The Effect of Confidence on Dropout Rate and Outcomes in Online Cognitive Bias Modification

Nicola Hohensee¹ · M. Joseph Meyer² · Bethany A. Teachman²

Received: 22 September 2019 / Revised: 3 January 2020 / Accepted: 4 February 2020 / Published online: 20 February 2020
© The Author(s) 2020

Abstract

Cognitive bias modification training for interpretation bias (CBM-I) trains less threat-oriented interpretation patterns using basic learning principles and can be delivered completely online. Thus, CBM-I may increase accessibility of treatment options for anxiety problems. However, online interventions are often affected by pronounced dropout rates, and findings regarding the effectiveness of CBM-I, especially when delivered online, are mixed. Therefore, it is important to identify further predictors of dropout rate and intervention outcomes. The current study uses data from an exclusively online, multi-session CBM-I feasibility trial to investigate the effects of perceived confidence in the likely helpfulness of a CBM-I program (rated at baseline) on subsequent dropout rate, and change in interpretation bias and anxiety symptoms for $N = 665$ individuals high in trait anxiety. Results showed that higher baseline confidence ratings predicted lower dropout rate, as expected. Partially in line with hypotheses, there was some evidence that higher confidence ratings at baseline predicted greater changes on some interpretation bias measures, but results were not consistent across all measures and fit indices. Inconsistent with hypotheses, confidence did not predict change in anxiety symptoms. Possible explanations for why the nonsignificant findings may have occurred, as well as implications of confidence in online interventions being an early predictor of high risk for dropout, are discussed.

Keywords Interpretation · Cognitive bias · Confidence · Anxiety · Attrition

Anxiety disorders are the most common form of mental illness in the USA, with approximately 18% of the US population affected by an anxiety disorder every year (Kessler et al. 2005). However, only a little over 20% of anxious individuals receive mental health treatment in response to their anxiety problems (Wang et al. 2005), suggesting that more accessible treatment options are needed. Cognitive bias modification for interpretation bias (CBM-I), which can be offered online, may be one such option. CBM-I uses basic learning principles to shift maladaptive interpretation biases (typically in a more benign direction), which, in turn, is theorized to decrease

anxiety symptoms (MacLeod and Mathews 2012). This approach can change both threat interpretation bias and anxiety symptoms (see meta-analyses and reviews in Hallion and Ruscio 2011; MacLeod and Mathews 2012); however, findings are mixed, with greater effects obtained for shifting cognitive biases than shifting symptoms (Menne-Lothmann et al. 2014). CBM-I is a relatively novel intervention and, as the mixed findings indicate, much remains unknown about the key predictors of when it works and when it does not, especially when it is delivered online. Furthermore, online interventions are especially affected by pronounced dropout rates, which usually range between 50 and 70% (Barak and Grolhol 2011) but are often higher than 90% (e.g., in MoodGYM, one of the most popular cognitive behavior therapy websites, only 10% of a sample of 82,000 completed two or more modules; Batterham et al. 2008). Given the large number of participants who drop out of online interventions even before they have tried them (e.g., they sign up and complete an assessment but never follow through with the intervention), it is important to identify predictors of dropout rate and intervention outcomes that can be assessed during the assessment phase, even before the active intervention has begun. Thus, the current study examined confidence in the likely helpfulness of a CBM-I

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s41347-020-00129-8>) contains supplementary material, which is available to authorized users.

✉ Nicola Hohensee
n_hohe01@uni-muenster.de

¹ Department of Psychology, Westfälische Wilhelms-Universität Münster, Fliegerstraße 21, 48149 Münster, Germany

² Department of Psychology, University of Virginia, Charlottesville, VA, USA

intervention, rated after hearing the intervention rationale but prior to intervention delivery, as a motivation-relevant predictor of dropout and cognitive bias and symptom outcomes.

Motivation-Relevant Predictors in Psychological Interventions

In the past, motivation-relevant predictors have often been associated with outcomes in psychological interventions. Two well-known concepts are credibility, defined as “how believable, convincing, and logical [a] treatment is,” and expectancy, defined as “improvements that clients believe will be achieved” (Deville and Borkovec 2000, p. 75). Both credibility and expectancy are conceptually related to general confidence in an intervention; thus, research addressing these concepts is informative for the current study (though note that credibility and expectancy are typically assessed *after* a portion of the intervention has been completed, while we assess confidence after exposure to the intervention rationale but prior to starting the intervention).

Higher credibility and expectancy have been associated with symptom improvement in online psychotherapy for anxiety problems, both when guided by written conversations (El Alaoui et al. 2015; Hedman et al. 2015) and when not guided (Nordgreen et al. 2012). In addition, both credibility and expectancy have been linked to greater engagement (defined as number of completed sessions or homework compliance) and lower dropout in online psychotherapy (Alfonsson et al. 2016; El Alaoui et al. 2015; Nordgreen et al. 2012; Snippe et al. 2015).

In the context of CBM-I, to our knowledge, three studies have examined the relationship between credibility/expectancy and outcome measures. For patients with social anxiety disorder, credibility was associated with pre-to-post change in social anxiety symptoms in a lab-based CBM-I training for attention and interpretation bias (Beard et al. 2011). For depressed individuals, credibility and expectancy were significantly associated with lower depression and clinical anger at the post-assessment in a mixed lab-based plus online CBM-I training for hostile interpretation bias (Smith et al. 2018, 2019). Interestingly, one study found a significant interaction effect between credibility/expectancy and the experimental conditions (Smith et al. 2018) and the other did not (Smith et al. 2019). More precisely, Smith et al. (2018) found that individuals with high expectancy reported lower postdepression symptoms in the active experimental condition compared to the control condition. Taken together, findings are mixed but suggest that motivation-related predictors might be important in the context of online CBM-I, particularly for active CBM-I in comparison to control conditions.

However, prior CBM-I studies focused only on the relation between motivation-related predictors and far-transfer

outcome measures like anxiety or depression symptoms. Thus, it remains unknown whether near-transfer outcome measures like interpretation bias can be affected as well. Furthermore, we need to better understand the moderating effect of experimental condition given findings in the literature have been mixed (Smith et al. 2018, 2019). Differential effects of motivation-related predictors on active experimental and control conditions might shed light on underlying mechanisms. Does confidence only boost nonspecific placebo effects or does pre-intervention confidence facilitates specific processes in CBM-I (e.g., contingency learning) as well?

Overview and Hypotheses

The present study uses data from an initial feasibility trial (Ji et al. 2018) to investigate perceived confidence in an online, multi-session CBM-I training and its effect on dropout rate, and change in interpretation bias and anxiety symptoms for individuals high in trait anxiety. Participants were assigned to one of three training conditions using a modified version of the ambiguous scenario paradigm (initially developed by Mathews and Macintosh 2000): (1) positive training with 90% positive CBM-I scenarios, (2) 50/50 training with 50% positive and 50% negative CBM-I scenarios, or (3) no CBM-I training. In addition, participants completed either an imagery exercise with anxiety primes or neutral primes, creating six-nested groups in total. Ji et al. (2018) generally showed that positive CBM-I training (especially in combination with anxiety primes) could successfully decrease negative interpretation bias and anxiety symptoms and increase positive interpretation bias in comparison to the 50/50 training condition and the no-CBM-I condition, though results were not consistent across all measures. The current paper builds on these group-level findings to consider how individual differences in confidence in the program may explain the variability in results.

Based on past findings from the broader literature, we hypothesize that higher perceived confidence will be associated with a lower dropout rate (*hypothesis 1*). In addition, we expect higher confidence at baseline to predict greater reduction in interpretation bias (*hypothesis 2.1*) and anxiety symptoms (*hypothesis 2.2*). Further, we expect that confidence will be the strongest predictor for the positive CBM-I condition, relative to the 50/50 and control conditions (*hypothesis 3.1*), because confidence will boost nonspecific placebo effects as well as specific CBM-I processes during positive CBM-I training. We further hypothesize that confidence will be the weakest (but still significant) predictor for the control condition without CBM-I (*hypothesis 3.2*) because confidence can only boost nonspecific placebo effects.

Before data analysis, the present study was preregistered at Open Science Framework, <https://osf.io/y4fw8/>.

Method

Participants

Participants were recruited through online advertisements, word of mouth (e.g., sharing Facebook posts), a press release, and from a link at the Project Implicit Mental Health site. They participated voluntarily in the online study, with no compensation provided. The study was conducted at MindTrails, a web-based CBM-I research infrastructure (see <https://mindtrails.virginia.edu/>). When visiting the website, potential participants first completed the depression, anxiety, stress scales – short form: anxiety subscale (DASS-AS; Lovibond and Lovibond 1995). Individuals who scored > 10 (suggesting moderate to extremely severe anxiety) were invited to participate after completing the screening. Overall, the initial sample reported an average score of 23.44 ($SD = 8.08$)¹ on the DASS-AS, indicating extremely severe anxiety.

Six hundred and sixty-five participants consented and began the first online training session between May 2016 and March 2018. All participants were included in the subsequent analyses as an intent-to-treat sample. By session six, $N = 572$ participants had dropped out (dropout rate 86%), leaving a sample of $N = 93$. Due to power concerns based on the large dropout rate, we decided prior to conducting analyses that only outcome measures through training session six would be included in the analyses (i.e., we did not include data up to session 8, following Ji et al. 2018). The reported gender of the sample was female (71.92%), male (25.38%), or other nonbinary category (2.70%). The mean age of the sample was 35 years ($SD = 13.61$, range = 19–92)², and the reported race was White (77.63%), Asian (7.36%), Black (3.45%), and other (11.56%). Ethnicity was reported as 83.76% non-Hispanic, 7.82% Hispanic, and 8.42% as other.

Materials

Only those materials that are directly relevant to the main questions assessed in this paper are listed here. A full list of measures administered as part of the larger, parent study is available from the last author and further details on the larger trial are reported at [ClinicalTrials.gov](https://clinicaltrials.gov) Identifier: NCT03498651.

Training Paradigm

After pre-assessment, participants were randomly assigned to one of six conditions, which were based on the type of CBM-I

training received (positive, 50/50, or the control group, which received no training), and the type of imagery exercise administered (anxiety or neutral).

In the CBM-I training conditions, participants completed up to eight sessions of CBM-I training adapted from Mathews and Macintosh (2000). Scenarios depicted everyday situations that were potentially anxiety provoking (e.g., concern a boss is evaluating you negatively) in which one or two letters of the last word in the scenario were missing. By completing the word fragment, the scenarios could be resolved in either a threatening manner (negative) or nonthreatening manner (positive). In addition, participants answered a comprehension question after completing the word fragment in each trial to reinforce the now-resolved emotional meaning of the scenario (scenario examples are presented in Appendix A). Participants in the positive and the 50/50 conditions completed 40 training scenarios per session. In the positive conditions, the word fragment was assigned a benign meaning in 90% of the cases, and in the 50/50 conditions, the emotionally ambiguous scenarios were resolved positively in 50% of the cases and negatively in 50% of the cases. Participants in the no-CBM-I training control condition completed anxiety symptom assessments and the imagery prime (see below) on the same schedule as the other participants, but were not assigned any interpretation training stimuli.

All conditions (including the no-CBM-I control) completed an imagery exercise prior to their assigned training condition that involved imagining either a personal upcoming situation that makes them feel anxious (anxiety primes) or an upcoming mundane task (neutral primes). The full imagery prime script is presented in Appendix B.

Interpretation Bias: Recognition Ratings

At pre-assessment, and after sessions three and six, participants were presented with nine emotionally ambiguous-titled scenarios modified from Mathews and Macintosh (2000). Participants were asked to complete a word fragment at the end of each scenario and answer a comprehension question afterwards. In contrast to the CBM-I training, the word fragment and the comprehension question were designed to preserve rather than resolve the emotional ambiguity of the scenario. Afterwards, only the titles of the original scenarios were presented, together with four versions of a new sentence. Each version of the sentence represented a different valenced interpretation of the scenario (anxiety-relevant positive ending, anxiety-relevant negative ending, anxiety-irrelevant positive ending, anxiety-irrelevant negative ending). The participants were asked to rate how similar in meaning each sentence was to the original scenario from (0) “Very different” to (3) “Very similar,” with more similar ratings indicating that the meaning was assigned to the ambiguous situation. To assess change in both negative and positive interpretation bias, average scores

¹ The score is based on $N = 611$ due to technical problems. Although precise screening scores were not recorded for $N = 54$ participants, all participants met eligibility criteria before they were invited for participation.

² The value is based on $N = 641$ due to exclusion of answers that were likely not valid.

were created separately for the anxiety-relevant negative and the anxiety-relevant positive endings.

Interpretation Bias: BBSIQ

At pre-assessment, and after sessions three and six, participants completed the Brief Body Sensations Interpretations Questionnaire (BBSIQ; Clark et al. 1997). The BBSIQ is a 14-item questionnaire that assesses endorsement of threatening, or positive and neutral, interpretations for seven ambiguous panic/body sensation situations and seven external, potentially threatening events from (0) “Not at all likely” to (4) “Extremely likely.” Usually, the BBSIQ is administered on a scale from (0) to (8), but we shortened it to align with other rating scales in the study. The negative interpretation bias was calculated by computing the mean of all threat interpretation ratings.

Anxiety: OASIS

The Overall Anxiety Severity and Impairment Scale (OASIS; Norman et al. 2006) was assessed at pre-assessment and after each session. The OASIS contains five items, which are rated on a 5-point Likert-scale from (0) to (4) with different anchors depending on each question. Items assess frequency and severity of anxiety symptoms, as well as impairment caused by anxiety. Item ratings are added for a total severity score with higher numbers indicating greater severity.

Confidence

At pre-assessment, before the first training session but after reading information about the rationale of CBM-I (see Appendix C), participants rated their confidence in the likely helpfulness of the training. Items were modified from Borkovec and Nau (1972) and rated on a 5-point Likert-scale from (0) “Not at all” to (4) “Extremely.” The exact wording was “How confident are you that an online training program will be successful in reducing your anxiety?” and “How confident are you that an online training program that is designed to change how you think about situations will be successful in reducing your anxiety?”. Given the two items were moderately to highly correlated ($r = .63$), the average of the two responses was used in all analyses to simplify the analyses and reduce the number of tests.

Data Analysis

We assessed the relation between confidence and dropout rate (*hypothesis 1*) with a binary logistic regression analysis, which is used to estimate the odds that an outcome occurs. For this analysis, we used the confidence score to predict a

binary variable that had a value of one if participants did not drop out of the study before session six, and zero otherwise.

We assessed all other hypotheses (2.1 to 3.2) by constructing latent growth curve models (LGM). Data were collapsed across the imagery prime variable due to power concerns and because we did not have unique hypotheses based on imagery prime. Thus, we constructed multigroup models for each outcome measure (Negative Recognition Ratings, Positive Recognition Ratings, BBSIQ, and OASIS), whereby each group within each model corresponds to one of the CBM-I conditions (positive, 50/50, and no-CBM-I). Latent intercepts and slopes were extracted from the manifest outcome measures across each measured time point. More precisely, for Recognition Ratings and BBSIQ, intercepts and slopes were extracted across pre-assessment, session three, and session six. For models that were performed using the OASIS scores, intercept and slope were extracted across all time points from pre-assessment to session six. In these models, slopes represent estimates of the development of outcome measures across different time points.

Before conducting the primary data analyses, linear and quadratic models were compared for all outcome measures. The model with the lowest Akaike information criterion (AIC) and the lowest Bayesian information criterion (BIC) was chosen, though other important fit indices were also reported: chi-square test, comparative fit index (CFI), Tucker Lewis Index (TLI), and root mean square error of approximation (RMSEA). Thus, paths from slope to outcome measures were either fixed to values describing a linear or a quadratic positive trend, depending on AIC and BIC. Furthermore, because of the increasing dropout toward the end of the study, the intercept was chosen to be an estimate of ending values of outcome measures at session six. This decision was made to maximize the number of observations used in the estimation of the slope and minimize estimation bias due to attrition (following a personal communication with Steve M. Boker). The covariance between slope and intercept was estimated in all models as well. To accommodate the high number of missing values without dropping cases, we used full information maximum likelihood estimation. This approach produces less biased parameter estimates than alternatives in the presence of missing data by estimating maximum likelihood based on the raw data, instead of covariances (Enders and Bandalos 2001). Means of intercept, slope, and independent variables were estimated by adding a constant to the model with paths to each of the variables, and the confidence score was included as the independent variable. Furthermore, OASIS score at the pre-training assessment and the imagery prime condition (dummy coded) were added as covariates. Paths were thus estimated to the intercept and slope for confidence, baseline OASIS, and imagery prime condition (though baseline OASIS was not included as a covariate in the OASIS outcome models given, it is already modeled as part of the change process). Note,

baseline OASIS was first standardized due to the large difference in variance compared to confidence and imagery prime condition. To ensure that there were no misspecifications in the models, we restricted the lower bounds of all variances to 0.001. An exemplary model is presented in Fig. 1.

For each outcome measure, the multigroup model with freely varying paths for the independent variables (*All-Diff-Model*) was compared to more restricted models via a likelihood ratio test. At first, the *All-Diff-Model* was compared to a multigroup model where paths from confidence to the intercept and slope were restricted to be the same values across training condition groups, indicating the same impact of confidence across conditions (*All-Equal-Model*). Second, a multigroup model was fitted where paths from confidence to the intercept and slope were restricted to zero, indicating no impact of confidence on outcomes across conditions (*All-Zero-Model*). The *All-Zero-Model* was compared to both the *All-Diff-Model* and the *All-Equal-Model*, resulting in three likelihood ratio tests for each outcome measure. Significant p values indicate a reliable gain of fit for the less restricted model. In addition, the AIC and BIC were compared for each of the three multigroup models. For the best-fitting multigroup model, single path estimates from confidence to intercept and slope were tested for significance.

Results

Attrition

Our total intent-to-treat sample ($N = 665$) reported an average confidence of 1.83 ($SD = 0.76$, range = 0–4), indicating that participants found the online training and its rationale

somewhat credible. Furthermore, dropout did not vary across CBM-I conditions ($p = .31$).

The binary logistic regression revealed a significant positive coefficient of .37 for confidence ($p = .01$), indicating that an increase of one unit in confidence increased the log odds of no-dropout (versus dropout) by .37 units. The resulting odds ratio for confidence of $OR = 1.44$ [1.08, 1.94] showed that increasing confidence by one unit led to a 44% increase in the odds of not dropping out. Taken together, the results indicate a smaller dropout rate (through session 6) for participants with higher confidence at baseline, as expected.

Outcome Measures

For all outcome measures, the correlation between confidence ratings and the baseline assessment of the outcome variable was assessed. Only Positive Recognition Ratings at baseline was significantly correlated with confidence ($r = .11$, $p < .05$), and the effect was very small. All other correlations were not reliable, suggesting that confidence is generally not associated with severity of interpretation bias or anxiety symptoms at baseline.

Negative Recognition Ratings

For change in Negative Recognition Ratings, a quadratic model, describing greater change from pre-assessment to session three than from session three to session six, showed the best fit with acceptable values for the chi-square statistic, CFI, and RMSEA, with only TLI having a value below the recommended guideline of .90 (see Table 1).

All likelihood ratio tests revealed no significant differences between multigroup models (all $p > .05$). In addition, the AIC

Fig. 1 OASIS = overall anxiety severity and impairment scale; RRNegative = Negative Recognition Ratings. Sample latent growth curve model for one CBM condition and Negative Recognition Ratings as outcome measure. Note, that values from slope to outcome measure indicate here a quadratic trend

Table 1 Fit indices for all outcome measures

	χ^2	CFI	TLI	RMSEA	AIC	BIC
Negative Recognition Ratings (quadratic)						
All-Diff-Model	0.19	0.92	0.73	0.02	5596.87	5862.97
All-Equal-Model	—	—	—	—	5593.68	5842.01
All-Zero-Model	—	—	—	—	5593.19	5832.66
Positive Recognition Ratings (quadratic)						
All-Diff-Model	0.21	0.94	0.81	0.02	5404.50	5670.58
All-Equal-Model	—	—	—	—	5401.73	5650.06
All-Zero-Model	—	—	—	—	5406.24	5645.71
BBSIQ (quadratic)						
All-Diff-Model	0.41	0.99	0.98	0.01	6196.92	6466.72
All-Equal-Model	—	—	—	—	6197.35	6449.17
All-Zero-Model	—	—	—	—	6197.86	6440.69
OASIS (linear)						
All-Diff-Model	0.69	1.03	1.17	0.00	11,934.78	12,204.22
All-Equal-Model	—	—	—	—	11,932.07	12,183.55
All-Zero-Model	—	—	—	—	11,931.62	12,174.12

BBSIQ brief body sensations interpretations questionnaire, *OASIS* overall anxiety severity and impairment scale, χ^2 Chi-square test, *CFI* comparative fit index, *TLI* Tucker Lewis index, *RMSEA* root mean square error of approximation, *AIC* Akaike information criterion, *BIC* Bayesian information criterion; χ^2 should be non-significant, CFI and TLI are acceptable above .90 and RMSEA is acceptable below 0.08

and BIC were lowest for the *All-Zero-Model* (see Table 1). Thus, the results indicated no significant effect of confidence on change in Negative Recognition Ratings for the three conditions.

Positive Recognition Ratings

For change in Positive Recognition Ratings, a quadratic model again had the best fit with acceptable values for chi-square test, CFI, and RMSEA. The TLI was below the recommended guideline of 0.90 (see Table 1).

Likelihood ratio tests did indicate a significant gain in fit for the *All-Equal-Model* ($p = .01$) and the *All-Diff-Model* ($p = .03$) in comparison to the *All-Zero-Model*. The *All-Diff-Model* did not show a significant gain in fit compared to the *All-Equal-Model* ($p = .27$), which indicates that a moderation effect by CBM-I condition is unlikely. In line with the likelihood ratio tests, AIC was lowest for the *All-Equal-Model* (see Table 1), which differed by more than four units to the AIC of the *All-Zero-Model*, indicating support for the *All-Equal-Model* model, which assumes an equal impact of confidence for all conditions (Burnham and Anderson 2004). With that being said, the BIC was lowest for the *All-Zero-Model* rather than the *All-Equal-Model* (see Table 1), and path estimates from confidence to the intercept and slope in the *All-Equal-Model* were nonsignificant (see Table 2). Thus, the *All-Equal-Model* that assumes an equal impact of confidence across conditions seemed to be the best fitting model for most indices, indicating that higher confidence ratings at baseline predicted more gains on the measure of Positive Recognition

Ratings over time; however, given the nonsignificant individual paths in the *All-Equal-Model*, this confidence effect should be interpreted with caution. (Note, the variable results may follow from the different tests and indices that were used. While the intercept and slope for confidence can be nonsignificant when examined individually, their combination can lead to significant findings in likelihood ratio tests.)

BBSIQ

Similar to the models for the Recognition Ratings, a quadratic model showed the best fit for change in BBSIQ scores with acceptable values for all fit indices (see Table 1).

Likelihood ratio tests indicated that the *All-Diff-Model* did not show a significant gain in fit compared to the *All-Equal-Model* ($p = .08$), but the *All-Diff-Model* did show a significant gain in fit compared to the *All-Zero-Model* ($p = .04$). The *All-Equal-Model* showed no significant gain in fit in comparison to the *All-Zero-Model* ($p = .10$). The likelihood ratio tests thus provide some support for the *All-Diff-Model*. However, the AIC values were very similar across models and the BIC was lowest for the *All-Zero-Model* (see Table 1), pointing to some discrepant results across the various fit indices.

To understand the variable findings, path estimates were examined separately by CBM-I condition. As evident in Table 2, confidence only had a significant influence on the slope in the positive CBM-I condition. Also, follow-up likelihood ratio tests showed that the estimate for the slope in the positive CBM-I condition was significantly different to the

Table 2 Path estimates for the best-fitting multigroup model from confidence to intercept and slope

		Estimate path	95% confidence interval
Positive Recognition Ratings			
All-Equal-Model			
	Intercept	0.04	[− 0.07, 0.16]
	Slope	− 0.01	[− 0.04, 0.02]
BBSIQ			
All-Diff-Model			
Positive	Intercept	− 0.22	[− 0.45, 0.00]†
	Slope	− 0.07	[− 0.13, − 0.02]*
50/50	Intercept	− 0.15	[− 0.35, 0.06]
	Slope	− 0.05	[− 0.10, 0.00]†
No-CBM	Intercept	− 0.03	[− 0.25, 0.19]
	Slope	0.02	[− 0.03, 0.07]

BBSIQ brief body sensations interpretations questionnaire, *OASIS* overall anxiety severity and impairment scale. Path estimates are not presented for Negative Recognition Ratings and *OASIS* because the best-fitting model was the *All-Zero-Model* where path estimates for intercept and slope were restricted to zero anyway

*indicates that confidence interval did not overlap zero; †indicates that confidence interval reached zero

estimate in the no-CBM-I condition ($p = .03$) but not to the estimate in the 50/50 CBM-I condition ($p = .85$). Estimates in the no-CBM-I condition and the 50/50 CBM-I condition did not differ significantly ($p = .07$).

Taken together, these results indicate that confidence's effect on change in BBSIQ scores depends on the CBM-I condition; specifically, the effect of confidence is strongest in the positive CBM-I condition, indicating a greater decline in threat interpretations for people with higher confidence. However, due to the differing results across fit indices, this confidence effect should again be interpreted with some caution.

OASIS

In contrast to the other outcome measures, a linear model showed the best fit for change in *OASIS* scores with acceptable values for all fit indices. It should be noted that fit indices also point toward a possible overestimation of the model (see Table 1).

For the *OASIS*, all likelihood ratio tests revealed no significant differences between multigroup models (all $p > .05$). In addition, the AIC and BIC were lowest for the *All-Zero-Model* (see Table 1), suggesting that confidence has no significant effect on the change in *OASIS* scores.

Discussion

The current study examined the effect of pre-treatment confidence in an online CBM-I program on dropout rate, and change in interpretation bias and anxiety symptoms for individuals high in anxiety. In line with hypotheses, participants

with higher confidence ratings at baseline had lower dropout rates. Support for hypotheses regarding confidence's effects on change in interpretation bias was more mixed. There was some evidence that higher confidence at baseline predicted greater changes on some interpretation bias measures (Positive Recognition Ratings and BBSIQ scores), especially for participants in the positive CBM-I condition (on the BBSIQ), but results were not consistent across all fit indices or individual path estimates. Finally, inconsistent with hypotheses, confidence did not predict change in anxiety symptoms (*OASIS* scores).

The nonsignificant findings for confidence predicting symptom change are surprising because these contradict Beard et al. (2011) and Smith et al. (2018, 2019), who found a significant positive relation between credibility/expectancy, which are closely related concepts to confidence, and change in far-transfer outcome measures like social anxiety and depression symptoms following CBM-I trainings. One possible explanation might be that a patient-therapist alliance is indeed a necessary mediator between confidence and its effects on symptom improvement. All previous CBM-I studies examining concepts related to confidence were at least partly lab-based and thus, at a minimum, superficial participant-researcher relationships were established. However, this explanation is not wholly satisfactory because Nordgreen et al. (2012) found a relation between credibility and symptom improvement in non-guided online CBT as well. Thus, the missing patient-therapist alliance is unlikely to be the only explanation for the nonsignificant findings.

Another difference between the current study and prior ones is the timing of the confidence measurement. Beard et al.

(2011), Smith et al. (2018, 2019), and Nordgreen et al. (2012), all assessed perceived credibility and expectancy *after* the first training session or after the first week of treatment. Thus, assessment of perceived credibility was likely affected by the participant's first training experience. Participants who had a positive first training experience (and perhaps made some early gains) most likely rated credibility higher than participants who perceived their first session as less helpful. In the current study, participants rated their general confidence in the training at the pre-treatment assessment, because we wanted to separate perceived confidence in CBM-I (based on the rationale explaining the intervention) from initial training effects. Also, given that dropout rates are high for online programs even before the intervention has begun, we elected to place the confidence assessment very early in the program. The difference in our findings raises the possibility that confidence's predictive effect for symptom change requires some initial experience with the intervention, at least in the context of interventions like CBM-I that involve little to no personal interaction. Alternatively, it may be that part of the earlier studies' observed credibility/expectancy effects actually reflects early treatment response. Future studies that systematically vary when confidence is assessed as a function of exposure to the intervention contents could help differentiate these possibilities.

Notably, confidence in the current study was predictive of dropout rate and some interpretation bias measures, supporting the importance of considering the impact of confidence even for exclusively online-based trainings. Given high dropout rates and limited reach are major challenges for most online interventions (e.g., Batterham et al. 2008), finding baseline predictors like confidence that predict lower dropout and more benefits on near-transfer outcome measures is valuable. It is especially useful to the extent that perceived confidence is malleable (e.g., based on the information provided about the intervention and its rationale). Participants in the current study indicated, before starting the intervention but after brief information about the rationale of CBM-I was provided, that they were somewhat confident regarding the likely success of the CBM-I training. Thus, perceived confidence could still be improved. Offering participants more detailed information about the mechanisms behind CBM-I, and showing them examples of training materials might help to increase the perceived confidence and associated benefits. For example, participants could be provided with vivid examples that emphasize the relation between less-threatening interpretations and less anxious feelings.

In this study, we intentionally presented the same general rationale to all participants, regardless of condition, so as not to create different expectations as a function of condition at baseline. Yet, we still saw some interactive effect of condition, with confidence predicting change on the BBSIQ interpretation bias measure for those in the positive training condition. With that being said, the

moderating effect of condition was not evident across most measures. This may be because we assessed perceived confidence before participants had been exposed to any differences in the program as a function of condition, or it could be that confidence's predictive effects are not augmented by specific intervention influences and operate instead on more general, nonspecific factors. Future work comparing moderation by different CBM-I and control conditions that vary whether the same or different rationales are provided will be helpful to determine the critical moderators of confidence's predictive effects.

It should be noted that the present study has important limitations. The dropout rate was very high, which decreased statistical power. Although we tried to address this with full information maximum likelihood and collapsing across imagery prime conditions, we cannot rule out the possibility that the high attrition rate affected our findings and might have contributed to the nonsignificance of certain results (e.g., confidence intervals in Positive Recognition Ratings). In addition, the TLI and CFI were greater than one, and RMSEA was estimated to be negative (and bounded at 0) in some models. It seems likely that these models may have been over-specified given the number of parameters, but reducing the complexity of the models would have made it difficult to ask our research questions. Therefore, we focused on model comparisons instead of fitting the data perfectly into single models, especially given the high level of missing information. Nevertheless, future studies with higher statistical power are needed for replication.

Notwithstanding the limitations, the current study represents an initial feasibility trial which indicates that confidence may be an important factor, especially for predicting dropout, even in exclusively online CBM-I trainings. Targeted future research is needed to better understand the relation between confidence and symptom change, and which aspects of this relationship are moderated by experimental condition.

Acknowledgments The authors thank Cynthia L. Grotz for her edits on a previous version of the paper, as well as Dan Funk, Diheng Zhang, Sam Portnow and the Teachman PACT lab, Barnes lab, and MindTrails team for their feedback and work developing the MindTrails platform. We are thankful to Emily Holmes, Simon Blackwell, Bundy Mackintosh, Andrew Mathews and their research teams for sharing some of their training and testing materials. Some of their scenarios were included, typically in a modified form, as part of our materials.

Funding Information Open Access funding provided by Projekt DEAL. This work was supported in part by NIMH grants (R34MH106770 and R01MH113752) to B. Teachman ([ClinicalTrials.gov](https://www.clinicaltrials.gov) Identifier: NCT03498651).

Compliance With Ethical Standards

Ethical Statement The research complies with ethical standards and ethical approval and was granted for the research by the Department of Psychology, University of Virginia.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alfonsson, S., Olsson, E., & Hursti, T. (2016). Motivation and treatment credibility predicts dropout, treatment adherence, and clinical outcomes in an internet-based cognitive behavioral relaxation program: a randomized controlled trial. *Journal of Medical Internet Research*, 18, e52. <https://doi.org/10.2196/jmir.5352>.
- Barak, A., & Grohol, J. M. (2011). Current and future trends in internet-supported mental health interventions. *Journal of Technology in Human Services*, 29, 155–196. <https://doi.org/10.1080/15228835.2011.616939>.
- Batterham, P. J., Neil, A. L., Bennett, K., Griffiths, K. M., & Christensen, H. (2008). Predictors of adherence among community users of a cognitive behavior therapy website. *Patient Preference and Adherence*, 2, 97–105.
- Beard, C., Weisberg, R. B., & Amir, N. (2011). Combined cognitive bias modification treatment for social anxiety disorder: a pilot trial. *Depression and Anxiety*, 28, 981–988. <https://doi.org/10.1002/da.20873>.
- Borkovec, T. D., & Nau, S. D. (1972). Credibility of analogue therapy rationales. *Journal of Behavior Therapy and Experimental Psychiatry*, 3, 257–260. [https://doi.org/10.1016/0005-7916\(72\)90045-6](https://doi.org/10.1016/0005-7916(72)90045-6).
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods and Research*, 33, 261–304. <https://doi.org/10.1177/0049124104268644>.
- Clark, D. M., Salkovskis, P. M., Öst, L. G., Breitholtz, E., Koehler, K. A., Westling, B. E., et al. (1997). Misinterpretation of body sensations in panic disorder. *Journal of Consulting and Clinical Psychology*, 65, 203–213. <https://doi.org/10.1037/0022-006X.65.2.203>.
- Deville, G. J., & Borkovec, T. D. (2000). Psychometric properties of the credibility/expectancy questionnaire. *Journal of Behavior Therapy and Experimental Psychiatry*, 31, 73–86. [https://doi.org/10.1016/S0005-7916\(00\)00012-4](https://doi.org/10.1016/S0005-7916(00)00012-4).
- El Alaoui, S., Ljotsson, B., Hedman, E., Kaldö, V., Andersson, E., Rück, C., et al. (2015). Predictors of symptomatic change and adherence in internet-based cognitive behaviour therapy for social anxiety disorder in routine psychiatric care. *PLoS One*, 10, e0124258. <https://doi.org/10.1371/journal.pone.0124258>.
- Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, 8, 430–457. https://doi.org/10.1207/S15328007SEM0803_5.
- Hallion, L. S., & Ruscio, A. M. (2011). A meta-analysis of the effect of cognitive bias modification on anxiety and depression. *Psychological Bulletin*, 137, 940–958. <https://doi.org/10.1037/a0024355>.
- Hedman, E., Andersson, E., Lekander, M., & Ljotsson, B. (2015). Predictors in Internet-delivered cognitive behavior therapy and behavioral stress management for severe health anxiety. *Behaviour Research and Therapy*, 64, 49–55. <https://doi.org/10.1016/j.brat.2014.11.009>.
- Ji, J. L., Zhang, D., Baee, S., Meyer, M. J., ... Teachman, B. A. (2018). Multi-session online interpretation bias training for anxiety in a community sample. Manuscript under review.
- 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, 617–627. <https://doi.org/10.1001/archpsyc.62.6.617>.
- 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. [https://doi.org/10.1016/0005-7967\(94\)00075-U](https://doi.org/10.1016/0005-7967(94)00075-U).
- MacLeod, C., & Mathews, A. (2012). Cognitive bias modification approaches to anxiety. *Annual Review of Clinical Psychology*, 8, 189–217. <https://doi.org/10.1146/annurev-clinpsy-032511-143052>.
- Mathews, A., & Macintosh, B. (2000). Induced emotional interpretation bias and anxiety. *Journal of Abnormal Psychology*, 109, 602–615. <https://doi.org/10.1037/0021-843X.109.4.602>.
- Menne-Lothmann, C., Viechtbauer, W., Höhn, P., Kasanova, Z., Haller, S. P., Drukker, M., van Os, J., Wichers, M., & Lau, J. Y. (2014). How to boost positive interpretations? A meta-analysis of the effectiveness of cognitive bias modification for interpretation. *PLoS One*, 9, e100925. <https://doi.org/10.1371/journal.pone.0100925>.
- Nordgreen, T., Havik, O. E., Öst, L. G., Furmark, T., Carlbring, P., & Andersson, G. (2012). Outcome predictors in guided and unguided self-help for social anxiety disorder. *Behaviour Research and Therapy*, 50, 13–21. <https://doi.org/10.1016/j.brat.2011.10.009>.
- Norman, S. B., Hami Cissell, S., Means-Christensen, A. J., & Stein, M. B. (2006). Development and validation of an overall anxiety severity and impairment scale (OASIS). *Depression and Anxiety*, 23, 245–249. <https://doi.org/10.1002/da.20182>.
- Smith, H. L., Dillon, K. H., & Cougle, J. R. (2018). Modification of hostile interpretation Bias in depression: a randomized controlled trial. *Behavior Therapy*, 49, 198–211. <https://doi.org/10.1016/j.beth.2017.08.001>.
- Smith, H. L., McDermott, K. A., Carlton, C. N., & Cougle, J. R. (2019). Predictors of symptom outcome in interpretation bias modification for dysphoria. *Behavior Therapy*. <https://doi.org/10.1016/j.beth.2018.10.001>.
- Snippe, E., Schroevers, M. J., Tovote, K. A., Sanderman, R., Emmelkamp, P. M., & Fleer, J. (2015). Patients' outcome expectations matter in psychological interventions for patients with diabetes and comorbid depressive symptoms. *Cognitive Therapy and Research*, 39, 307–317. <https://doi.org/10.1007/s10608-014-9667-z>.
- Wang, P. S., Lane, M., Olfson, M., Pincus, H. A., Wells, K. B., & Kessler, R. C. (2005). Twelve-month use of mental health services in the United States: results from the National Comorbidity Survey Replication. *Archives of General Psychiatry*, 62, 629–640. <https://doi.org/10.1001/archpsyc.62.6.629>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.