed being diagnosed with depression in a previous survey independently conducted by CloudResearch. Second, we allowed workers at all levels of experience to complete our screener so as not to overly recruit experienced workers that might have already come in contact with similar study procedures. Third, we only allowed workers with an approval rating above 80% to take our screener as these workers generally engage in tasks as they are meant to be engaged. Fourth, we screened participants who have passed CloudResearch's attention and engagement measures. Fifth, we avoided those who came from suspicious geocode locations and only screened those whose country (i.e., the United States) had been verified. Sixth, we only screened those with unique IP addresses, so that workers could not do the study again from a different MTurk account. Finally, we posted our screening HIT at various times during the day. This technique of microbatching limits bias introduced from collecting data at one particular time of day and at one particular day of week. Overall, these procedures were employed in hopes of improve our data quality.

Several other recruitment sources were used. First, we recruited community members by posting fliers on community boards across the Reno/Tahoe metropolitan area and through word of mouth. Second, we recruited UNR students by emailing UNR staff about our study (e.g., department directors, instructors, and advisors), asking Counseling Services and Student Health Center to pass out fliers, and posting the study on SONA System's University Research Software (SONA). Third, we recruited participants using social media, such as Facebook and Reddit, by posting an online advertisement. Fourth, we used Craigslist to recruit participants across the United States by posting an advertisement on Craigslist's computer gig page. We also attempted to run a YouTube ad and contact Nevada State Agencies (e.g., Reno Housing Authority), but these recruitment sources did not bring in any participants. All recruitment material, except MTurk HITs directed perspective participants to a study specific landing page that outlined the study expectations in more detail (https://dashstudy.org). MTurk HITs outlined the study in more detail than other recruitment material.

Eligibility Criteria

After consenting, a screening tool was presented. To be eligible for this study, participants had to (1) be over the age of 18, (2) be fluent in English, (3) have regular access to a computer that afforded them privacy, (4) be experiencing significant symptoms of depression in the prior two weeks (i.e., score 10 or above on the Patient Health Questionnaire - 9; PHQ-9), and (5) have not changed their anti-depressant medications in the past month if they had a prescription. Participants were excluded if they did not meet these criteria and if they (1) were actively suicidality, (2) were in therapy for depression, (3) had experienced a psychotic episode in the past six months, and (4) had experienced a manic episode in the past six months. Those that did not meet criteria for this study were given a link to Psychology Today (https://psychologytoday.com/us) so they could search for a therapist in their local area, a link to MoodGym (https://moodgym.com.au/) so they could access a comparable online self-help application, and a phone number and link to the Suicide Prevention Lifeline (1-800-273-8255; https://suicidepreventionlifeline.org/chat/).

These criteria were developed for several reasons. First, a score of 10 or above on the PHQ-9 has shown to have excellent sensitivity and specificity for a MDD diagnosis (Kroenke & Spitzer, 2002). We excluded those already in therapy and those not stable on anti-depressant medications because we know that these treatments can impact outcomes (Petersen et al., 2007) and we did not want to introduce extraneous variance into our data, thereby improving the internal validity of our claims. We chose not to completely exclude those taking anti-depressant medications because these medications are commonly taken (Pratt, Brody, & Qiuping, 2011) and there is evidence to suggest that depressive symptoms can remain stable on anti-depressant medications (Petersen et al., 2007). Additionally, since those with active suicidality and elevated symptoms of psychosis and mania require special considerations that CBT for depression does not address, these individuals were excluded. Given the high comorbidity between depressive and anxiety disorders (Kessler, Chiu, Demler, & Walters, 2005), those with comorbid anxiety disorders were not excluded.

Participants were compensated for their time. MTurk participants received compensation for various parts of the study and were paid for each individual HIT (e.g., session 1 was called Session 1 HIT). The total that they could have received was $10.50. SONA participants were given a choice to receive SONA credits and/or an Amazon gift card for their participation. SONA participants could have received up to 11 SONA credits used towards extra credit in their courses. All other participants received a $3.00 Amazon gift card for completing the pre program measures, a $5.00 Amazon gift card for completing the remainder of the study, and a $2.00 Amazon gift card if they completed a post PHQ-9 after four weeks of not completing a session. Finally, they received a chance to win one of three $50.00 Visa gift cards. For every depression measure they completed, they gained one chance to win, five chances to win total. They were told that the odds of winning a Visa gift card were roughly 1 in 60. Finally, they were told that the free self-help program was valued at $27.00 based on a similar self-help program (i.e., MoodGym yearly subscription cost).

Measures

Demographics and Technology

Several questions were asked to screen potential participants and gain a better understanding of our sample. We collected participants' age, gender, ethnicity, and education and whether they were fluent in English. We asked a few questions to understand their access to and comfort with technology: (1) "Do you have access to a computer with the Internet that you could regularly use in privacy?" (2) "On average, how many hours do you spend using the Internet a week?" (3) "On a scale from 0 to 10, 10 being very confident, how confident do you feel in your ability to use the Internet?" (Wade, Cassedy, Sklut, et al., 2019). We automatically collected two technology specifications, which were predominately used to account for any variability in our behavioral tasks (e.g., emotional Stroop task). First, we measured participants Internet speed. We did this by measuring how long it took to upload a photo to their browser. Second, we measured the height and width of participants' Internet browser. When participants got to the behavioral measures of the pre program measures, they were asked to switch over to a desktop or laptop if they were using their phone or tablet. At the next screen, we triggered a JavaScript function to make their browser full screen and then measured the height and width of their browser.

Depression Background

We then asked questions about participants' depression, including their age of onset, course of their depression (e.g., episodic, continuous), number of episodes (i.e., if episodic), and historical and current treatment status (e.g., psychotherapy, psychopharmacology). We attempted to assess for depression subtypes, particularly atypical and melancholic depression, because different treatment responses have been observed among different subtypes (Fava et al., 1997). These subtypes appear to be distinguished by reactivity to positive events (Bentley, Pagalilauan, & Simpson, 2014). Hence, we asked participants, "In your most recent period of depression, did you ever feel better, even temporarily, when something good happened or there was the possibility of something good happening?" These questions gave us a better understanding of our sample and to whom our results generalized.

Outcome Measures

Patient Health Questionnaire - 9. A common and brief measure of depression is the Patient Health Questionnaire - 9 (PHQ-9). This measure has been well validated to screen for MDD and measure depression severity (Martin, Rief, Klaiberg, & Braehler, 2006; Kroenke & Spitzer, 2002). It consists of nine items that are responded to within a retrospective context of two-weeks. For instance, "Over the last 2 weeks, how often have been bothered by any of the following problems? Little interest or pleasure in doing things." Each item is responded to on a scale ranging from "Not at all" = 0 to "Nearly every day" = 3. Scores range from 0 to 27.

Researchers have reported high internal consistency (Cronbach's alpha = 0.89; Rief, Nanke, Klaiberg, & Braehler, 2004) and test-retest reliability (Intraclass correlation

= 0.81; Löwe, Unützer, Callahan, Perkins, & Kroenke, 2004). These psychometric properties make the PHQ-9 a good measure for tracking depressive symptoms across time. For the current study, we obtained a slightly larger Guttman's lambda 6 (.86) than Cronbach's alpha (.85), which suggests unequal item loadings for our PHQ-9 measure (McDonald, 1999; Zinbarg, Revelle, Yovel, & Li, 2005). The factor loadings ranged from .28 to .80, also suggesting that the assumption of equal item loadings (i.e., tau equivalence) was violated. In these cases, omega is a better measure of general factor saturation (Flora, 2020), which is the proportion of variance in the total score accounted for by the general factor. Using a confirmatory factor analysis (CFA) to specify a latent factor of depression and after specifying several correlated residuals above .1, we obtained a CFI of .96, TLI of .93, RMSEA of .08, and SRMR of .04, which are generally acceptable. Omega for this model was .79.

We added a question to assess for active suicidality. Question 9 of the PHQ-9 assess for any suicidal ideation, passive or active. If a participated endorsed any value except "Not at all," then they were asked, "A lot of people have passive thoughts of suicide that they will not act on. Is there a chance that you will act on these thoughts and commit suicide in the near future?" Those that clicked "Yes" were excluded from the study, provided the National Suicide Prevention Lifeline (NSPL) phone number (1-800-273-8255), and directed to the NSPL online chat service (https://suicidepreventionlifeline.org/chat).

Program Usage/Engagement. A few outcome variables were derived from participants' usage of the self-help program: (1) whether they started the program, (2) the number of sessions completed, (3) the number of homework assignments completed, (4) whether they finished the program, and (5) how many emails they sent to study personnel. Whether they started the program or not was used in a preliminary analysis, while the other four outcomes were used in the primary analyses. Participants did not have to complete a homework assignment for their session to be considered complete. However, completing homework is important for the effectiveness of self-help and hence was also analyzed. We calculated the total number of emails by adding up all emails sent and subtracting this by the number of successive emails before a reply from study personnel (i.e., some participants sent multiple emails in a row before receiving a reply). This prevented overinflating participants scores if they sent multiple emails that could have otherwise been sent in one email. These outcomes were assessed once the study was completed. The study was marked as completed after all participants either completed their final measure of depression or ceased to complete a session after 28 days of it becoming available.

Symptom Rating Measures

Generalized Anxiety Disorder - 7. Given that we did not exclude those with comorbid anxiety, we measured anxiety to control for any effect that anxiety may have on our outcomes (Karyotaki et al., 2015). The Generalized Anxiety Disorder - 7 (GAD-7) was used as this measure has been well validated to measure anxiety severity (Spitzer, Kroenke, Williams, & Löwe, 2006). It consists of seven items that are responded to within a retrospective context of two-weeks. For instance, “Over the last 2 weeks, how often have been bothered by any of the following problems? Feeling nervous, anxious or on edge.” Each item is responded to on a scale ranging from “Not at all” = 0 to “Nearly every day” = 3. Scores range from 0 to 21. Spitzer and colleagues (2006) reported high internal consistency (Cronbach's alpha = 0.92) and test-retest reliability (Intraclass correlation = 0.83). These psychometric properties make the GAD-7 a good measure for measuring baseline anxiety symptoms. Using the same internal consistency procedures as above, we obtained an omega value of .82, which is within an acceptable range.

Mood Disorder Questionnaire. Isometsä and colleagues (2003) developed the Mood Disorder Questionnaire, which is made up of twelve "Yes" or "No" questions with the following stem, "Has there ever been a period of time when you were not your usual self and…" For the purpose of this study, we framed this stem to be "In the past six months." For example, "In the past six months has there ever been a period of time when you were not your usual self and you were so irritable that you shouted at people or started fights or arguments?" Another question assessed for whether these symptoms have occurred during the same period of time, while another assessed for how much of a problem these symptoms have caused. A recent meta-analysis by Wang and colleagues (2019) found that endorsing (1) seven out of twelve symptoms, (2) "Yes" to the symptoms occurring together, and (3) "Moderate" or "Severe" problems demonstrated a sensitivity of .8 and specificity of .7 for a bipolar diagnosis. This measure was only used to screen out those with active mania in the past six months. Those that met the criteria by Wang and colleagues (2019) were excluded from the study.

WHO WMH-CIDI - Psychosis Screen. The World Health Organization originally developed this measure as a semi-structured interview (World Health Organization, 1990). It includes six major sections assessing the major symptoms of schizophrenia (e.g., visual hallucinations, auditory hallucinations, thought insertion, thought broadcasting, delusions of reference, and delusions of persecution). The instructions read, “The next six questions have to do with unusual experiences. We are not talking about unusual experiences when you were dreaming, half asleep, or under the influence of alcohol or drugs. We are asking if you have experienced any of these things within the last six months. Take your time to read each question carefully.” For instance, “The first experience is about seeing visions that other people who were there could not see, such as seeing objects, people, lights, or patterns that were not actually present. We are not talking about religious or spiritual visions or seeing things when you were dreaming, half asleep, or under the influence of alcohol or drugs. In the past six months, did you ever see a vision that other people could not see?” This measure was only used to screen out those with active psychosis in the past six months. If any of the six questions were answered with a “Yes”, the perspective participant was excluded from the study.

Self-Regulatory Measures

Nijmegen Motivation List 2. Keijsers and colleagues (1999) refined the initial NML as a measure of motivation for therapy. Their refinements led to greater psychometric properties. The Nijmegen Motivation List 2 has three subscales: preparedness, distress, and doubt. The preparedness subscale was the only subscale that predicted dropout in an outpatient clinic performing CBT (Keijsers et al., 1999) and has greater predictive validity than the full scale (Kampman, Keijsers, Hoogduin, & Hendriks, 2008). As a result, the preparedness subscale was used in the current study. This scale has 10-items to measure how willing/ready someone is to engage in treatment (e.g., “I will do anything to get rid of my problems.”). Each item is rated on a 6-point scale ranging from “Not at all applicable” = 1 to “Very applicable” = 6. The internal consistency of this scale is acceptable (Cronbach’s alpha = 0.81). Using the same internal consistency procedures as above, we obtained an omega value of .91, which is within an acceptable range.

Goal Specificity Task. The Goal Specificity Task has been developed to quantify the specificity of spontaneously generated idiographic goals (Dickson & MacLeod, 2004). In this task, participants were first given a definition of goals ("Goals refer to future experiences that individuals think they will be trying to accomplish or trying to avoid"). Participants were then told that they would have to think of goals that are important to them and are plausible to achieve in the next week, next month, or next year. We added instructions to pull for varying degrees of specificity (e.g., "Some goals may be well thought out, while some goals may be less thought out. We want both if you have both."). They were then provided with two prompts displayed in counterbalanced order. The first prompt asked participants to list approach goals ("In the future it will be important for me to accomplish…"), while the second prompt asked participants to list avoidance goals ("In the future it will be important for me to avoid…").

A coding scheme was devised to enumerate the specificity respondents used in framing their goals. This coding scheme drew, in part, on the work of Locke et al. (1989) and Williams et al. (1996) and had been successfully used in Dickson and MacLeod's (2004) study. Goals can vary with respect to specificity by including either particular targets or global aspirations. A particular target refers to a single action event that lasts less than a day that people want to have happen. A global aspiration refers to a "type of event," or the sort of thing that people hope to have happen. Goals were coded as either general, moderate, or specific. General goals describe global aspirations rather than particular targets (e.g., "be happy"). Moderate goals describe specific targets without specification of place, time, or people (e.g., “meditate on a regular basis”). Specific goals describe specific targets with at least one specification of place, time, or people (e.g., “exercise with my partner once a week”). Specific goals had to make reference to when, where, or with whom someone was going to do the activity to be considered specific. General references to time (e.g., “regularly”), place (e.g., “outside”), or people (e.g., “others”) did not make a goal specific. This coding system reached acceptable levels of interrater reliability in a previous study (Cohen’s kappa = 0.97; Dickson & MacLeod, 2004). For the current study, we obtained an interrater reliability of .82, which is within an acceptable range.

Once coded, responses were converted into a scale of goal specificity, where general goals were coded as one point, moderate goals as two points, and specific goals as three points. Goals that were not consistently coded between coders were resolved before scoring. Duplicate goals were removed before scoring. Scores were added up and averaged by the number of goals generated. Separate scores were also calculated for approach and avoidance goals. These procedures have been successfully implemented in multiple studies on depression (e.g., Belcher & Kangas, 2014; Dickson & MacLeod, 2004).

Emotional Stroop Task. This task is a common measure of response inhibition found in the literature (MacLeod, 1991; Williams et al., 1996). Participants were presented with twenty-two depressive (e.g., “despair”) and twenty-two neutral (“arrange”) words printed in blue (#1E90FF), green (#77E30E), yellow (#FFFF00), and red (#F91100) colors in front of a black background (#000000). Words were chosen from a list of words matched for lexical word frequency, pronounceability, and word length.

(Mitterschiffthaler, et al., 2008). Participants were asked to ignore the meaning of the words and name the colors that the words were printed as fast as they could. Participants were asked to make responses on the keyboard: “B” for blue print, “G” for green print, “Y” for yellow print, and “R” for red print. Participants had eight trials to practice this keyboard placement (e.g., %%%% printed in #77E30E). A reminder was placed on the bottom of he screen that read, “Respond to the color of the print.” After this practice, there were a total of 176 trials. Words were presented in a pseudo-randomized block design, which produces larger interference effects than a purely randomized design (Epp et al., 2012). The first block of 16 trials were for practice, such that participants were given a reminder of the instructions (“Ignore the meaning of the word and respond to the color of the print.”) and feedback on their performance (e.g., “Correct”). This left 10 blocks of trials used to calculate the interference score.

A trial entailed the onset of a word stimulus and the moment a participant made a response or if the 2000m.s. display time had elapsed. Before each word was presented, a white crosshair (#FFFFFF) was presented at the center of the screen for 750ms. Responses were made within the 2000ms word presentation. This timeframe was chosen because response inhibition in depression generally occurs in later time intervals (e.g., > 1000ms; Gibb, McGeary, & Beevers, 2016). A JavaScript function was developed to collect reaction times on participants’ browsers and was triggered by keystrokes. Reaction times were collected right when a participant made a response, which reflected the duration between stimulus onset and participant response. Reaction times were not collected after 2000ms. Trials without a response during this time were marked as incorrect. Responses that did not match the correct printed color were also marked as incorrect. After removing incorrect trials, we calculated interference scores by taking the average reaction time for depressed words and subtracting it by the average reaction time for neutral words. Higher scores indicate a longer reaction time to depressed words compared to neutral words and hence more attentional capture and lower response inhibition. Our mean reaction time (RT) for depressive words was 830.6ms (*SD* = 192.6ms), and 830.84ms for neutral words (*SD* = 189.83ms). Our average depressive word RT was slightly higher than previous research and our average neutral word RT was quite a bit higher than previous research among clinically validated MDD patients (e.g., depressed RT = ~775ms, neutral RT = ~675ms; Mitterschiffthaler, et al., 2008). Our mean interference score was -0.24 (*SD* = 44.94), which is considerable lower than expected. See Table 6 for descriptive statistics of incorrect data. See discuss section below for explanations of why we might have observed such low scores and implications for interpreting our results.

Monetary Choice Task. This task is a common measure of delay discounting found in the literature (Madden & Bickel, 2010). In this task (Du, Green, & Myerson, 2002), participants were presented with a series of hypothetical choices between two monetary values with different delays. One value was paid out immediately and this amount varied from trial to trial based on the participant's choice. We set this initial value at $100 for every block of six trials. The other amount of money remained fixed at $200 and its delay also remained fixed for a single block. This value was chosen as prior research has suggested that those with depression discount long-term rewards at around this value (Pulca et al., 2014). The delay time increased as blocks of six trials increased (1 month, 3 months, 9 months, 2 years, 5 years, 10 years, and 20 years). We used an adjustment procedure as this has been successfully validated in other studies (Koffarnus & Bickel, 2014; Prencipe et al., 2011). Within each block of six trials, if a participant selected the immediate value, this value decreased by 50%. If a participant selected the delayed value, the immediate value increased by 150%. Participants were told to please make decisions as if the choices were real. These procedures have been used to identify the relationship between depression and discounting (Moody, Franck, & Bickel, 2016).

The area under the curve was chosen over the rate of the discounting curve (i.e., K) because it tends to be normally distributed (Myerson, Green, & Warusawitharana, 2001). First, indifference points (i.e., subjective value of delay amounts) were calculated for each block by taking the difference in immediate values between the last immediate reward chosen and the last immediate reward rejected (Du et al., 2002). With indifference points on the y-axis and delayed times on the x-axis, the area under the curve was used as an index of delay discounting. Areas under the curve were calculated by converting indifference points and delayed times into proportions. Each delayed time was expressed as a proportion of the maximum delayed time (e.g., 1 month / 240 months), while each indifference point was expressed as a proportion of the nominal amount (e.g., $75 / $200). These values were then used to construct the area under the curve. The area under the curve equals the sum of equation 1 for each paired amount (e.g., 1 month and 3 months; 3 months and 9 months) where x_1 and x_2 are successive delay amounts and y_1 and y_2 are indifference points associated with those delays (Odum, 2011). Values range between 0 and 1, where 1 equals no discounting. We obtained a mean AUC value of .25 (*SD* = .35), which is consistent with prior research (Imhoff et al., 2104).

$$(x_2 - x_1)\,[(y_1 + y_2) / 2] \qquad (1)$$

Intervention

The self-help program included five modules of CBT for depression delivered through an interactive web application. CBT procedures were adapted from Young and colleagues (2014) chapter on *Cognitive Therapy for Depression* in the fifth edition of *Clinical Handbook of Psychological Disorders: A Step-By-Step Treatment Manual* (Barlow, 2014). Although most patients experience significant depression relief by the twelfth session of face-to-face CBT (Hayes, Laurenceau, Feldman, Strauss & Cardaciotto, 2007; Lutz, Stulz, & Köck, 2009), self-help is a little different. Maximum improvement typically occurs by the 4^{th} or 6^{th} session (Delgadillo et al., 2014; Donker et al., 2013). It is recommended that those that do not improve by this point should be “stepped-up” to a more intensive intervention (e.g., face-to-face therapy; Delgadillo et al., 2014). We chose to include five sessions of CBT as this is within the recommended length of self-help and nicely lines up measurement time-points in our study design (see study design section below).

We designed the self-help program to be completed within a month. Since the program content was cumulative, participants were required to complete a session before the next one was available. Subsequent sessions were available five days after the prior one had been completed to allow participants time to do offline homework assignments. We encouraged participants to complete each session within a week of it becoming available. However, participants could take up to a month to complete a session before they were considered a non-completer. Being allowed to complete sessions at their leisure resembles how self-help is generally used in the public.

Self-Help Sessions

Sessions included unique content that built on top of prior sessions. Each session, with the exception of the first and last, included four components: (1) review homework from prior week, (2) cover additional educational material, (3) practice exercises that highlight the key principles discussed, and (4) assign subsequent homework. All homework assignments were downloadable interactive .pdf files that participants could use to complete exercises offline.

Session 1. The first session focused on psychoeducation, goal setting, and activity scheduling. Topics for this session included: (1) the nature of depression, (2) how depression becomes MDD, (3) principles of CBT, (4) research on the effectiveness of CBT, and (5) expectations moving forward (e.g., completing assignments outside of each online). Participants were then introduced to the importance of goals (e.g., increase accomplishment and enjoyment). A list of the goals they created during the Goal Specificity Task was presented and participants were asked to choose which goals were most related to alleviating their depression. If no goal was related to their depression, they were asked to generate one (e.g., "get out of the house more"). Their most depression-related goals were selected, and participants were coached on how to make these goals more specific (i.e., specific targets with at least one specification of place, time, or people; e.g., "exercise with my partner once a week"). Participants chose one activity that would feasibly get them closer to one of these goals and asked to monitor their feelings of accomplishment and pleasure before, during, and after this activity for homework.

Session 2. The second session focused on processing the outcomes of activity scheduling, barriers to activity scheduling, and the role of self-talk in depression. Participants were asked to review their homework assignment and pay particular attention to how the scheduled activity made them feel accomplishment and enjoyment. Barriers to activity scheduling were discussed (e.g., tasks that are not feasible given time, finances, and motivation). Another feasible task that was related to their refined goal (see session 1 section above) was planned for the following week. After this, participants were reminded of how some “self-talk” (i.e., automatic thoughts) can impact depression. They were asked to think about the last week and times in which they felt the most depressed. They were asked to type what was going on during these times and type the thoughts they had during these situations. If they had a hard time remembering, this was validated through psychoeducation about memory and depression. A thought monitoring exercise was assigned for the following week. Participants were instructed on how to overcome barriers to thought monitoring (e.g., making note of thoughts as soon as possible).

Session 3. The third session focused again on reactions to activity scheduling and patterns of unhelpful thinking. A review of their scheduled activity was conducted, with emphasis on how they felt before and after the task. Another activity was scheduled for homework. A shift to patterns of thinking was made. They were shown a list of unhelpful patterns of thinking commonly associated with depression (e.g., black and white thinking). They were asked to select the common patterns that they noticed in the past week. If they did not do this exercise for homework, they were asked to complete it online. Common ways of correcting unproductive thinking were introduced (e.g., thinking in shades of gray). They were asked to practice these alternative and more productive ways of thinking about the situations that they noted from the past week (e.g., “What kind of thinking would capture the accuracy of the situation?”). Participants were then assigned a homework assignment, in which they were to monitor their unhelpful thinking and practice productive thinking throughout the week.

Session 4. The fourth session focused on efforts to reappraise automatic and unproductive thoughts and discuss how these thoughts can reflect core beliefs. Before reflecting on thoughts, participants were asked to note how they felt before and after they completed their scheduled activity. Another activity was scheduled. Participants were asked to reflect on the unproductive thoughts that came up for them in the past week along with their efforts to reappraise. The cognitive triad was discussed (e.g., “I’m a failure”). They were asked to think about a moment when they felt depressed and do a sentence completion task to uncover core beliefs (e.g., “I am ______”). They were asked to think about how the alternative thoughts they generated in the last week impact their beliefs about themselves, others, and their future. For instance, participants that believed, “I’m a failure” were instructed to note the evidence in the past week that was in opposition to this. Participants were guided to formulate more nuanced beliefs such as, “I don’t always achieve the things I want to, but sometimes I do.” The homework for this session was to practice productive thinking again but notice how these alternative thoughts impact their core beliefs.

Session 5. The last session focused on relapse prevention. Participants reviewed the prior week’s homework. They first reviewed how they felt after completing their scheduled activity and then how their productive thinking impacts their self-concept. They then reviewed the three major skills they learned (i.e., goal setting, activity scheduling, cognitive reappraisal) and how practicing these skills have impacted their life. Psychoeducation around relapse prevention was given (e.g., practicing skills, having a plan to practice). Their identified goals related to alleviating depression were presented and specific tasks that they can continue to do were typed into the application. They were then asked to check the cognitive reappraisal strategies that they want to practice moving forward. Finally, they were asked to note any other strategies that they could think of that might increase practice of skills and decrease relapse. A relapse prevention document was generated from their responses and downloadable for their record.

Study Conditions

Guided Condition. Half of our participants were given access to an additional guided feature of the self-help program (Berger, Hämmerli, et al., 2011). These participants were told that they could contact study personnel at any time via e-mail to help them throughout the self-help program (e.g., understanding content, resolving barriers to access). Study personnel responded to these emails as soon as possible and no later than 24-hours after they received a message. Participants in this condition received a number of emails asking if they needed help, such as if they completed a session, did not do a homework assignment, did not complete a session after seven and then fourteen days after the session became available. Unlike Berger and colleagues (2011), we did not provided feedback on symptom improvements in our initial emails to guided participants. Feedback has a small effect on symptoms in the literature (Knaup, Koesters, Schoefer, Becker, & Puschner, 2009) and may account for some of the benefits of guided self-help. As a result, we moved all feedback on symptoms to the self-help application, which all participants could access. Here is an initial email sent to guided participants after they completed a session:

> "You completed session <insert number>! Your next session will be available on <insert date>. We will send you another email when your next session is available. Until then, is there any part of the program that you would like to discuss, perhaps a difficult section you would like help with? Our team of researchers is available to reach out to you within the next 24 hours if you need. As a reminder, your login ID is <insert ID>."

Unguided Condition. The other half of our participants did not receive the guided component of self-help (e.g., encourage engagement, problem solve problems). However, given that contact with study personnel and the additional treatment dose that comes with contact itself could account for the incremental benefits of guided self-help (Berger, Hämmerli, et al., 2011; Richards & Richardson, 2012; Schippers, Adam, Smolenski, Wong, & de Wit, 2017), we added messaging to increase equivalence between our experimental groups and control for contact and dose. For instance, participants in the unguided group were told that to help them with understanding the content of the sessions or how to do the exercises, there were helpful tips located within each session and in the "Skills" tab of the intervention. They were emailed during similar times as the guided group (e.g., completed a session, did not complete a homework assignment). For instance:

> "It looks like you had a hard time doing one of the offline activities. This program has some helpful tips for how to do the offline activities. Do you have a moment

to go to dashstudy.org/study.html and look under the “Skills” tab for more information? As a reminder, your login ID is <insert ID>.”

Waitlist Condition. A group of participants were randomized to a waitlist control condition. In this condition, participants were asked to fill out the PHQ-9 every other week while they wait for a month. Week three assessment was only available for nine days so that week three was not conflated with week five. They received reminder emails to complete this assessment, one at the time the assessment was available and after a week if they did not complete the assessment.

> You only have two more days to complete an assessment of your depression. We do this to monitor how you are feeling throughout the study. Each assessment you complete will give you a chance to win one of three $50.00 gift cards, which we will draw at the end of the study. This assessment will only take a few minutes of your time. Do you have a moment to go to dashstudy.org/study.html and login with your login ID to complete this assessment? Your login ID is <insert ID>.

At the fifth week, they were randomized to either treatment group and informed that they could start the self-help program. See design section below for more information. Depression data from week one to week five were used to check the efficacy of our intervention. Data during the active treatment phase (i.e., week five to week nine) were used to test our remaining hypotheses.

Additional Messages. All participants were emailed a copy of their consent form and welcomed into the study upon enrollment. If participants did not do session one within a day of enrolling in the study, a reminder email was sent to them indicating that the self-help program was ready for them. This email also re-iterated the incentives to increase engagement in the study. When participants completed the self-help program, they were thanked for participating. If they were in the immediate-start group, they were informed that they would be asked to complete a couple follow-up depression assessments, week seven and week nine. These participants received additional reminders to complete week seven and week nine depression assessments. All emails were followed by a text message sent to participants' phones directing them to check their e-mail and look for the aforementioned messages.

Self-Help Application

The application was structured to maximize efficiency and effectiveness. For instance, there were three higher-order tabs for participants to navigate: "Dashboard", "Sessions", and "Skills". The "Dashboard" tab included basic information on participant engagement (i.e., sessions completed, when subsequent sessions would become available) and participant progress (i.e., depressive symptoms over time). The "Sessions" tab had a left-floated list of all completed sessions with the current session that participants had not completed at the bottom open and ready to complete. Given that each session built off of the prior session and that sufficient time was needed to complete offline homework assignments, new sessions were available five days after the previous session had been completed. New sessions were completed in the right-portion of the Sessions tab. Each of the four session components was traversed one at a time using "Next" buttons. Participants were able to go back to prior components if needed. For instance, the first component, review homework, included inputs for participants to react to their homework assignments (e.g., note patterns of thinking they engage in) and then a "Next" button for when they were ready to move on. Once a session was completed, content participants entered during the session became static content. The “Skills” tab outlined the skills that participants develop throughout the self-help program. For instance, once they finish the first session, an outline of goal setting and activity scheduling and the accompanied worksheet were provided. This tab allowed participants to quickly review the skills that they had learned without having to navigate to the appropriate sections of the appropriate sessions.

Technology

The consent form, measurements, and CBT intervention were programmed with R (R Core Team, 2019) version 3.6.2 utilizing the shiny package (Chang, Cheng, Allaire, Xie, & McPherson, 2020) version 1.5.0. The shiny package is a web application framework for R that integrates HTML, CSS, and JavaScript functions. JavaScript (JS) has millisecond timing capabilities, which makes it desirable for collecting data from tests dependent on collecting accurate reaction times (Crump, McDonnell, & Gureckis, 2013). The current study did not utilize Flash Player (FP) because FP became unsupported in major Internet browsers by the end of 2020 and JS has shown to produce similar results in reaction time tests as more recent versions of Flash Player (Reimers & Steward, 2015). JS also integrates seamlessly with R shiny.

Reaction timed tests administered online at home come with systematic error in measured reaction times across trials; however, we do not believe this was a problem for the current study. Reimer and Steward (2015) compared the difference between apparent and actual reactions times using a basic reaction time test programmed in JS and FP and compared this difference across Internet browsers and computer systems. They noted that the average difference between apparent and actual reaction times was around 30ms. This variability was not influenced by browser type but was influenced by computer system. The within-system variability was low (*SD* = 10ms). Although two studies have found different reaction times between the Stroop task when administered on a desktop in the lab, online in the lab, and online at home utilizing FP and JS, this difference was consistent for within-subjects conditions (e.g., congruent, incongruent trials) and there was no difference in interference scores (Crump et al., 2013; Linnman, Carlbring, Åhman, Andersson, & Andersson, 2006). These studies suggest that there is systematic error in measured reaction times that is generally distributed across trials and hence within-subject conditions. Our within-subjects randomization procedure in our emotional Stroop task should have balanced any error across within-subjects conditions. Since we had a neutral condition for our Emotional Stroop Task, we effectively accounted for this error in our interference score.

Data was securely collected and stored. The web application was hosted on a remote server rented from Digital Ocean (DO). DO operates with the highest standards for data security and have ISO/IEC 27001:2013 and EU-U.S. and Swiss-U.S. Privacy Shield certifications. The application communicated with participant interactions behind a Secure Sockets Layer (SSL), which is a cryptographic protocol designed to provide a secure communication channel between computers. Once data were collected, they were stored in a MySQL database in a remote server managed by the main author of this study. Access to this server required a Secure Shell (SSH) connection, which is another cryptographic network protocol. SSL and SSH ensured that data transfer was done in a secure manner at all times. According to DO's privacy policy and privacy notice documentation, DO personnel do not access data from rented servers unless under the strict permission of the renter. No identifiable information (e.g., email addresses) was stored on this server thereby creating a firewall between identifiable data and data collected for the purpose of this study. These procedures minimized the risk and harm of a data breach.

Study Design

This study utilized a three-arm, randomized, delayed-start, longitudinal design (D'Agostino, 2009). In this design, two thirds of our sample was immediately randomized to the guided or unguided self-help condition, while the other one third started the self-help program at the fifth week. This design allowed us to determine the efficacy of our two treatment delivery methods against a waitlist control by analyzing depression symptoms within the first five weeks of the study. It also maximized our power to test more complex analyses reliant on larger sample sizes by ensuring that all participants went through active treatment. At the fifth week, those in the waitlist control were randomized to one experimental group. Those in the immediate-start group were followed-up until the ninth week of the study so that everyone was in the study for the same amount of time. Active treatment time points were obtained on the first, third, and fifth weeks of the program for the immediate-start group and weeks five, seven, and nine for the delayed-start group (i.e., waitlist group). See Table 1 for a depiction of this design. This design has been successfully implemented in prior treatment studies (e.g., Papa, Sewell, Garrison-Diehn, & Rummel, 2013). Since our time points were tied to particular sessions and we anticipate a fair amount of dropout, we decided to consider someone dropped out if they did not complete a session within 28 days of it being available to him or her. At that time, participants were asked to complete a final measure of depression. Any PHQ-9 measurement after that time was used for the post program time point.

Some have suggested that this design increases a studies Type-2 error rate when the intervention during the delayed period has a significant effect on outcomes, thereby reducing the overall treatment effect (Spineli, Jenz, Großhennig, & Koch, 2017). We do not believe that this was an issue for the current study. In a meta-analysis of RCTs for depression, authors observed a significant medium effect size of waitlist groups (Rutherford, Mori, Sneed, Pimontel, & Roose, 2012). However, the studies in this meta-analysis that observed significant effects had waitlist duration at or above 8 weeks, which is twice as long as the current study. Furthermore, the expected four-week remission rate for adults in waitlist control groups is around 10% (Whiteford et al., 2012). Given that we expected to randomize 45 people to our waitlist control group (see data plan section below), we expected that no more than five participants would remit before starting treatment (i.e., PHQ-9 < 10). Those who did remit before starting the treatment phase were excluded from relevant analyses so as not to reduce our treatment effect (i.e., Spineli et al., 2017) and preserve the external validity of our study (e.g., generalize to patients that screen positively for MDD). This suggests that our four-week, delayed-start phase could adequately serve as our waitlist control without compromising our ability to adequately reject our null hypotheses, all while also increasing our power to detect an effect.

Procedures

Study procedures were conducted online. This provides generalizability for future implementations (i.e., stratified care delivered online). Once recruited, prospective participants were directed to an official study website, which detailed the purpose of the study, eligibility requirements, and expectations if enrolled. MTurk participants were provided this information on the main enrollment HIT. Participants were told that the purpose of the study was to understand how an online self-help program might be used to reduce depression. All eligibility requirements were provided to perspective participants. Participants were told that the study would last two months and that during this time they would: (1) complete several pre-intervention measures (e.g., monetary choice task), which would take around 20 minutes to complete, (2) either start the program right away or start the program after a one month delay, (3) engage in a four-week online self-help program and spend around 30-60 minutes doing cognitive-behavioral exercises online and offline, (4) contact a study therapist if you need additional help, depending on which group you are assigned, and (5) complete a measure of depression every other week for the duration of the study.

An area on the website and enrollment HIT directed participants to the study consent form. They were then asked to read the informed consent form. During this time, they were told that this intervention was not intended to replace psychotherapy or medical treatment for depression, they could contact study personnel via e-mail if they have any questions before committing to the study, and they could withdraw their participation at any time. Once participants consented, they were asked questions to confirm their eligibility (see participants and measures section above). Those that were eligible were allowed to proceed. Those that were not eligible were given a link to Psychology Today so they could search for a therapist in their local area, a link to MoodGym, and information on the National Suicide Prevention Lifeline (see above).

Participants that consented into the study and were eligible were asked to enter their email address and phone number so that the application could send them a unique login ID. MTurk participants were asked to enter their worker ID as use it for their login ID. IP addresses were automatically collected to identify people that attempted to enroll in the study multiple times. For non-MTurk participants, the login ID was generated on the remote server and consisted of alternating numbers and letters (e.g., a1e3r5). Participant login IDs, email addresses, phone numbers, and IP addresses were sent to a designated email address created specifically for this study. Study personnel manually enter this information in a secure database within UNR's NevadaBox service, which is a secure cloud-based data-storage service provided to UNR personnel. Only members of this study had access to this database. Identifiable information was not stored with study data (e.g., PHQ-9), but their unique login ID was. This effectively created a link between identifiable data and study-specific data, but also created a firewall between these two types of data. This information allowed study personnel to check in with participants if they were not adhering to study procedures (e.g., troubleshoot problems with homework assignments) and provide participants with their login ID if they lost it. Participants used their unique login ID to access study content throughout the course of the study.

Participants were then randomized to either the guided, unguided, or waitlist groups using a predefined random generation of 1s, 2s, and 3s, where "1" corresponded to the guided group, "2" corresponded to the unguided group, and "3" corresponded to the waitlist group. This ensured that each group would have an equal number of participants for our manipulation check. Given that we only needed our waitlist group to check if our program was effective, we stopped randomizing participants into that group once we obtained enough data (i.e., 45 participants per group). Those in the immediate-start group were asked to complete the remaining pre program measures (e.g., Emotional Stroop Task). At the fifth week of the study, those in the waitlist control group were randomized to one of the two experimental groups and then asked to complete the remaining pre program measures. Before the behavioral tasks (e.g., Goal specificity task), all participants were reminded that they would need to be on a desktop or laptop to proceed. Behavioral tasks were completed in full screen mode within their web browser so that they would ample space to interact. When they were done with these measures, they were directed to the self-help portion of the application.

Before sessions one, three and five, all participants were asked to complete the PHQ-9 to monitor depression throughout the program. All participants were told that during active treatment, they were to go online once a week to start each session. They were asked to try to complete each session within seven days of it being available to him or her. Subsequent sessions were available five days after the previous one had been completed to allow for ample time to complete offline homework assignments. If a participant did not complete a session within 28 days of it being available to them, they were considered dropped out and asked to complete a final PHQ-9. At the end of the ninth weeks, participants were given a link to Psychology Today so they could search for a therapist in their local area, a link to MoodGym if they would like additional self-help, and information for the National Suicide Prevention Lifeline.

Data Plan

Given that we were studying how people respond to a self-help program, only those that started the program were included in our primary analyses. Multilevel linear modeling (MLM) and generalized linear modeling were used to statistically test our hypotheses. Before we analyzed our primary hypotheses, we tested the efficacy of our two self-help conditions in comparison to our waitlist control condition using MLM. Our outcome variable was depression severity with three time points. Our time variable was numeric and consisted of 0's, 1's and 2's corresponding to the three time points during active treatment. Group assignment was dummy coded where the waitlist control was our reference group. Interactions terms between time and group terms were created. Interactions are best analyzed using a nested model comparison approach (Cohen et al., 2003). In this approach, primary variables (i.e., main effects) are entered in the first block, while two-way interaction terms are added in the second block and the blocks are compared. This nested model comparison approach allowed us to see how much more variance in our outcome would be accounted for by our interaction effect above and beyond our main effects. We used a likelihood ratio test (a.k.a., chi-square difference test) and two fit indices to compare nested models, Akaike information criterion (AIC) and Bayesian information criterion (BIC). Given that people vary in their baseline depression scores and rate of depression over time (Lutz et al., 2009), we specified two random effects, a random intercept and random time effect. For this model and all other multilevel models, we assumed a first-order autoregressive covariance structure because this structure best fits data in which the covariance between time-points decreases as the distance between time-points increases (e.g., 1 to 2, 1 to 3; Mansour, Nordheim, & Rutledge, 1985) and this type of longitudinal pattern is observed in depression scores measured at multiple time points, including the PHQ-9 (Hedeker, Gibbons, & Waternaux, 1999; Patten & Schopflocher, 2009). These procedures allowed us to check if our experimental groups had a different treatment response than our waitlist control group.

We used these same procedures to see if our experimental groups were significantly different in their rate of depression during the active treatment and follow-up phases of the study. However, due to sparse follow-up data, we only included completers and applied a Kenward-Roger correction for this particular analysis (Kenward & Roger, 2009). Small sample sizes come with bias in Level 1 standard errors and are difficult to approximate the degrees of freedom for the denominators of hypothesis tests, thereby inflating a tests Type-1 error (Kenward & Roger, 2009). The Kenward-Rogers correction has shown to produce non-biased results with Level 2 cluster sizes of 14 (McNeish & Stapleton, 2016). The lowest cell of data obtained for the current study was 14, which was for week 9 follow-up data among unguided participants. This suggests that a Kenward-Rogers correction could be adequately used for our follow-up analysis.

Although the measurement error for our PHQ-9 measure was within an acceptable range (omega = .79), some have noted an increase in false conclusions due to measurement error at these levels (Cole & Preacher, 2014; Flora, 2020). As result, we tested our manipulation check using multilevel structural equation modeling (SEM), which allowed us to model a latent factor of depression and account for measurement error. Model fit indices were used to test how well our model fit the data. Comparative fit index (CFI) and Tucker-Lewis index (TLI) values of .95 or greater are considered adequate fits of the data (Hu & Bentler, 1999). Root mean square error of approximation (RMSEA) values less than or equal to .05 are indicative of good model fit, while values above .08 are indicative of poor model fit (Finch & French, 2015). Finally, standardized root mean square residual (SRMR) values less than or equal to .08 are indicative of food model fit (Finch & French, 2015). Multilevel SEM was only meant to see if our results would be consistent across the two modeling techniques. We did not analyze all of our multilevel models with multilevel SEM because this modeling technique is not fully implemented in R (e.g., lacking diagnostic tests) and our sample size was too low to adequately test our models (Kline, 2012).

Our main aim was to examine if variables derived from theories of self-regulation moderate the effect of guided and unguided self-help on depression. This created a three-way interaction between time (Level 1), group (Level 2), and process (Level 2). Process variables (e.g., response inhibition) were grand mean centered before creating interaction terms to minimize multicollinearity that can accrue with interaction terms (Cohen et al., 2003). Main effects were specified in the first block, two-way interactions were specified in a second block, and three-way interactions were specified in a third block. As before, a random intercept and random time effect were also specified. The two-way interaction model allowed us to compare the effectiveness of our guided and unguided groups while controlling for process, while the three-way interaction model allowed us to test our first primary hypothesis. Separate analyses were performed for each process variable. If we included all four processes variables and respective interaction terms, this would likely have led to overfitting in our regression model, which makes results hard to interpret (Babyak, 2004). This led to four separate nested model comparisons.

The MLM regression equations included, with Y_{ij} being depression score at a particular time (i) for a given participant (j):

$$\text{Level 1: } Y_{ij} = \beta_{0j} + \beta_{1j}\text{time}_{ij} + e_{ij} \tag{2}$$

$$\text{Level 2: } \beta_{0j} = \Upsilon_{00} + \Upsilon_{01}\text{group}_j + \Upsilon_{02}\text{process}_j + \Upsilon_{03}\text{groupXprocess}_j + \upsilon_{oj} \tag{3}$$

$$\beta_{1j} = \Upsilon_{10} + \Upsilon_{11}\text{group}_j + \Upsilon_{12}\text{process}_j + \Upsilon_{13}\text{groupXprocess}_j + \upsilon_{1j} \tag{4}$$

We checked all assumptions of our multilevel models (e.g., normality of level-1 and level-2 residuals; Luke, 2004). Although we anticipated an unbalanced dataset, MLM with maximum likelihood estimation is robust up to 50% missing data, provided that missing data are missing at random (MAR; Black, Harel, & McCoach, 2011). MAR means that the missing observations are conditional on observable variables in the dataset (e.g., group assignment). Although we anticipated more missing data towards the end of treatment and in the unguided self-help condition, which would have violated the more strict assumption of missing completely at random, we modeled time and group assignment, which preserved the MAR assumption of maximum likelihood estimation. We also attempted to make up for any missing data by soliciting a post measure of depression for those that had dropped out.

We could not confirm the diagnostic status of our participants because our study was conducted online and we used a diagnostic screener. This may have lead to participants misrepresenting themselves and/or false positives for a depression diagnosis. Since outlying cases are more likely to not come from the population to which researchers are trying to generalize (Cohen et al., 2003), population of MDD in our case, we examined and removed outlying cases. For our multilevel models, outliers were checked and removed according to procedures outlined in Langford and Lewis (1998) and Chen (2008). Overall, these procedures allowed us to test if our groups and moderators influenced depression over time.

For our remaining hypotheses, we performed a series of generalized linear modeling (e.g., odds of completing program). Analysis of time was irrelevant for these analyses so MLM was not performed. For our hypothesized models predicting the odds of completing the self-help program, we applied a logit link function to account for the binary nature of this outcome variable. We also applied a logit link function for our preliminary analysis predicting the odds of starting the program. This preliminary analysis helped us understand how the data used in our primary analyses may have been skewed towards certain study characteristics due to participants not starting their allocated intervention. Areas under the curve (AUC) and Akaike Information Criterion (AIC) were reported as estimates of model fit. AUC is an indicator of how well a model classifies cases in an outcome, leading to intuitive interpretations (e.g., AUC of .5 suggests the model is as good as a coin toss). Pseudo R^2 values were not reported, as these are generally hard to interpret and the nine pseudo R^2 values can wildly differ (e.g., .04 to .63; Walker & Smith, 2016).

For our other hypothesized models (i.e., sessions completed, homework assignments completed, emails sent), we applied a logarithm link function to account for the Poisson distributions observed in our data. A nested model comparison approach was also employed to test the interactions between group assignment and self-regulation. For the same reasons as above, we performed four separate nested model comparisons. All assumptions of generalized linear modeling were checked (e.g., Vuong test to check equal dispersion for number of sessions completed). We conducted zero-inflated Poisson regressions for outcomes that were zero-stacked. We also observed all dispersion parameters and conducted negative-binomial regressions when the assumption of equal-dispersion was not met. We considered removing outlying cases using recommendations from Cohen and colleagues (2003) and Tabachnick and Fidell (2007). These procedures allowed us to test the influence of our self-regulatory variables on our other study outcomes.

These analyses required a sample size of 180 to sufficiently power our study. We powered the study for the hypothesis that required the largest sample size. The auxiliary hypothesis that our program was effective is a cross-level interaction between time and group in a MLM framework. However, given that our first primary hypothesis tested moderators of this effect, more participants were needed. Since sample size requirements for MLM are notoriously difficult to calculate (Snijders, 2001), we sought suggestions derived from simulation studies. In estimating sample size, we considered the perspective effect size, number of time points, covariance structure, inclusion of random effects, and occurrence of missing data (Hedeker et al., 1999). We powered the study to detect a large effect size ($d = 0.8$) because larger effects generally have more clinical significance (Kraemer & Kupfer, 2006). As noted earlier, the number of time points that were included in our MLM analyses was three. Also noted earlier, a first-order autoregressive covariance structure was specified (Hedeker et al., 1999; Mansour et al., 1985; Patten & Schopflocher, 2009) and we assumed a random intercept and random effect of time (Lutz et al., 2009). Given these parameters, and assuming a two-tailed significance criterion of .05 and power to detect an effect of 0.8, a sample size of 45 per group appeared reasonable to test our group by time, two-way interaction models accounting for missing data (Hedeker et al., 1999). However, given that our third hypothesis was to test moderation of guided and unguided self-help, and given that moderation typically needs twice as many participants to be adequately powered (Cohen et al., 2003), 180 participants appeared adequate to test our hypotheses.

All statistical analyses were conducted using R (R Core Team, 2019) version 3.6.2. The nlme package (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2019) version 3.1-142 was used to analyze our multilevel linear models. We chose this package because it allowed us to specify an AR(1) autoregressive structure. We used the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) version 1.1-21 to check outliers of our MLMs. The pbkrtest package (Halekoh, & Højsgaard, 2012) version 0.5.1 was used to apply the Kenward-Roger correction for our follow-up data. The lavaan package (Rossell, 2012) version 0.6-8 was used to analyze our structural equation models. The stats package (R Core Team, 2019) version 3.6.2 was used to model our logistic and Poisson regression models. Finally, we used the pscl package (Zeileis, Kleiber, & Jackman, 2008) version 1.5.5 for our zero-inflated Poisson models and zero-inflated negative binomial models.

Results

Enrollment Numbers

928 people across the United States were recruited from a variety of sources. Participants were recruited from MTurk (54.5%), Craigslist (31.8%), social media (6.9%), UNR (2.8%), and the Reno/Tahoe community (1.8%). 2.2% of people were recruited from unknown sources due to not initially collecting these data. Only 391

(42.1%) people were eligible. After removing unknown sources and collapsing UNR, social media, and community into an "other" category, the relationship between recruitment source and odds of being eligible was trending significant, $\chi^2(2, 907) = 5.87$, $p = .05$. Those from Craigslist had a 41% decrease in the odds of being eligible compared to the "other" group, $p < .05$, 95% CI [.38, .91], while those from MTurk had a 37% decrease in the odds of being eligible compared to the "other" group, $p < .05$, 95% CI [.42, .96]. People were ineligible for a variety of reasons, and often multiple reasons. Four people were under the age of 18; one was not fluent in English; six did not have a computer; 218 had a PHQ-9 score below 10; 37 endorsed active suicidal ideation; 105 were already in therapy for depression; 13 had changed their depression medication(s) within the last month; 105 were above cut off for mania in the past six months; 198 were above cut off for psychosis in the past six months; and five were excluded for unknown reasons due to not initially collecting these data. Our large values for mania and psychosis may be due to these conditions being particularly difficult to report on in self-report formats.

Of the 391 that were eligible, 336 enrolled in the study. 121 were assigned to the guided group, 116 to the unguided group, and 99 to the waitlist group. The mean age of our sample was 34.15 years old (*SD* = 10.9). The sample was predominately women (56.8%), with 23% men, 1.8% transgender, 0.6% other, and 17.9% without data due to not completing the pre program measures. The sample was also predominately White (59%), with 5.1% Asian American, 5.1% Black or African American, 3.6% Hispanic/Latinx, 0.6% American Indian, 8% mixed ethnicity, 1.2% other, and 17.4% without data due to not completing the pre program measures. The sample consisted of

1.2% without a High School degree, 19.3% with a High School or GED degree, 16.1% with a trade school certificate or Associate's degree, 36% with a Bachelor's degree, 8.9% with a Master's degree, 0.9% with a Doctoral degree, and 17.6% without data due to not completing the pre program measures. For all analyses, factor levels were collapsed so that cells had at least 10% data (e.g., ethnicity included two levels, white and other).

Only those that started their allocated intervention were including in our primary analyses because our study primarily focused on moderators of our treatment effect. 184 participants completed our pre program measures and started their allocated intervention. This included 77 guided participants, 62 unguided participants, and 45 waitlist participants. Of these, 96 completed the self-help program. This included 39 guided participants, 31 unguided participants, and 26 waitlist participants. Post PHQ-9 data were obtained from an additional 25 participants that dropped out before completing the program. Week seven and nine follow-up PHQ-9 data were obtained from some participants. See Figure 1 for a flowchart of participation and sample sizes for each group throughout the study.

Odds of Starting Program

Since a sizable portion of enrolled participants did not start their allocated intervention, we attempted to predict the odds of starting the program to elucidate if our primary results may have been skewed towards any study characteristics. First, depression comes with changes to goal-directed behavior (e.g., energy, motivation, etc.), which might be why we observed that as pre program depression increased, the odds of starting the program decreased by 5%, $p < .05$, 95% CI [.90, .99]. The model with pre program depression was an improvement over an intercept only model, $\chi^2(1, 334) = 4.84$,

$p < .05$, AUC = .57. This suggests that the sample used in our primary analyses was less depressed than could have otherwise been.

Several socio-demographic variables predicted the odds of starting the program. Among those that enrolled in our study, recruitment source predicted the odds of starting the pre program measures, $\chi^2(2, 297) = 15.83$, $p < .001$. The area under the curve was .6. The Craigslist group had a 60% decrease in the odds of starting compared to the “other” group, $p < .05$, 95% CI [.17, .88], while the MTurk group had a 232% increase in the odds of starting compared to the Craigslist group, $p < .001$, 95% CI [1.83, 6.21]. No other comparison was significant. Group assignment predicted the odds of starting the pre program measures, $\chi^2(2, 333) = 7.43$, $p < .05$. The area under the curve (AUC) was .58. The guided group had a 110% increase in the odds of starting compared to the waitlist group, $p < .001$, 95% CI [1.22, 3.63]. No other comparison was significant. After retaining males and females, the relationship between gender and odds of starting was trending, $\chi^2(1, 267) = 3.46$, $p < .10$, AUC = .56, such that males were 40% less likely to start than females, $p < .10$, 95% CI [.35, 1.03]. Depression reactivity (“no” or “yes”) significantly predicted the odds of starting, $\chi^2(1, 243) = 4.18$, $p < .05$, AUC = .55, such that those with reactivity were 52% less likely to start the program than those without reactivity, $p = .05$, 95% CI [.21, .98]. However, this result is tenuous as only 10 out of 245 (4%) participants were not reactive and did not start the program, leaving open the possibility of separation in this model. These results suggest that our primary analyses may be skewed by recruitment source (i.e., more MTurk participants), group assignment (i.e., more guided participants), and to a lesser extent, gender (i.e., more female participants).

A few self-regulatory variables predicted the odds of starting the program. Removing two outlying cases, the relationship between autonomous motivation and odds of starting was significant, $\chi^2(1, 269) = 5.29$, $p < .05$, AUC = .58, such that for every one-unit increase in motivation there was a 3.2% decrease in the odds of starting, $p < .05$, 95% CI [.94, .995]. Counter to expectations that could be derived from theories of self-regulation, those with high motivation were less likely to start the program. Removing two outlying cases, the relationship between goal specificity and odds of starting was significant, $\chi^2(1, 180) = 4.85$, $p < .05$, AUC = .63, such that for every one-unit increase in specificity there was a 78% decrease in the odds of starting, $p < .05$, 95% CI [.06, .85]. Approach goal specificity was significant, $exp(b) = 0.37$, $p < .05$, 95% CI [0.14, 0.98], but avoidance goal specificity was not significant. This suggests that the effect of goal specificity in general may be weighted by the effect of approach goal specificity. The relationship between response inhibition and odds of starting was significant, $\chi^2(1, 219) = 4.47$, $p < .05$, AUC = .62. The odds of starting decreased by 1%, $p < .05$, 95% CI [.98, .999]. Given that higher scores actually represent lower response inhibition, we can take this to mean that those with lower response inhibition were more likely to not start the program. No other relationships were significant. These results suggest that our primary analyses may be skewed by autonomous motivation (i.e., lower motivated participants), goal specificity (i.e., lower specified participants), and response inhibition (i.e., higher inhibited participants).

Group Differences

Of those that started the program and were included in the primary analyses, 77 were in the guided group, 62 in the unguided group, and 45 in the waitlist group. Groups did not differ on any socio-demographic or self-regulatory variables with one caveat. Delay discounting did appear to differ between groups, $F(2, 181) = 3.5$, $p < .05$. However, the assumption of normality was not met (e.g., z-skew = 21.86) and equal variance was questionable (e.g., Levene = 2.88, $p < .10$). Applying a log transformation to our discounting variable improved our model assumptions (e.g., z-skew = 1.13; Levene = 0.76, $p > .10$) and our model was still significant, $F(2, 181) = 3.20$, $p < .05$. The R^2 value was .03, suggesting that three percent of the variance in discounting was accounted for by our groups. Both experimental groups were different from our waitlist group, with lower discounting scores observed in our waitlist group: waitlist vs. unguided group, $b = -0.48$, $t(181) = -2.27$, $p < .05$, 95% CI [-0.86, -0.06]; waitlist vs. guided group, $b = -0.46$, $t(181) = -2.26$, $p < .05$, 95% CI [-0.9, -0.06]. The experimental groups did not differ. It is likely that this did not impact our primary analyses. Collapsing our waitlist group into our experimental groups, we do not observe a group effect on discounting, $R^2 = .0007$, $F(1, 182) = 0.14$, $p > .10$, suggesting that the lower scores in our waitlist group were equally distributed across our experimental groups. In general, these results suggest that our randomization was effective and that our groups did not become biased after losing participants. See Table 2a and Table 2b for descriptive statistics of all socio-demographic and self-regulatory variables by group among those that started the program and included in our primary analyses.

Manipulation Checks

Comparing Experimental Groups to Waitlist Control

Two auxiliary hypotheses were tested to check the experimental manipulations. The first hypothesis was that our unguided and guided groups would show a greater decrease in depression over time compared to our waitlist group. Using our model building approach, we specified (1) an intercept only model to serve as a baseline model, (2) a random intercept model to examine how much variance was accounted for by our participants, (3) a random effects model, in which we added the random effect of time and the AR(1) autoregressive structure, (4) a main effect model, and (5) an interaction effect model, in which the interactions between time and groups were included.

Models met the assumptions of normality and equal variance for both levels. The time and two group terms significantly accounted for variance in our missing data, suggesting that the MAR assumption was also met. For the first step, we compared a random intercept model to an intercept only model. Adding the random intercept led to a better fit, $\chi^2(3) = 75.81$, $p < .001$, AIC = 2775.76, BIC = 2788.14. The interclass correlation was .45, suggesting that 45% of the variance in depression scores was accounted for by the between-subjects effect. Adding the random effect of time in the second step also led to a better fit, $\chi^2(6) = 55.17$, $p < .001$, AIC = 2726.59, BIC = 2751.35, indicating that participants varied in their rate of depression over time. The main effect model, in which the fixed effects of time and group were added was also a significant improvement, $\chi^2(9) = 92.35$, $p < .001$, AIC = 2640.23, BIC = 2677.38. Only the intercept and time effects were significant, $b = 16.38$, $t(273) = 26.13$, $p < .001$, 95% CI [15.15, 17.61] and $b = -2.38$, $t(273) = -11.02$, $p < .001$, 95% CI [-2.80, -1.96], respectively. The group terms were not significant. These results tell us that (1) participants started the study at an elevated level of depression, (2) the waitlist group significantly improved over time, and (3) groups had similar depression scores at the start of the study. Finally, adding the two interaction terms led to a better fit, $\chi^2(11) = 12.18$, p

< .001, AIC = 2632.06, BIC = 2677.45. Both interaction terms were significant. Depression decreased more over time in the unguided group than the waitlist group, b = -1.73, t(271) = -3.28, p < .01, 95% CI [-2.77, -0.70], and also more in the guided group than the waitlist group, b = -1.40, t(271) = -2.82, p < .01, 95% CI [-2.38, -0.43]. See Figure 2 for a visual depiction of these relationships. See Table 3 for the sample sizes, means, and standard deviations of these cells as well as F-tests for each time point.

As noted earlier, high measurement error for our PHQ-9 measure was observed. Multilevel SEM was used to see if the MLM results changed after specifying a latent variable of depression to account for the error in our PHQ-9 measure. After specifying all residual correlations above .10, the main effect model fit the data well, CFI = .98, TLI = .96, RMSEA = .03. The within-subjects covariance and between-subjects covariance matrices both had acceptable SRMR values, .04 and .05, respectively. For this model, time significantly predicted decreased depression, b = -0.86, z(99) = -10.6, p < .001, 95% CI [-1.02, -0.7]. The unguided group did not significantly differ from the waitlist group, b = -0.21, z(99) = -0.90, p > .10, 95% CI [-0.68, 0.25], nor did the guided group, b = -0.02, z(99) = -0.07, p > .10, 95% CI [-0.46, 0.43]. These results are consistent with the main effect results observed with MLM. After specifying all residual correlations above .10, the interaction effect model fit the data well, CFI = .98, TLI = .97, RMSEA = .03. The within-subjects covariance and between-subjects covariance matrices both had acceptable SRMR values, .04 and .06, respectively. As before, both interaction terms were significant. Depression decreased more over time in the unguided group than the waitlist group, b = -0.72, z(117) = -4.31, p < .001, 95% CI [-1.04, -0.39], and also more in the guided group than the waitlist group, b = -0.55, z(117) = -4.31, p < .001, 95% CI [-0.86, -

0.24]. These interaction results are also consistent with the findings obtained with MLM and provided confidence in our ability to analyze our data with MLM.

Comparing Unguided and Guided Groups During Active Treatment

The second auxiliary hypothesis was that our guided group would have a larger treatment response compared to our unguided group. Active treatment data from the waitlist group were used. Eight participants (17.8%) in the waitlist group fell below a 10 on the PHQ-9 before starting the self-help program and were therefore removed. This made it so that all participants, even those who started in the waitlist group, experienced an elevated level of depression before starting the program. We used the same model building approach as described in the prior section. For all models, the unguided group served as our reference group. The main effect model was a significant improvement over our random effects model, $\chi^2(8) = 89.97$, $p < .001$, AIC =2380.572, BIC = 2412.76. The group term was not significant, $b = 0.35$, $t(174) = 0.54$, $p > .10$, 95% CI [-0.93, 1.62]. This suggests that our two groups had a similar level of depression before starting the program. The interaction effect model did not fit the data better than our main effect model, $\chi^2(9) = 1.23$, $p > .10$, AIC = 2381.34, BIC = 2417.55. Regressing individual treatment time points (i.e., week 1, week 3, and week 5) onto our group term revealed that depression at any time point did not differ between groups. This suggests that our two experimental groups had similar depression scores across the active treatment phase of our study (see Figure 3). More specifically, it suggests that providing guidance did not improve our outcomes. See Table 4 for the sample sizes, means, and standard deviations of these cells as well as *F*-tests for each time point.

Given that the waitlist group was collapsed into guided and unguided groups for these analyses, the waitlist group's impact on depression was analyzed separately. We specified an additional three-way interaction model, in which a time by group by waitlist ("no" or "yes") was included. Neither the two-way interaction model nor the three-way interaction model were significant improvements, $\chi^2(12) = 3.55$, $p > .10$, AIC =2383.97, BIC = 2437.02 and $\chi^2(13) = 1.26$, $p > .10$, AIC = 2384.71, BIC = 2437.02, respectively. However, it is important to note that very few people were in some cells and provided post program data. For instance, only 10 participants started in the waitlist group, where then assigned to the guided group, and ultimately provided post program data. Such sparse data may have made it difficult to find unbiased and reliable estimates. As a result, we further explored the potential impact of being in the waitlist condition by removing the experimental group term from our model and just analyzing waitlist status. Removing four Level 1 observations and six Level 2 clusters that were deemed to have undue influence on our main effect model, our main effect model was significant, $\chi^2(8) = 90.78$, $p < .001$, AIC = 2379.75, BIC = 2411.94. In addition to time, waitlist status was significant, $b = -1.71$, $t(167) = -2.02$, $p < .05$, 95% CI [-3.36, -0.05]. We take this to mean that those who started in the waitlist group had significantly lower depression scores upon starting the self-help program than those who did not start in the waitlist group. The interaction effect model was not significant, $\chi^2(9) = 2.06$, $p > .10$, AIC = 2379.69, BIC = 2415.91. This suggests that waitlist group status did not impact our observed treatment response.

Another factor that we suspected could have impacted our treatment effects was whether participants completed the program or not. That is, the role of dropout was examined. Similar to the above paragraph, an additional three-way interaction model, with a time by group by completed (“no” or “yes”) term, was added. As before, the main effect model was significant, but only time was significant. Neither the two-way interaction model nor the three-way interaction model were significant improvements, $\chi^2(12) = 2.93$, $p > .10$, AIC = 2385.157, BIC = 2433.44 and $\chi^2(13) = 0.15$, $p > .10$, AIC = 2387.01, BIC = 2439.31, respectively. Similar to the previous analyses, there were very few data in some cells. For instance, seven participants whom where assigned to the unguided group, did not complete the program, and provided post depression data. As before, we removed the group term and analyzed completion status separately. The main effect model was significant, $\chi^2(8) = 90.11$, $p < .001$, AIC = 2382.29, BIC = 2418.49. However, only time was significant, $b = -2.70$, $t(236) = -11.15$, $p < .001$, 95% CI [-3.17, -2.22], suggesting that even participants that did not finish the program improved. The interaction effect model was not significant, suggesting that the rate of improvement was similar for those that did and did not complete the program.

Comparing Unguided and Guided Groups During Follow-Up

We sought to examine if our treatment effect retained during our follow-up period. Those that were initially assigned to the waitlist group were not included in this analysis, as they did not provide follow-up data. Of these 139 participants, 49 (35%) were missing week 5 data, 89 (64%) were missing week 7, and 103 (74%) were missing week 9 data. These percentages were well above the 50% missing data cut-off for MLM. As a result, we tested this model with only completers ($N = 36$). This included 15 participants in our unguided group with 45 depression scores total and 21 participants in our guided group with 63 depression scores total. Due to such as small sample size, we tested our

MLM with a Kenward-Roger correction. The main effect model was not significant, $F(2, 45.3) = 1.9$, $p > .10$, nor was our interaction effect model, $F(1, 34) = 0.32$, $p > .10$. These results suggest that participants that provided follow-up data maintained their depression levels a month after the self-help program ended.

Hypothesis 1: Self-Regulation Will Moderate the Effect of Group on Symptom Reduction Over Time

Our first hypothesis focused on how autonomous motivation, goal specificity, response inhibition, and delay discounting would moderate the effect treatment on depression. For each model, we tested our originally proposed three-way interaction and then re-tested our model without our grouping variable. We did this because (1) our unguided and guided groups did not significantly differ in treatment response, (2) removing this variable provided additional power to detect an effect, and (3) a two-way interaction model is easier to interpret than a three-way interaction model.

The Impact of Autonomous Motivation

For our motivation analyses, a time by group by motivation interaction term was specified, with the unguided group serving as a reference and motivation being grand mean centered. Including the main effects was a significant improvement, $\chi^2(9) = 95.92$, $p < .001$, AIC = 2376.61, BIC = 2412.82, as was the two-way interaction model, $\chi^2(12) = 8.14$, $p < .05$, AIC = 2374.48, BIC = 2422.76. However, the three-way interaction model did not fit the data better than our two-way interaction model, $\chi^2(13) = 0.40$, $p > .10$, AIC = 2376.07, BIC = 2428.38. The fit indices for our three-way interaction model were larger than the fit indices for our two-way interaction model, also suggesting that our three-way interaction model was a poorer fit.

For the main effect model, time and motivation were both significant. The effect of time was similar across analyses and has been covered in other sections. For the relationship of motivation, as motivation increased, so did pre program depression, b = 0.08, $t(173) = 2.50$, $p < .05$, 95% CI [0.02, 0.15]. This could be understood as those with elevated depression at the start of self-help had greater motivation to change. For our two-way interaction model, the interaction between time and group continued to not be significant, b = 0.64, $t(234) = 1.33$, $p > .1$, 95% CI [-0.29, 1.57], but the interaction between time and motivation was significant, b = -0.07, $t(234) = -2.6$, $p < .001$, 95% CI [-0.11, -0.02]. These results suggest that motivation predicts the rate of depression regardless of group status.

For ease of interpretation, a simple slopes analysis was performed after removing our group term. After removing our group term from the model, the results were similar. All assumptions were still met after removing this variable. The main effect of motivation was still significant, b = 0.08, $t(174) = 2.53$, $p < .05$, 95% CI [0.02, 0.15], as was the interaction between time and motivation, b = -0.06, $t(235) = -2.49$, $p < .05$, 95% CI [-0.11, -0.01]. The simple slopes analysis suggested that those with low (-1SD) motivation had a significant decrease in depression over time, b = -2.09, $t(235) = -5.98$, $p < .001$, 95% CI [-2.77, -1.40]. Those with high (+1SD) motivation also had a significant decrease in depression over time, albeit a stronger relationship, b = -3.29, $t(235) = -9.95$, $p < .001$, 95% CI [-3.93, -2.64]. See Figure 4 for a plot of these simple slopes. These results suggest that autonomous motivation could serve as an important factor in predicting response to self-help programs.

The Impact of Goal Specificity

Similar to our motivation analysis, we specified a time by group by specificity interaction term. The main effect model was significant, $\chi^2(9) = 81.62$, $p < .001$, AIC = 1891.98, BIC = 1926.12. The main effect of time was significant and the main effect of specificity was trending, $b = 2.54$, $t(136) = 1.83$, $p < .10$, 95% CI [-0.19, 5.27]; the latter suggesting a possible relationship between specificity and pre program depression. Adding the two-way interactions did not improve the models fit, $\chi^2(12) = 6.04$, $p > .11$, AIC = 1891.94, BIC = 1937.46, nor did adding the three-way interaction, $\chi^2(13) = 0.03$, $p > .10$, AIC = 1893.91, BIC = 1943.22. The relatively high likelihood ratio value (6.04) and *p*-value (.11) of our two-way interaction model may suggest that we were underpowered to detect our two-way interaction effects.

Removing the group term revealed slightly different results. The estimates from the main effect model were almost identical. However, the interaction effect model was trending, $\chi^2(9) = 3.59$, $p < .10$, AIC = 1888.45, BIC = 1922.59. The interaction between specificity and time was trending, $b = -2.03$, $t(187) = -1.9$, $p < .10$, 95% CI [-4.12, 0.06]. Even though this effect was trending, we sought to better understand this relationship by conducting a simple slopes analysis. Time significantly predicted depression at low levels of specificity, $b = -2.43$, $t(187) = -6.25$, $p < .001$, 95% CI [-3.19, -1.67] and at high levels of specificity, albeit a large effect, $b = -3.51$, $t(187) = -8.78$, $p < .001$, 95% CI [-4.29, -2.74]. See Figure 5 for a plot of these simple slopes. Overall, these results suggest that those with high goal specificity might respond better to self-help than those with low goal specificity. However, the trending effect limits our ability to make strong conclusions at this time.

The goal specificity task consists of two goal types, avoidance and approach goals. We explored whether these goal types moderate the relationship between time and depression. For these models, the group term was removed to improve power. A log transformation was applied to the approach goal specificity variable because it was Poisson distributed and a log transformation provided more variance to test the interaction model. More specifically, because the original variable was Poisson distributed, the low approach specificity value (i.e., -1SD below the mean) was not found in the data; it was below the minimum value. For the approach goal type, the main effect model was significant, $\chi^2(8) = 81.01$, $p < .001$, AIC = 1856.73, BIC = 1886.93, and the main effect of approach specificity was trending, $b = 2.55$, $t(135) = 1.69$, $p < .10$, 95% CI [-0.41, 5.52]. The interaction effect model was significant, $\chi^2(9) = 4.04$, $p < .05$, AIC = 1854.69, BIC = 1888.66, and so was the interaction between time and approach specificity, $b = -2.50$, $t(183) = -2.01$, $p < .05$, 95% CI [-4.92, -0.07]. Simple slopes analysis revealed that time significantly predicted depression at low levels of approach specificity, $b = -2.39$, $t(183) = -5.80$, $p < .001$, 95% CI [-3.19, -1.58] and at high levels of approach specificity, albeit a larger effect, $b = -3.61$, $t(183) = -8.78$, $p < .001$, 95% CI [-4.41, -2.80]. Those with specified goals related to approaching meaningful activities were more likely to see improvements in their depression; see Figure 6. Overall, these results suggest that the specificity of approach goals is likely important when predicting treatment responses to self-help programs for depression.

For the avoidance goal type, the results were different. Similar to approach goal specificity, avoidance specificity was also Poisson distributed. An inverse transformation provided more variance to test the interaction model. The main effect model was significant, $\chi^2(8) = 79.96$, $p < .001$, AIC = 1891.64, BIC = 1921.99, but only the effect of time was significant. Additionally, the interaction effect model was not significant, $\chi^2(9) = 1.14$, $p > .1$, AIC = 1892.5, BIC = 1926.63. One potential reason why avoidance specificity did not predict depression but approach specificity did is that approach goals were more specified than avoidance goals. Using MLM with a logarithmic link function, goal type significantly predicted specificity, $exp(b) = 1.06$, $t(136) = 2.95$, $p < .01$, 95% CI [1.02, 1.11]. Approach goal specificity had a higher mean (1.34) and standard deviation (0.37) than avoidance goal specificity ($M = 1.24$, $SD = 0.29$). This suggests that the variance of specificity for avoidance goal was lower than for approach goals and may have limited our ability to detect an effect of avoidance goal specificity.

The Impact of Response Inhibition

Response inhibition scores were derived from the emotional Stroop tasks. In this task, higher scores indicate a harder time pulling away from prepotent stimuli and hence reflect lower response inhibition. The only model that was significant was our main effect model, $\chi^2(9) = 90.07$, $p < .001$, AIC = 2382.47, BIC = 2418.68. Only time was significant. After removing our group term, the interaction between time and inhibition remained non-significant, $\chi^2(9) = 1.43$, $p > .10$, AIC = 2381.31, BIC = 2417.52. However, removing univariate outliers ($N = 4$) from the response inhibition variable, our interaction effect model was trending, $\chi^2(9) = 3.41$, $p < .10$, AIC = 2303.774, BIC = 2339.72, with a beta coefficient also trending, $b = 0.12$, $t(227) = 1.86$, $p < .10$, 95% CI [-0.0006, 0.03]. This suggests that response inhibition might impact treatment response.

Although our interaction effect was only trending, we attempted to explore this relationship more. For our simple slopes analysis, we flipped high and low status to reflect the intended meaning of our response inhibition construct (i.e., higher scores on the emotional Stroop task reflect lower inhibition). Removing the four univariate outliers discussed in the previous paragraph, simple slopes analysis revealed that depression significantly improved for those with low inhibition, $b = -2.45$, $t(227) = -7.18$, $p < .001$, 95% CI [-0.004, -1.76], and also improved, albeit more, for those with high inhibition, $b = -2.96$, $t(227) = -9.25$, $p < .001$, 95% CI [-3.59, -2.33]. See Figure 7 for a visual depiction of these simple slopes. These results suggest that response inhibition might be an important factor when predicting response to self-help programs for depression. However, one thing to consider is that our mean inhibition score for this analysis, in its original form, was only 3.17 (*SD* = 37.01), which suggests that on average our participants did not have an interference effect and in some cases had a facilitation effect (i.e., quicker to respond to depression words than neutral words). As a result, the interpretation may be more that those with high facilitation are more likely to respond to self-help. However, the trending interaction effect limits our ability to make a strong interpretation either way.

Impact of Delay Discounting

Delay discounting scores were obtained from calculating the area under the curve for indifference points obtained using the monetary choice task. High values indicate greater discounting, which reflects a greater tendency towards instant gratification decision-making. Higher scores ultimately suggest a tendency away from delayed gratification decision-making, a key self-regulatory process. A log transformation was applied to the discounting variable because it was Poisson distributed and a log transformation gave provided more variance to test the interaction models. More specifically, because the original variable was Poisson distributed, the low discounting value (i.e., -1SD below the mean) was not found in the data; it was below the minimum value. As before, a three-way interaction between time, group, and discounting was added. The main effect model was significant, $\chi^2(9) = 92.65$, $p < .001$, AIC = 2379.89, BIC = 2416.10, but the two-way interaction model was not, $\chi^2(12) = 5.42$, $p > .10$, AIC = 2380.47, BIC = 2428.75, nor was the three-way interaction model was not significant, $\chi^2(13) = 0.10$, $p > .10$, AIC = 2382.38, BIC = 2434.68. Time was the only significant main effect, but the main effect of discounting was trending, $b = -0.50$, $t(173) = -1.66$, $p < .10$, 95% CI [-1.09, 0.09]. Although the directionality of this relationship is not what we would expect to observe, it is unclear if this finding would replicate in future research.

As before, we removed our group term and re-fit our models. Removing six observations and three clusters that were deemed to have undue influence on the results, the main effect model was significant, $\chi^2(8) = 92.38$, $p < .001$, AIC = 2410.35, BIC = 2410.35, but the interaction effect model was not, $\chi^2(9) = 0.61$, $p > .1$, AIC = 2305.47, BIC = 2341.46. The main effect of discounting was not significant, $b = -0.49$, $t(170) = -1.51$, $p > .10$, 95% CI [-1.12, 0.15]. See Figure 8 for a visualization of these results. Overall, these results suggest that discounting likely did not impact how people responded to self-help for depression in this study.

Hypothesis 2: Self-Regulation Will Moderate the Effect of Group on the Odds of Complete the Program

Our second hypothesis focused on how autonomous motivation, goal specificity, response inhibition, and delay discounting would moderate the odds of completing the program. So, are those with self-regulatory abilities more likely to stick out an unguided self-help program than those without such abilities? We expected that the relationship between self-regulation and odds of completing the program would be less significant in the guided group. For all analyses, we first specified an intercept-only model that served as our baseline for comparison. We analyzed our originally proposed two-way interaction and then re-analyzed our models without out our group variable to increase our power to detect an effect. We calculated and reported odds ratios for ease of interpretation.

The Impact of Autonomous Motivation

We started by retaining our group term in the model and specifying a group by motivation term. All assumptions of logistic regression were met and no outlying cases were identified. We did not find any significant results. The main effect model was not significant, $\chi^2(2, 181) = 0.08$, $p > .10$, AUC = .52, AIC = 260.65, nor was the interaction effect model, $\chi^2(1, 180) = 0.003$, $p > .10$, AUC = .52, AIC = 262.65. Even after removing the group term, motivation still did not predict the odds of completing the program, $\chi^2(1, 182) = 0.03$, $p > .1$, AUC = .51, AIC = 258.7. Observing the AIC values for the models with and without our group term suggests that we have a slightly better model fit without the group term. This finding is consisted with the results obtained from hypothesis one; the group term does not seem to impact the results and generally contributes to a poorer model fit. These results also suggest that autonomous motivation did not influence the odds of completing a self-help program in this study.

The Impact of Goal Specificity

For our goal specificity models, we did not observe any significant findings. The main effect model was not significant, $\chi^2(2, 136) = 0.43$, $p > .10$, AUC = .53, AIC = 198.2, nor was the interaction effect model, $\chi^2(1, 135) = 0.12$, $p > .10$, AUC = .55, AIC =

200.08. Even after removing the group term, specificity did not predict the odds of completing the program, $\chi^2(1, 137) = 0.43$, $p > .10$, AUC = .54, AIC = 192.2. Specificity of approach goals did not predict the odds of completing, nor did it moderate the impact of our treatment groups on the odds of completing. Removing seven multivariate outliers, the model in which we specified the interaction between group and avoidance goal specificity was significant, $\chi^2(1, 128) = 4.32$, $p < .05$, AUC = .65, AIC = 179.38. The interaction term was trending significant, *exp(b)* = 59.27, $z(128) = 1.93$, $p = .054$, 95% CI [1.24, 5449.81]. However, these parameters (e.g., high odds ratio, high standard error, large CI bands) and an observed standard error of 2.11 suggest that this model may have suffered from problems with separation. The limited variance in this variable that we mentioned in the previous hypothesis might have contributed to this outcome. As a result, we did not further unpack this finding. Overall, these results suggest that average goal specificity and approach goal specificity did not impact the odds of completing our self-help program for depression. The impact of avoidance goal specificity is currently unknown.

The Impact of Response Inhibition

The relationship between response inhibition and odds of completing the self-help program revealed different results. When we retained the group term, the main effect model was not significant, $\chi^2(2, 181) = 0.07$, $p > .10$, AUC = .5, AIC = 260.66. Removing the four previously discussed univariate outliers and five additional multivariate outliers, the interaction effect model fit the data better than the main effect model, $\chi^2(1, 171) = 7.12$, $p < .01$, AUC = .58, AIC = 240.25. The interaction between group and inhibition was significant, *exp(b)* = 0.98, $z(171) = -2.51$, $p < .05$, 95% CI [0.95, 0.99]. Post-hoc tests revealed that the odds of completing were higher in the guided group than the unguided group among those with high inhibition (or high facilitation depending on how we interpret these findings; see discussion above), *exp(b)* = 2.58, $z(171) = 1.90$, $p = .057$, 95% CI [1.002, 7.14] and lower in the guided group than the unguided group among those with low inhibition, *exp(b)* = 0.38, $z(171) = -1.94$, $p = .052$, 95% CI [0.14, 0.98]. Although the *p*-values were only trending, the confidence intervals were significant. However, the ranges of the confidence intervals obtained were relatively large, suggesting that the true effect size is largely uncertain. On the surface, these results suggest that response inhibition could impact whether people complete a self-help program for depression, with higher inhibition contributing to an increase in the likelihood of completing a guided self-help program. However, since the raw mean discounting score was around zero, these results suggest that those with high facilitation are more likelihood to complete a guided self-help program.

The Impact of Delay Discounting

For the discounting models, we included our log transformed discounting variable as previously described. The main effect model was not significant, $\chi^2(2, 181) = 0.47$, $p > .1$, AUC = .53, AIC = 260.27, nor was the interaction effect model, $\chi^2(1, 180) = 1.8$, $p > .1$, AUC = .56, AIC = 260.46. Removing the group term did not impact the results. This suggests that discounting did not impact whether someone completed the current self-help program or not.

Hypothesis 3: Self-Regulation Will Moderate the Effect of Group on the Rate of Sessions and Homework Assignments Completed

Engagement with an assigned intervention is important for its effectiveness. Two variables that reflect treatment engagement include sessions completed and homework assignments completed. The mean number of sessions completed was 3.72 (SD = 1.57). A majority of participants completed all five sessions (52.2%), with 11.4% completing four sessions, 9.8% completing three sessions, 9.8% completing two sessions, and 16.8% completing one session. The mean number of homework assignments completed was 3.49 (SD = 2.58). A majority of our participants completed zero assignments (19.6%), with the next highest completing all seven assignments (17.9%). Additionally, 13.5% completed one assignment, 5.4% completed two assignments, 10.9% completed three assignments, 7.1% completed four assignments, 14.7% completed five assignments, and 10.9% completed six assignments. See Table 5 for descriptive and inferential statistics of these outcomes by group.

Given these distributions, we chose to analyze our sessions completed data with a Poisson regression, and analyze our assignments completed data with a zero-inflated Poisson regression. We checked for dispersion. For the sessions completed outcome, the mean and variance were not considered to be different, $z(183) = -4.65$, $p > .10$. For the assignments completed, the mean and variance were also not considered to be different, $z(183) = -0.52$, $p > .10$. As a result, we retained the default dispersion parameter. For all models, we calculated and reported incidence rate ratios for ease of interpretation.

The Impact of Autonomous Motivation

Similar to hypothesis two, we included a group by motivation interaction term. We subsequently removed the group term to examine if motivation alone predicted the outcomes. For the sessions completed outcome, the main effect model was not a significant improvement over the intercept-only model, $\chi^2(2, 181) = 0.65$, $p > .10$, AIC = 716.64. Similarly, the interaction effect model was not a significant improvement over the main effect model, $\chi^2(1, 180) = 0.02$, $p > .10$, AIC = 718.62. Upon removing the group term, the results remained the same, $\chi^2(1, 182) = 0.59$, $p > .10$, AIC = 714.70. These results suggest that autonomous motivation did not predict the number of sessions participants complete.

For the assignments completed analysis, we found a non-significant main effect, $\chi^2(4, 178) = 4.09$, $p > .10$, AIC = 835.31, and a non-significant interaction effect, $\chi^2(2, 176) = 0.57$, $p > .10$, AIC = 838.19. Even after removing the group term, the main effect model remained non-significant, $\chi^2(2, 180) = 3.14$, $p > .10$, AIC = 832.25. Overall, these results suggest that autonomous motivation did not predict engagement in self-help.

The Impact of Goal Specificity

For the sessions completed outcome, the main effect model was not a significant improvement over our intercept-only model, $\chi^2(2, 136) = 0.59$, $p > .10$, AIC = 542.54. Similarly, the interaction effect model was not a significant improvement over the main effect model, $\chi^2(1, 135) = 0.05$, $p > .10$, AIC = 544.49. Upon removing the group term, the results remained the same, $\chi^2(1, 137) = 1.04$, $p > .10$, AIC = 540.55. Similar results were obtained for approach goal specificity and avoidance goal specificity.

A similar pattern of results was obtained for assignments completed. The main effect model was not significant, $\chi^2(4, 133) = 1.27$, $p > .10$, AIC = 630.8, nor was the interaction effect model, $\chi^2(2, 131) = 0.14$, $p > .1$, AIC = 634.66. Similar results were obtained for approach goal specificity and avoidance goal specificity. Overall, these results suggest that goal specificity did not predict engagement with self-help, as measured by sessions completed and homeworker assignments completed.

The Impact of Response Inhibition

For the response inhibition models, the main effect model and interaction effect model were both not significant, $\chi^2(2, 181) = 0.05$, $p > .1$, AIC = 717.24 and $\chi^2(1,180) = 0.15$, $p > .10$, AIC = 719.09, respectively. We obtained similar results for our assignments completed outcome. Our main effect model was not significant, $\chi^2(4, 178) = 1.04$, $p > .10$, AIC = 838.36, nor was our interaction effect model, $\chi^2(2, 179) = 0.84$, $p > .10$, AIC = 841.51. Similar results were obtained for both outcomes when we removed the group term. These results suggest that response inhibition did not impact engagement with self-help, as measured by sessions completed and homework assignments completed.

The Impact of Delay Discounting

For the discounting models, we included our log-transformed area under the curve version of our variable. The main effect model was not significant, $\chi^2(2, 181) = 0.05$, $p > .10$, AIC = 717.24, nor was the interaction effect model, $\chi^2(1, 180) = 0.15$, $p > .10$, AIC = 719.09. Even removing the group term did not improve our model fit, $\chi^2(1, 182) = 0.03$, $p > .10$, AIC = 715.26. For the assignments completed models, we also did not observe any significant findings. The main effect model was not significant, $\chi^2(4, 178) = 0.49$, $p > .10$, AIC = 835.99, nor was the interaction effect model, $\chi^2(2, 176) = 0.51$, $p > .10$, AIC = 839.48. Overall, these results suggest that delay discounting did not influence engagement with self-help.

Hypothesis 4: Self-Regulation Will Moderate the Effect of Group on the Amount of Emails Sent to Study Personnel

Self-help can be difficult to do. Guided self-help provides additional support through out a self-help program. We collected data on how much support participants received as measured by how many emails they sent to study personnel. A majority of our participants sent zero emails (60%). The mean number of emails sent was 1.16 (*SD* = 2.35, *Var* = 5.56). One participant sent 17 emails. It is important to note that a majority of the emails we received were requesting help with technological and payment issues. When we asked participants if they needed help (e.g., they did not do a homework assignment), participants would often responded by saying they did not need help. Very few participants responded to our automated emails or independently reached out for help with the self-help material ($N = 5$). As result, our measure of help received extends beyond what we expected during the planning and onset of this study.

Using a Vuong test, we observed a significant difference between a standard Poisson distribution and a zero-inflated Poisson (ZIP) distribution, $z(183) = -3.56$, $p < .001$, and a significant difference between a standard ZIP distribution and an overdispersed ZIP distribution, $z(183) = -2.32$, $p < .05$. As a result, we analyzed these data using a zero-inflated, negative-binomial regression. We reported incidence rate ratios for the count portion of our models and odds ratios for the zero-inflation portion of our models for ease of interpretation. We obtained confidence intervals using a bootstrap procedure and bias-corrected confidence intervals were reported to correct for bias and skew that comes with this procedure.

The Impact of Autonomous Motivation

For the motivation models, we observed a non-significant main effect model, $\chi^2(4, 177) = 2.77$, $p > .10$, AIC = 527.12. This suggests that the groups did not differ with the number of emails sent and motivation did not predict the number of emails sent. The interaction effect model was trending, $\chi^2(2, 175) = 5.48$, $p < .10$, AIC = 525.64. The interaction effect for the count portion of the model (number of emails sent aside from zero) was significant, *exp(b)* = 1.10, $z(175) = 2.77$, $p < .01$, 95% CI [1.01, 1.18]. Although a small effect, these results suggest that motivation might moderate the group effect. Simple slopes analysis revealed a significant group effect within the count portion of the model among those high in motivation, *exp(b)* = 3.13, $z(175) = 2.50$, $p < .05$, 95% CI [1.13, 7.19]. No other effect was significant. These results suggest that more emails were sent to study personnel in guided self-help than unguided self-help among those with high autonomous motivation. However, due to the trending interaction effect and large confidence interval range, the true value of this effect is uncertain.

The Impact of Goal Specificity

For our specificity models, we did not observe any significant main effects or interaction effect. Even after removing the group term, specificity continued to not be significant. We separately examined approach and avoidance goal specificity. We first analyzed approach goal specificity. Removing nine multivariate outliers, the main effect model was significant, $\chi^2(4, 121) = 18.23$, $p < .01$, AIC = 335.29. The group term was not significant for either portions of the model. The main effect of approach specificity appeared significant for the zero-inflation portion of the model, but the beta estimate (14.79), odds ratio (2630047.41), and standard error (8.64) were unreasonably high and questionable. This suggests that the assumption of separation may have been violated for this portion of the model. When we analyzed specificity without group and after removing eight multivariate outliers, the main effect model was significant, $\chi^2(2, 126) = 12.89$, $p < .01$, AIC = 350.02. However, as before, we found unreasonably high model parameter estimates. All avoidant specificity models were not significant. Overall, these results suggest that average goal specificity and avoidance goal specificity did not predict the number of emails sent during self-help. We do not really know if approach goal specificity predicts the number of emails sent due to our very high model parameters.

The Impact of Response Inhibition

For the response inhibition models, removing five multivariate outliers, the main effect model was not significant, nor was the interaction effect model. Removing the group term did not impact the results. Overall, these results suggest that response inhibition did not play a role in how much help participants requested, as measured by emails sent to study personnel, in our self-help program.

The Impact of Delay Discounting

For our discounting models, we again used the log-transformed variable. Removing five multivariate outliers, the main effect model was trending significant, $\chi^2(4, 172) = 9.31$, $p = .054$, AIC = 470.68. The relationship between discounting and emails sent for the count portion of our model appeared significant, $exp(b) = 0.76$, $z(172) = -2.02$, $p < .05$, 95% CI [0.59, 1.006]. Although we observed a significant *p*-value, the bias corrected confidence intervals suggested otherwise, which indicates that this relation likely does not exit. No other relationship was significant.

When we removed the group term, the main effect of discounting for the count portion of the model was significant, $exp(b) = 0.70$, $z(174) = -2.69$, $p < .01$, 95% CI [0.56, 0.9]. As discounting increased, the rate of emails sent decreased. Because higher scores on the monetary choice task suggest a greater tendency towards immediate-gratification decision-making, these findings suggest that those with impulsive tendencies reached out for help less during self-help for depression.

Discussion

The aim of this study was to investigate if self-regulation would moderate the difference between unguided and guided self-help for depression on various self-help outcomes, including treatment response, odds of completing, rate of sessions completed, rate of homework assignments completed, and number of emails sent. Although meta-analyses demonstrate that unguided self-help is less effective and has higher dropout than guided self-help, people do still benefit from unguided programs (e.g., Cuijpers et al., 2019), as they did in this study. Since unguided self-help leans on the individual to initiate and maintain change, and those with self-regulation are more likely to initiate and maintain change on their own (e.g., Hofmann et al., 2012; Lally & Garnder, 2013), it was reasonable to hypothesize that those with low self-regulation in the unguided group would have the hardest time. However, significant interaction effects were seldom observed and significant main effects of self-regulation were more common. When we did observe significant interaction effects, they were not always as expected.

Implications of Our Group Effects

We tested two auxiliary hypotheses that have implications for our primary analyses. Consistent with the extant literature on self-help for depression (e.g., Cuijpers et al., 2019), participants in our two self-help groups reported a significant reduction in depression compared to participants in our waitlist group. However, we did not expect waitlist participants to report such a significant reduction in depression symptoms after a month, and in many cases after two weeks. As noted earlier, in a meta-analysis on spontaneous improvement, only one of the studies with a waitlist period of eight weeks showed a significant reduction in depression, while four of the studies with a waitlist period of eight weeks or less did not show a reduction (Rutherford et al., 2012). We also did not expect so many cases to remit after a month (N = 8, 17.8%). A systematic review on remission rates suggests that a reasonable four-week remission rate for adults in waitlist groups is around 10% (Whiteford et al., 2012). The expected improvements in our waitlist group are likely due to typical causes of waitlist effects, such as regression to the mean, natural fluctuations in disorder severity, spontaneous recovery, and/or hope from knowing support was soon available. However, the differences between our expected and observed waitlist effect implies additional improvements and may suggest that our sample was different in some way to the previously cited reviews.

Participants in these reviews generally had a diagnosis of MDD that was obtained from clinical interviews, with some being derived from symptom rating cut-offs in Whiteford and colleagues (2012). We noted earlier that the diagnostic status of our participants could not be confirmed due to our study being conducted online and using a diagnostic screener. This could have lead to participants misrepresenting themselves and/or more false MDD positives. Given the time during which we collected data (i.e., late Fall, Winter, and early Spring during the COVID-19 pandemic), it is very possible that we enrolled participants with seasonal affective disorder (SAD) and/or adjustment disorder with depressed mood (AD-DM). These disorders are more likely to improve and in many cases remit on their own as circumstances change (e.g., brighter weather in the case of SAD, restrictions lifting in the case of AD-DM; Bachem & Maercker, 2016; Ben-Ezra, Hou, & Goodwin, 2021). In addition, around 60% of our sample indicated that their most recent episode of depression was characterized by reactivity to positive events. Although reactivity to positive events is only one characteristic that distinguishes reactive and melancholic depression subtypes, this characteristic on its own implies more fluctuation in depression severity. The possibility that our sample included other dysphoric disorders than MDD and we may have over recruited those with reactive depression impacts how we interpret our results and to whom we generalize our findings.

The second auxiliary hypothesis that we tested was that our experimental groups would differ, with the guided group showing better outcomes. We did not observe any significant main effects of group. That is, guided self-help did not fair better than unguided self-help on any metric. This is generally counter to the previously cited literature, as guided self-help appears to be more effective and retain more participants than unguided self-help, albeit a small effect (e.g., Cuijpers et al., 2019). Although Berger and colleagues (2011) did not find a significant difference in depression outcomes, they did find a small effect in favor of guided self-help; this study may have been underpowered to detect a small effect. The difference between our findings and the aforementioned literature likely has to do with deviations in our protocol and unexpected behavior from our participants.

First, the three head-to-head comparison studies we cited in our literature review (i.e., Berger, Hämmerli, et al., 2011; Kenter et al., 2016; Wade et al., 2019) included an unguided condition with no human contact during the active treatment portion of these studies. We reasoned that contact with study personnel and the additional intervention that comes with contact could be accounting for the incremental benefits of guided self-help (Berger, Hämmerli, et al., 2011; Richards & Richardson, 2012; Schippers, Adam, Smolenski, Wong, & de Wit, 2017). As a result, we added email messages to our unguided group to control for contact. These messages directed participants to a help page in our self-help application and asked participants to read such material. This also controlled for the effect of dose. Additionally, since technical and payment issues were important to resolve for all participants, administrative support was introduced. Computer-based treatments do fair better with administrative support than with no support (Richards & Richardson, 2012). These deviations may have contributed to the medium to large effect size we obtained between our unguided group and waitlist group (d = 0.75), which is comparable to the between-subjects effect size of guided self-help observed in the literature (e.g., d = 0.81; Cuijpers et al., 2019). Unfortunately, this may have restricted the variance in our group term.

Second, our guided protocol was different than Berger and colleagues' (2011) protocol in a couple ways. Berger and colleagues (2011) initial email to participants was little more personalized than the current study. In these emails, study personnel commented on any symptom improvements that their participants might have made. The current study's initial emails only provided feedback on participants' work (e.g., "It looks like you had a hard time doing one of the offline activities."); however, this feedback was provided to all participants. The current study did provide feedback on symptom improvements in the form of a time-series graph within the self-help application itself, just not in the initial emails; however, this too was provided to all participants. Since feedback is a unique intervention in it and of itself (Knaup et al., 2009), we wanted to ensure that our groups were equivalent in this way. Berger and colleagues also included in their initial emails, "recognition and reinforcement of the participant's independent work." Although we provided recognition of participants' work (e.g., "You completed session 1!"), this recognition was provided to all participants and we did not praise guided participants for working independently unless they replied back looking for help. Here is an example of the feedback Berger and colleagues (2011) provided:

> "I was very impressed that last week you worked intensively on your negative thoughts. I was also pleased to see that according to the mood barometer, you are doing better. Very good. Go on like this. If you have any questions, please contact me."

Even though all of our guided emails specifically asked participants if they needed help from study personnel, the lack of feedback related to symptoms and lack of praise for independent work may have contributed to why we received fewer emails back from participants than we expected. Berger and colleagues (2011) observed a mean number of emails sent by their guided participants to be 3.57 (SD = 3.92), while we observed a smaller number of emails sent (M = 1.34, SD = 2.60). It is possible that these deviations contributed to the lower than expected within-subjects effect of our guided self-help group (d = 0.53) and further restricted the variance in our group term.

The content of the emails received from participants may have also impacted our results. The content of the emails received in Berger and colleagues (2011) appeared to be largely related to the self-help material (e.g., encouraging engagement, problem solving barriers). Because of this, we anticipated that contact in the guided group was going to be related to the self-help material, not other study related issues. However, a majority of the emails we did receive were related to technical or payment issues and not related to the self-help material. It is likely that this further lowered our expected within-subjects effect and further restricted the variance in our group term. It is possible that these three issues, (1) controlling for contact and feedback in our unguided group, (2) not providing personalized feedback and praise in our initial emails to guided participants, and (3) the lower than expected therapeutic support in our guided group, contributed to why we did not observe any group effects on our outcomes.

Impact of Self-Regulation

Much of our argument for examining these four self-regulatory processes, autonomous motivation, goal specificity, response inhibition, and delay discounting included research that cited how these processes are (1) important for goal-directed behavior, (2) important for the regulation of depressive habits, and (3) are generally lacking in those with MDD. Because of this, we expected these factors to impact our results relatively similarly. However, these processes represent different aspect of self-regulation, which may be why we observed different relationships between these processes and our various study outcomes. See Table 6 for Pearson correlations between depression time points and self-regulatory variables.

Autonomous Motivation

Autonomous motivation emerges out of the initial and ongoing discrepancies between actual and desired states (Ning & Downing, 2012). Since motivation kick starts the change process, it is not surprising then that motivation predicted whether participants started the self-help program or not in the current study. Unexpectedly, however, as pre program motivation increased, the odds of starting the program decreased. Although this was a small relationship (OR = 0.97), it was significant. Although this was unexpected, this finding can still be loosely interpreted within a self-regulatory framework. It is possible that those with high motivation were able to make changes on their own or find alternative sources of help. An alternative explanation is that since motivation and pre program depression were related, it is possible that those with more depression were more impaired and hence had a harder time starting the program or were more pessimistic about the program and whether it could help. As noted earlier, as pre program depression increased, the odds of starting the program decreased. This supports the idea that those with elevated depression may be less inclined to start self-help, but this relationship might not be do to a lack of motivation, as those with elevated depression were highly motivated.

Those with MDD are known to have higher levels of amotivation and lower levels of autonomous motivation at earlier stages of change (Vancampfort et al., 2016). As a result, we initially expected to observe a negative relationship between autonomous motivation and pre program depression. However, we found that as motivation increased, pre program depression decreased. It is not likely that motivation increases depression and we take this to mean that those with depression whom consider doing a self-help program are motivated to change. Since our sample became less depressed before they started their intervention and we could not guarantee a diagnosis of MDD, our sample may have developed more non-clinical characteristics. For instance, in non-clinical contexts, discrepancies between actual and ideal states can increase depression and this discrepancy is a major source of motivation (Carver, 2004). Although amotivation is a characteristic of depression, it tends to be observed more in those with melancholic depression (Fava et al., 1997; Mizushima et al., 2013). Our limited assessment of depression reactivity suggests we may have sampled more from a reactive depression population than a melancholic population and this might have contributed to these results. This suggests that the relationship between motivation and depression is more complicated and may depend on different moderating factors (e.g., depression subtypes).

Multiple studies looking at the impact of motivation on treatment response to CBT for depression suggest that autonomous motivation early in treatment (e.g., session 3) predicts lower post depression scores and greater remission (McBride et al., 2010; Zuroff et al., 2007). Although, there is some evidence to suggest that this relationship is specific to those with low levels of recurrent depression (i.e., two or fewer previous episodes of depression; McBride et al., 2010). We observed a similar relationship; motivation significantly predicted the rate of depression over time. Since we observed this effect for both groups and prior research has found a similar relationship in traditional face-to-face therapy, it is possible that autonomous motivation is a general predictor of treatment response across levels of care and cannot be used to stratify care.

However, since our unguided group also included contact with participants, it is still unknown if motivation would moderate self-help programs with typical self-help protocols (i.e., unguided self-help typically includes no contact at all). Furthermore, whether this effect is due to benefits afforded by having high motivation or just that those with high motivation, who were also more depression, had more room for improvement is still unclear. In a post hoc-analysis, we did observe a significant interaction between pre program depression and time. This suggests that those with elevated pre program depression did respond better, which supports the idea that they had more room for improvement.

We expected that the effect of motivation on depression would be because motivation would contribute to more engagement in the self-help program. We generally did not observe a significant relationship between motivation and our engagement outcomes (e.g., number of sessions completed). However, we did observe a trending relationship in which our guided group sent more emails than our unguided group among those with high motivation. This finding is consistent with the alternative version of this hypothesis; those with self-regulation, motivation in this case, would be better able to seek help if it were provided to them. If we observed that our guided group had a larger reduction in symptoms among highly motivated participants, it would be easy to point to the increase in help seeking behavior observed within this portion of our sample as an explanation for our treatment effect. However, this was not the case. Setting this aside, since the interaction between group and motivation on emails sent was trending, it is harder to know if this effect would replicate. Finally, there is still the possibility that other forms for treatment engagement (e.g., cognitively processing self-help material offline) could account for the influence of motivation on treatment response to self-help.

Goal Specificity

Goals represent desired outcomes states and specific goals can help guide behaviors towards these states (Seijts et al., 2004). Similar to autonomous motivation, goal specificity also predicted the odds of starting the self-help program and in the opposite direction of what we expected. That is, as specificity increased, the odds of starting decreased. However, not all goals are the same. Although this effect was observed for specificity in general, when we examined goal subtypes, we only observed a relationship with approach goals and not avoidance goals. It is possible that those with more specific approach goals already have a sense of what they needed to do to improve their depression and do not need self-help programs. This unfortunately meant that the specificity scores in our primary analyses were lower than they could have been otherwise, and this could have limited the variance in this variable. However, we do not believe this was a problem as we observed a very similar mean and standard deviation of approach goal specificity to prior research in the area (Dickson & MacLeod, 2004).

This study was the first of its kind to observe that specificity of approach goals predicts the treatment response of self-help. That is, those with well-thought-out goals related to approaching meaningful activities showed a larger reduction in depression than those without specific goals. The pathway between approach specificity and the observed treatment response could not be explained by our engagement data, unfortunately. Since we encouraged participants to choose activities that would get them closer to their long-term goals and assumed that those with higher approach goal specificity would be better able to plan their activities accordingly, we expected that those with higher specificity would complete more homework assignments. However, we did not observe a significant relationship between approach goal specificity and assignments completed, or any other indicators of treatment engagement. Nonetheless, the importance of clearly defining goals related to approaching meaningful activities is not lost. Replacing avoidance behavior with meaningful activities is one of the key mechanisms of change in CBT (Young et al., 2014) and specificity of future oriented thinking related to positive events is correlated with depression (Gamble, Moreau, Tippett, & Addis, 2019). Perhaps the pathway between specificity of approach goals and treatment response is something more nuanced than typical indicators of treatment engagement.

It should be noted that the trending main effect of approach goal specificity on depression indicates that those with high approach specificity had slightly higher pre program depression scores. It may be the case that the better treatment response among those with high approach specificity was due in part to having a larger room for improvement than those with low approach specificity. This trending main effect is counter to what we would expect, as those with elevated depression have lower goal specificity than healthy controls (Dickson & MacLeod, 2004). However, a group different of approach specificity between healthy controls and those with elevated depression is different than a linear relationship between approach specificity and pre program depression within a depressed sample. The previous literature on depression and goal specificity was not done within a treatment context. Since we previously observed that those with high pre program depression were more motivated to start our self-help program, they might have also been motivated to think more about their goals for how to change. At the very least, the relationship between depression and goal specificity might be more complicated than previously thought.

Although it makes theoretical sense why we observed specificity of approach goals and not specificity of avoidance goals to be related to our treatment response, we should note a couple statistical anomalies with our avoidance variable that might have contributed to null findings. As we noted earlier, avoidance goals were significantly less specific than approach goals. This is counter to prior research examining goal specificity, as those with MDD have the same level of specificity between approach goals and avoidance goals (Bletcher & Kangas, 2014; Dickson & MacLeod, 2004). We observed a lower mean and standard deviation of avoidance goal specificity than prior research. The mean for the current study was 1.24 (*SD* = 0.29), while Dickson and MacLeod (2004) reported a mean of 1.53 (*SD* = 0.43). This restricted variance in our avoidance specificity variable might have made it difficult to find a meaningful effect. Additionally, this along with the restricted variance in our group term that we previously discussed may have contributed to the potential separation we observed when we attempted to predict the odds of completing the program with the interaction between group and avoidance specificity. For these reasons, we are cautions to say that avoidance goal specificity does not impact outcomes in self-help.

Response Inhibition

We measured response inhibition using the emotional Stroop task, where higher scores are supposed to represent greater interference from depression-related information and hence lower inhibition of this attentional capture response. Our raw inhibition score is worth discussing as it has implications for interpreting our results. We expected to observe an interference effect of around 100ms, which has been previously observed in participants with MDD (e.g., Mitterschiffthaler et al., 2008). We actually observed a mean of -0.24ms (*SD* = 44.94). Since we used a diagnostic screener and not a clinical interview, our sample may represent those from a general dysphoric population as opposed to a MDD population. In a meta-analysis on the emotional Stoop effect, Epp and colleagues (2012) did not find a within-subjects effect comparing negative and neutral words among studies using diagnostic screeners. Using a diagnostic screener along with the timing of our enrollment also raises the possibility that we could have enrolled those with SAD and AD-DM. It is generally unknown if there is an emotional Stroop effect in these populations and if there is not, this may have watered down our effect and hence resulted in lower scores.

Delivering this measure online might have made the task difficult to do for those with depression. Although researchers have observed similar Stroop effects when administered on a desktop in the lab, online in the lab, and online at home (Crump et al., 2013; Linnman et al., 2006), these studies included participants from normal populations. Our sample was derived from people with elevated depressive symptoms. The cognitive deficits evident in depression (Ajilchi & Nejati, 2017) may have compounded the error that was introduced when doing a reaction-timed test online at home. The number of incorrect trials observed in Table 7 suggests that some people either had a hard time doing the task or did not put an adequate amount of effort forward. Despite this, we did observe a few significant relationships, which suggest that we had enough variability to find significant effects. However, since our mean was near zero, our interpretations should change; “low response inhibition” reflects an interference effect, while “high response inhibition” reflects a facilitation effect.

We observed a few findings with response inhibition that are worth discussing. Response inhibition predicted the odds of starting self-help, the treatment response of self-help, and the odds of completing self-help. We interpret this to mean that those with lower response inhibition (i.e., high interference) fair worse in self-help. These findings are consistent with theories of self-regulation. There were a few caveats to these conclusions, however. First, the interaction effect of response inhibition and time on depression was only trending, which limits our ability to make a strong conclusion at this time. Furthermore, the impact of response inhibition on the odds of completing the program was actually a two-way interaction between group and response inhibition; one of the few significant interactions with our group term that we observed. We initially reasoned that those with low inhibition would need more help, but because they would not get it in the unguided group, they would be more inclined to disengage from the program. However, we observed that the odds of completing the program were higher for the unguided group than the guided group among those with low inhibition (i.e., interference effect). It is unclear why we might have observed this finding.

It is also difficult to interpret the other portion of this effect; those in the guided group were more likely to complete the program than the unguided group if they had high inhibition (i.e., facilitation effect). It raises the question of what does it mean for those with high facilitation to be more likely to complete the program in the guided group than in the unguided group? We initially speculated that this might have to do with depression severity. One could reason that those with a facilitation effect may be less depressed than those with an interference effect and those with less depression would complete the program at higher rates in the guided group because they (1) would be less impaired and (2) have the additional support provided by the guided intervention. However, this does not explain why those with an interference effect were more likely to complete the unguided self-help program. Another problem with this explanation is that we would have expected to observe a significant relationship between response inhibition and pre program depression severity, which we did not observe.

An alternative explanation is that this might be a statistical anomaly of repeated significance testing, which can inflate a studies Type-1 error rate (Bender & Lange, 2001). Although this could apply to all other conclusions we drew based on a significance criteria of .05, this could be more of a potential issue for our response inhibition finding. Unlike other findings in this study that have previous research to fall back on, this is the first study of its kind to observe such an interaction effect. One way to resolve this issue would be to see if this finding replicates before making strong inferences to the larger population. Nonetheless, this interaction is difficult to explain and future research could potentially help account for this finding.

Delay Discounting

Delay discounting is the process of devaluing the subjective value of outcomes as time goes on (Madden & Bickel, 2010). High delay discounting represents a tendency towards immediate-gratification decision-making, while low discounting represent a tendency towards delayed-gratification decision-making. Since delayed-gratification decision-making is important for moving towards long-term goals, we expected that those with low delay discounting (i.e., high self-regulation) would show better outcomes, especially in the unguided group, than those with high delay discounting. However, we only observed one significant relationship with discounting, the amount of emails sent to study personnel.

We initially expected that those with high discounting would need more help and hence seek more help from study personnel. However, our results were in favor of our alternative explanation; as discounting increased, the rate of emails sent decreased across our sample. This could be for a couple reasons. There is a difference between needing help and asking for help. Those with more impulsivity, and hence less self-regulation, may not have the skills to actually seek help. Those with more self-regulation may be better adept at utilizing their social resources to bolster their own change. It is also possible that those with more impulsivity were more inpatient and did not want to wait for a reply from study personnel. These explanations are largely speculative, as theories around help seeking do not account for the impact of impulsivity and self-regulation. Since this hypothesis was exploratory, this finding should be replicated before making strong inferences to the larger population.

A final explanation is that since our sample was largely from MTurk, participants who are quite motivated to do tasks properly, they may have been particularly keen to ask for administrative support, especially the more conscientious workers. In a post-hoc analysis, we did not observe a main effect of recruitment source on emails sent, but did observe a significant interaction between group assignment and recruitment source, such that MTurk participants did send more emails in the guided than unguided group than those recruited from "other" sources. Although there was no difference in discounting scores between recruitment sources, it is possible that a three way group by source by discounting interaction, such that conscientious MTurk participants, might have been significant. However, we were likely underpowered to run such a model and future research could help further clarify the role of discounting on help requested in self-help.

With respect to our null findings, it is possible that discounting does not impact self-help outcomes among those with elevated depression. One might postulate that our sample could have contributed to these null findings. However, our discounting scores were consistent with prior research and even subclinical depression is associated with a diminished valuation of long-term rewards (e.g., Imhoff et al., 2014; Stringaris et al., 2015). However, a meta-analysis on discounting rates between depressed and non-depressed individuals has yet to be performed. With such a comprehensive analysis, it is possible that delay discounting as measured by the monetary choice task would not be seen in subclinical/dysphoric samples. With the evidence we had available to us, however, it seemed reasonable to predict that delay discounting would have impacted our outcomes.

Another possible explanation for these null findings is that there may be a difference between discounting in a hypothetical monetary choice task and disengaging from goal efforts related to depression in the real world. Although research on non-depressed populations suggests that (1) discounting rates are the same for hypothetical and real rewards (Johnson & Bickel, 2002; Madden, Begotka, Raiff, & Kastern, 2003), (2) discounting rates between reward types are highly correlated (Levy & Glimcher, 2011), and (3) discounting rates, even from hypothetical discounting tasks, predict impulsive reward-seeking behaviors years later (Odum, 2011), this may not translate to those with depression. Previous research on discounting typically focuses on conditions under appetitive control (e.g., smoking, gamboling, etc.), where depression is associated with a diminished anticipatory reward response. However, this possibility only rests on the fact that research on delay discounting and depression is sparse. With the exception of the results reported in the previous paragraph, the predictive utility of delay discounting in depression has yet to be seen.

Study Implications

Several implications of this study should be noted. First, we identified several variables that could be further studied to predict the odds of starting self-help. Starting a psychological service is challenging and if interventions could be developed to target those who struggle the most to start services, we could potentially reach more people and further reduce the treatment gap. Second, the mechanisms of guided self-help should be investigated. When we controlled for non-specific factors that typically come with guided self-help, we did not observe a difference between guided and unguided self-help. Additionally, the mere addition of administrative support brought our unguided group to a level of effectiveness comparable to typical guided self-help program. Third, this was the first study to use theoretically derived variables to predict self-help outcomes. This study could open the door for future research in stratified care.

Limitations

A few limitations of the current study should be noted. To recap on some of the limitations we have already mentioned, since we did not conduct a clinical interview to confirm a diagnosis, despite the established validity of the PHQ-9, it is not 100% clear if our participants met criteria for MDD, which limits our ability to generalize to a MDD population. Even though our results cannot be generalized to this population, self-help programs for depression are used in similar populations to our sample (e.g., people during quarantines; Fischer et al., 2020) and the current results have implications for these populations. Second, since we performed our pre program survey online, unexplained error could have been introduced into our data. This might have contributed to the measurement error we observed in our PHQ-9 measure and the results we obtained with our emotional Stroop task. Third, since we added email contacts to our unguided group, our results may not generalize to more traditional forms of unguided self-help, which eliminate contact with study personnel or other administrative staff.

Another limitation of the current study is that our sample predominately included MTurk workers (66.8%). MTurk workers are socialized to complete tasks and a major motivator for completing tasks is money. Since users of self-help in the real world are not incentivized by money, the level of engagement we observed in our study might be different than what we would expect to see in the public. Since we paid MTurk participants differently than the rest of our sample (i.e., MTurk participants were incentivized more regularly throughout the program), this gave us an opportunity to see if different incentive structures impacted our outcomes. As noted earlier, MTurk participants were more likely to start the self-help program than Craigslist participants. However, in a subsequent post-hoc analysis, recruitment source did not predict the rate of sessions or homework assignments completed or the odds of completing the program, but did interact with group assignment to predict emails sent. MTurk participants sent more emails in the guided group than the unguided group, suggesting that they may been more motivated to seek administrative support. The amount of money participants received did predict the odds of completing, albeit in an unexpected way. As payment increased, the odds of completing the program decreased. With this finding, it is possible that we would see different completion rates in the real world. Overall, these results suggest that MTurk workers are more likely to start online self-help programs but not more likely to complete them, qualities that may be appealing for future research on self-help. Furthermore, MTurk is not a clinical context and our results may not generalize to those in clinical settings, such as private practice or primary care.

Since our sample was predominately female (72.7%), European American (72.3%), and in their mid thirties (*M* = 34.53; *SD* = 10.71), our results might not generalize to those outside of these demographics. Although demographic variables, such as gender and age, are mixed as to whether they predict treatment response to self-help, they may be proxy variables for other important factors. For instance, research suggests that age predicts comfort with technology (Czaja & Sharit, 1998) and those with lower comfort with technology respond worse to unguided self-help programs (Wade, Cassedy, McNally, et al., 2019). Our own post-hoc analyses found that age predicted the odds of completing our self-help program, but comfort with technology did not predict the odds of completing, possibly because of the ceiling effect observed in this variable. A similar variable, Internet use (hours per week), did predict. Although age remained significant after controlling for Internet use and we did not observe a significant interaction between age and use, our study was not designed for these analyses and future research could further unpack the influence of age on self-help. Nonetheless, as younger populations who are socialized with the Internet grow older, this may become less of an issue for online self-help programs. Until then, online self-help programs, and research that uses them, may not generalize to populations less inclined to use the Internet.

Another limitation is that our self-help program was largely text based. Although we allowed for idiosyncratic text responses to online exercises, which Berger and colleagues (2011) did not allow, we did not provide any audio clips of the text or videos to demonstrate the content. A couple participants noted in a feedback question that they had a hard time reading all of the online content and would have liked additional features. However, some praised the text-based nature of our program. We did not measure different learning styles and hence might have missed an important covariate in our analyses. It is possible that we would have observed different results with this variable in our models. At the very least, these results may only generalize to those who are inclined to learn self-help material through reading text.

Future Research

Future research could improve upon the aforementioned limitations to make stronger conclusions and generalizations. For instance, recruiting from private practices or community health care centers could promote recruitment of MDD patients and provide more generalizability to contexts with patient populations. Additionally, conducting clinical interviews and administering study measures in a lab before allocating study interventions would provide more control over the enrolled sample and limit measurement error from being introduced into the study. This study could be replicated with these changes to see if similar results are obtained.

It is interesting that we found no difference between our unguided and guided self-help groups. This is the first study of its kind to control for contact, feedback, and inevitably dose in a head-to-head comparison between unguided and guided self-help for depression and raises questions about the mechanisms of change of guided self-help. The observed finding should be replicated before making strong claims about the mechanism of guided self-help, however. Dismantling studies could further clarify why guided self-help is effective. Is it due to the accountability that comes with knowing someone is monitoring outcomes? Or is it merely the addition of some extra intervention? Future research could help answer these questions.

It is possible that self-regulatory variables cannot be used to stratify patients in to traditional levels of care. For instance, autonomous motivation appears to predict depression outcomes across levels of care, including self-help in the current study. However, our self-help groups had similar levels of treatment dose. It is unknown if self-regulatory processes moderate the impact of dose on CBT outcomes. Those with self-regulation may be able to learn and apply CBT more quickly. It is possible that self-regulation could explain early treatment responding in CBT for depression (Lutz et al., 2009). For instance, we could hypothesize that those with high self-regulation would need less of a dose of CBT than those with low self-regulation, which could have implications for treatment planning.

Conclusion

The present study demonstrated that (1) adding administrative support to self-help can improve outcomes, (2) self-regulation is important for predicting self-help outcomes, and (3) the relationship between self-regulation and self-help outcomes does not appear to be explained by treatment engagement, at least in the ways we measured. Self-regulation is multi-faceted and some processes predict better at early stages of self-help, while others predict in more nuanced ways. These findings could be generalized to a larger dysphoric population, but we refrain from generalizing to a MDD population.

Future research could strengthen our inferences by improving upon the limitations of the current study (e.g., adding clinical interviews).

References

Ajilchi, B., & Nejati, V. (2017). Executive functions in students with depression, anxiety, and stress symptoms. *Basic and Clinical Neuroscience*, *8*(3), 223-232.

Alloy, L. B., & Ahrens, A. H. (1987). Depression and pessimism for the future: Biased use of statistically relevant information in predictions for self versus others. *Journal of Personality and Social Psychology*, *52*(2), 366.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders (4th ed.)*. Washington DC, USA: American Psychiatric Press.

Amlung, M., Marsden, E., Holshausen, K., Morris, V., Patel, H., Vedelago, L., ... & McCabe, R. E. (2019). Delay Discounting as a Transdiagnostic Process in Psychiatric Disorders: A Meta-analysis. *JAMA Psychiatry*, *76*(11), 1176-1186.

Andrade, L., Caraveo-Anduaga, J. J., Berglund, P., Bijl, R. V., Graaf, R. D., Vollebergh, W., ... & Kawakami, N. (2003). The epidemiology of major depressive episodes: results from the International Consortium of Psychiatric Epidemiology (ICPE) Surveys. *International Journal of Methods in Psychiatric Research*, *12*(1), 3-21.

Andrews, J. D. (1989). Psychotherapy of depression: A self-confirmation model. *Psychological Review, 96*(4), 576.

Adriaanse, M. A., Kroese, F. M., Gillebaart, M., & De Ridder, D. T. (2014). Effortless inhibition: Habit mediates the relation between self-control and unhealthy snack consumption. *Frontiers in Psychology, 5*, 444.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*, 179–211.

Ajzen, I. (2011). The theory of planned behaviour: reactions and reflections. *Psychology and Health, 26*(9), 1113-1127.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215.

Bandura, A. (2015). On deconstructing commentaries regarding alternative theories of self-regulation. *Journal of Management, 41*(4), 1025-1044.

Barlow, D. H. (2014). *Clinical handbook of psychological disorders: A step-by-step treatment manual.* New York, NY: Guilford Publications.

Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Bauer, D. & Cox, W.M. (1998). Alcohol-related words are distracting to both alcohol abusers and non-abusers in the Stroop colour naming task. *Addiction, 93*, 1539–42.

Babyak, M. A. (2004). What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine, 66*(3), 411-421.

Bachem, R., & Maercker, A. (2016). Self-help interventions for adjustment disorder problems: A randomized waiting-list controlled study in a sample of burglary victims. *Cognitive Behaviour Therapy*, *45*(5), 397-413.

Beatty, L., & Binnion, C. (2016). A systematic review of predictors of, and reasons for, adherence to online psychological interventions. *International Journal of Behavioral Medicine*, *23*(6), 776-794.

Beevers, C. G. (2005). Cognitive vulnerability to depression: A dual process model. *Clinical Psychology Review, 25*(7), 975-1002.

Bender, R., & Lange, S. (2001). Adjusting for multiple testing-when and how? *Journal of clinical epidemiology, 54*(4), 343-349.

Bennett, S. D., Cuijpers, P., Ebert, D. D., McKenzie Smith, M., Coughtrey, A. E., Heyman, I., ... & Shafran, R. (2019). Practitioner Review: Unguided and guided self-help interventions for common mental health disorders in children and adolescents: A systematic review and meta-analysis. *Journal of Child Psychology and Psychiatry, 60*(8), 828-847.

Bentley, S. M., Pagalilauan, G. L., & Simpson, S. A. (2014). Major depression. *Medical Clinics*, *98*(5), 981-1005.

Berger, T., Caspar, F., Richardson, R., Kneubühler, B., Sutter, D., & Andersson, G. (2011). Internet-based treatment of social phobia: a randomized controlled trial comparing unguided with two types of guided self-help. *Behaviour Research and Therapy*, *49*(3), 158-169.

Berger, T., Hämmerli, K., Gubser, N., Andersson, G., & Caspar, F. (2011). Internet-based treatment of depression: a randomized controlled trial comparing guided with unguided self-help. *Cognitive Behaviour Therapy*, *40*(4), 251-266.

Belcher, J., & Kangas, M. (2014). Reduced goal specificity is associated with reduced memory specificity in depressed adults. *Cognition & Emotion, 28*(1), 163-171.

Ben-Ezra, M., Hou, W. K., & Goodwin, R. (2021). Investigating the relationship between COVID-19-related and distress and ICD-11 adjustment disorder: two cross-sectional studies. *BJPsych Open*, *7*(e21), 1-7.

Biglan, A. (1987). A behavior-analytic critique of Bandura's self-efficacy theory. *The Behavior Analyst, 10*(1), 1-15.

Black, A. C., Harel, O., & Betsy McCoach, D. (2011). Missing data techniques for multilevel data: Implications of model misspecification. *Journal of Applied Statistics, 38*(9), 1845-1865.

Bower, P., & Gilbody, S. (2005). Stepped care in psychological therapies: access, effectiveness and efficiency: narrative literature review. *The British Journal of Psychiatry*, *186*(1), 11-17.

Bradley, B. P., Mogg, K., Millar, N., & White, J. (1995). Selective processing of negative information: Effects of clinical anxiety, concurrent depression, and awareness. *Journal of Abnormal Psychology*, *104*(3), 532.

Brown, T. C., & Latham, G. P. (2002). The effects of behavioural outcome goals, learning goals, and urging people to do their best on an individual's teamwork behaviour in a group problem-solving task. *Canadian Journal of Behavioural Science*, *34*(4), 276.

Carvalho, J. P., & Hopko, D. R. (2011). Behavioral theory of depression: Reinforcement as a mediating variable between avoidance and depression. *Journal of Behavior Therapy and Experimental Psychiatry, 42*(2), 154-162.

Carver, C. S., & Scheier, M. F. (1982). Control theory: A useful conceptual framework for personality–social, clinical, and health psychology. *Psychological Bulletin*, *92*(1), 111.

Carver, C. S. (2004). Negative affects deriving from the behavioral approach system. *Emotion 4*(1), 3-22.

Carver, C. S. (2018). Control theory, goal attainment, and psychopathology. *Psychological Inquiry*, *29*(3), 139-144.

Chang, W., Cheng, J., Allaire, J., Xie, Y, & McPherson, J. (2020). shiny: Web application framework for R. R package version 1.5.0. https://CRAN.R-project.org/package=shiny.

Cohen J., Cohen P., West S., Aiken L. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences (3rd ed.).* Mahwah, NJ: Lawrence Erlbaum Associates.

Cole, D., & Preacher, K. (2014). Manifest variable path analy- sis: Potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods*, *19*, 300–315.

Collins, M. D., & Jackson, C. J. (2015). A process model of self-regulation and leadership: How attentional resource capacity and negative emotions influence constructive and destructive leadership. *The Leadership Quarterly, 26*(3), 386-401.

Cooper, A. A., & Conklin, L. R. (2015). Dropout from individual psychotherapy for major depression: A meta-analysis of randomized clinical trials. *Clinical Psychology Review, 40*, 57-65.

Corrigan, P. W., Druss, B. G., & Perlick, D. A. (2014). The impact of mental illness stigma on seeking and participating in mental health care. *Psychological Science in the Public Interest, 15*(2), 37-70.

Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PloS One, 8*(3), e57410.

Cuijpers, P., Donker, T., van Straten, A., Li, J., & Andersson, G. (2010). Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. *Psychological Medicine, 40(*12), 1943-1957.

Cuijpers, P., Noma, H., Karyotaki, E., Cipriani, A., & Furukawa, T. A. (2019). Effectiveness and acceptability of cognitive behavior therapy delivery formats in adults with depression: a network meta-analysis. *JAMA Psychiatry, 76*(7), 700-707.

D'Agostino Sr, R. B. (2009). The delayed-start study design. *New England Journal of Medicine, 361*(13), 1304-1306.

Dear, B. F., Fogliati, V. J., Fogliati, R., Johnson, B., Boyle, O., Karin, E., ... & Titov, N. (2018). Treating anxiety and depression in young adults: a randomized controlled trial comparing clinician-guided versus self-guided Internet-delivered cognitive behavioural therapy. *Australian & New Zealand Journal of Psychiatry*, *52*(7), 668-679.

Deci, E. L., & Ryan, R. M. (1980). Self-determination theory: When mind mediates behavior. *The Journal of Mind and Behavior, 1*(1), 33-43.

Deci, E. L., & Ryan, R. M. (2012). Self-determination theory in health care and its relations to motivational interviewing: a few comments. *International Journal of Behavioral Nutrition and Physical Activity*, *9*(1), 1-24.

Delgadillo, J., McMillan, D., Lucock, M., Leach, C., Ali, S., & Gilbody, S. (2014). Early changes, attrition, and dose–response in low intensity psychological interventions. *British Journal of Clinical Psychology, 53*(1), 114-130.

Devilly, G. J., & Borkovec, T. D. (2000). Psychometric properties of the credibility/expectancy questionnaire. *Journal of Behavior Therapy and Experimental Psychiatry*, *31*(2), 73-86.

Dickson, J. M., & MacLeod, A. K. (2004). Approach and avoidance goals and plans: Their relationship to anxiety and depression. *Cognitive Therapy and Research, 28*(3), 415-432.

Dickson, J. M., Moberly, N. J., O'Dea, C., & Field, M. (2016). Goal fluency, pessimism and disengagement in depression. *PloS One, 11*(11), e0166259.

Donker, T., Bennett, K., Bennett, A., Mackinnon, A., van Straten, A., Cuijpers, P., ... & Griffiths, K. M. (2013). Internet-delivered interpersonal psychotherapy versus internet-delivered cognitive behavioral therapy for adults with depressive symptoms: randomized controlled noninferiority trial. *Journal of Medical Internet Research, 15*(5), e82.

Du, W., Green, L., & Myerson, J. (2002). Cross-cultural comparisons of discounting delayed and probabilistic rewards. *The Psychological Record*, *52*(4), 479-492.

Ebert, D. D., Van Daele, T., Nordgreen, T., Karekla, M., Compare, A., Zarbo, C., ... & Kaehlke, F. (2018). Internet-and mobile-based psychological interventions: applications, efficacy, and potential for improving mental health. *European Psychologist*, *23*(2), 167-187.

Epp, A. M., Dobson, K. S., Dozois, D. J., & Frewen, P. A. (2012). A systematic meta-analysis of the Stroop task in depression. *Clinical Psychology Review*, *32*(4), 316-328.

Epstein, L. H., Salvy, S. J., Carr, K. A., Dearing, K. K., & Bickel, W. K. (2010). Food reinforcement, delay discounting and obesity. *Physiology & Behavior,* 100(5), 438-445.

Fava, M., Uebelacker, L. A., Alpert, J. E., Nierenberg, A. A., Pava, J. A., & Rosenbaum, J. F. (1997). Major depressive subtypes and treatment response. *Biological Psychiatry*, *42*(7), 568-576.

Ferster, C. B. (1973). A functional analysis of depression. *American Psychologist, 28*(10), 857-870.

Finch, W. H., & French, B. F. (2015). *Latent variable modeling with R*. New York, NY: Routledge.

Fisher, A. J. (2015). Toward a dynamic model of psychological assessment: Implications for personalized care. *Journal of Consulting and Clinical Psychology*, *83*(4), 825-836.

Fischer, R., Bortolini, T., Karl, J. A., Zilberberg, M., Robinson, K., Rabelo, A., ... & Mattos, P. (2020). Rapid review and meta-meta-analysis of self-guided interventions to address anxiety, depression, and stress during COVID-19 social distancing. *Frontiers in Psychology*, *11* (563876).

Flora, D. B. (2020). Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates. *Advances in Methods and Practices in Psychological Science*, *3*(4), 484-501.

Galla, B. M., & Duckworth, A. L. (2015). More than resisting temptation: Beneficial habits mediate the relationship between self-control and positive life outcomes. Journal of *Personality and Social Psychology, 109*(3), 508-525.

Gamble, B., Moreau, D., Tippett, L. J., & Addis, D. R. (2019). Specificity of future thinking in depression: A meta-analysis. *Perspectives on Psychological Science*, *14*(5), 816-834.

Garratt, G., Ingram, R. E., Rand, K. L., & Sawalani, G. (2007). Cognitive processes in cognitive therapy: Evaluation of the mechanisms of change in the treatment of depression. *Clinical Psychology: Science and Practice, 14*(3), 224-239.

Gibb, B. E., McGeary, J. E., & Beevers, C. G. (2016). Attentional biases to emotional stimuli: Key components of the RDoC constructs of sustained threat and loss. *American Journal of Medical Genetics*, *171*(1), 65-80.

Gwaltney, C. J., Metrik, J., Kahler, C. W., & Shiffman, S. (2009). Self-efficacy and smoking cessation: A meta-analysis. *Psychology of Addictive Behaviors, 23*(1), 56-66.

Hagger, M. S., & Chatzisarantis, N. L. (2009). Integrating the theory of planned behaviour and self-determination theory in health behaviour: A meta-analysis. *British Journal of Health Psychology, 14*(2), 275-302.

Halekoh, U., & Højsgaard, S. (2014). A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models–the R package pbkrtest. *Journal of Statistical Software*, *59*(9), 1-30.

Hayes, A. M., Laurenceau, J. P., Feldman, G., Strauss, J. L., & Cardaciotto, L. (2007). Change is not always linear: The study of nonlinear and discontinuous patterns of change in psychotherapy. *Clinical Psychology Review, 27*(6), 715-723.

Hedeker, D., Gibbons, R. D., & Waternaux, C. (1999). Sample size estimation for longitudinal designs with attrition: comparing time-related contrasts between two groups. *Journal of Educational and Behavioral Statistics, 24*(1), 70-93.

Hofmann, S. G., Curtiss, J., & McNally, R. J. (2016). A complex network perspective on clinical science. *Perspectives on Psychological Science, 11*(5), 597-605.

Hofmann, W., Friese, M., & Strack, F. (2009). Impulse and self-control from a dual-systems perspective. *Perspectives on Psychological Science, 4*(2), 162-176.

Hofmann, W., Schmeichel, B. J., & Baddeley, A. D. (2012). Executive functions and self-regulation. *Trends in Cognitive Sciences, 16*(3), 174-180.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55.

Imhoff, S., Harris, M., Weiser, J., & Reynolds, B. (2014). Delay discounting by depressed and non-depressed adolescent smokers and non-smokers. *Drug and Alcohol Dependence*, *135*(1), 152-155.

Isometsä, E., Suominen, K., Mantere, O., Valtonen, H., Leppämäki, S., Pippingsköld, M., & Arvilommi, P. (2003). The mood disorder questionnaire improves recognition of bipolar disorder in psychiatric care. *BMC psychiatry*, *3*(8), 1-4.

Jacob, K. L., Christopher, M. S., & Neuhaus, E. C. (2011). Development and validation of the cognitive-behavioral therapy skills questionnaire. *Behavior Modification, 35*(6), 595-618.

Johnson, M. W., & Bickel, W. K. (2002). Within-subject comparison of real and hypothetical money rewards in delay discounting. *Journal of the Experimental Analysis of Behavior*, *77*(2), 129-146.

Kampman, M., Keijsers, G. P., Hoogduin, C. A., & Hendriks, G. J. (2008). Outcome prediction of cognitive behaviour therapy for panic disorder: Initial symptom severity is predictive for treatment outcome, comorbid anxiety or depressive disorder, cluster C personality disorders and initial motivation are not. *Behavioural and Cognitive Psychotherapy, 36*(1), 99-112.

Kan, I. P., & Drummey, A. B. (2018). Do imposters threaten data quality? An examination of worker misrepresentation and downstream consequences in Amazon's Mechanical Turk workforce. *Computers in Human Behavior*, *83*, 243-253.

Karyotaki, E., Kleiboer, A., Smit, F., Turner, D. T., Pastor, A. M., Andersson, G., ... & Christensen, H. (2015). Predictors of treatment dropout in self-guided web-based interventions for depression: An ‘individual patient data’ meta-analysis. *Psychological Medicine, 45*(13), 2717-2726.

Kees, J., Berry, C., Burton, S., & Sheehan, K. (2017). An analysis of data quality: Professional panels, student subject pools, and Amazon's Mechanical Turk. *Journal of Advertising*, *46*(1), 141-155.

Keijsers, G. P., Schaap, C. P., Hoogduin, C., Hoogsteyns, B., & de Kemp, E. C. (1999). Preliminary results of a new instrument to assess patient motivation for treatment in cognitive-behaviour therapy. *Behavioural and Cognitive Psychotherapy, 27*(2), 165-179.

Kenter, R. M. F., Cuijpers, P., Beekman, A., & van Straten, A. (2016). Effectiveness of a web-based guided self-help intervention for outpatients with a depressive disorder: Short-term results from a randomized controlled trial. *Journal of Medical Internet Research*, *18*(3), e80.

Kenward, M. G., & Roger, J. H. (2009). An improved approximation to the precision of fixed effects from restricted maximum likelihood. *Computational Statistics & Data Analysis*, *53*(7), 2583-2595.

Kessler, R. C., Chiu, W. T., Demler, O., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry, 62*(6), 617-627.

Kiby, K. N., Petry, N. M., & Bickel, W. K. (1999). Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *Journal of Experimental Psychology: General, 128*(1), 78-87.

Kline, R. B. (2012). *Principles and practice of structural equation modeling (3rd ed.).* New York, NY: The Guilford Press.

Knaup, C., Koesters, M., Schoefer, D., Becker, T., & Puschner, B. (2009). Effect of feedback of treatment outcome in specialist mental healthcare: meta-analysis. *The British Journal of Psychiatry*, *195*(1), 15-22.

Koffarnus, M. N., & Bickel, W. K. (2014). A 5-trial adjusting delay discounting task: Accurate discount rates in less than one minute. *Experimental and Clinical Psychopharmacology, 22*(3), 222-228.

Kohn, R., Saxena, S., Levav, I., & Saraceno, B. (2004). The treatment gap in mental health care. *Bulletin of the World Health Organization*, *82*(11), 858-866.

Kraemer, H. C., & Kupfer, D. J. (2006). Size of treatment effects and their importance to clinical research and practice. *Biological Psychiatry, 59*(11), 990-996.

Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric Annals, 32*(9), 509-515.

Lally, P., & Gardner, B. (2013). Promoting habit formation. *Health Psychology Review, 7*(sup1), S137-S158.

Landers, R. N., & Behrend, T. S. (2015). An inconvenient truth: Arbitrary distinctions between organizational, Mechanical Turk, and other convenience samples. *Industrial and Organizational Psychology*, *8*(2), 142-164.

Langford, I. H., & Lewis, T. (1998). Outliers in multilevel data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *161*(2), 121-160.

Levy, D. J., & Glimcher, P. W. (2011). Comparing apples and oranges: using reward-specific and reward-general subjective value representation in the brain. *Journal of Neuroscience*, *31*(41), 14693-14707.

Lewin, A. B., Peris, T. S., Bergman, R. L., McCracken, J. T., & Piacentini, J. (2011). The role of treatment expectancy in youth receiving exposure-based CBT for obsessive compulsive disorder. *Behaviour Research and Therapy*, *49*(9), 536-543.

Linnman, C., Carlbring, P., Åhman, Å., Andersson, H., & Andersson, G. (2006). The Stroop effect on the Internet. *Computers in Human Behavior, 22*(3), 448-455.

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433-442.

Locke, E. A. (1991). Goal theory vs. control theory: Contrasting approaches to understanding work motivation. *Motivation and Emotion*, *15*(1), 9-28.

Locke, E. A., Chah, D. O., Harrison, S., & Lustgarten, N. (1989). Separating the effects of goal specificity from goal level. *Organizational Behavior and Human Decision Processes*, *43*(2), 270-287.

Locke, E. A., & Latham, G. P. (2006). New directions in goal-setting theory. *Current Directions in Psychological Science*, *15*(5), 265-268.

Loeb, K. L., Wilson, G. T., Gilbert, J. S., & Labouvie, E. (2000). Guided and unguided self-help for binge eating. *Behaviour Research and Therapy*, *38*(3), 259-272.

Lombardi, D. R., Button, M. L., & Westra, H. A. (2014). Measuring motivation: Change talk and counter-change talk in cognitive behavioral therapy for generalized anxiety. *Cognitive Behaviour Therapy*, *43*(1), 12-21.

Lorenzo-Luaces, L., German, R. E., & DeRubeis, R. J. (2015). It's complicated: The relation between cognitive change procedures, cognitive change, and symptom change in cognitive therapy for depression. *Clinical Psychology Review, 41*, 3-15.

Löwe, B., Unützer, J., Callahan, C. M., Perkins, A. J., & Kroenke, K. (2004). Monitoring depression treatment outcomes with the patient health questionnaire-9. *Medical Care, 42*(12), 1194-1201.

Luke, D. A. (2004). *Quantitative Applications in the Social Sciences Series. Multilevel Modeling.* Thousand Oaks, CA: Sage Publications.

Lutz, W., Stulz, N., & Köck, K. (2009). Patterns of early change and their relationship to outcome and follow-up among patients with major depressive disorders. *Journal of Affective Disorders, 118*(1-3), 60-68.

MacKillop, J., Weafer, J., Gray, J. C., Oshri, A., Palmer, A., & de Wit, H. (2016). The latent structure of impulsivity: Impulsive choice, impulsive action, and impulsive personality traits. *Psychopharmacology*, *233*, 3361-3370.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*(2), 163-203.

Madden, G. J., & Bickel, W. K. (2010). *Impulsivity: The behavioral and neurological science of discounting*. Washington DC: American Psychological Association.

Madden, G. J., Begotka, A. M., Raiff, B. R., & Kastern, L. L. (2003). Delay discounting of real and hypothetical rewards. *Experimental and Clinical Psychopharmacology*, *11*(2), 139-145.

Mansour, H., Nordheim, E. V., & Rutledge, J. J. (1985). Maximum likelihood estimation of variance components in repeated measures designs assuming autoregressive errors. *Biometrics, 41,* 287-294.

Marks, I. M., Mataix-Cols, D., Kenwright, M., Cameron, R., Hirsch, S., & Gega, L. (2003). Pragmatic evaluation of computer-aided self-help for anxiety and depression. *The British Journal of Psychiatry, 183*(1), 57-65.

Martin, A., Rief, W., Klaiberg, A., & Braehler, E. (2006). Validity of the brief patient health questionnaire mood scale (PHQ-9) in the general population. *General Hospital Psychiatry, 28*(1), 71-77.

McBride, C., Zuroff, D. C., Ravitz, P., Koestner, R., Moskowitz, D. S., Quilty, L., & Bagby, R. M. (2010). Autonomous and controlled motivation and interpersonal therapy for depression: Moderating role of recurrent depression. *British Journal of Clinical Psychology*, *49*(4), 529-545.

McCabe, S. B., & Gotlib, I. H. (1995). Selective attention and clinical depression: performance on a deployment-of-attention task. *Journal of abnormal psychology, 104*(1), 241-245.

McDonald, R. P. (1999). Test theory: A unified treatment. L. Mahway, NJ: Erlbaum Associates.

McNeish, D., & Stapleton, L. M. (2016). Modeling clustered data with very few clusters. *Multivariate behavioral research*, *51*(4), 495-518.

Menchola, M., Arkowitz, H. S., & Burke, B. L. (2007). Efficacy of self-administered treatments for depression and anxiety. *Professional Psychology: Research and Practice, 38*(4), 421-429.

Meyer, B., Berger, T., Caspar, F., Beevers, C., Andersson, G., & Weiss, M. (2009). Effectiveness of a novel integrative online treatment for depression (Deprexis): Randomized controlled trial. *Journal of Medical Internet Research, 11*(2), e15.

Milyavskaya, M., Inzlicht, M., Hope, N., & Koestner, R. (2015). Saying "no" to temptation: Want-to motivation improves self-regulation by reducing temptation rather than by increasing self-control. *Journal of Personality and Social Psychology, 109*(4), 677.

Mitterschiffthaler, M. T., Williams, S. C. R., Walsh, N. D., Cleare, A. J., Donaldson, C., Scott, J., & Fu, C. H. Y. (2008). Neural basis of the emotional Stroop interference effect in major depression. *Psychological Medicine, 38*(2), 247-256.

Mizushima, J., Sakurai, H., Mizuno, Y., Shinfuku, M., Tani, H., Yoshida, K., ... & Mimura, M. (2013). Melancholic and reactive depression: a reappraisal of old categories. *BMC Psychiatry*, *13*(1), 1-6.

Mogg, K., & Bradley, B. P. (2005). Attentional bias in generalized anxiety disorder versus depressive disorder. *Cognitive Therapy and Research, 29*(1), 29-45.

Moody, L., Franck, C., & Bickel, W. K. (2016). Comorbid depression, antisocial personality, and substance dependence: Relationship with delay discounting. *Drug and Alcohol Dependence*, *160*(1), 190-196.

Moorey, S. (2010). The six cycles maintenance model: growing a "vicious flower" for depression. *Behavioural and Cognitive Psychotherapy, 38*(2), 173-184.

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology, 38*(1), 30-38.

Myerson, J., Green, L., & Warusawitharana, M. (2001). Area under the curve as a measure of discounting. *Journal of the Experimental Analysis of Behavior, 76*(2), 235-243.

Ning, H. K., & Downing, K. (2012). Influence of student learning experience on academic performance: The mediator and moderator effects of self-regulation and motivation. *British Educational Research Journal, 38*(2), 219-237.

Odum, A. L. (2011). Delay discounting: I'm a k, you're a k. *Journal of the Experimental Analysis of Behavior, 96*(3), 427-439.

Ottenbreit, N. D., & Dobson, K. S. (2004). Avoidance and depression: the construction of the Cognitive–Behavioral Avoidance Scale. *Behaviour Research and Therapy, 42*(3), 293-313.

Ouellette J. A., & Wood W. (1998). Habit and intention in everyday life: The multiple processes by which past behavior predicts future behavior. *Psychological Bulletin, 124*, 54–74.

Papa, A., Sewell, M. T., Garrison-Diehn, C., & Rummel, C. (2013). A randomized open trial assessing the feasibility of behavioral activation for pathological grief responding. *Behavior Therapy, 44*(4), 639-650.

Patten, S. B., & Schopflocher, D. (2009). Longitudinal epidemiology of major depression as assessed by the Brief Patient Health Questionnaire (PHQ-9). *Comprehensive Psychiatry, 50*(1), 26-33.

Pessoa, L., Kastner, S., & Ungerleider, L. G. (2002). Attentional control of the processing of neutral and emotional stimuli. *Cognitive Brain Research, 15*(1), 31-45.

Peterson, T. J., Feldman, G., Harley, R., Fresco, D. M., Graves, L., Holmes, A., ... & Fava, M. (2007). Extreme response style in recurrent and chronically depressed patients: Change with antidepressant administration and stability during continuation treatment. *Journal of Consulting and Clinical Psychology*, *75*(1), 145-153.

Pinheiro J, Bates D, DebRoy S, Sarkar D, & R Core Team (2019). nlme: Linear and nonlinear mixed effects models. R package version 3.1-142, https://CRAN.R-project.org/package=nlme.

Powers W. T. (1991). Commentary on Bandura's "human agency." *American Psychologist, 46*, 151–153.

Pratt, L. A., Brody, D. J., & Gu, Q. (2011). *Antidepressant use in persons aged 12 and over: United States, 2005-2008*. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.

Prencipe, A., Kesek, A., Cohen, J., Lamm, C., Lewis, M. D., & Zelazo, P. D. (2011). Development of hot and cool executive function during the transition to adolescence. *Journal of Experimental Child Psychology, 108*(3), 621-637.

Pulcu, E., Thomas, E. J., Trotter, P. D., McFarquhar, M., Juhász, G., Sahakian, B. J., ... & Elliott, R. (2015). Social-economical decision making in current and remitted major depression. *Psychological Medicine, 45*(6), 1301-1313.

Quilty, L. C., McBride, C., & Bagby, R. M. (2008). Evidence for the cognitive mediational model of cognitive behavioural therapy for depression. *Psychological Medicine, 38*(11), 1531-1541.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1-36.

Rachlin, H. (1995). Self-control: Beyond commitment. *Behavioral and brain sciences, 18*(1), 109-121.

Reimers, S., & Stewart, N. (2015). Presentation and response timing accuracy in Adobe Flash and HTML5/JavaScript Web experiments. *Behavior Research Methods, 47*(2), 309-327.

Renner, F., Lobbestael, J., Peeters, F., Arntz, A., & Huibers, M. (2012). Early maladaptive schemas in depressed patients: Stability and relation with depressive symptoms over the course of treatment. *Journal of Affective Disorders, 136*, 581–590.

Richards, D., & Richardson, T. (2012). Computer-based psychological treatments for depression: a systematic review and meta-analysis. *Clinical Psychology Review, 32*(4), 329-342.

Richardson, T., Stallard, P., & Velleman, S. (2010). Computerized cognitive behavioural therapy for the prevention and treatment of depression and anxiety in children and adolescents: A systematic review. *Clinical Child and Family Psychology Review*, *13*(3), 275-290.

Rief, W., Nanke, A., Klaiberg, A., & Braehler, E. (2004). Base rates for panic and depression according to the Brief Patient Health Questionnaire: A population-based study. *Journal of Affective Disorders, 82*(2), 271-276.

Riper, H., Blankers, M., Hadiwijaya, H., Cunningham, J., Clarke, S., Wiers, R., ... & Cuijpers, P. (2014). Effectiveness of guided and unguided low-intensity internet interventions for adult alcohol misuse: a meta-analysis. *PloS One, 9*(6).

Riso, L. P., du Toit, P. L., Blandino, J. A., Penna, S., Dacey, S., Duin, J. S., et al. (2003). Cognitive aspects of chronic depression. *Journal of Abnormal Psychology, 112*(1), 72–80.

Robiner, W. N. (2006). The mental health professions: Workforce supply and demand, issues, and challenges. *Clinical Psychology Review*, *26*(5), 600-625.

Robinson, J., Rosenzweig, C., Moss, A. J., & Litman, L. (2019). Tapped out or barely tapped? Recommendations for how to harness the vast and largely unused potential of the Mechanical Turk participant pool. *PloS One*, *14*(12), e0226394.

Rüsch, N., Todd, A. R., Bodenhausen, G. V., & Corrigan, P. W. (2010). Biogenetic models of psychopathology, implicit guilt, and mental illness stigma. *Psychiatry Research*, *179*(3), 328-332.

Rutherford, B. R., Mori, S., Sneed, J. R., Pimontel, M. A., & Roose, S. P. (2012). Contribution of spontaneous improvement to placebo response in depression: A meta-analytic review. *Journal of Psychiatric Research, 46*(6), 697-702.

Schippers, M., Adam, P. C. G., Smolenski, D. J., Wong, H. T. H., & De Wit, J. B. F. (2017). A meta-analysis of overall effects of weight loss interventions delivered via mobile phones and effect size differences according to delivery mode, personal contact, and intervention intensity and duration. *Obesity Reviews, 18*(4), 450-459.

Seijts, G. H., Latham, G. P., Tasa, K., & Latham, B. W. (2004). Goal setting and goal orientation: An integration of two different yet related literatures. *Academy of Management Journal, 47*(2), 227-239.

Singer, J. D., Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* New York, NY: Oxford University Press.

Sitzmann, T., & Yeo, G. (2013). A meta-analytic investigation of the within-person self-efficacy domain: Is self-efficacy a product of past performance or a driver of future performance? *Personnel Psychology, 66*(3), 531-568.

Sheldon, K. M., & Elliot, A. J. (1999). Goal striving, need satisfaction, and longitudinal well-being: The self-concordance model. *Journal of Personality and Social Psychology, 76*(3), 482.

Shi, L., & Chen, G. (2008). Detection of outliers in multilevel models. *Journal of Statistical Planning and Inference, 138*(10), 3189-3199.

Smith, P. & Waterman, M. (2003). Processing bias for aggression words in forensic and nonforensic samples. *Cognition and Emotion, 17*, 681-701.

Snijders, T. A. (2001). Sampling. In A. H. Leyand & H. Goldstein (Eds.), *Multilevel modeling of health statistics* (pp. 159-174). New York: John Wiley.

Söchting, I., Tsai, M., & Ogrodniczuk, J. S. (2016). Patients' perceptions of treatment credibility and their relation to the outcome of group CBT for depression. *Archives of Psychiatry and Psychotherapy*, *4*, 7-15.

Spek, V., Cuijpers, P. I. M., Nyklíček, I., Riper, H., Keyzer, J., & Pop, V. (2007). Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. *Psychological Medicine*, *37*(3), 319-328.

Spineli, L. M., Jenz, E., Großhennig, A., & Koch, A. (2017). Critical appraisal of arguments for the delayed-start design proposed as alternative to the parallel-group randomized clinical trial design in the field of rare disease. *Orphanet Journal of Rare Diseases, 12*(140), 1-7.

Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin, 124*(2), 240-261.

Stringaris, A., Vidal-Ribas Belil, P., Artiges, E., Lemaitre, H., Gollier-Briant, F., Wolke, S., ... & Fadai, T. (2015). The brain's response to reward anticipation and depression in adolescence: Dimensionality, specificity, and longitudinal predictions in a community-based sample. *American Journal of Psychiatry*, *172*(12), 1215-1223.

Sumner, J. A., Griffith, J. W., & Mineka, S. (2010). Overgeneral autobiographical memory as a predictor of the course of depression: A meta-analysis. *Behaviour Research and Therapy, 48*(7), 614-625.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics (7th ed.)*. Boston, MA: Pearson.

Takahashi, T., Oono, H., Inoue, T., Boku, S., Kako, Y., Kitaichi, Y., ... & Tanaka, T. (2008). Depressive patients are more impulsive and inconsistent in intertemporal choice behavior for monetary gain and loss than healthy subjects - An analysis based on Tsallis' statistics. *Neuro Endocrinolology Letters, 29*(3), 351-358.

Thompson-Hollands, J., Bentley, K. H., Gallagher, M. W., Boswell, J. F., & Barlow, D. H. (2014). Credibility and outcome expectancy in the unified protocol: Relationship to outcomes. *Journal of Experimental Psychopathology*, *5*(1), 72-82.

Vancampfort, D., Moens, H., Madou, T., De Backer, T., Vallons, V., Bruyninx, P., ... & Probst, M. (2016). Autonomous motivation is associated with the maintenance stage of behaviour change in people with affective disorders. *Psychiatry Research*, *240*(30), 267-271.

Vancouver, J. B., Thompson, C. M., Williams, A. A. (2001). The changing signs in the relationships between self-efficacy, personal goals, and performance. *Journal of Applied Psychology, 86*, 605–620.

Wade, S. L., Cassedy, A. E., McNally, K. A., Kurowski, B. G., Kirkwood, M. W., Stancin, T., & Taylor, H. G. (2019). A Randomized Comparative Effectiveness Trial of Family-Problem–Solving Treatment for Adolescent Brain Injury: Parent Outcomes From the Coping with Head Injury through Problem Solving (CHIPS) Study. *The Journal of Head Trauma Rehabilitation*, *34*(6), 1-9.

Wade, S. L., Cassedy, A. E., Sklut, M., Taylor, H. G., McNally, K. A., Kirkwood, M. W., ... & Kurowski, B. G. (2019). The relationship of adolescent and parent preferences for treatment modality with satisfaction, attrition, adherence, and efficacy: The coping with Head Injury through Problem-Solving (CHIPS) study. *Journal of Pediatric Psychology, 44*(3), 388-401.

Walker, D. A., & Smith, T. J. (2016). Nine pseudo R2 indices for binary logistic regression models. *Journal of Modern Applied Statistical Methods*, *15*(1), 848-854.

Wang, Y. Y., Xu, D. D., Liu, R., Yang, Y., Grover, S., Ungvari, G. S., ... & Xiang, Y. T. (2019). Comparison of the screening ability between the 32-item Hypomania Checklist (HCL-32) and the Mood Disorder Questionnaire (MDQ) for bipolar disorder: A meta-analysis and systematic review. *Psychiatry Research*, *273*, 461-466.

Watkins, E. R., & Nolen-Hoeksema, S. (2014). A habit-goal framework of depressive rumination. *Journal of Abnormal Psychology*, *123*(1), 24-34.

Weitz, E. S., Hollon, S. D., Twisk, J., Van Straten, A., Huibers, M. J., David, D., ... & Faramarzi, M. (2015). Baseline depression severity as moderator of depression outcomes between cognitive behavioral therapy vs. pharmacotherapy: an individual patient data meta-analysis. *JAMA Psychiatry*, *72*(11), 1102-1109.

Westra, H. A. (2011). Comparing the predictive capacity of observed in-session resistance to self-reported motivation in cognitive behavioral therapy. *Behaviour Research and Therapy, 49*(2), 106-113.

Whiteford, H. A., Harris, M. G., McKeon, G., Baxter, A., Pennell, C., Barendregt, J. J., & Wang, J. (2013). Estimating remission from untreated major depression: A systematic review and meta-analysis. *Psychological Medicine*, *43*(8), 1569-1585.

Williams, J. M. G. (1996). Depression and the specificity of autobiographical memory. In D. C. Rubin (Ed.), *Remembering our past: Studies in autobiographical memory* (pp. 244–267). Cambridge University Press.

Williams, J. M. G., Mathews, A., & MacLeod, C. (1996). The emotional Stroop task and psychopathology. *Psychological Bulletin*, *120*(1), 3-24.

World Health Organization. (1990). Composite international diagnostic interview. Geneva, Switzerland: World Health Organization.

Young, J. E., Rygh, J. L., Weinberger, A. D., & Beck, A. T. (2014). Cognitive therapy for depression. In D. H. Barlow (Ed.), *Clinical handbook of psychological disorders: A step-by-step treatment manual* (pp. 275–331). The Guilford Press.

Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software 27*(8).

Zinbarg, R. E., Revelle, W., Yovel, I., and Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ωH: Their relations with each other and two alternative conceptualizations of reliability. Psychometrika, 70(1), 123-133.

Zuroff, D. C., Koestner, R., Moskowitz, D. S., McBride, C., Marshall, M., & Bagby, M. R. (2007). Autonomous motivation for therapy: A new common factor in brief treatments for depression. *Psychotherapy Research, 17*(2), 137-147.

Appendix

Table 1

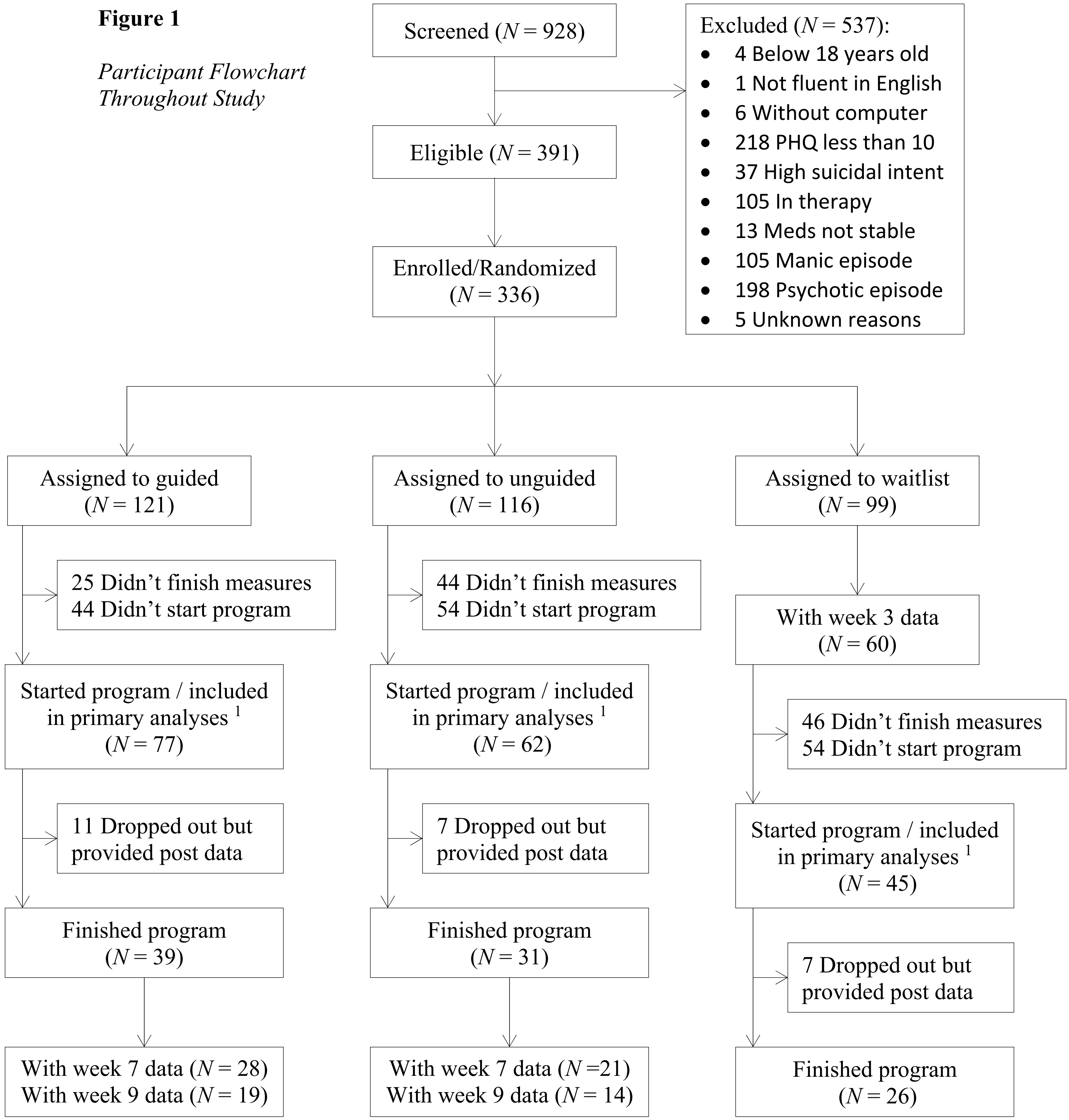
Three-arm, randomized, delayed-start, longitudinal design

Week 1 Time 1	Week 2	Week 3 Time 2	Week 4	Week 5 Time 3	Week 6	Week 7 Time 4	Week 8	Week 9 Time 5
Guided	Guided	Guided	Guided	Guided	-	-	-	-
Unguided	Unguided	Unguided	Unguided	Unguided	-	-	-	-
Waitlist	Waitlist	Waitlist	Waitlist	Guided Unguided	Guided Unguided	Guided Unguided	Guided Unguided	Guided Unguided

Note. The PHQ-9 was administered on the weeks indicated with times (e.g., Time 1). Active treatment for those assigned to the waitlist group included weeks five, seven, and nine. These time points were renamed to “Week One”, “Week Three”, and “Week Five”, respectively, in primary analyses.

Figure 1

Participant Flowchart Throughout Study



Note. [1] Some participants were excluded from different analyses. See primary analyses for more details.

Table 2a

Study Characteristics: Counts and Percentages of Factor Variables

Characteristic	Waitlist		Unguided		Guided		Total	
	N	%	*N*	%	*N*	%	*N*	%
Source								
Craigslist	9	20	6	9.5	5	6.4	20	10.9
MTurk	25	55.6	43	69.4	55	71.4	123	66.8
Social	5	11.1	7	11.3	6	7.8	18	9.8
UNR	0	0	3	4.8	3	3.9	6	3.3
Unknown [1]	6	13.3	3	4.8	8	10.3	17	9.2
Gender								
Female	36	80	41	66.1	56	72.7	133	72.2
Male	8	17.8	18	29	19	24.7	45	24.5
Trans Female	0	0	1	1.6	1	1.3	2	1.1
Trans Male	0	0	2	3.2	0	0	2	1.1
Other	1	2.2	0	0	1	1.3	2	1.1
Ethnicity								
Asian American	3	6.7	4	6.4	6	7.8	13	7.1
Black/African American	5	11.1	2	3.2	2	2.6	9	4.9
Hispanic/Latinx	3	6.6	3	4.8	4	5.2	10	5.4
American Indian	0	0	0	8.1	1	1.3	1	0.5
White	26	57.8	47	75.8	60	77.9	133	72.3

Characteristic	Waitlist		Unguided		Guided		Total	
	N	%	*N*	%	*N*	%	*N*	%
Other	0	0	1	1.6	2	2.6	3	1.6
Mixed	8	17.8	5	8.1	2	2.6	15	8.1
Education								
Less than HS	1	2.2	1	1.6	0	0	2	1.1
High School	11	24.4	17	27.4	16	20.8	44	23.9
Trade / Associates	9	20	14	22.6	14	18.2	37	20.1
Bachelor's	16	35.6	26	41.9	36	46.7	78	42.4
Master's	7	15.5	4	14.3	11	14.3	22	11.9
Doctorate	1	2.2	0	0	0	0	1	.5
Course								
Episodic	7	15.5	8	12.9	16	20.8	31	16.8
Continuous	31	68.9	37	59.7	46	59.7	114	61.9
Both	7	15.5	17	27.4	15	19.5	39	21.2
Reactivity								
Yes	32	71.1	47	75.8	49	63.6	128	69.5
No	7	15.5	12	19.3	20	26	39	21.2
Unknown [1]	6	13.3	3	4.9	8	10.4	17	9.2
Treatment History								
Therapy	9	20	8	12.9	8	10.4	25	13.6
Medication	4	8.9	7	11.2	13	16.9	24	13

Characteristic	Waitlist		Unguided		Guided		Total	
	N	%	*N*	%	*N*	%	*N*	%
Other	0	0	0	0	1	1.3	1	.5
Mixed	21	46.7	35	56	35	45.4	91	49.4
Nothing	11	24.4	12	19.3	20	26	43	23.4
Medication Status								
No	12	26.7	18	29	19	24.7	49	26.6
Yes	33	73.3	44	71	58	75.3	135	73.4

Note. [1] Unknown data means that we did not initially collect these data.

Table 2b

Study Characteristics: Means and Standard Deviations of Numeric Variables

Characteristic	Waitlist		Unguided		Guided		Total	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age	31.58	10.23	36.05	10.97	35.04	10.58	34.53	10.71
Internet Hours [1]	35.6	22.85	32.84	22.01	36.75	22.97	35.15	22.56
Internet Confidence	9.38	1.61	9.38	1.57	9.46	1.37	9.41	1.49
Internet Speed [2]	7.16	4.34	5.61	3.91	5.86	5.09	6.08	4.56
Depression Onset [3]	16.22	7.09	16.92	9.99	17.03	7.23	16.79	8.19
Depression Episodes	13.80	12.82	21.88	26.23	13.74	18.15	16.62	20.62
Anxiety Symptoms	10.31	4.62	11.76	5.29	11.89	5.06	11.47	5.05
Mania Symptoms	2.38	2.43	2.89	1.96	2.96	2.26	2.79	2.21
Autonomous Motivation	43.62	10.95	44.68	10.43	46.34	8.64	45.11	9.87
Goal Specificity	1.28	0.09	1.28	0.27	1.31	0.27	1.29	0.27
Response Inhibition [4]	8.37	36.71	-7.53	51.60	0.59	43.12	-0.24	44.94
Delay Discounting [5]	0.36	0.41	0.19	0.21	0.23	0.38	0.25	0.35

Note. [1] Hours per week; [2] Megabytes per second; [3] Age of onset; [4] Scores above 0 on the emotional Stroop task reflect an interference effect, while scores below 0 reflect a facilitation effect; [5] Values are raw area under the curve scores before log transformation.

Figure 2

Interaction Between Time and Group (Waitlist, Unguided, and Guided)

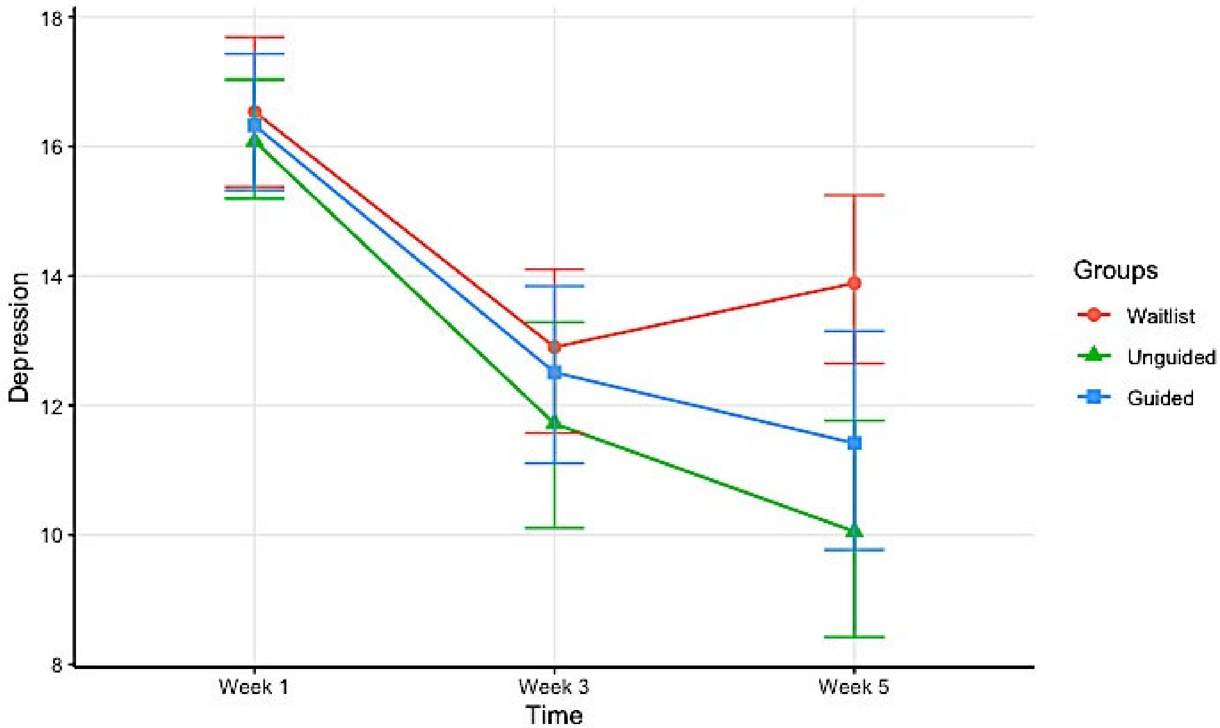


Table 3

Depression Across Time and By Group (Waitlist, Unguided, and Guided)

Depression	Waitlist			Unguided			Guided			F	p	R^2
	N	M	SD	N	M	SD	N	M	SD			
Week 1	45	16.53	4.07	66	16.08	3.98	77	16.32	4.71	0.17	.86	.002
Week 3	40	12.90	4.45	46	11.71	5.71	57	12.51	5.34	0.58	.56	.01
Week 5	45	13.89	4.59	38	10.05	5.36	48	11.47	6.35	5.25	.006	.06

Note. Interferential statistics are derived from multiple linear regression models that were compared to an intercept-only model. R^2 values are reported in decimals, not percentages.

Figure 3

Interaction Between Time and Group (Unguided and Guided)

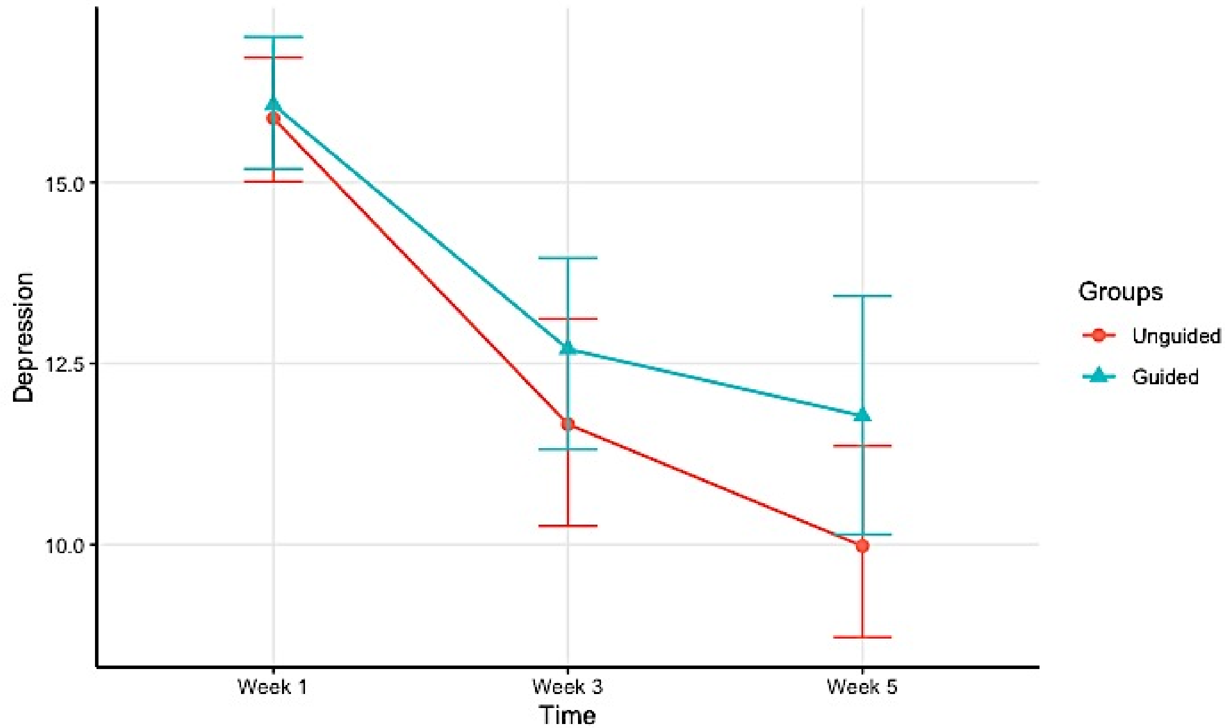


Table 4

Depression Across Active Treatment and By Group (Unguided and Guided)

Depression	Unguided			Guided			*F*	*p*	R^2
	N	*M*	*SD*	*N*	*M*	*SD*			
Week 1	80	15.89	3.98	96	16.07	4.62	0.08	.78	.0004
Week 3	59	11.66	5.33	70	12.70	5.54	1.16	.28	.009
Week 5	50	9.86	4.71	58	11.47	6.44	3.03	.08	.03

Note. Interferential statistics were derived from multiple linear regression models that were compared to an intercept-only model. R^2 values are reported in decimals, not percentages.

Table 5

Numeric Outcomes By Group (Unguided and Guided)

Outcomes	Unguided		Guided		χ^2	*p*
	M	*SD*	*M*	*SD*		
Sessions	3.75	1.56	3.70	1.58	0.04	.85
Assignments	3.49	2.57	3.50	2.60	0.82	.66
Emails	0.95	2.03	1.34	2.60	1.35	.51

Note. Interferential statistics were derived from different regression models based on the respective distributions (e.g., zero-inflated negative binomial regression). All models were compared to an intercept-only model.

Figure 4

Simple Slopes of Time by Autonomous Motivation

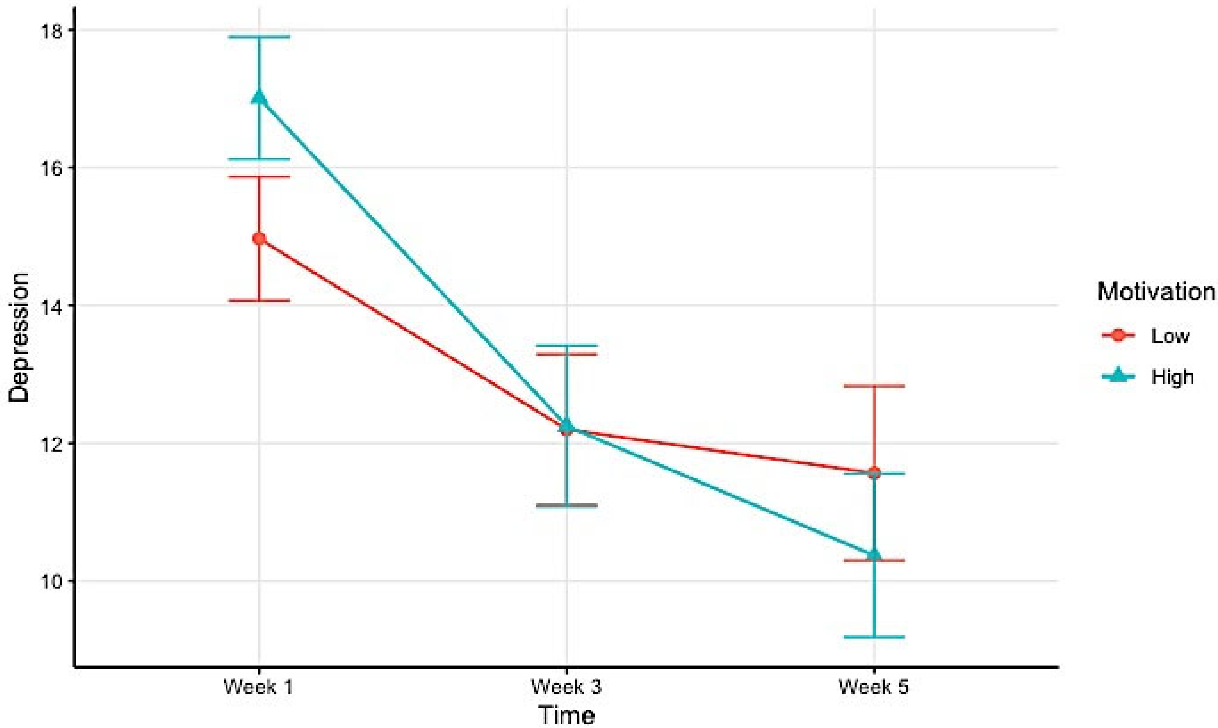


Figure 5

Simple Slopes of Time by Goal Specificity

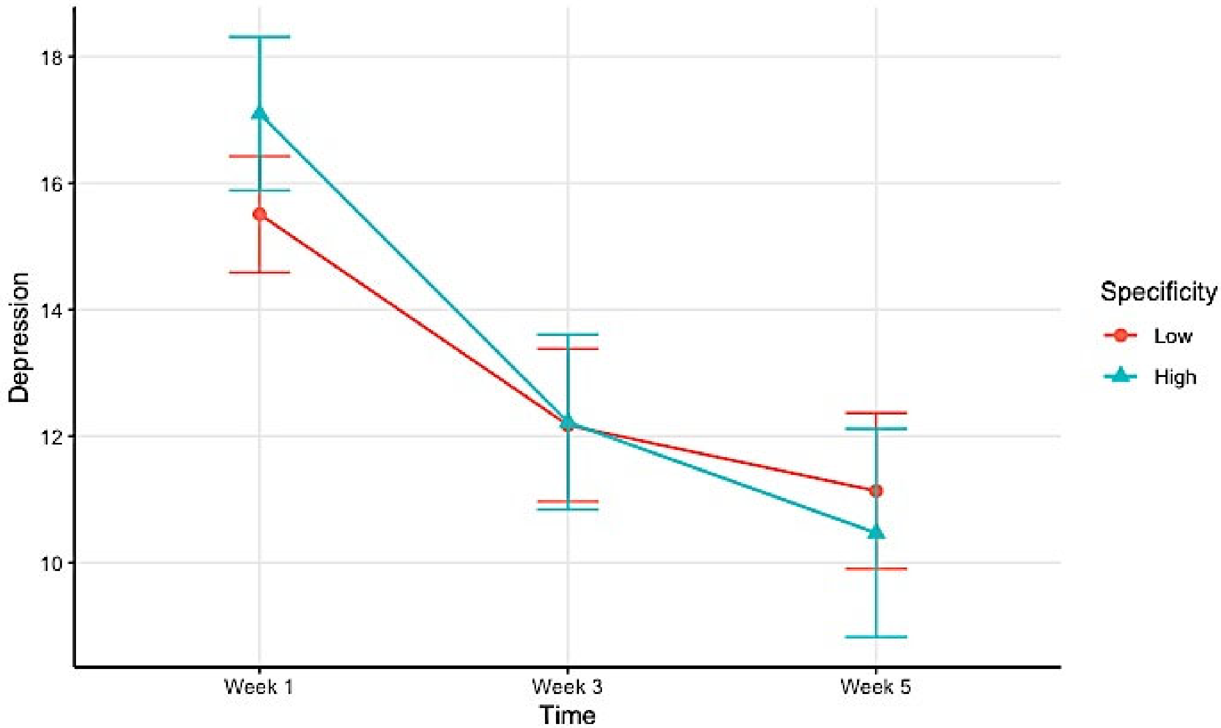
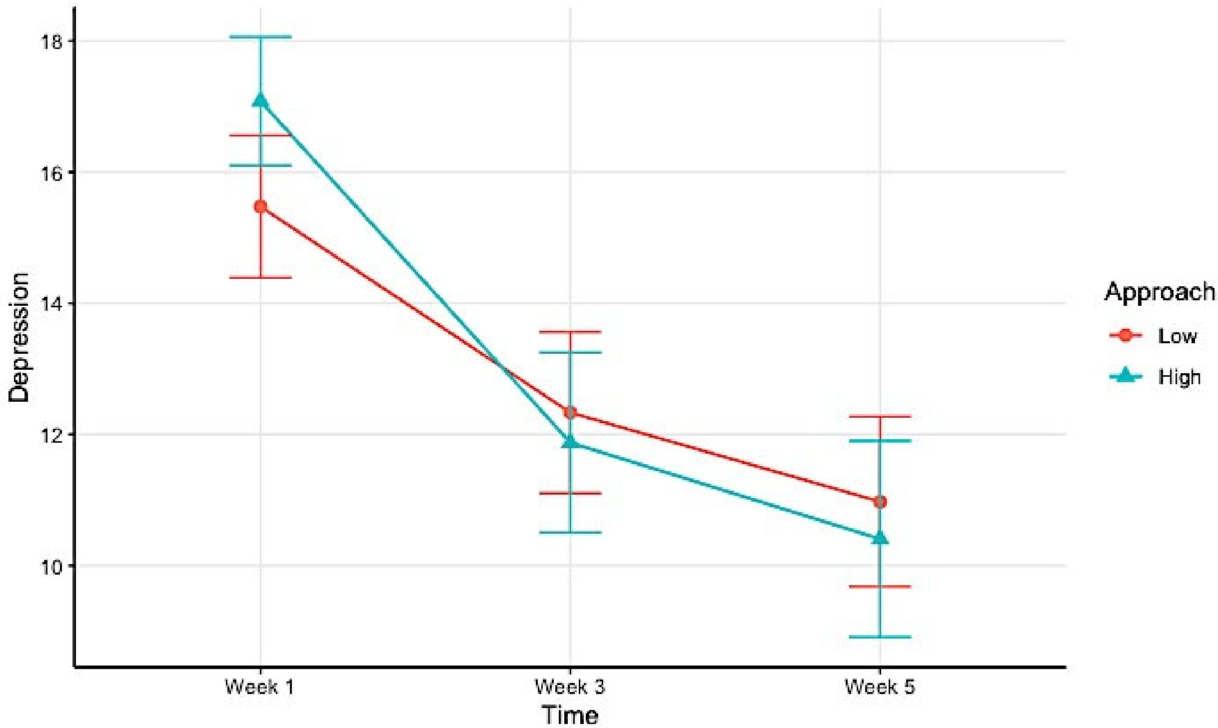


Figure 6

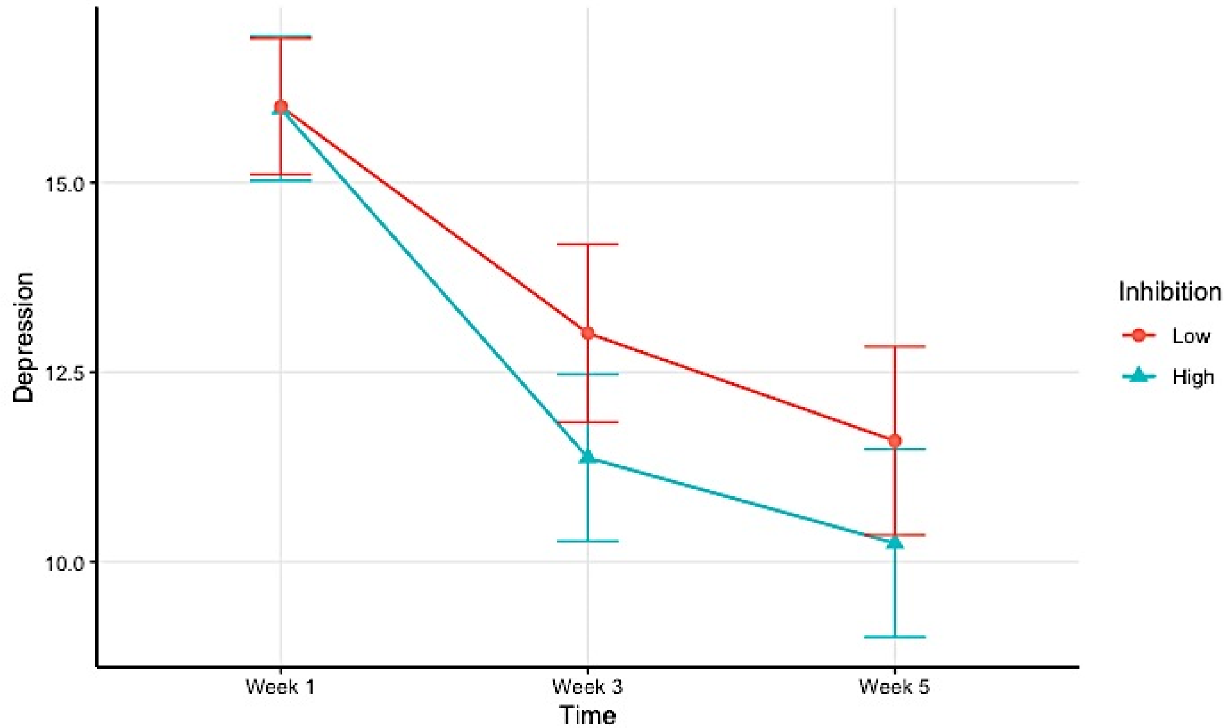
Simple Slopes of Time by Approach Goal Specificity



Note. Approach goal specificity was log transformed so that low (-1SD) and high (+1SD) values were found in the data.

Figure 7

Simple Slopes of Time by Response Inhibition



Note. Inhibition labels were switched for ease of interpretation. Higher scores on the emotional Stroop task reflect more interference and less inhibition. Univariate outliers from this variable were removed from this analysis.

Table 6

Pearson Correlations Among Depression and Self-Regulation

	W1	W3	W5	AM	GS	RI	DD
W1	-						
W3	.62***	-					
W5	.54***	.75***	-				
AM	.19*	.04	-.02	-			
GS	.18*	.00	-.14	-.04	-		
RI	.02	-.13	-.11	-.02	.07	-	
DD	-.14	.00	-.06	-.04	-.08	-.11	-

Note. W1 = week 1, W3 = week 3, W5 = week 5, AM = autonomous motivation, GS = goal specificity, RI = response inhibition, DD = delay discounting. Correlations are during active treatment. * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 7

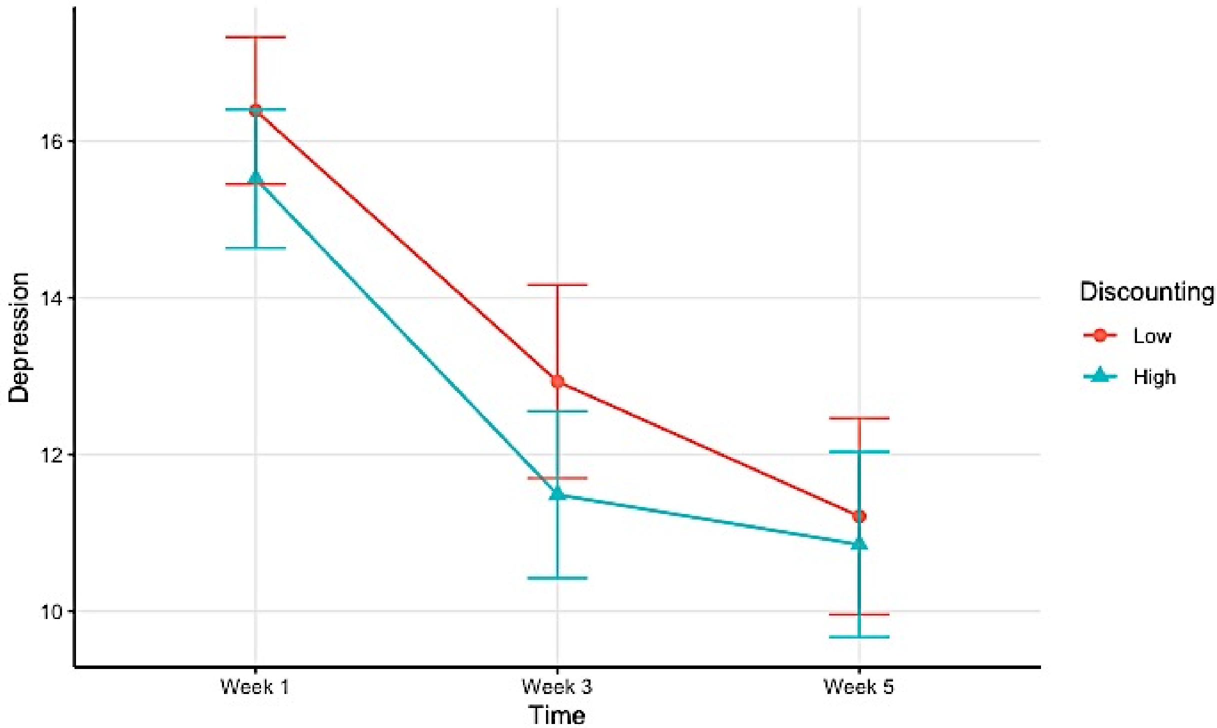
Incorrect Data for the Emotional Stroop Task

Type	Waitlist			Unguided			Guided			Total		
	M	*SD*	%	*M*	*SD*	%	*M*	*SD*	%	*M*	*SD*	%
Trials Incorrect	14.02	28.37	8.6	26.40	44.13	16.5	23.74	40.99	14.8	22.28	39.57	13.9
Trials Missing	8.47	22.94	5.3	16.63	34.30	10.4	15.74	32.93	9.8	14.28	31.32	8.9
Trials Mismatched	5.55	17.47	3.5	9.76	20.50	6.1	8.00	16.13	5.0	8.01	18.02	5.0

Note. Incorrect trials are the combination of missing trials and mismatched trials. Missing trials are trials where participants did not respond. Mismatched trials are trials in which participants responded with the wrong color.

Figure 8

Simple Slopes of Time by Delay Discounting



Note. High delay discounting reflects a greater tendency towards instant gratification decision-making. The area under the curve discounting values was log transformed. Multivariate outliers were removed from this analysis.