MANUSCRIPT PUBLISHED IN JOURNAL OF CONTEXTUAL BEHAVIORAL SCIENCE

A multiple-baseline evaluation of brief RNT-focused acceptance and commitment therapy for moderate emotional disorders

Francisco J. Ruiz[1]

Cindy L. Flórez[1]

María B. García-Martín[1]

Andrea Monroy-Cifuentes[1]

Katherine Barreto-Montero[1]

Diana M. García-Beltrán[1]

Diana Riaño-Hernández[1]

Marco A. Sierra[1]

Juan C. Suárez-Falcón[2]

Verónica Cardona-Betancourt[3]

Bárbara Gil-Luciano[4,5]

[1]Fundación Universitaria Konrad Lorenz, [2]Universidad Nacional de Educación a Distancia, [3]Private practice, [4]Universidad de Almería, [5]Madrid Institute of Contextual Psychology

Correspondence address: Francisco J. Ruiz (franciscoj.ruizj@konradlorenz.edu.co), Fundación Universitaria Konrad Lorenz, Carrera 9 bis, Nº 62-43. Bogotá (Colombia)

Abstract

Repetitive negative thinking (RNT) in the form of worry and rumination has been identified as a particularly counterproductive experiential avoidance strategy implicated in the onset and maintenance of emotional disorders. The current study analyzes the effect of an individual, 2-session, RNT-focused, acceptance and commitment therapy (ACT) protocol in the treatment of moderate emotional disorders. Ten adults suffering from moderate to severe emotional symptoms according to the Depression Anxiety and Stress Scale-21 (DASS-21) and the General Health Questionnaire-12 (GHQ-12) participated in the study. Participants completed 5- to 7-week baselines without showing improvement trends in the DASS-21 or the GHQ-12. Afterwards, they received the ACT protocol, and a 3-month follow-up was conducted. A Bayesian approach to analyze clinically significant changes (CSC) for single-case experimental designs (SCED) was conducted, which required at least substantial evidence of the intervention effect and scores in the nonclinical range. Nine of the 10 participants showed CSC in the GHQ-12, and 7 participants in the DASS-21. The standardized mean difference effect sizes for SCED were computed, which facilitates comparison and integration of the results with group designs. Very large effect sizes were found for emotional symptoms (d = 2.44 and 2.68), pathological worry (d = 3.14), experiential avoidance (d = 1.32), cognitive fusion (d = 2.01), repetitive thinking (d = 2.51), and valued living (d = 1.54 and 1.41). No adverse events were found. RNT-focused ACT protocols deserve further empirical tests.

Key words: Acceptance and commitment therapy; Relational frame theory; Depression; Generalized anxiety disorder; Repetitive negative thinking.

1. Introduction

The last decade has seen a growing interest in the identification and analysis of transdiagnostic processes involved in emotional disorders such as experiential avoidance (Hayes, Wilson, Gifford, Follette, & Strosahl, 1996; Luciano & Hayes, 2001), repetitive negative thinking (Ehring & Watkins, 2008; Harvey, Watkins, Mansell, & Shafran, 2004), emotional disturbances (Kring, 2008), selective attention (Harvey, Watkins et al., 2004), etc. The identification of these processes has led to the proposal of transdiagnostic psychological treatments with acceptance and commitment therapy (ACT; Hayes, Strosahl, & Wilson, 1999), metacognitive therapy (MCT; Wells, 2009), the unified protocol for transdiagnostic treatment of emotional disorders (Barlow et al., 2010), and rumination-focused cognitive-behavioral therapy (RF-CBT; Watkins, 2016), among others.

The research has been mostly focused on specific transdiagnostic processes, but not so much on their interrelation. Accordingly, it is difficult to state which transdiagnostic processes are more core aspects than others in explaining the onset and maintenance of emotional disorders (Harvey, Watkins, et al., 2004). One recent research (Ruiz, Riaño-Hernández, Suárez-Falcón, & Luciano, 2016) has analyzed the link between two of the most analyzed transdiagnostic processes in emotional disorders from a functional contextual standpoint: experiential avoidance (EA) and repetitive negative thinking (RNT).

EA is a central construct in ACT and was conceptualized according to relational frame theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001), a functional-contextual approach to human language and cognition. EA is a pattern of verbal regulation based on deliberate efforts to either avoid or escape from discomfiting private experiences even when doing so leads to actions that are inconsistent with one's values and goals (Hayes et

al., 1996). EA has been identified as playing a relevant role in the onset and maintenance of emotional disorders (Boulanger, Hayes, & Pistorello, 2010; Ruiz, 2010).

RNT has been identified as a core feature of emotional disorders (Ehring & Watkins, 2008; Harvey, Watkins, et al., 2004) such as depression (Nolen-Hoeksema, 2004), posttraumatic stress disorder (Michael, Halligan, Clark, & Ehlers, 2007), social anxiety (Kashdan & Roberts, 2007), and generalized anxiety disorder (Borkovec, 1994). Although RNT might have some adaptive functions (see a review in Watkins, 2008), worry and rumination have been robustly identified in prospective and experimental studies as common factors in the onset and maintenance of emotional disorders (e.g., Ehring & Watkins, 2008; Harvey, Watkins, et al., 2004; Nolen-Hoeksema, 2000). According to Watkins (2008), RNT becomes especially counterproductive when it is characterized by reduced concreteness (i.e., abstract level of construal) and its main purpose is to reduce fear, sadness, or uncertainty (i.e., experiential avoidant functions).

Based on an RFT account of the self (Luciano, 2017; Luciano, Ruiz, & Törneke, submitted), Ruiz, Riaño-Hernández, et al. (2016) have highlighted several interrelations between EA and RNT: (a) triggers of RNT are built in the individual's learning history and usually become hierarchically related to the extent that one of the strongest triggers (i.e., the thought/emotion at the top of the hierarchy) symbolically contains the remaining ones; (b) unconstructive RNT is an especially maladaptive experiential avoidance strategy; (c) RNT tends to be the first reaction to fear, unattained goals, and incoherence; (d) RNT tends to prolong negative affect; which usually leads to (e) engagement in additional experiential avoidance strategies in an attempt to finally reduce discomfort; and (f) the repetition of this reinforcing cycle generates an inflexible and maladaptive repertoire in reaction to triggers. The practical implication of this account is that ACT protocols primarily focused on

disrupting unconstructive RNT in response to the trigger at the top of the hierarchy should produce quick changes and be particularly effective for the treatment of emotional disorders (see further details in Ruiz, Riaño-Hernández, et al., 2016).

A first step in developing a RNT-focused ACT protocol was conducted by Ruiz, Riaño-Hernández, et al. (2016). Specifically, these authors investigated whether a one-session, RNT-focused, ACT protocol could be sufficient to significantly reduce high levels of worry and rumination. This seemed a logical first step, as RNT-focused ACT protocols would only make sense if they reduce RNT rapidly. A two-arm, randomized multiple-baseline design with 11 participants suffering from mild to moderate emotional symptoms was implemented. The RNT-focused ACT protocol was designed following the RFT account of psychological flexibility (Luciano et al., 2011; Luciano, Valdivia-Salas, & Ruiz, 2012; Ruiz & Perete, 2015; Törneke, Luciano, Barnes-Holmes, & Bond, 2016; see further details in Ruiz, Riaño-Hernández, et al., 2016). The results showed significant reductions in at least three out of the four RNT measures during the 6-week follow-up. Effect sizes were large in all RNT-related measures and in emotional symptoms.

Given that the one-session protocol was shown to be highly effective in reducing RNT, a second step could be to analyze the effect of a brief RNT-focused ACT protocol with participants suffering from emotional disorders. Accordingly, the aim of this study was to analyze the effect of a 2-session, RNT-focused, ACT protocol in participants suffering from moderate to severe emotional symptoms. A nonconcurrent multiple baseline design was conducted where the effect of the ACT protocol was directly replicated in 10 participants who showed stable moderate emotional symptoms during the 5 to 7 weeks of baseline and who previously reported being entangled with thoughts, memories, or worries

for at least the last 6 months. The SCRIBE statement (Tate et al., 2016) was followed to guide the reporting of this single-case experimental design.

2. Method

2.1. Participants

Participants were recruited through advertisements in social media beginning with the questions: “Do you spend too much time distressed about the past or future? Do you want to be more focused on the things that are important to you?” Seventy-six individuals showed interest in the study and were asked to respond to an online survey. Initial inclusion criteria were: (a) over 18 years old; (b) at least 6 months entangled with thoughts, memories, and worries; (c) significant interference of thoughts, memories, and worries in at least 2 life domains; (d) not showing extremely severe scores on depression and/or anxiety in the Depression, Anxiety, and Stress Scale-21 (see the Outcome Measures section); and (e) showing scores between 15 and 25 on the General Health Questionnaire-12 (see the Outcome Measures section). The initial exclusion criterion was current psychological or psychiatric treatment, including taking psychotropic medication.

The application of the initial inclusion and exclusion criteria led to the rejection of 55 potential participants: 5 individuals were younger than 18 years, 13 were entangled with thoughts, memories, and worries for less than 6 months, 7 were receiving psychological or psychiatric treatment, 26 showed extremely severe scores on depression and/or anxiety (they were invited to an alternative study), and 4 showed scores below 15 on the GHQ-12. Of the remaining 21 potential participants, 10 did not respond to emails or did not attend the informative session. In summary, 11 participants met the initial inclusion criteria and attended an interview conducted by the second author. All individuals agreed to participate

and provided informed consent. Participants were remunerated with 25,000 Colombian pesos (approximately 8 US dollars) for concluding the study as compensation for the intensive measurement carried out in the study.

One participant showed significant improvement trends on the outcome measures across the baseline according to the Theil-Sen slope (see Data Analysis section). Accordingly, this participant was excluded from the study. In conclusion, the final sample consisted of 10 participants (4 men, mean age = 23.2, *SD* = 4.24). Table 1 shows demographic data of the participants, details of the problem, and the main triggers to engage in RNT and experiential avoidance strategies. Participants showed a range of affected life areas between 2 and 7 (among 8 areas; *M* = 3.9, *SD* = 1.45). Six participants had received psychological treatment in the past: 2 participants for general anxiety (P3 and P9), 1 participant for social anxiety (P8), 1 for insomnia and couple therapy (P1), and 2 for depression (P6 and P10).

INSERT TABLE 1 ABOUT HERE

2.2. Design and Variables

A non-concurrent, multiple-baseline across participant design was conducted in which the effect of a 2-session, RNT-focused, ACT protocol was evaluated. Participants were randomly assigned to one of five therapists using the web-based tool Research Randomizer (Urbaniak & Plous, 2013). Following recent guidelines, the minimum number of data points for baseline was set at 5 (Kratochwill & Levin, 2014). All participants provided baseline data for 5 to 7 weeks depending on their availability to initiate the intervention. The sessions of the protocol were separated by one month so that the effect of the first session could be analyzed. A 12-week follow-up was conducted after implementing the first session of the protocol.

Dependent variables were divided into primary outcome and secondary measures. As the main aim of this study was to explore the effect of the ACT protocol on treating moderate emotional disorders, the primary outcome measures were scores on emotional symptoms and psychological distress. Secondary measures were scores on process variables such as experiential avoidance, cognitive fusion, pathological worry, repetitive negative thinking, and valued living. All measures were applied on a weekly basis across the study until the 8-week follow-up. A last measurement was conducted at the 12-week follow-up.

The ACT protocol was implemented in an individual format by 5 therapists. All therapists had received at least 40 hours of ACT training and were trained in the application of the protocol for approximately 12 hours. Blinding procedures were not implemented because the study only involved one intervention, and the dependent measures were applied by means of automatic emails through the Internet.

2.3. Primary outcomes

Depression Anxiety and Stress Scales – 21 (DASS-21; Lovibond & Lovibond, 1995; Spanish version by Daza, Novy, Stanley, & Averill, 2002). The DASS-21 is a 21-item, 4-point Likert-type scale (3 = *applied to me very much. or most of the time*; 0 = *did not apply to me at all*) consisting of sentences describing negative emotional states experienced during the last week. It contains three subscales (Depression, Anxiety, and Stress) and has shown good internal consistency and convergent and discriminant validity. The usual cutoffs for the Depression subscale are: 0-4 nonclinical, 5-6 mild, 7-10 moderate, 11-13 severe, and 14 or above extremely severe. Cutoffs for the Anxiety subscale are: 0-3 nonclinical, 4-5 mild, 6-7 moderate, 8-9 severe, and 10 or above extremely severe. Lastly, cutoffs for the Stress subscale are: 0-7 nonclinical, 8-9 mild, 10-12 moderate, 13-16 severe,

and 17 or above extremely severe. The DASS-21 has good psychometric properties (alpha of .93 in the total scale) in Colombian samples and a factor structure consisting of three correlated factors corresponding to the above-mentioned subscales and a general, second-order factor (Ruiz, García-Martín, Suárez-Falcón, & Odriozola-González, 2017). The hierarchical structure of the DASS-21 allows obtaining a global score on emotional symptoms by summing all items. Due to the transdiagnostic nature of the current sample, the DASS-total score was considered a primary outcome. Mean scores on the DASS-total for a nonclinical sample of 894 participants was 19.36 (*SD* = 12.48), whereas for a sample of 245 clinical participants, it was 26.87 (*SD* = 14.53).

General Health Questionnaire-12 (Goldberg & Williams, 1988; Spanish version by Rocha, Pérez, Rodríguez-Sanz, Borrell, & Obiols, 2011). The GHQ-12 is a 12-item, 4-point Likert-type scale that is frequently used as screening for psychological disorders. Respondents are asked to indicate the degree to which they have recently experienced a range of common symptoms of distress, with higher scores reflecting greater levels of psychological distress. The Likert scoring method was used in this study, with scores ranging from 0 to 3 assigned to each of the four response options. The GHQ-12 has good psychometric properties in Colombia (Ruiz, García-Beltrán, & Suárez-Falcón, 2017). Specifically, Cronbach's alpha was .91 both in a nonclinical sample (*N* = 372) and a clinical sample (*N* = 344). Mean scores for the nonclinical and clinical samples were 11.87 (*SD* = 7.47) and 16.54 (*SD* = 7.86), respectively. According to the receiver operating characteristic (ROC) curves, a threshold score of 11/12 was optimal to identify emotional disorders.

2.4. Secondary outcomes

Acceptance and Action Questionnaire – II (AAQ-II; Bond et al., 2011; Spanish version by Ruiz, Langer, Luciano, Cangas, & Beltrán, 2013). The AAQ-II is a general measure of experiential avoidance. It consists of 7 items that are rated on a 7-point Likert-type scale (7 = *always true*; 1 = *never true*). The items reflect unwillingness to experience unwanted emotions and thoughts and the inability to be in the present moment and behave according to value-directed actions when experiencing psychological events that could undermine them. The Spanish version of the AAQ-II has shown good psychometric properties and a one-factor structure in Colombia (Cronbach's alpha of .91 in general population; Ruiz, Suárez-Falcón, et al., 2016). The mean score on the AAQ-II in a Colombian clinical sample (*N* = 277) was 29.67 (*SD* = 10.3), whereas the score in a general nonclinical sample was 22.86 (*SD* = 9.51).

Cognitive Fusion Questionnaire (Gillanders et al., 2014; Spanish version by Ruiz, Suárez-Falcón, Riaño-Hernández, & Gillanders, 2017). The CFQ is a 7-item, 7-point Likert-type scale (7 = *always*; 1 = *never true*), consisting of sentences describing instances of cognitive fusion. This scale has been validated in English for a wide variety of clinical and nonclinical populations. The Spanish version by Ruiz et al. (2017) has shown similar psychometric properties and factor structure to the original version (alpha of .93 in general population). The mean score on the CFQ in a Colombian clinical sample (*N* = 277) was 31.53 (*SD* = 10.86), whereas the score in a general nonclinical sample was 23.80 (*SD* = 9.51).

Penn State Worry Questionnaire – 11 (PSWQ-11; Meyer, Miller, Metzeger, & Borkovec, 1990; Spanish version by Sandín, Chorot, Valiente, & Lostao, 2009). The PSWQ is a 16-item, 5-point Likert (5 = *very typical of me*; 1 = *not at all typical of me*), self-report instrument that was designed to evaluate the permanent and unspecific degree of

worry that characterizes GAD. A reduced, 11-item version was used in this study as recommended by Sandín et al. (2009) because PSWQ reverse-scored items are difficult to understand for Spanish-speaking participants, which worsen the psychometric properties of the instrument. The PSWQ-11's internal consistency is excellent and it shows good test-retest reliability and discriminant validity. The PSWQ-11 has excellent internal consistency in Colombia, with a mean Cronbach's alpha of .95. The mean score in a Colombian nonclinical sample (N = 710) was 27.47 (SD = 10.44), whereas in a clinical sample (N = 335), it was 36.26 (SD = 10.13). A threshold score of 37/38 was considered adequate for identifying severe GAD, whereas a threshold of 32/33 was adequate for moderate GAD.

Perseverative Thinking Questionnaire (PTQ; Ehring et al., 2011). The PTQ is a 15-item, 5-point Likert (4 = *almost always*; 0 = *never*), self-report instrument that was designed to evaluate the tendency to engage in RNT when facing negative experiences or problems. Unlike other measures, the PTQ is a content-independent self-report of RNT. It has shown a hierarchical factor structure, with a higher-order factor representing RNT in general and three lower-order factors: (a) Core features of RNT, (b) Unproductiveness of RNT, and (c) Mental capacity captured by RNT. The PTQ has shown excellent internal consistency, high test-retest reliability, and convergent and predictive validity. As there is no Spanish translation of the PTQ, we back-translated it. Preliminary data from our laboratory indicate that the PTQ possesses excellent internal consistency in Colombia (mean Cronbach's alpha of .96). The mean score in a Colombian nonclinical sample (N = 583) was 22.72 (SD = 14.40), whereas in a clinical sample (N = 126), it was 37.06 (SD = 13.88).

Valuing Questionnaire (VQ; Smout, Davies, Burns, & Christie, 2014; Spanish version by Ruiz, Suárez-Falcón, & Riaño-Hernández, submitted). The VQ is a 10-item, 6-

point Likert (6 = *completely true*; 0 = *not at all true*), self-report instrument designed to assess general valued living during the past week. The VQ has two subscales: Progress (i.e., enactment of values, including clear awareness of what is personally important and perseverance) and Obstruction (i.e., disruption of valued living due to avoidance of unwanted experience and distraction from values). The Spanish version has shown good psychometric properties. Mean scores obtained on the VQ in Colombia for general population were 19.5 (*SD* = 6.43) for Progression and 11.7 (*SD* = 6.88) for Obstruction, whereas mean scores for a clinical sample (*N* = 235) were 17.23 (*SD* = 6.63) and 15.42 (*SD* = 7.12), respectively.

2.5. ACT protocol

The protocol consisted of two 60-min sessions and it was based on the RFT definition of psychological flexibility and formation of the self (Luciano, 2017; Luciano et al., 2012; Törneke et al., 2016). Specifically, the protocol aimed at developing the ability of framing the main ongoing triggers for worry/rumination through a hierarchical relation with the deictic I so as to provoke a reduction of their discriminative avoidant functions and allow the derivation of augmental rules that specify abstract, delayed, probabilistic, and positively reinforcing consequences and behavior in coordination with them. In less technical words, we aimed at developing the ability to discriminate ongoing triggers for worry/rumination, take distance from them (i.e., defusion), and behave according to what is most important in that moment for the individual in the long term (i.e., values).

Table 2 presents the content of the two protocol sessions (a complete description of the protocol in English and Spanish can be found at https://osf.io/ and in the first author's institutional webpage). The aims of Session 1 were: (a) to present the intervention rationale, (b) to identify the main triggers to engage in RNT and other related experiential avoidance

strategies, (c) to promote the realization of the counterproductive effect of engaging in RNT and related experiential avoidance strategies, and (d) to identify the RNT process and defusion training. The aims of Session 2 were: (a) To review the advances in disengaging from RNT and engaging in valued actions, (b) multiple-exemplar training in identifying triggers for RNT and defusing from them, and (c) to identify further valued actions to engage in instead of RNT.

At the end of Session 2, participants were given 5 audio files (30 minutes approximately) in order to practice what was worked on during the protocol on a daily basis. The aim of the exercises was to facilitate the identification of valued actions in which participants could engage during the day and defusing from the triggers for RNT that could surface and impede valued actions.

INSERT TABLE 2 ABOUT HERE

2.6. Procedure

The study was conducted in the Clinical Psychology laboratory of a Colombian university. The procedures of the study were approved by the Internal Ethics Committee. All measures were administered online through Typeform (www.typeform.com). Participants who showed interest in the research and met the initial inclusion criteria were invited to an assessment and informative session led by the second author. If individuals were eligible, the study functioning was presented, and all informed consents were signed (all individuals agreed to participate). Lastly, the first baseline evaluation was conducted.

At the end of baseline collection, participants' scores on the outcome measures (i.e., total scores on the DASS-21 and GHQ-12) were analyzed with the Theil-Sen slope (Sen, 1968; see the Data Analysis section) to explore whether there were any statistically significant improvement tendencies during baseline. After confirming that there were no

significant improvement trends, participants were scheduled in order to initiate the protocol implementation.

2.7. Data analysis

As stated above (see Participants section), we only present the data from participants who did not show a statistically significant improvement trend during the baseline in the primary outcome variables (DASS-Total and GHQ-12). To assess the presence of significant trends in the baseline, the Theil-Sen slope (Sen, 1968; Vannest, Parker, Davis, Soares, & Smith, 2012) was computed before introducing the intervention. The Theil-Sen slope is a nonparametric linear regression slope that does not assume any particular data distribution. It has stronger power/precision than the Koenig and Tukey nonparametric slopes. The Theil-Sen slope approximates the efficiency of linear regression when data meet all parametric assumptions and it significantly exceeds efficiency when data are very non-normal and skewed (Vannest et al., 2012). Accordingly, although it is not very frequent in psychology studies, the Theil-Sen slope is the method of choice in medicine and physical sciences for making decisions with time series data. The Theil-Sen slope was computed using the on-line calculator provided by Vannest, Parker, and Gonen (2011).

Only one participant showed statistically significant improvement trends during baseline (in both the DASS-Total and GHQ-12); therefore, her results were excluded from the current study (the intervention was also implemented for ethical reasons and she showed the stabilization of near zero-level of emotional symptoms during the 3-month follow-up). Two participants showed statistically significant deteriorating trends in the GHQ-12 (Participant 1 and 7). Although this is not a perfect situation, we maintained both participants in the study because the presence of a deteriorating trend does not necessarily

impede establishing the effect of the intervention effect (Hayes, 1981), and the participants showed them only in one of the primary outcome measures.

Following a bottom-up analysis of single-case experimental designs (SCED) (Parker & Vannest, 2012), the results were first graphed and, subsequently, statistical analyses for SCED were selected and computed. In general, although some participants showed considerable variability in emotional symptoms during baseline, the data showed baselines with no significant trends. The presence of variability is not ideal, but it is frequent in clinical datasets. As Hayes (1981) stated, variability is only an issue in this setting if it is so extreme that it would impede the observation of the intervention effect. This was not the case in the current study. At the follow-up, participants' scores usually reached stability at the last three follow-up observations, which are the most relevant ones in terms of clinical significance of the findings. Accordingly, we decided to focus the statistical analysis of each participant on all baseline data and the last three follow-up points (see a similar rationale in Au et al., 2017; Parker & Vannest, 2012).

As we decided to focus the statistical analysis on the last three follow-up points where scores usually reached stability, we selected the JZS+AR Bayesian hypothesis testing for single-subject designs (de Vries & Morey, 2013, 2015). Although the nonparametric *Tau-U* statistic (Parker, Vannest, Davis, & Sauber, 2011) was used to measure the intervention effect in self-registers in the related study by Ruiz, Riaño-Hernández, et al. (2016), it was not used in this study because it would show a ceiling effect due to few overlaps between baseline and follow-up data.

The JZS+AR Bayesian model (de Vries & Morey, 2013) is an adaptation of the JZS *t*-test (Rouder, Speckman, Sun, Morey, & Iverson, 2009) that accounts for the serial dependence typical of single-subject designs with an autoregressive (AR(1)) model. It

provides a *Bayes factor* (B_{ar}), which quantifies the relative evidence in the data for the hypothesis of intervention effect (i.e., the true means of both phases differ: $B_{ar} > 1$) and for the hypothesis of no intervention effect (i.e., the true mean in the baseline equals the true mean in the intervention phase: $B_{ar} < 1$). The Bayes factor can be also seen as the extent to which a rational person should adjust their beliefs, expressed as relative odds, in favor of the hypothesis of intervention effect according to the data (de Vries & Morey, 2013). Bayes factors were interpreted according to the guidelines provided by Jeffreys (1961) and Wagenmakers, Wetzels, Borsboom, and van der Maas (2011): 1 = No evidence of treatment effect; 1-3 = Anecdotal evidence of treatment effect; 3-10 = Substantial evidence of treatment effect; 10-30 = Strong evidence of treatment effect; 30-100 = Very strong evidence of treatment effect; and >100 = Extreme evidence of treatment effect (note that $B_{ar} < 1$ are interpreted in the same way, but favoring the hypothesis of no treatment effect).

One of the distinctive features of Bayesian statistics is that they include prior expectations of the parameters (e.g., the intervention effect). These prior expectations are expressed by prior distributions that receive high density at plausible parameter values and low density at implausible parameter values (Lee, 2004). Prior distributions can be determined based on previous research, expert knowledge, scale boundaries, and statistical considerations (de Vries & Morey, 2013).

To propose prior distributions, the JZS+AR model uses an estimation of two relevant parameters: (a) an effect size of the intervention effect, termed δ, consisting of standardizing the difference in true means between phases; and (b) a parameter for the lag 1 (p) autocorrelation termed b. De Vries and Morey (2013) suggested three prior distributions for δ in which it is located at 0 and follows a Cauchy distribution that differ in width according to a factor termed r (suggested r values of 0.5, 1.0, and 2.0), which is equal to

half the inter-quartile range of the distribution. When $r = 0.5$, $r = 1.0$, or $r = 2.0$, the Cauchy distribution posits higher density at Cohen's *d* values of 0.5, 1.0, and 2.0, respectively. The authors advocated using $r = 1$ by default because in SCED, effect sizes tend to be larger than in group studies (e.g., Beeson & Robey, 2006; Parker & Vannest, 2009). Additionally, the authors suggested three prior distributions for the lag 1 autocorrelation ($b = 1$, $b = 5$, $b = 15$) and advocated for the use of $b = 5$, which reflects the expectation of positive but low autocorrelations, while also considering values of .4 or .5 plausible. This prior distribution is consistent with the literature in SCED showing that autocorrelation in this type of studies is reasonably low (e.g., Parker et al., 2005).

Following the guidelines of de Vries and Morey (2013) and the results obtained in a similar study (Ruiz, Riaño-Hernández, et al., 2016), we selected a value of $r = 1$ for the prior distribution of δ. However, we also conducted a Bayesian sensitivity analysis that investigated the robustness of the results with *r* values of 0.5 and 2.0, which posit higher density in the Cauchy distribution at, respectively, medium and very large effect sizes. Conducting sensitivity analyses is frequently suggested by Bayesian statisticians to investigate whether the results obtained are excessively dependent on the selected prior distribution (Gelman et al., 2014). Regarding the prior distribution of the autocorrelation, we followed the suggestion provided by de Vries and Morey (2013) of choosing $b = 5$. All analyses with the JZS+AR model were conducted in the BayesSingleSub R package (de Vries & Morey, 2015). Due to prior evidence showing the effect of RNT-focused ACT protocols (Ruiz, Riaño-Hernández, et al., 2016), we conducted one-sided Bayes factor testing the hypothesis that $\delta = 0$ against the alternative that $\delta > 0$.

Clinically significant change (CSC) was computed following the suggestion of de Vries, Meijer, van Bruggen, and Morey (2016). These authors discussed the logical

problems of usual approaches to compute CSC based on *p* values such as the reliable change index (RCI) advocated by Jacobson and Truax (1991). Additionally, these approaches are usually computed in SCED by taking into account only one datum for the baseline and intervention phases, which weakens the conclusions reached. The Bayesian approach used in this study required: (a) at least substantial evidence of intervention effect according to the JZS+AR test ($B_{ar} > 3$), and (b) crossing a cutoff point in the last follow-up measure (i.e., the 12-week follow-up) that places the participant closer to the mean of the functional population than to the clinical one. To test the latter criterion, we used the data for clinical and nonclinical participants on each measure as shown in the measures description.

To obtain an overall estimate of the effect size of the intervention, the design-comparable effect size for multiple-baseline designs developed by Pustejovsky, Hedges, and Shadish (2014) was computed. This standardized mean difference effect size for SCED shares the same metric as the Cohen's *d* typically used in group designs, which facilitates the direct comparison and integration through meta-analysis of the results obtained in both types of designs. This *d*-statistic has a formal mathematical development, requires at least three cases for computation, and corrects for small sample bias using Hedges' *g*. It is an extension of the standardized mean difference advocated by Hedges, Pustejovsky, and Shadish (2012, 2013) that uses restricted maximum likelihood estimation and offers the possibility of obtaining the *d*-statistic by controlling for baseline trend and taking into account change in slope. The R package scdhlm was used to compute this *d*-statistic (Pustejovsky, 2016) following the guidelines provided by Valentine, Tanner-Smith, and Pustejovsky (2016). According to the global visual inspection of the dataset, we modelled baselines without trends including both fixed and random effects for level. The treatment

phase was modelled with linear trends with both fixed and random effects at level and slope. We computed the *d*-statistic to estimate the overall effect sizes at the 4-week follow-up (i.e., just before implementing the second session of the protocol) and at the 12-week follow-up. Additionally, the results obtained in the study by Ruiz, Riaño-Hernández, et al. (2016) were reanalyzed with the statistical procedure advocated by Pustejovsky et al. (2014) in order to compare the effect sizes obtained in both studies by observing the 83.4% confidence intervals (CI). If the 83.4% CI do not overlap, a statistically significant difference between the studies can be stated (Cumming, 2012).

3. Results

3.1. Within-participant results

Figure 1 shows the scores' evolution on the primary outcome measures (DASS-21 total and GHQ-12 scores). Visual inspection shows that the ACT protocol was effective in decreasing emotional symptoms to some degree in all participants.

INSERT FIGURE 1 ABOUT HERE

Table 3 shows the effect sizes and B_{ar} on the *JZS+AR* Bayesian model. Seven participants showed at least substantial evidence of intervention effect according to Bayes factors in the DASS-total scores (emotional symptoms): extreme evidence for treatment effect (i.e., $B_{ar} > 100$) was found for P1, P5, and P10; very strong evidence for P4 (i.e., $B_{ar} = 30\text{-}100$), and strong evidence for P3 and P6 (i.e., $B_{ar} = 10\text{-}30$). Anecdotal evidence was found for P2, P7, and P9 (i.e., $B_{ar} = 1\text{-}3$). Nine participants showed at least substantial evidence of intervention effect for the GHQ-12 (psychological distress): extreme evidence was found for P1, P4, P5, P6, and P10; very strong evidence for P8; substantial evidence (i.e., $B_{ar} = 3\text{-}10$) for P3, P7, and P9. Anecdotal evidence of intervention effect was found for P2.

INSERT TABLE 3 ABOUT HERE

With regard to secondary outcomes (process measures), all participants showed at least substantial evidence of intervention effect for the AAQ-II (EA), PSWQ (pathological worry), and PTQ (RNT), 9 participants for the CFQ (cognitive fusion), 8 participants for values obstruction (VQ-Obstruction) and 7 participants for values progress (VQ-Progress). No participant showed evidence of deterioration for any primary or secondary outcome.

Overall, the Bayesian sensitivity analysis conducted with alternative prior distributions showed that the results were relatively robust considering that the Bayes factors did not vary in a way that made the interpretation of the results significantly different (see Appendix A).

Table 3 also shows the results in clinically significant change. With respect to primary outcome measures, 9 and 7 participants showed CSC in GHQ-12 and DASS-Total at the 12-week follow-up, respectively. CSC rates in process measures were also very high for the AAQ-II (9/10), CFQ (9/10), PSWQ (9/10), PTQ (10/10), VQ-Progress (7/10), and VQ-Obstruction (8/10).

3.2. Between-participant results

Figure 2 shows the mean results across participants in all variables of the study. After the intervention, participants showed better scores than nonclinical samples in all measures. Mean scores were below nonclinical level by the 2-week follow-up for cognitive fusion (CFQ), EA (AAQ-II) and emotional symptoms (DASS-21); by the 4-week follow-up for valued living (VQ) and psychological distress (GHQ-12); and by the 6-week follow-up for pathological worry (PSWQ) and RNT (PTQ).

Table 4 shows the mean scores in all measures during baseline, at the 4-, 8-, and 12-week follow-ups, and the *d*-statistics for SCED. According to the *d*-statistic, most of the

changes took place before the implementation of the second session (4-week follow-up), but after it, participants continued improving, although to a lesser degree. The *d*-statistics at the 12-week follow-up were very large both for primary outcome measures (DASS-total: *d* = 2.4, 95% CI [1.5, 3.5]; GHQ-12: *d* = 2.7, 95% CI [1.7, 3.8]) and process measures (AAQ-II: *d* = 1.3, 95% CI [0.6, 2.2]; CFQ: *d* = 2.0, 95% CI [1.0, 3.1]; PSWQ: *d* = 3.1, 95% CI [2.0, 4.3]; PTQ: *d* = 2.5, 95% CI [1.5, 3.5]; VQ-Progress: *d* = 1.5, 95% CI [0.7, 2.5]; VQ-Obstruction: *d* = 1.4, 95% CI [0.8, 2.2]). The model suggested by Pustejovsky et al. (2014) showed convergence problems for the PSWQ and PTQ at the 12-week follow-up, so we decided to report the results with a previous version of the procedure suggested by Hedges et al. (2012, 2013) taking into account only the last three points of follow-up (this procedure is not appropriate for datasets with linear trends).

INSERT FIGURE 2 ABOUT HERE

INSERT TABLE 4 ABOUT HERE

3.3. Comparison with Ruiz, Riaño-Hernández, et al. (2016)

Figure 3 depicts the effect sizes obtained in Ruiz, Riaño-Hernández, et al. (2016) and the ones obtained in the current study for both the 4- and 12-week follow-up. The results of the 4-week follow-up seem more comparable because only one session with a 6-week follow-up was implemented in Ruiz, Riaño-Hernández, et al. Overall, no statistically significant differences were found between the studies although effect sizes for DASS-Total, GHQ-12, and PSWQ were approximately twice as high for the current study. Regarding the comparison with the 12-week follow-up, the effect sizes of this study were significantly higher for DASS-Total, GHQ-12, and PSWQ. Few differences were observed with respect to the remaining measures.

INSERT FIGURE 3 ABOUT HERE

4. Discussion

This study followed the previous RFT analysis by Ruiz, Riaño-Hernández, et al. (2016) that linked the literature about two widely investigated transdiagnostic processes: RNT and EA. These authors suggested that ACT protocols focused on disrupting RNT might show greater efficacy than standard ACT protocols for emotional disorders. To initiate the investigation of this topic, Ruiz, Riaño-Hernández, et al. explored whether a 1-session, RNT-focused, ACT protocol was sufficient to significantly reduce RNT levels in participants suffering from mild to moderate emotional symptoms. The protocol showed large effect sizes in reducing RNT levels and emotional symptoms, which warranted further research. The current study constituted a second step in developing and analyzing the effect of RNT-focused ACT protocols by testing the effect of a 2-session protocol in participants suffering from moderate to severe emotional disorders.

Ten participants who were entangled with thoughts, memories, and worries for at least 6 months and who showed interference in at least two life areas participated in this non-concurrent multiple baseline design. Participants showed moderate and stable levels of emotional symptoms during the 5 to 7 weeks of baseline. Specifically, mean scores on depression and anxiety during baseline were in the moderate range as measured by the DASS-21, whereas the mean score for stress was in the severe range. The mean scores on the GHQ-12 were also in the clinical range. Afterward, participants received a 2-session, RNT-focused, ACT protocol. The protocol sessions were separated by one month to allow the observation of the effect of the first session.

Clinical significant changes were analyzed according to a Bayesian framework for SCED (de Vries et al., 2016), which is stricter than the method suggested by Jacobson and Truax (1991) and does not incur in the logical problems of *p* values. This analysis required

at least substantial evidence of the intervention effect ($B_{ar} > 3$) and final scores in the range of nonclinical participants. Of the 10 participants, 7 and 9 showed CSC in the DASS-21 and GHQ-12, respectively. Most participants also showed CSC in pathological worry (9/10), repetitive thinking (9/10), experiential avoidance (9/10), cognitive fusion (8/10), progress in values (7/10) and obstruction in values (8/10).

The standardized mean difference effect sizes for SCED were very large at both the 4- and 12-week follow-up (DASS-total: d = 1.9 and 2.4; GHQ-12: d = 2.2 and 2.7; PSWQ: d = 2.8 and 3.1; AAQ-II: d = 1.0 and 1.3; CFQ: d = 1.5 and 2.0; PTQ: d = 1.9 and 2.5; VQ-Progress: d = 1.3 and 1.5; VQ-Obstruction: d = 0.8 and 1.4). Importantly, these effect sizes are in the same metric as the between-group Cohen's d, which promotes the comparison and integration of the current findings with empirical evidence regarding the effect of ACT protocols in emotional disorders.

The RNT-focused ACT protocol was implemented in two sessions that were interspersed by one month in order to observe the degree to which the first session of the protocol had a significant effect. Overall, the results showed that, after the first session, participants began to show significant and gradual improvements in all measures. Indeed, the majority of changes were produced before the implementation of the second session of the protocol. This seems to indicate that a 1-session, RNT-focused, ACT protocol might be enough to obtain very large effect sizes for participants with moderate to severe emotional symptoms. However, this finding should be taken with caution because the design employed in this study did not permit observing the long-term effects of the first protocol session. Further research might analyze the long-term effects of a 1-session, RNT-focused, ACT protocol in treating moderate emotional disorders.

The comparison of the results of the current study with the pioneer study by Ruiz, Riaño-Hernández, et al. (2016) is tricky because the designs are not identical. The most reasonable comparison seems to be between the 6-week follow-up of Ruiz, Riaño-Hernández, et al. and the 4-week follow-up of the current study (i.e., before the implementation of the second session). Overall, the effect sizes were larger for the protocol of the current study in the primary outcomes although the difference was not statistically significant. The results of the 12-week follow-up of the current study were significantly higher for the DASS-21, GHQ-12, and PSWQ. This could be due to several reasons related to the study designs: (a) the protocol of the current study contained 2 sessions; (b) the follow-up was longer in the current study, which gave participants more time to show additional improvements; (c) the participants of the current study showed higher scores on emotional symptoms at baseline; and (d) following our clinical experience, the protocol of this study put more emphasis on identifying the triggers for RNT at the top of the hierarchy and in promoting awareness and distancing from the RNT process than the study of Ruiz, Riaño-Hernández, et al. (2016). Accordingly, further studies might explore: (a) the longer term effects of RNT-focused ACT protocols; (b) whether the RNT-focused ACT protocols are better suited for individuals with high emotional symptoms; and (c) whether putting more emphasis on identifying the main triggers for RNT and promoting awareness and defusion from the RNT process may lead to better outcomes.

Although the results of this study are very promising and encourage the development of brief RNT-focused ACT protocols for treating emotional disorders, some limitations are worth noting. Firstly, as opposed to concurrent multiple baseline designs, the non-concurrent multiple baseline design used in this study cannot control for history or maturation effects that might occur simultaneously with the application of the intervention

(Harvey, May, & Kennedy, 2004). To reduce this methodological weakness, we established the inclusion criteria of experiencing emotional difficulties for at least the last 6 months and showing no improvement trends in the primary outcome measures across the 5 to 7 weeks of baseline. Meeting these inclusion criteria makes it less plausible for the intervention effects to be due to history or maturation effects. Additionally, two characteristics of the current study increase the internal and external validity of the experimental design (Kratochwill & Levin, 2014): the replication of similar effects of the RNT-focused ACT protocol in 10 participants and the randomization of the participants to five therapists. Secondly, a general limitation of usual multiple baseline designs is that they lack active control conditions that control for the non-specific effects of therapy. Thirdly, the current study relied solely on self-report measures. Further studies should evaluate the intervention effect including independent clinician-administered assessments. Fourthly, the age range of the participants was relatively narrow (from 19 to 34 years) and most of them were college students, which limits the external validity of the study. Accordingly, further studies should apply RNT-focused ACT protocols to a more diverse adult population. Furthermore, it could be argued that the very large effect sizes found in this study are due to the fact that psychological therapy could have more effect in college students than in more diverse samples of adults suffering from emotional disorders. However, this does not seem to be the case according to the recent meta-analysis by Cuijpers, Cristea, Ebert, et al. (2016), which has shown that psychological treatments of depression in college students have effect sizes comparable to trials conducted among depressed adults. Fifthly, this study does not have a control group, and the sample size is small. Nevertheless, it should be noted that the sequence in analyzing the effect of relatively new treatments usually begins with conducting an open trial with a limited sample (e.g., implementing the intervention to 10-15

participants after conducting only a pretreatment and posttreatment measurement). In this sense, the current study is methodologically better than the usual open trials because it consists of replications of a single-case experimental design where the pretreatment and posttreatment scores do not rely on a single measurement.

The effect sizes obtained in this study are unusually large. For instance, the meta-analysis conducted by Cuijpers, Cristea, Karyotaki, Reijnders, and Huibers (2016) found that cognitive behavior therapy (CBT) yields weighted effect sizes of d = .75 for major depression and d = .80 for GAD. This contrasts with the effect sizes obtained in the current study in terms of emotional symptoms (d = 2.4 and 2.7 for DASS-total and GHQ-12, respectively). With regard to the effect of CBT in GAD worry as measured by the PSWQ, the weighted effect size for CBT found in the meta-analysis by Hanrahan, Field, Jones, and Davey (2013) was 1.81, which contrasts with the effect sizes found in this study (d = 3.1). However, the experimental design of this study cannot explain why the ACT protocol reached these unusually large effect sizes. Following Ruiz, Riaño-Hernández, et al. (2016), this could be due to three main reasons: (a) the protocol simultaneously addressed the three angles to promote psychological flexibility (Törneke et al., 2016) during the sessions; (b) the protocol was focused on disrupting the first and most pervasive reaction to triggers (i.e., worry/rumination), which extends discomfort and supports further EA strategies; and (c) the protocol emphasized identifying and working with the trigger for RNT at the top of the self-hierarchy.

In conclusion, this study constitutes an initial and very promising step in the analysis of brief RNT-focused ACT protocols for the treatment of emotional disorders. The results of this study are more generalizable to individuals with moderate to severe emotional symptoms. Further studies might conduct randomized controlled trials to

compare the effect of the ACT protocol with waiting-list control conditions or brief versions of empirically established treatments such as behavioral activation or cognitive therapy.

Statement. This article reports all the variables, conditions, and results collected in the study described. The primary and secondary outcomes were decided when designing the study and prior to data collection. The recruitment of participants was extended during three months, after which no further participants were accepted. The ACT protocol used in this study will be available in Spanish and English at https://osf.io/. The Spanish translations of the self-reports used in this study will also be available at the institutional webpage of the first author and at https://osf.io/. The dataset of the study will be available at https://osf.io/ or as supplementary files if the journal accepts this."

References

Au, T. M., Sauer-Zavala, S., King, M. W., Petrocchi, N., Barlow, D. H., & Litz, B. T. (2017). Compassion-based therapy for trauma-related shame and post-traumatic stress: Initial evaluation using a multiple baseline design. *Behavior Therapy, 48*, 207-221.

Barlow, D. H., Farchione, T. J., Fairholme, C. P., Ellard, K. K., Boisseau, C. L., Allen, L. B., & Ehrenreich-May, J. T. (2010). *Unified protocol for transdiagnostic treatment of emotional disorders: Therapist guide*. New York: Oxford University Press.

Beeson, P. M., & Robey, R. R. (2006). Evaluating single-subject treatment research: Lessons learned from the aphasia literature. *Neuropsychology Review, 16*, 161-169.

Bond, F. W., Hayes, S. C., Baer, R. A., Carpenter, K. M., Guenole, N., Orcutt, H. K., . . . Zettle, R. D. (2011). Preliminary psychometric properties of the Acceptance and Action Questionnaire – II: A revised measure of psychological inflexibility and experiential avoidance. *Behavior Therapy, 42*, 676-688.

Borkovec, T. D. (1994). The nature, functions and origins of worry. In G. Davey, & F. Tallis (Eds.), *Worrying: Perspectives on theory, assessment and treatment* (pp. 5-33). Sussex, England: Wiley & Sons.

Boulanger, J. L., Hayes, S. C., & Pistorello, J. (2010). Experiential avoidance as a functional contextual concept. In A. Kring, & D. Sloan (Eds.), *Emotion regulation and psychopathology* (pp. 107-134). New York: Guilford Press.

Cuijpers, P., Cristea, I. A., Ebert, D. D., Koot, H. M., Auerbach, R. P., Bruffaerts, R., & Kessler, R. C. (2016). Psychological treatment of depression in college students: A meta-analysis. *Depression and Anxiety, 33*, 400-414.

Cuijpers, P., Cristea, I. A., Karyotaki, E., Reijnders, M., & Huibers, M. J. (2016). How effective are cognitive behavior therapies for major depression and anxiety disorders? A meta-analytic update of the evidence. *World Psychiatry, 15*, 245-258.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge.

Daza, P., Novy, D. M., Stanley, M., & Averill, P. (2002). The Depression Anxiety Stress Scale-21: Spanish translation and validation with a Hispanic sample. *Journal of Psychopathology and Behavioral Assessment, 24*, 195-205.

de Vries, R. M., Meijer, R. R., van Bruggen, V., & Morey, R. D. (2016). Improving the analysis of routine outcome measurement data: what a Bayesian approach can do - for you. *International Journal of Methods in Psychiatric Research, 25*, 155-167.

de Vries, R. M., & Morey, R. D. (2013). Bayesian hypothesis testing for single-subject designs. *Psychological Methods, 18*, 165-185.

de Vries, R. M., & Morey, R. D. (2015). A tutorial on computing Bayes factors for single-subject designs. *Behavior Therapy, 46*, 809-823.

Ehring, T., & Watkins, E. R. (2008). Repetitive negative thinking as a transdiagnostic process. *International Journal of Cognitive Therapy*, *1*, 192-205.

Ehring, T., Zetsche, U., Weidacker, K., Wahl, K., Schönfeld, S., & Ehlers, A. (2011). The Perseverative Thinking Questionnaire (PTQ): Validation of a content-independent measure of repetitive negative thinking. *Journal of Behavior Therapy and Experimental Psychiatry, 42*, 225-232.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (Vol. 2). Boca Raton, FL: CRC press.

Gillanders, D. T., Bolderston, H., Bond, F. W., Dempster, M., Flaxman, P. E., Campbell, L., … Remington, B. (2014). The development and initial validation of the Cognitive Fusion Questionnaire. *Behavior Therapy, 45*, 83-101.

Goldberg, D., & Williams, P. (1988). *A user's guide to the General Health Questionnaire*. Windsor, UK: NFER-Nelson.

Hanrahan, F., Field, A. P., Jones, F. W., & Davey, G. C. (2013). A meta-analysis of cognitive therapy for worry in generalized anxiety disorder. *Clinical Psychology Review, 33*, 120-132.

Harvey, A. G., Watkins, E. R., Mansell, W., & Shafran, R. (2004). *Cognitive behavioural processes across disorders*. Oxford, UK: Oxford University Press.

Harvey, M. T., May, M. E., & Kennedy, C. H. (2004). Nonconcurrent multiple baseline designs and the evaluation of educational systems. *Journal of Behavioral Education, 13*, 267-276.

Hayes, S. C. (1981). Single case experimental design and empirical clinical practice. *Journal of Consulting and Clinical Psychology, 49*, 193-211.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational frame theory. A post-Skinnerian account of human language and cognition*. New York: Kluwer Academic Press.

Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (1999). *Acceptance and commitment therapy. An experiential approach to behavior change*. New York: Guilford Press.

Hayes, S. C., Wilson, K. G., Gifford, E. V., Follette, V. M., & Strosahl, K. D. (1996). Experiential avoidance and behavioral disorders: A functional dimensional approach to diagnosis and treatment. *Journal of Consulting and Clinical Psychology, 64*, 1152.

Hedges, L. G., Pustejovsky, J. E., & Shadish, W. R. (2012). A standardized mean difference effect size for single-case designs. *Research Synthesis Methods, 3*, 224-239.

Hedges, L. G., Pustejovsky, J. E., & Shadish, W. R. (2013). A standardized mean difference effect size for multiple baseline designs across individuals. *Research Synthesis Methods, 4*, 324-341.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*, 12-19.

Jeffreys, H. (1961). *Theory of probability*. Oxford, England: Oxford University Press.

Kashdan, T. B., & Roberts, J. E. (2007). Social anxiety, depressive symptoms, and post-event rumination: Affective consequences and social contextual influences. *Journal of Anxiety Disorders*, *21*, 284-301.

Kratochwill, T. R., & Levin, J. R. (2014). *Single-case intervention research: Methodological and statistical advances*. Washington, DC: American Psychological Association.

Kring, A. M. (2008). Emotion disturbances as transdiagnostic processes in psychopathology. In M. Lewis, J. M. Haviland-Jones, & L. Feldman (Eds.), *Handbook of emotions* (pp. 691-708). New York: Guilford Press.

Lee, P. M. (2004). *Bayesian statistics: An introduction* (3rd ed.). New York, NY: Wiley.

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, *33*, 335-343.

Luciano, C. (2017). The self and responding to the own's behavior. Implications of coherence and hierarchical framing. *International Journal of Psychology and Psychological Therapy, 17*, 267-275.

Luciano, C., & Hayes, S. C. (2001). Trastorno de evitación experiencial [Experiential avoidance disorder]. *International Journal of Clinical and Health Psychology, 1*, 109-157.

Luciano, C., Ruiz, F. J., Vizcaíno-Torres, R., Sánchez, V., Gutiérrez-Martínez, O., & López-López, J. C. (2011). A relational frame analysis of defusion interactions in acceptance and commitment therapy. A preliminary and quasi-experimental study with at-risk adolescents. *International Journal of Psychology and Psychological Therapy, 11*, 165-182.

Luciano, C., Valdivia-Salas, S., & Ruiz, F. J. (2012). The self as the context for rule-governed behavior. In L. McHugh & I. Stewart (Eds.), *The self and perspective taking: Research and applications* (pp. 143-160). Oakland, CA: Context Press.

Meyer, T. J., Miller, M. L., Metzeger, R. L., & Borkovec, T. D. (1990). Development and validation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy, 28*, 487-495.

Michael, T., Halligan, S. L., Clark, D. M., & Ehlers, A. (2007). Rumination in posttraumatic stress disorder. *Depression and Anxiety*, *24*, 307-317.

Nolen-Hoeksema, S. (2000). The role of rumination in depressive disorders and mixed anxiety/depressive symptoms. *Journal of Abnormal Psychology*, *109*, 504-511.

Nolen-Hoeksema, S. (2004). The response styles theory. In C. Papageorgiou, & A. Wells (Eds.), *Depressive rumination*: *Nature, theory and treatment* (pp. 107-123). Chichester, UK: Wiley.

Parker, R. I., Brossart, D. F., Vannest, K. J., Long, J. R., Garcia de Alba, R., Baugh, F. G., & Sullivan, J. R. (2005). Effect sizes in single case research: How large is large? *School Psychology Review, 34*, 116-132.

Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy, 40*, 357–367.

Parker, R. I., & Vannest, K. J. (2012). Bottom-up analysis of single-case research designs. *Journal of Behavioral Education, 21*, 254-265.

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy, 42*, 284-299.

Pustejovsky, J. E. (2016). scdhlm: A web-based calculator for between-case standardized mean differences (Version 0.3.1) [Web application]. Retrieved from: https://jepusto.shinyapps.io/scdhlm

Pustejovsky, J. E., Hedges, L. V., Shadish, W. R. (2014). Design-comparable effect sizes in multiple baseline designs: A general modeling framework. *Journal of Educational and Behavioral Statistics, 39*, 368-393.

Rocha, K., Pérez, K., Rodríguez-Sanz, M., Borrell, C., & Obiols, J. E. (2011). Propiedades psicométricas y valores normativos del General Health Questionnaire (GHQ-12) en población española [Psychometric properties and normative scores of the General Health Questionnaire (GHQ-12) in Spanish population]. *International Journal of Clinical and Health Psychology, 11*, 125-139.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t-tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*, 225-237.

Ruiz, F. J. (2010). A review of acceptance and commitment therapy (ACT) empirical evidence: Correlational, experimental psychopathology, component and outcome studies. *International Journal of Psychology and Psychological Therapy, 10*, 125-162.

Ruiz, F. J., García-Beltrán, D. M., & Suárez-Falcón, J. C. (2017). General Health Questionnaire-12 validity in Colombia and factorial equivalence between clinical and nonclinical participants. *Psychiatry Research, 256*, 53-58.

Ruiz, F. J., García-Martín, M. B., Suárez-Falcón, J. C., & Odriozola-González, P. (2017). The hierarchical factor structure of the Depression Anxiety and Stress Scale-21. *International Journal of Psychology and Psychological Therapy, 17*, 97-105.

Ruiz, F. J., Langer, A. I., Luciano, C., Cangas, A. J., & Beltrán, I. (2013). Measuring experiential avoidance and psychological inflexibility: The Spanish translation of the Acceptance and Action Questionnaire. *Psicothema*, *25,* 123-129.

Ruiz, F. J., & Perete, L. (2015). Application of a relational frame theory account of psychological flexibility in young children. *Psicothema*, *27*, 114-119.

Ruiz, F. J., Riaño-Hernández, D., Suárez-Falcón, J. C., & Luciano, C. (2016). Effect of a one-session ACT protocol in disrupting repetitive negative thinking: A randomized multiple-baseline design. *International Journal of Psychology and Psychological Therapy, 16*, 213-233.

Ruiz, F. J., Suárez-Falcón, J. C., Cárdenas-Sierra, S., Durán, Y., Guerrero, K., & Riaño-Hernández, D. (2016). Psychometric properties of the Acceptance and Action Questionnaire – II in Colombia. *The Psychological Record, 66*, 429-437.

Ruiz, F. J., Suárez-Falcón, J. C., Riaño-Hernández, D., & Gillanders, D. (2017). Psychometric properties of the Cognitive Fusion Questionnaire in Colombia. *Revista Latinoamericana de Psicología, 49*, 80-87.

Sandín, B., Chorot, P., Valiente, R. M., & Lostao, L. (2009). Validación española del cuestionario de preocupación PSWQ: estructura factorial y propiedades psicométrica [Spanish validation of the PSWQ Worry Questionnaire: Factor structure and psychometric properties]. *Revista de Psicopatología y Psicología Clínica, 14*, 107-122.

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association, 63*, 1379-1389.

Smout, M., Davies, M., Burns, N., & Christie, A. (2014). Development of the Valuing Questionnaire (VQ). *Journal of Contextual Behavioral Science, 3*, 164-172.

Tate, R. L., Perdices, M., Rosenkoetter, U., McDonald, S., Togher, L., Shadish, W., ... Sampson, M. (2016). The Single-Case Reporting Guideline In BEhavioural Interventions (SCRIBE) 2016: Explanation and elaboration. *Archives of Scientific Psychology, 4*, 10-31.

Törneke, N., Luciano, C., Barnes-Holmes, Y., & Bond, F. W. (2016). Relational frame theory and three core strategies in understanding and treating human suffering. In R. D. Zettle, S. C. Hayes, D. Barnes-Holmes, & A. Biglan (Eds.), *The Wiley Handbook of Contextual Behavioral Science* (pp. 254-272). Oxford, UK: Wiley-Blackwell.

Urbaniak, G. C., & Plous, S. (2013). Research Randomizer, Version 4.0 (Web-based application. Available at www.randomizer.org

Valentine, J. C., Tanner-Smith, E. E., & Pustejovsky, J. E. (2016). Between-case standardized mean difference effect sizes for single-case designs: A primer and

tutorial using the scdhlm web application. Oslo, Norway: The Campbell Collaboration.

Vannest, K. J., Parker, R. I., Davis, J. L., Soares, D. A., & Smith, S. L. (2012). The Theil-Sen slope for high stakes decision from progress monitoring. *Behavioral Disorders, 37*, 271-280.

Vannest, K. J., Parker, R. I., & Gonen, O. (2011). Single-case research: Web-based calculator for SCR analysis, Version 1.0 (Web-based application). College Station: Texas A&M University. Available from www.singlecaseresearch.org

Wagenmakers, E. J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of Psi: Comment on Bem (2011). *Journal of Personality and Social Psychology, 3*, 426-432.

Watkins, E. R. (2008). Constructive and unconstructive repetitive thought. *Psychological Bulletin, 134*, 163-206.

Watkins, E. R. (2016). *Rumination-focused cognitive-behavioral therapy for depression*. New York: Guilford Press.

Wells, A. (2009). *Metacognitive therapy for anxiety and depression*. New York: Guilford Press.

Table 1

Demographical Data, Problem Details, and Scores on Depression and Generalized Anxiety Disorder

	Sex	Age	Study/ Employ	Life areas affected	Main triggers for RNT	Experiential avoidance strategies
P1	M	34	Technician	7	Fear of failure; Feeling of incomprehension	Worry/Rumination, listening to music, cleaning, exercising, drinking alcohol, sleeping
P2	F	19	Undergr.	4	Fear of being rejected; Why do I disappoint my family?	Worry/Rumination, sleeping, watching videos, talking to friends and family, doing activities that look productive
P3	F	26	Graduate	4	Fear of failure at job; Why am I frustrated?	Worry/Rumination, distraction, sleeping, eating
P4	F	22	Undergr.	5	Fear of failure; Why am I not doing things well?	Worry/Rumination, calling boyfriend, sleeping, talking to family, chatting with friends
P5	F	22	Undergr.	4	Fear of failure; Why I can't do things well?	Worry/Rumination, watching films, playing with family, smoking, listening to music
P6	F	21	Undergr.	3	Fear of failure; What caused my depression?	Worry/Rumination, studying, exercising, sleeping, walking, drinking coffee, talking to friends, resting
P7	M	21	Technician	2	Fear of disappointing other people; Why am I not doing something important?	Worry/Rumination, Working, listening to music, calling friends.
P8	M	24	Undergr.	4	Fear of failure; Why didn't anyone support me?	Worry/Rumination
P9	M	21	Undergr.	2	Fear of failure; Why did that person leave me?	Worry/Rumination, reading, talking to friends, psychedelic drugs, listening to music, studying.
P10	F	22	Undergr.	4	Fear of failure; Why do I think so much?	Worry/Rumination, reading, listening to music, watching TV, talking to friends and mother, playing with the dog.

Table 2
Summary of the ACT protocol

Phase	Aims	Therapeutic interactions
Session 1 (60 min)	1. Presentation of the intervention rationale	▪ Work proposal: develop the skill to identify entanglement with our thoughts and learn to focus on what is really important in our lives.
	2. Identification of the main triggers to initiate worry/rumination and other experiential avoidance strategies related to them	▪ Worry begins when fear of a future event appears: "What is the fear that is the "daddy" of all your fears?" ▪ Rumination begins when one needs an explanation about something that happened: "What is the explanation you need that is the "daddy" of all of them?" ▪ Explore the consequences of worry/rumination and experiential avoidance strategies connected to them.
	3. Promoting discrimination of the counterproductive effect of engaging in worry/rumination and other experiential avoidance strategies	▪ Socratic dialogue: (a) In which direction are you going when you worry/ruminate and you try to avoid/control your thoughts? (b) Are they helpful at the short term? (c) And at the long term? and (d) Are the thoughts even stronger than before? ▪ Pushing triggers away metaphor. The experimenter writes the participant's triggers on a piece of paper and puts it near the participant's face. When participants begin to push the piece of paper away with their hands, the experimenter resists. ▪ Questions: (a) How much strength do your thoughts have when you push? (b) Can you do anything important while pushing? (c) How much stronger would they be if you pushed one more year? (d) And 5 more years?
	4. Identification of the RNT process and defusion training	▪ Go around exercise: While reading a book, the therapist shows a trigger for RNT on a card and the participant stops reading and begins the RNT process going around a chair in circles. Every time the participant makes a loop, she says the next thought of the chain and chooses making another loop (the same process is repeated 10 times). Then, the participant is invited to read again and choose just to observe the triggers for RNT and go back to reading without entangling with them.
Session 2 (60 min)	1. Review	▪ Exploration of RNT and valued actions during the past week.
	2. Multiple-exemplar training in identifying triggers for RNT and defusion exercises	▪ Centering/defusion exercise: focusing on breathing and the noises around while noticing who is choosing to do that. Then, practicing putting ongoing thoughts in balloons and letting them be. ▪ Free association exercise (based on Wells, 2009): The therapist reads 12 words separated by 7 s. The participant has to notice what thought comes to her mind and chooses between entangling with it or putting it in a balloon and letting it be. ▪ Daydreaming and worrying exercise (based on Wells, 2009): The participant is invited to daydream for 2 minutes. Each 20 s, the therapist asks the participant to notice what she was thinking and how she could choose between following or stopping the process. The same process was repeated with worry. ▪ Eye-contact exercise (Hayes et al., 1999): The participant and therapist look into each other's eyes for 2 minutes while noticing every thought and emotion and choosing to continue. ▪ Subtracting exercise: The participant is invited to count backwards from 100 to 0 by subtracting 3, while noticing the sensation of being wrong and choosing to continue. ▪ Writing with the nondominant hand: The participant writes for 3 minutes with her nondominant hand while noticing the discomfort, not entangling with it and choosing to continue. ▪ Pink elephant exercise (based on Hayes et al., 1999): The participant is invited to avoid thinking about a pink elephant.
	3. Identification of valued actions	▪ The therapist asks the participant to list some valued actions that the participant could do instead of being entangled with her thoughts ("things that make her proud at the end of the day").

Table 3

Results in the JZS+AR Analysis and Clinical Significant Change for each Participant and Measure with a Prior Distribution with $r = 1$

		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	%
PRIMARY OUTCOME MEASURES												
DASS – Total (emotional symptoms)	δ	6.50	1.02	2.13	3.69	6.15	2.80	0.72	1.80	0.81	9.37	
	B_{ar}	**>100**	2.31	**11.7**	**54.3**	**>100**	**29.3**	1.45	**7.57**	1.65	**>100**	70%
	CSC	**YES**	NO	**YES**	**YES**	**YES**	**YES**	NO	**YES**	NO	**YES**	70%
GHQ-12 – Total (psychological distress)	δ	6.50	1.05	1.89	4.49	7.46	4.08	1.49	2.95	0.37	4.91	
	B_{ar}	**>100**	2.51	**8.62**	**>100**	**>100**	**>100**	**4.97**	**33.5**	**3.55**	**>100**	90%
	CSC	**YES**	NO	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	90%
SECONDARY OUTCOME MEASURES												
PSWQ (pathological worry)	δ	14.8	2.62	1.85	5.77	8.47	1.88	2.60	6.00	5.71	10.5	
	B_{ar}	**>100**	**15.3**	**8.77**	**>100**	**>100**	**8.23**	**20.5**	**>100**	**>100**	**>100**	100%
	CSC	**YES**	**YES**	NO	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	90%
AAQ-II (experiential avoidance)	δ	3.22	2.03	2.64	6.06	3.77	4.26	1.87	2.60	1.36	1.43	
	B_{ar}	**46.8**	**10.0**	**22.9**	**>100**	**87.7**	**>100**	**9.23**	**23.0**	**3.62**	**4.44**	100%
	CSC	**YES**	**YES**	NO	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	90%
CFQ (cognitive fusion)	δ	7.61	1.13	2.65	5.78	3.06	3.94	1.15	2.23	7.34	5.48	
	B_{ar}	**>100**	2.77	**24.1**	**>100**	**37.9**	**>100**	**3.05**	**14.3**	**>100**	**>100**	90%
	CSC	**YES**	NO	NO	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	80%
PTQ (perseverative thinking)	δ	12.7	1.85	1.42	12.8	6.42	1.51	1.17	7.00	5.23	6.24	
	B_{ar}	**>100**	**7.20**	**4.62**	**>100**	**>100**	**5.31**	**3.16**	**>100**	**>100**	**>100**	100%
	CSC	**YES**	**YES**	NO	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	90%
VQ – Progress values	δ	3.08	2.15	3.70	1.80	2.36	1.29	0.12	1.25	0.07	0.55	
	B_{ar}	**36.3**	**11.3**	**74.7**	**5.77**	**17.9**	**3.56**	0.41	**3.41**	0.42	1.09	70%
	CSC	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	NO	**YES**	NO	NO	70%
VQ – Obstruction vales	δ	2.94	1.48	3.64	0.92	3.64	2.13	0.76	2.70	2.57	3.38	
	B_{ar}	**36.0**	**4.91**	**69.5**	1.89	**44.8**	**11.5**	1.53	**22.9**	**15.0**	**72.8**	80%
	CSC	**YES**	**YES**	**YES**	NO	**YES**	**YES**	NO	**YES**	**YES**	**YES**	80%

Note. B_{ar} = Bayes Factors of the JZS+AR model. $B_{ar} > 1$ supports the hypothesis of intervention effect. $B_{ar} > 3$ are in bold to highlight where at least substantial evidence of treatment effect was found.

Table 4

Means and Standard Deviations in each Self-Report Measure at Baseline, Posttreatment, 6-Week and 12-Week Follow-Ups, and Effect Sizes at the 4-Week, 8-Week, and 12-Week Follow-Ups

	Baseline	4-week F-U	8-week F-U	12-week F-U	*d*-statistic for SCED 4-w F-U		*d*-statistic for SCED 12-w F-U	
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*d (SE)*	*95% CI*	*d (SE)*	*95% CI*
PRIMARY OUTCOMES								
DASS-Total: Emotional symptoms	27.71 (5.73)	10.70 (6.68)	7.30 (6.93)	7.80 (6.05)	1.93 (0.52)	[1.0, 3.0]	2.44 (0.53)	[1.5, 3.5]
DASS – Depression	7.67 (3.10)	3.00 (2.11)	1.60 (2.12)	1.60 (2.01)	1.10 (0.46)	[0.3, 2.0]	1.51 (0.49)	[0.7, 2.5]
DASS – Anxiety	5.96 (2.86)	2.10 (2.23)	2.00 (2.75)	1.50 (2.27)	1.02 (0.34)	[0.4, 1.7]	1.21 (0.37)	[0.6, 2.0]
DASS – Stress	13.95 (3.05)	5.60 (3.50)	3.70 (2.83)	4.70 (3.16)	1.72 (0.54)	[0.7, 2.8]	2.22 (0.52)	[1.3, 3.3]
GHQ-12: Psychological distress	19.16 (3.10)	7.70 (3.53)	6.30 (4.83)	5.50 (3.60)	2.24 (0.54)	[1.2, 3.3]	2.68 (0.56)	[1.7, 3.8]
SECONDARY OUTCOMES (PROCESS MEASURES)								
PSWQ: Pathological worry	42.11 (2.75)	28.80 (10.50)	22.90 (7.98)	22.90 (8.14)	2.81 (0.87)	[1.2, 4.6]	3.14 (0.37)	[2.0, 4.3]
AAQ-II: Experiential avoidance	28.77 (7.67)	18.70 (9.94)	15.50 (7.37)	14.50 (7.07)	0.98 (0.42)	[0.2, 1.8]	1.32 (0.40)	[0.6, 2.2]
CFQ: Cognitive fusion	35.24 (5.74)	20.30 (9.70)	16.50 (8.49)	15.40 (7.83)	1.52 (0.58)	[0.5, 2.7]	2.01 (0.55)	[1.0, 3.1]
PTQ: Perseverative thinking	39.26 (4.99)	22.60 (11.85)	17.00 (12.91)	14.70 (11.06)	1.87 (0.73)	[0.5, 3.3]	2.51 (0.28)	[1.5, 3.5]
VQ: Valued living - Progress	15.69 (4.22)	22.20 (4.42)	22.40 (6.60)	23.70 (5.14)	1.27 (0.39)	[0.6, 2.1]	1.54 (0.46)	[0.7, 2.5]
VQ: Valued living - Obstruction	16.69 (5.45)	11.00 (5.52)	6.40 (4.95)	8.20 (4.87)	0.81 (0.38)	[0.1, 1.6]	1.41 (0.37)	[0.8, 2.2]

Note. AAQ-II = Acceptance and Action Questionnaire – II; CFQ = Cognitive Fusion Questionnaire; DASS = Depression, Anxiety, and Stress Scales-21; GHQ-12 = General Health Questionnaire-12; PSWQ = Penn State Worry Questionnaire; PTQ = Perseverative Thinking Questionnaire; VQ = Valuing Questionnaire.

Figure 1. Scores on emotional symptoms (DASS-21) and psychological distress (GHQ-12) for every participant.

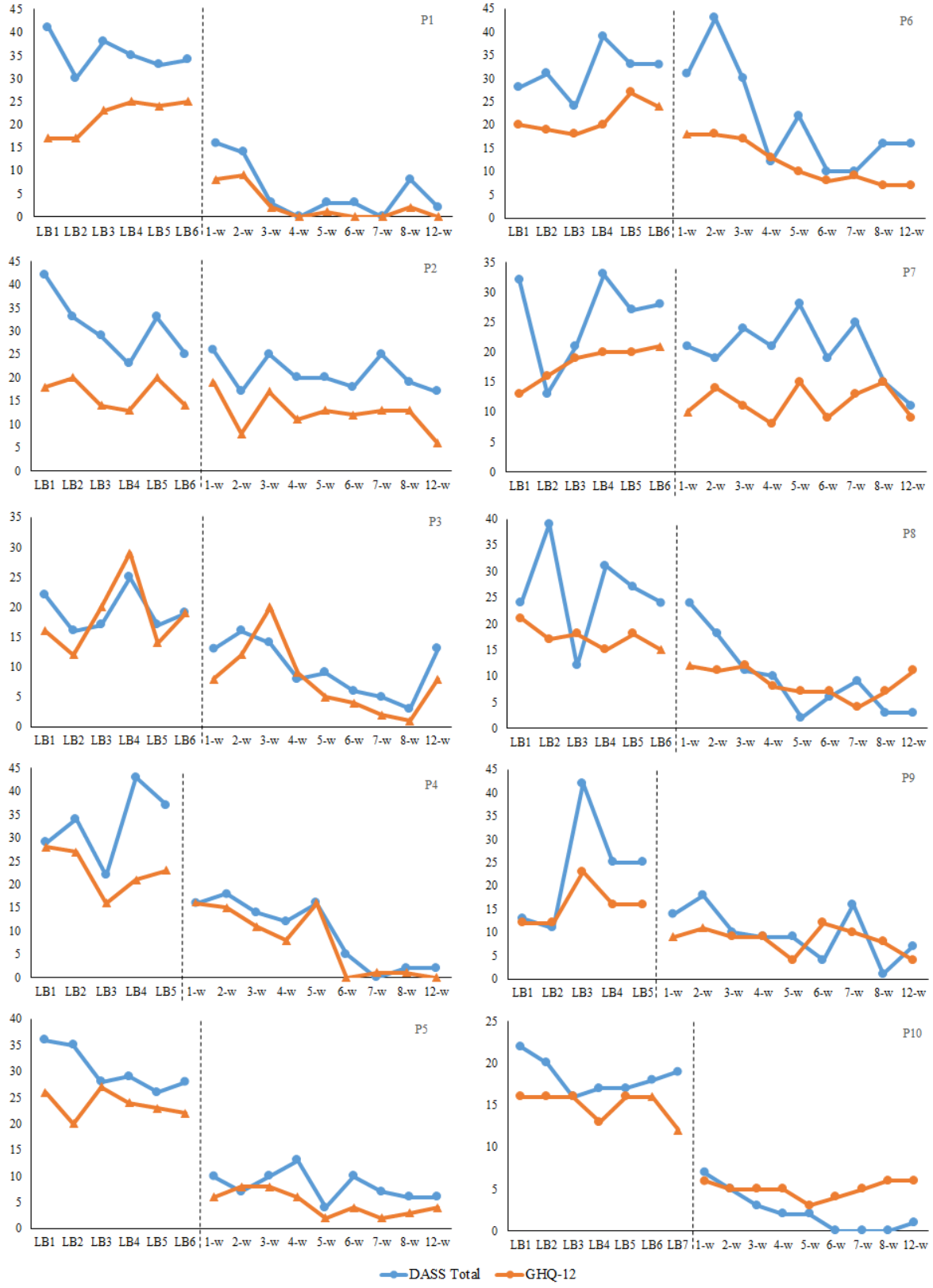


Figure 2. Mean scores' evolution on outcome and process measures. Bars represent 95% confidence intervals

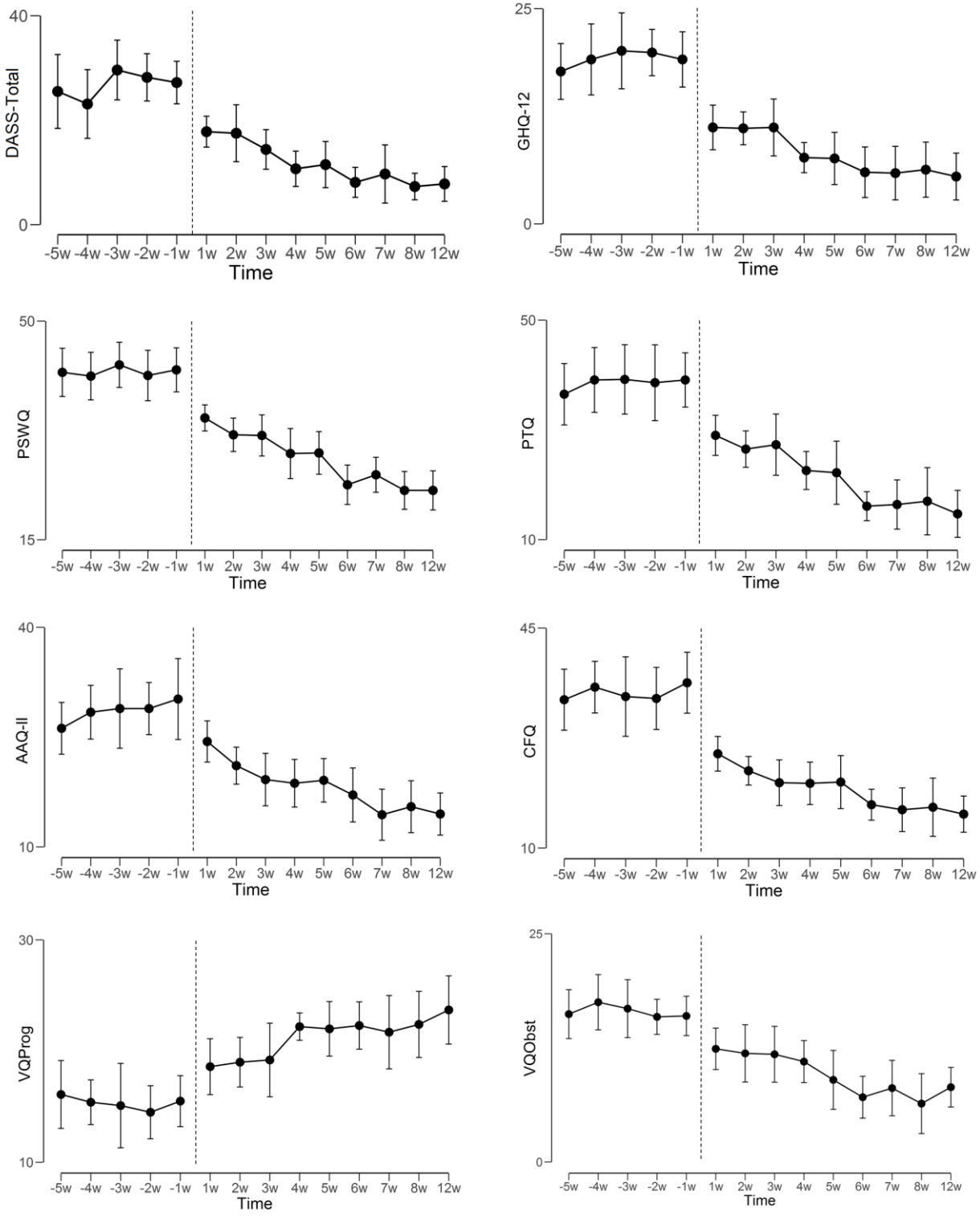
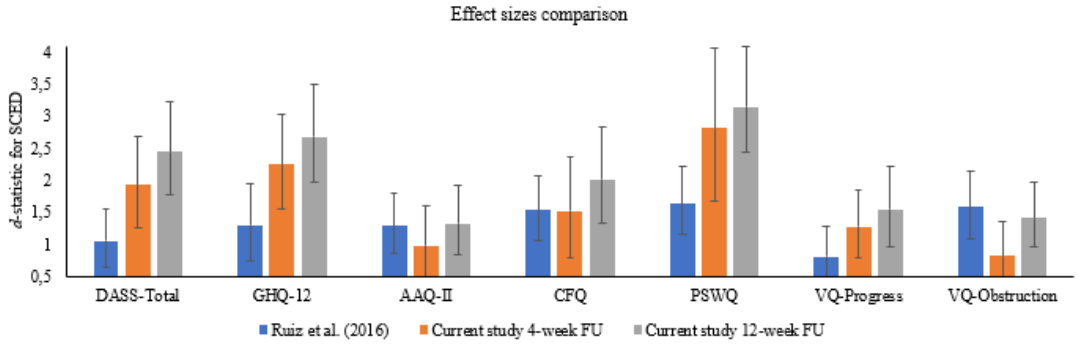


Figure 3. Comparison of effect sizes between the current study and the study by Ruiz et al. (2016)



Appendix A. Results in the JZS+AR Analysis for each Participant and Measure with Prior Distributions of $r = 0.5$, $r = 1.0$, and $r = 2.0$

			P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
PRIMARY OUTCOMES												
DASS – Total	$r = 0.5$	δ	6.37	0.71	1.76	3.31	6.03	2.48	0.49	1.42	0.54	9.39
		B_{ar}	**>100**	2.14	**8.25**	**25.9**	**>100**	**17.1**	1.52	**5.29**	1.63	**>100**
	$r = 1$	δ	6.50	1.02	2.13	3.69	6.15	2.80	0.72	1.80	0.81	9.37
		B_{ar}	**>100**	2.31	**11.7**	**54.3**	**>100**	**29.3**	1.45	**7.57**	1.65	**>100**
	$r = 2$	δ	6.68	1.30	2.46	3.91	6.39	3.12	0.95	2.12	1.06	9.49
		B_{ar}	**>100**	2.10	**15.9**	**72.2**	**>100**	**45.4**	1.17	**8.92**	1.34	**>100**
GHQ-12 – Total	$r = 0.5$	δ	6.07	0.74	1.55	4.31	7.33	3.88	1.13	2.64	0.97	4.85
		B_{ar}	**>100**	2.31	**6.12**	**47.8**	**>100**	**57.7**	**3.86**	**19.7**	2.89	**>100**
	$r = 1$	δ	6.50	1.05	1.89	4.49	7.46	4.08	1.49	2.95	0.37	4.91
		B_{ar}	**>100**	2.51	**8.62**	**>100**	**>100**	**>100**	**4.97**	**33.5**	**3.55**	**>100**
	$r = 2$	δ	6.36	1.29	2.20	4.80	7.49	4.28	1.76	3.18	1.62	5.06
		B_{ar}	**>100**	2.20	**10.2**	**>100**	**>100**	**>100**	**4.78**	**45.9**	**3.43**	**>100**
PROCESS MEASURES												
PSWQ (Pathological worry)	$r = 0.5$	δ	14.8	2.35	1.56	5.66	8.57	1.54	2.35	5.90	5.62	10.5
		B_{ar}	**>100**	**10.5**	**6.31**	**>100**	**>100**	**5.99**	**13.7**	**>100**	**>100**	**>100**
	$r = 1$	δ	14.8	2.62	1.85	5.77	8.47	1.88	2.60	6.00	5.71	10.5
		B_{ar}	**>100**	**15.3**	**8.77**	**>100**	**>100**	**8.23**	**20.5**	**>100**	**>100**	**>100**
	$r = 2$	δ	14.7	2.98	2.19	6.01	8.70	2.19	2.91	6.17	5.90	10.5
		B_{ar}	**>100**	**23.1**	**10.6**	**>100**	**>100**	**9.54**	**27.3**	**>100**	**>100**	**>100**
AAQ-II (Experiential avoidance)	$r = 0.5$	δ	2.92	1.70	2.36	5.99	3.60	4.09	1.53	2.27	1.03	1.08
		B_{ar}	**23.5**	**7.01**	**14.4**	**>100**	**53.2**	**79.2**	**6.03**	**12.2**	**3.01**	**3.51**
	$r = 1$	δ	3.22	2.03	2.64	6.06	3.77	4.26	1.87	2.60	1.36	1.43
		B_{ar}	**46.8**	**10.0**	**22.9**	**>100**	**87.7**	**>100**	**9.23**	**23.0**	**3.62**	**4.44**
	$r = 2$	δ	3.47	2.34	2.99	6.29	4.05	4.52	2.18	2.87	1.68	1.74
		B_{ar}	**63.3**	**11.6**	**33.8**	**>100**	**>100**	**>100**	**10.5**	**30.6**	**3.74**	**4.45**
CFQ (Cognitive fusion)	$r = 0.5$	δ	7.49	0.83	2.46	5.61	2.75	3.65	0.84	1.90	7.34	5.37
		B_{ar}	**>100**	2.53	**17.1**	**>100**	**19.4**	**40.1**	2.72	**8.80**	**>100**	**>100**
	$r = 1$	δ	7.61	1.13	2.65	5.78	3.06	3.94	1.15	2.23	7.34	5.48
		B_{ar}	**>100**	2.77	**24.1**	**>100**	**37.9**	**>100**	**3.05**	**14.3**	**>100**	**>100**
	$r = 2$	δ	7.67	1.40	3.00	5.98	3.29	4.12	1.41	2.53	7.55	5.64
		B_{ar}	**>100**	2.53	**35.3**	**>100**	**46.6**	**>100**	2.76	**17.5**	**>100**	**>100**
PTQ (Perseverative thinking)	$r = 0.5$	δ	12.6	1.49	1.11	12.8	6.23	1.15	0.83	6.9	5.01	6.21
		B_{ar}	**>100**	**5.08**	**3.73**	**>100**	**>100**	**4.07**	2.65	**>100**	**93.8**	**>100**
	$r = 1$	δ	12.7	1.85	1.42	12.8	6.42	1.51	1.17	7.00	5.23	6.24
		B_{ar}	**>100**	**7.20**	**4.62**	**>100**	**>100**	**5.31**	**3.16**	**>100**	**>100**	**>100**
	$r = 2$	δ	12.9	2.19	1.70	13.0	6.45	1.78	1.43	7.20	5.42	6.42
		B_{ar}	**>100**	**8.48**	**4.39**	**>100**	**>100**	**5.43**	2.87	**>100**	**>100**	**>100**
VQ (Progress values)	$r = 0.5$	δ	2.85	1.83	3.40	1.41	2.05	0.95	0.07	0.93	0.03	0.35
		B_{ar}	**24.7**	**7.91**	**35.2**	**4.33**	**10.8**	2.90	0.59	2.85	0.60	1.22
	$r = 1$	δ	3.08	2.15	3.70	1.80	2.36	1.29	0.12	1.25	0.07	0.55
		B_{ar}	**36.3**	**11.3**	**74.7**	**5.77**	**17.9**	**3.56**	0.41	**3.41**	0.42	1.09
		δ	3.35	2.48	3.89	2.15	2.64	1.58	0.17	1.55	0.07	0.68

	$r = 2$	B_{ar}	**51.3**	**14.5**	**>100**	**6.76**	**23.4**	**3.42**	0.25	**3.38**	0.25	0.76
VQ (Obstruction values)	$r = 0.5$	δ	2.72	1.16	3.44	0.64	3.35	1.76	0.51	2.35	2.21	3.20
		B_{ar}	**21.5**	**3.90**	**41.6**	1.89	**23.7**	**7.56**	1.58	**11.9**	**9.72**	**43.5**
	$r = 1$	δ	2.94	1.48	3.64	0.92	3.64	2.13	0.76	2.70	2.57	3.38
		B_{ar}	**36.0**	**4.91**	**69.5**	1.89	**44.8**	**11.5**	1.53	**22.9**	**15.0**	**72.8**
	$r = 2$	δ	3.26	1.76	3.85	1.15	3.99	2.42	0.97	2.98	2.94	3.59
		B_{ar}	**54.5**	**4.97**	**96.9**	1.59	**75.6**	**13.5**	1.21	**27.3**	**19.9**	**96.6**

Note. B_{ar} = Bayes Factors of the JZS+AR model. $B_{ar} > 1$ supports the hypothesis of intervention effect. $B_{ar} > 3$ are in bold to highlight where at least substantial evidence of treatment effect was found.